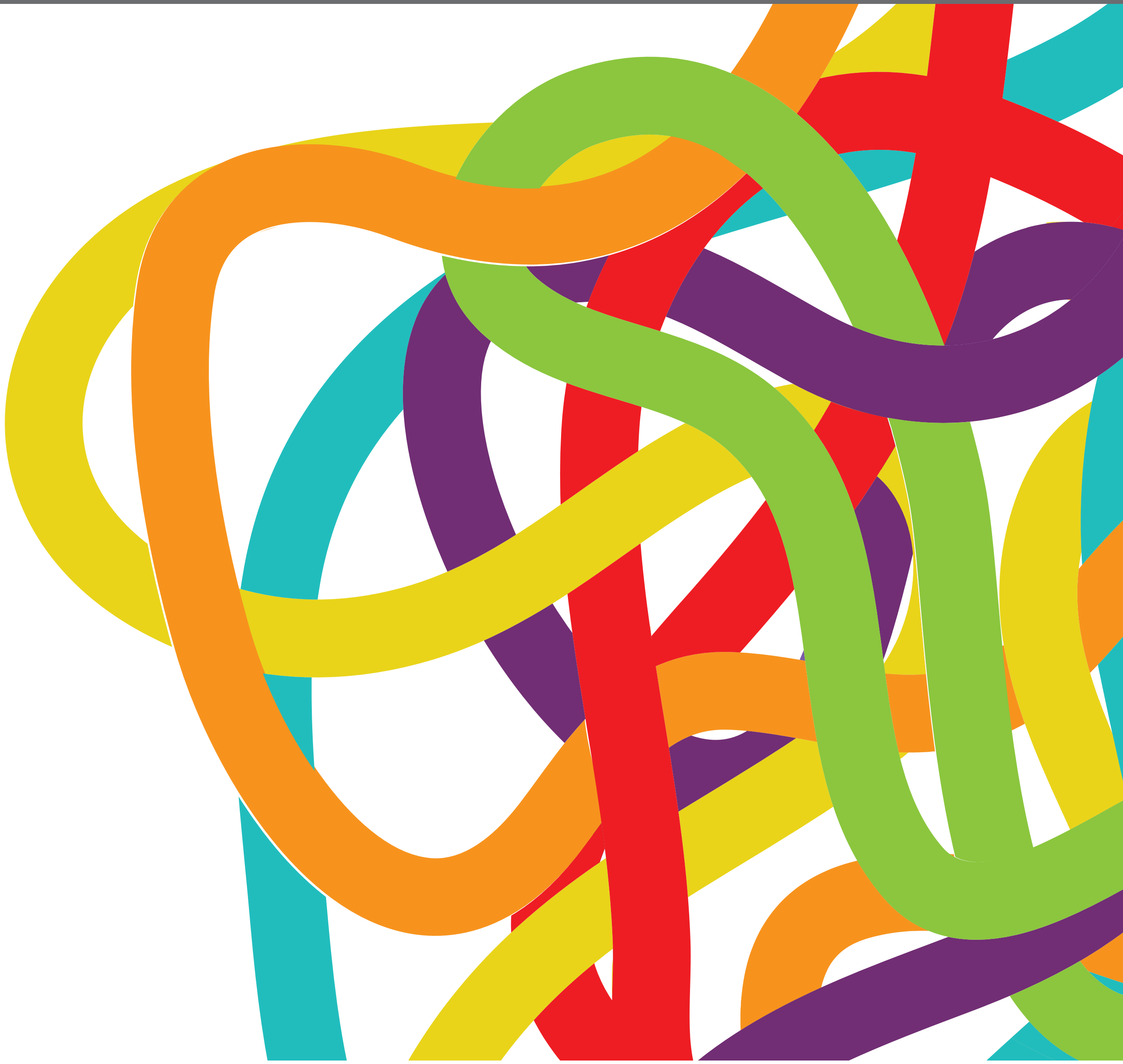
ULTRASOUND IN ONCOLOGY: APPLICATION OF BIG DATA AND ARTIFICIAL INTELLIGENCE

EDITED BY: Hui-Xiong Xu, Wei Wang, Richard Gary Barr and
Katsutoshi Sugimoto
PUBLISHED IN: Frontiers in Oncology



frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88974-311-7
DOI 10.3389/978-2-88974-311-7

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.
By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

ULTRASOUND IN ONCOLOGY: APPLICATION OF BIG DATA AND ARTIFICIAL INTELLIGENCE

Topic Editors:
Hui-Xiong Xu, Shanghai Tenth People's Hospital, Tongji University, China
Wei Wang, The First Affiliated Hospital of Sun Yat-sen University, China
Richard Gary Barr, Northeast Ohio Medical University, United States
Katsutoshi Sugimoto, Tokyo Medical University, Japan

Citation: Xu, H.-X., Wang, W., Barr, R. G., Sugimoto, K., eds. (2022). Ultrasound in Oncology: Application of Big Data and Artificial Intelligence. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88974-311-7

Table of Contents

frontiers
in Oncology

EDITORIAL
published: 22 December 2021
doi: 10.3389/fonc.2021.819487

Editorial: Ultrasound in Oncology: Application of Big Data and Artificial Intelligence

Yu-Ting Shen[1,2,3], *Wen-Wen Yue*[1,2,3] and *Hui-Xiong Xu*[1,2,3]*

[1] *Center of Minimally Invasive Treatment for Tumor, Department of Medical Ultrasound, Shanghai Tenth People's Hospital, Tongji University, Shanghai, China,* [2] *Ultrasound Research and Education Institute, Clinical Research Center for Interventional Medicine, School of Medicine, Tongji University, Shanghai, China,* [3] *Shanghai Engineering Research Center of Ultrasound Diagnosis and Treatment, National Clinical Research Center for Interventional Medicine, Shanghai, China*

Keywords: artificial intelligence, ultrasomics, ultrasound, big data, medicine

Editorial on the Research Topic

Ultrasound in Oncology: Application of Big Data and Artificial Intelligence

Edited and reviewed by:
Giuseppe Esposito,
MedStar Georgetown University Hospital, United States

Correspondence:
Hui-Xiong Xu
xuhuixiong@tongji.edu.cn

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *21 November 2021*
Accepted: *08 December 2021*
Published: *22 December 2021*

Citation:
Shen Y-T, Yue W-W and Xu H-X (2021) Editorial: Ultrasound in Oncology: Application of Big Data and Artificial Intelligence. Front. Oncol. 11:819487. doi: 10.3389/fonc.2021.819487

With the rapid development of science and technology, big data and artificial intelligence (AI) have ushered in a new era for medicine, especially the medical imaging. AI techniques are particularly applicable to imaging-based domains because the pixel values of the images themselves are quantifiable, which is the primary source of data for training and validating algorithms. Big data-based AI has attracted extensive attention for its superior performance and repeatability in medical image recognition. The application of AI technology can provide new clinical perspectives in complex medical imaging to improve diagnostic and surveillance accuracy (1, 2).

As a flexible imaging method, ultrasound (US), with its unique advantages including radiation-free, real-time imaging and ease of use, is expanding globally into various clinical areas as a first-line imaging modality. Ultrasomics is a field of image analysis with a bright future by extracting high-throughput quantitative data information from images and objectively characterizing and interpreting them for clinical analysis and diagnosis (3, 4). AI algorithm-based high-throughput data analysis of ultrasomics can reduce the high operator-dependent inherent nature of US technology, which clearly facilitates precision medicine with clinical decision systems support (Yin et al.; Kuang et al.). In our previous work (5), we summarized the relevant applications of US based on AI technology in different organs in recent years. The related studies also confirmed that this technology undoubtedly has a broad prospects in the field of US. For example, with a prospective and multicenter study design, we (6) developed an assembled convolutional neural network model for identifying molecular subtypes of breast cancer that could contribute to the clinical management of patients with breast disease. And the adoption of tele-US could improve the quality of breast US examinations for inexperienced US doctors (7). In addition, the machine learning-based US visual method (8) achieved an area under the curve (AUC) of 0.9 for the diagnosis of thyroid nodules, effectively reducing unnecessary fine needle aspiration biopsies in the clinical treatment of thyroid nodules. Significantly, the implementation of telemedicine technologies, including tele-US, has shown great value in protecting the health of patients and physicians while providing clinical care decisions for them during the coronavirus disease 2019 (COVID-19) pandemic (9–11). The advanced technologies such as big data analytics, AI technology, 5G networks and the internet of things undoubtedly provide invaluable opportunities for the development of ultrasomics. The main contributions of this Frontiers Research Topic are as follows:

Nowadays, ultrasonography has been endorsed as the first-line imaging modality for diagnosing thyroid diseases by many guidelines (12). Two of the researches Zhang et al. as well as Liang et al. focused on the application of computer-aided diagnosis (CAD) systems to significantly improve the diagnostic accuracy of thyroid nodules, which reduced the need for unnecessary fine needle punctures. It is worth noting that Hou et al. constructed a deep learning-based ultrasonomic model and achieved an AUC of 0.924 for the diagnosis of benign and malignant nodules in patients with Hashimoto's thyroiditis. Of particular interest is that the study's dedicated to develop a radiomic nomogram model to non-invasively and reliably predict extrathyroidal extension in patients with papillary thyroid cancer (Wang et al.). These researches all contribute to the response of clinical decision making for precision US medicine.

In recent years, US has been widely used as a significant supplementary modality for breast cancer screening. With the application of AI-based ultrasomics, the diagnostic efficacy of breast disease evaluation has been greatly improved. Xiao et al. studied and demonstrated that deep learning-based computer-aided diagnostic system could significantly outperform the experienced radiologists in terms of diagnostic specificity, sensitivity, and accuracy for patients with asymptomatic breast disease. Benefit from this, unnecessary biopsies in asymptomatic screening patients can be avoided, reducing the waste of medical resources. Another meaningful study designed an ultrasomics feature-based nomogram to predict axillary lymph node status for breast cancer patients with relative accuracy, which helps in clinical decision making (Luo et al.).

It is vital for us to accurately identify gallbladder polyps preoperatively, of which cholesterol polyps and adenomatous polyps are the two most common types, while only adenomatous polyps are true polyps that potentially tend to develop into gallbladder cancer. Yuan et al's ultrasomics analysis of spatial and morphological features extracted from raw US images can effectively improve the preoperative diagnosis of true and false gallbladder polyps and provide a reliable basis for clinical decisions related to gallbladder polyp surgery.

Considering that different subtypes of liver cancer determine different treatment modalities and the poor prognosis of hepatocellular-cholangiocarcinoma subtypes, precise identification of liver disease preoperatively is critical. Peng et al. developed a machine learning-based moderate radiomics model to extract high-throughput US image features and achieve superior differentiation of histopathological subtypes of primary liver cancer. In addition, Li et al. conduced a machine learning-based ultrasomics method for analyzing and processing US images, which achieved preoperative individualized diagnostic performance comparable to that of radiologists for patients with atypical hepatocellular carcinoma (aHCC) and focal nodular hyperplasia (FNH). Based on this, the combination of ultrasomics (Nishida and Kudo) approach with AI is gradually becoming an effective tool for the diagnosis and analysis of liver tumors demonstrating the great potential of AI-based ultrasomics applications for precision medicine (13).

It is well known that image feature extraction is the most vital step of data analysis for ultrasomics. To address this critical issue, Jin et al. investigated automatic segmentation algorithms based on the multiple U-net model for ultrasomics features extraction of ovarian cancer patients, which showed good operational reproducibility and reliability. Another research is worthy of our attention is a U-net based on gastric wall detection network developed by Sui et al. By automatically analyzing the hierarchical structure of the gastric wall in gastric ultrasomics, which enable to accurately identify gastric diseases. The data quality and privacy involved in the application of AI in clinical practice and the transparency of algorithms are two important issues that need to be addressed urgently. With the attention and the resolution of these troubles, in the future, AI-based ultrasomics will be considered as a part of routine US examinations that will benefit both patients and physicians. AI technology, encouraged by the demand for ultrasomics application from computer technologies and clinical practice, AI-based ultrasomics will undoubtedly have a broad future. The wide application of AI-based ultrasomics help us to acquire quantify diagnostic information of diseases as well as improve the accuracy and reproducibility of US diagnosis.

We hope that this Frontiers Research Topic will be an enrichment for US medicine, we give our acknowledgement to all authors for their efforts and commitments, as well as the reviewers who have corrected each of the inadequate contributions.

AUTHOR CONTRIBUTIONS

Y-TS drafted the Editorial. All authors conceived, designed and supervised this work and gave final approval to the published version.

REFERENCES

1. Chen JH, Asch SM. Machine Learning and Prediction in Medicine - Beyond the Peak of Inflated Expectations. *N Engl J Med* (2017) 376:2507–9. doi: 10.1056/NEJMp1702071
2. Yu KH, Beam AL, Kohane IS. Artificial Intelligence in Healthcare. *Nat BioMed Eng* (2018) 2:719–31. doi: 10.1038/s41551-018-0305-z
3. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More Than Pictures, They Are Data. *Radiology* (2016) 278:563–77. doi: 10.1148/radiol.2015151169
4. Li W, Huang Y, Zhuang BW, Liu GJ, Hu HT, Li X, et al. Multiparametric Ultrasomics of Significant Liver Fibrosis: A Machine Learning-Based Analysis. *Eur Radiol* (2019) 29:1496–506. doi: 10.1007/s00330-018-5680-z
5. Shen YT, Chen L, Yue WW, Xu HX. Artificial Intelligence in Ultrasound. *Eur J Radiol* (2021) 139:109717. doi: 10.1016/j.ejrad.2021.109717
6. Zhou BY, Wang LF, Yin HH, Wu TF, Ren TT, Peng C, et al. Decoding the Molecular Subtypes of Breast Cancer Seen on Multimodal Ultrasound Images Using an Assembled Convolutional Neural Network Model: A Prospective and Multicentre Study. *EbioMedicine* (2021) 74:103684. doi: 10.1016/j.ebiom.2021.103684

7. Sun YK, Li XL, Wang Q, Zhou BY, Zhu AQ, Qin C, et al. Improving the Quality of Breast Ultrasound Examination Performed by Inexperienced Ultrasound Doctors With Synchronous Tele-Ultrasound: A Prospective, Parallel Controlled Trial. *Ultrasonography* (2021). doi: 10.14366/usg.21081
8. Zhao CK, Ren TT, Yin YF, Shi H, Wang HX, Zhou BY, et al. A Comparative Analysis of Two Machine Learning-Based Diagnostic Patterns With Thyroid Imaging Reporting and Data System for Thyroid Nodules: Diagnostic Performance and Unnecessary Biopsy Rate. *Thyroid* (2021) 31:470–81. doi: 10.1089/thy.2020.0305
9. Shen YT, Chen L, Yue WW, Xu HX. Digital Technology-Based Telemedicine for the COVID-19 Pandemic. *Front Med (Lausanne)* (2021) 8:646506. doi: 10.3389/fmed.2021.646506
10. Li XL, Guo LH, Sun LP, Yue WW, Xu HX. Teleultrasound for the COVID-19 Pandemic: A Statement From China. *Adv Ultrasound Diagn Ther* (2020) 4:27–111. doi: 10.37015/AUDT.2020.200036
11. Lv F, Wang J, Yu X, Yang A, Liu J-B, Qian L, et al. Chinese Expert Consensus on Critical Care Ultrasound Applications at COVID-19 Pandemic. *Adv Ultrasound Diagn Ther* (2020) 2:27–42. doi: 10.37015/AUDT.2020.200022
12. Sawka AM, Gagliardi AR, Haymart MR, Sturgeon C, Bernet V, Hoff K, et al. A Survey of American Thyroid Association Members Regarding the 2015 Adult Thyroid Nodule and Differentiated Thyroid Cancer Clinical Practice Guidelines. *Thyroid* (2020) 30:25–33. doi: 10.1089/thy.2019.0486
13. Zhang H, Guo L, Wang D, Wang J, Bao L, Ying S, et al. Multi-Source Transfer Learning *Via* Multi-Kernel Support Vector Machine Plus for B-Mode Ultrasound-Based Computer-Aided Diagnosis of Liver Cancers. *IEEE J BioMed Health Inform* (2021) 25:3874–85. doi: 10.1109/JBHI.2021.3073812

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 07 July 2020
doi: 10.3389/fonc.2020.01070

Diagnostic Value of Breast Lesions Between Deep Learning-Based Computer-Aided Diagnosis System and Experienced Radiologists: Comparison the Performance Between Symptomatic and Asymptomatic Patients

*Mengsu Xiao, Chenyang Zhao, Jianchu Li, Jing Zhang, He Liu, Ming Wang, Yunshu Ouyang, Yixiu Zhang, Yuxin Jiang and Qingli Zhu**

Department of Ultrasound, Chinese Academy of Medical Sciences and Peking Union Medical College Hospital, Beijing, China

OPEN ACCESS

Edited by:
Hui-Xiong Xu,
Tongji University, China

Reviewed by:
Yuming Jiang,
Stanford University, United States
JIanhua Zhou,
Sun Yat-sen University Cancer Center (SYSUCC), China

***Correspondence:**
Qingli Zhu
zqlpumch@126.com

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions,
a section of the journal Frontiers in Oncology

Received: *03 February 2020*
Accepted: *28 May 2020*
Published: *07 July 2020*

Citation:
Xiao M, Zhao C, Li J, Zhang J, Liu H, Wang M, Ouyang Y, Zhang Y, Jiang Y and Zhu Q (2020) Diagnostic Value of Breast Lesions Between Deep Learning-Based Computer-Aided Diagnosis System and Experienced Radiologists: Comparison the Performance Between Symptomatic and Asymptomatic Patients.
Front. Oncol. 10:1070.
doi: 10.3389/fonc.2020.01070

Purpose: The purpose of this study was to compare the diagnostic performance of breast lesions between deep learning-based computer-aided diagnosis (deep learning-based CAD) system and experienced radiologists and to compare the performance between symptomatic and asymptomatic patients.

Methods: From January to December 2018, a total of 451 breast lesions in 389 consecutive patients were examined (mean age 46.86 ± 13.03 years, range 19–84 years) by both ultrasound and deep learning-based CAD system, all of which were biopsied, and the pathological results were obtained. The lesions were diagnosed by two experienced radiologists according to the fifth edition Breast Imaging Reporting and Data System (BI-RADS). The final deep learning-based CAD assessments were dichotomized as possibly benign or possibly malignant. The diagnostic performances of the radiologists and deep learning-based CAD were calculated and compared for asymptomatic patients and symptomatic patients.

Results: There were 206 asymptomatic screening patients with 235 lesions (mean age 45.06 ± 10.90 years, range 21–73 years) and 183 symptomatic patients with 216 lesions (mean age 50.03 ± 14.97 years, range 19–84 years). The sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy and area under the receiver operating characteristic curve (AUC) of the deep learning-based CAD in asymptomatic patients were 93.8, 83.9, 75.0, 96.3, 87.2, and 0.89%, respectively. In asymptomatic patients, the specificity (83.9 vs. 66.5%, $p < 0.001$), PPV (75.0 vs. 59.4%, $p = 0.013$), accuracy (87.2 vs. 76.2%, $p = 0.002$) and AUC (0.89 to 0.81, $p = 0.0013$) of CAD were all significantly higher than those of the experienced radiologists. The sensitivity (93.8 vs. 80.0%), specificity (83.9 vs. 61.8%,), accuracy (87.2 vs. 73.6%) and AUC (0.89 vs. 0.71) of CAD were all higher for asymptomatic patients than for symptomatic patients. If the BI-RADS 4a lesions diagnosed by the radiologists in asymptomatic patients were downgraded to BI-RADS 3 according to the CAD, then 54.8% (23/42) of the lesions would avoid biopsy without missing the malignancy.

Conclusion: The deep learning-based CAD system had better performance in asymptomatic patients than in symptomatic patients and could be a promising complementary tool to ultrasound for increasing diagnostic specificity and avoiding unnecessary biopsies in asymptomatic screening patients.

Keywords: computer-aided diagnosis, deep learning, breast, ultrasound, symptomatic

INTRODUCTION

Breast cancer is a leading cause of cancer-related mortality in women worldwide (1). As an important supplementary modality for mammography, ultrasound plays an important role in dense breast tissue. Ultrasound is more suitable for Asian women, most of whom have thinner and denser breast glands and a younger age of onset for breast cancer, than Western women. A multicenter randomized trial across China compared ultrasound and mammography for breast cancer screening in high-risk Chinese women and showed that ultrasound had a significantly higher sensitivity and accuracy than mammography (2). Currently, ultrasound is widely used as the primary screening modality for breast cancer in China (3). However, ultrasounds often lead to a certain number of false-positive lesions and unnecessary biopsies or surgeries because ultrasound has low specificity and positive predictive value (PPV) (4–6). This has become an urgent problem of ultrasound in breast cancer screening in China.

In recent years, a deep learning-based computer-aided diagnosis (CAD) system for breast ultrasound (S-Detect^{TM} for Breast in RS80A; Samsung Medison Co., Ltd., Seoul, Korea) has become commercially available (7). This system has good performance in diagnosing benign and malignant breast lesions and especially in improving the specificity of ultrasound (8). Our early study showed that the deep learning-based CAD had the same diagnostic accuracy as experienced radiologists, and the specificity of the CAD was higher than that of the radiologists, which helped to reduce the number of unnecessary biopsies (9). Our recent study also showed that the deep learning-based CAD had a better performance in the breast benign lesions than the radiologists, especially in fibroadenomas and adenosis (10).

Radiologists often consider clinical factors (such as age, high-risk factors, clinical symptoms, and surgical history) as well as the images to make comprehensive judgments; in contrast, the CAD only considers ultrasound images without any clinical factors. Thus, we believe that the deep learning-based CAD is better at diagnosing asymptomatic patients than symptomatic patients since it only analyzes imaging data. Currently, the major mode of achieving early detection for breast cancer in China is hospital-based opportunistic screening among asymptomatic self-referred women (3), so we proposed CAD may be more helpful in breast cancer asymptomatic screening. To the best of our knowledge, no reports have been published on this topic yet. This study prospectively analyzed the value of deep learning-based CAD in asymptomatic screening patients by comparing with symptomatic patients.

MATERIALS AND METHODS

Patients

From January to December 2018, a total of 409 consecutive patients were examined at the Peking Union Medical College Hospital. All lesions underwent biopsy, and the pathologies were obtained. This prospective study was approved by the institutional review board. Informed consent was obtained from all patients included in the study.

Inclusion criteria were listed as follows:

(1) Had breast lesions clearly visualized by ultrasound;
(2) Underwent biopsy of the lesions and had pathological results;
(3) Provided informed consent.

Exclusion criteria were listed as follows:

(1) Patients who were pregnant or lactating;
(2) Patients who had breast biopsy or were undergoing neoadjuvant chemotherapy or radiotherapy.

Among these patients, 8 women whose lesions can't be visualized by ultrasound, 5 women who were pregnant or lactating and 7 women who had breast biopsy or were undergoing neoadjuvant chemotherapy were excluded. Ultimately, a total of 451 breast lesions in 389 patients were included in this study. The patients were divided into symptomatic and asymptomatic groups. Patients with any clinical manifestations of the breast are classified as symptomatic group, including palpable breast masses, localized pain, nipple discharge, trauma, redness and swelling of the breast, skin changes, nipple retraction, and nipple eczematoid changes. The patients in the asymptomatic group had no symptoms in their breasts and had undergone ultrasound for breast cancer screening. **Figure 1** shows the flow chart of study.

Ultrasound Examination

The ultrasound examinations were performed using a 3–12 MHz linear transducer (RS80A with Prestige, Samsung Medison, Co. Ltd., Seoul, Korea). Two radiologists (QL Zhu and MS Xiao) with 17 and 12 years of experience in breast imaging bilaterally examined the whole breasts of all patients by using ultrasound.

Abbreviations: CAD, deep learning framework-based computer-aided diagnosis; SE, sensitivity; SP, specificity; PLR, positive likelihood ratio; NLR, negative likelihood ratio; 95% CI, 95% confidence interval.

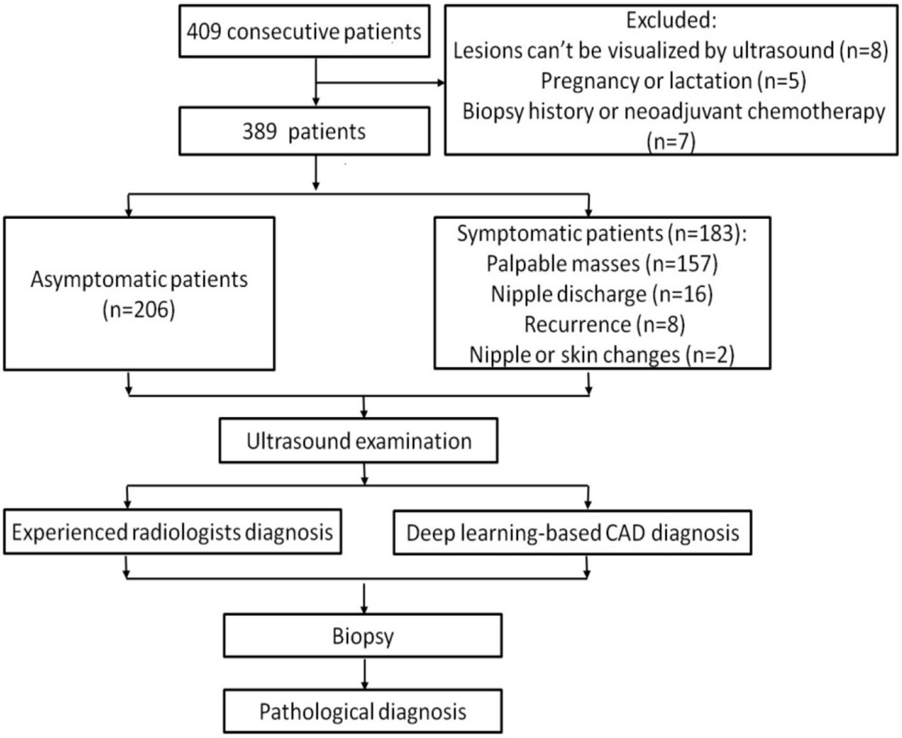


FIGURE 1 | The flow diagram of the study.

The radiologists were aware of the clinical information (history, symptoms, etc.), mammographic results, magnetic resonance imaging (MRI) results, and previous ultrasound results before performing the ultrasound examination. When a breast lesion was detected, two images of the longitudinal and transverse sections of the largest lesion diameter were routinely obtained, and still images were recorded. The lesions were diagnosed by the experienced radiologists based on fifth edition Breast Imaging Reporting and Data System (BI-RADS) by the American College of Radiology (11). The radiologists were blinded to the CAD results when they made the diagnosis for breast lesions. The final diagnosis was classified as follows: category 3, probably benign; category 4a, low suspicion for malignancy; category 4b, intermediate suspicion for malignancy; category 4c, moderate concern for malignancy; and category 5, highly suggesting malignancy. The radiologists were blinded to the pathologic results. The diagnostic cutoff was category 4a. Category 3 lesions were considered benign, while category 4a, 4b, 4c, and 5 lesions were considered malignant.

Deep Learning-Based CAD Examination

The CAD examination was performed by using deep learning-based CAD software (Samsung Healthcare, South Korea) by the same two radiologists who performed ultrasound examination. The CAD system utilizes large data sets collected from numerous breast exam cases and provides the characteristics of displayed lesion. The CAD applies a novel feature extraction technique and support vector machine classifier. By adopting a deep learning algorithm in the processes of lesion segmentation, analysis of characteristics and assessment, the CAD gives a dichotomized diagnosis whether a selected lesion is benign or malignant according to the proposed feature combinations integrated according to the BI-RADS.

On the maximum diameter section of the lesion, the radiologists started the CAD in the center of the lesion. If the maximum diameter of tumor was larger than the machine screen, we selected the most representative section (showing the most suspicious features) of the lesion for CAD to analyze. A region of interest (ROI) was automatically drawn along the border of the lesion. If the automatic outline of ROI was not considered accurate, the radiologists could manually modify the tumor boundary. Based on the given ROI, all of the data and information about the lesion were extracted and analyzed. The CAD system comprehensively analyzed the extracted information, provided a BI-RADS lexicon of the lesions including shape, orientation, margins, pattern and posterior acoustic features, and made a dichotomized diagnosis (possibly benign and possibly malignant) (**Figures 2–4**). The entire deep learning-based CAD process took only a few seconds.

Pathological Diagnosis

All of the breast lesions in our study underwent biopsy, and histopathological results were considered the gold standard,

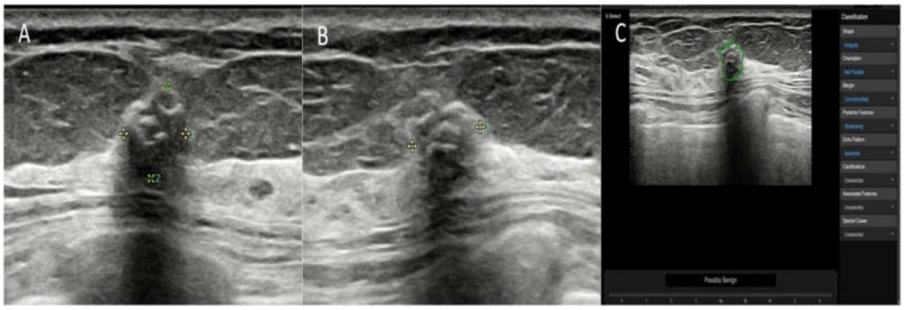


FIGURE 2 | The breast mass of an asymptomatic 57-year-old woman. **(A,B)** The longitudinal section and cross-section of the lesion showed a 6-mm mass with calcifications and posterior shadowing. The orientation is not parallel. The experienced radiologists diagnosed the lesion as BI-RADS 4a. **(C)** A ROI was automatically drawn along the margin of the mass (green line). The raw imaging data were automatically analyzed, and the final diagnosis of the deep learning-based CAD system was a possibly benign tumor. The mass was pathologically proven to be a fibroadenoma.

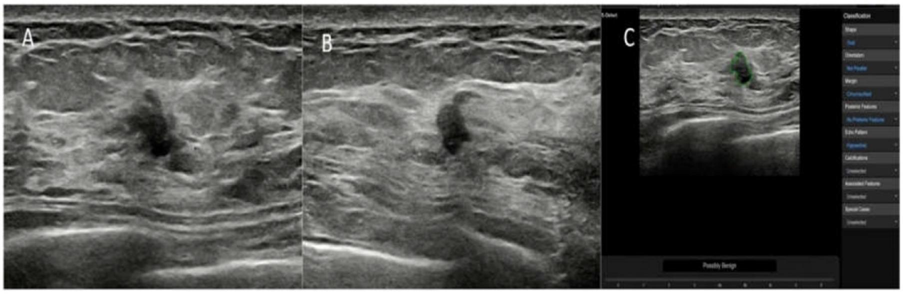


FIGURE 3 | The breast lesion of an asymptomatic 46-year-old woman. **(A,B)** The longitudinal section and cross-section of the lesion showed a 7-mm lesion with irregular shape and ill-defined margins. The diagnosis of the experienced radiologists was a BI-RADS 4b lesion. **(C)** The diagnosis of the deep learning-based CAD system was a possibly benign tumor. The pathological result was fat necrosis.

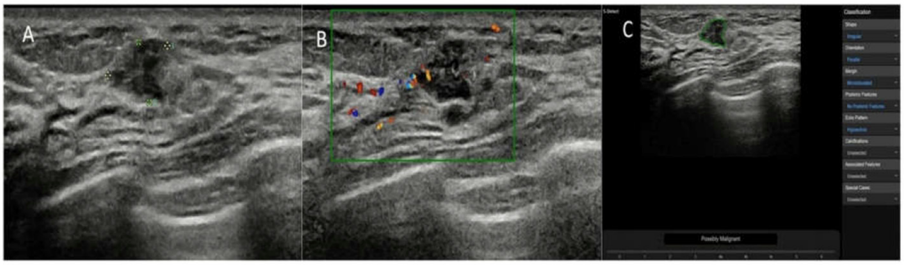


FIGURE 4 | The breast lesion of an asymptomatic 50-year-old woman. **(A)** The longitudinal section of a 9-mm lesion. **(B)** Color Doppler flow imaging of the lesion. The diagnosis of the experienced radiologists was a BI-RADS 3 lesion. **(C)** The diagnosis of the deep learning-based CAD system was a possibly malignant tumor. The pathological result was invasive ductal carcinoma.

including all of the category 3 lesions. The category 3 lesions underwent biopsy according to the patients' choices or patients with high risk factors, including family history and nipple discharge. Immunohistochemical examinations were performed when needed.

Statistical Analysis

Statistical analysis was performed using SPSS 21.0 (SAS Inc., Cary, NC, USA). The diagnostic performances of the physician and deep learning-based CAD system were analyzed and compared in terms of the sensitivity, specificity, positive likelihood ratio (PLR), negative likelihood ratio (NLR), PPV, negative predictive value (NPV) and accuracy. The 2 × 2 contingency table, chi-square test and McNemar test were used to compare the differences in performance. Receiver operating characteristic (ROC) curves were drawn, and the areas under the ROC curves (AUCs) were calculated. A $p < 0.05$ was regarded as statistically significant.

TABLE 1 | The pathological types of the 220 malignant lesions.

Malignant lesions	Number of asymptomatic lesions (ratio%)	Number of symptomatic lesions (ratio%)	Total number (ratio%)
Intraductal carcinoma	13 (16.25)	17 (12.14)	30 (13.64)
Invasive ductal carcinoma, not otherwise specified	56 (70)	98 (70)	154 (70)
Invasive lobular carcinoma	5 (6.25)	5 (3.57)	10 (4.55)
Apocrine carcinoma	1 (1.25)	1 (0.71)	2 (0.91)
Papillary carcinoma	0 (0)	7 (5)	7 (3.18)
Mucinous carcinoma	1 (1.25)	4 (2.86)	5 (2.23)
Neuroendocrine carcinoma	2 (2.5)	0 (0)	2 (0.91)
Malignant phyllodestumours	0 (0)	5 (3.56)	5 (2.23)
Metaplastic carcinoma	0 (0)	1 (0.71)	1 (0.45)
Medullary carcinoma	1 (1.25)	0 (0)	1 (0.45)
Tubular carcinoma	1 (1.25)	0 (0)	1 (0.45)
myofibroblastoma	0 (0)	1 (0.71)	1 (0.45)
Diffuse large B-cell lymphoma	0 (0)	1 (0.71)	1 (0.45)
Total	80	140	220

TABLE 2 | The pathological types of the 231 benign lesions.

Benign lesions	Number of asymptomatic lesions (ratio%)	Number of symptomatic lesions (ratio%)	Total number (ratio%)
Fibroadenoma	92 (59.35)	34 (44.74)	126 (54.55)
Adenosis	39 (25.16)	15 (19.74)	54 (23.38)
Intraductal papilloma	13 (8.39)	16 (21.05)	29 (12.55)
Phyllodestumour	1 (0.65)	2 (2.63)	3 (1.30)
Chronic inflammation	7 (4.52)	4 (5.26)	11 (4.76)
Granular inflammation	1 (0.65)	3 (3.95)	4 (1.73)
Hamartoma	0 (0)	1 (1.32)	1 (0.43)
Epidermoid cyst	0 (0)	1 (1.32)	1 (0.43)
Cyst	1 (0.65)	0 (0)	1 (0.43)
Fat necrosis	1 (0.65)	0 (0)	1 (0.43)
Total	155	76	231

RESULTS

In total, 206 asymptomatic screening patients had 235 lesions; the mean age was 45.06 ± 10.90 years (range 21–73 years), and the mean lesion size was 1.44 ± 0.60 cm (range 0.4–4.9 cm). A total of 183 symptomatic patients had 216 lesions, including 16 patients with nipple discharge, 157 with palpable masses, 8 with recurrence after malignant tumor resection, and 2 with nipple depressions or skin changes. The mean age was 50.03 ± 14.97 years (range 19–84 years), and the mean lesion size was 2.42 ± 1.26 cm (range 0.3–9.2 cm). The symptomatic patients were significantly older than the asymptomatic patients ($p < 0.001$). The lesions in the symptomatic patients were significantly larger than those in the asymptomatic patients ($p < 0.001$). The pathological results of the lesions are listed in **Tables 1**, **2**.

The diagnostic performances of the deep learning-based CAD system and radiologists (asymptomatic patients and symptomatic patients) are shown in **Table 3**. The diagnostic performances of the deep learning-based CAD system and radiologists for lesions <1 cm (asymptomatic patients and symptomatic patients) are shown in **Table 4**. The false-positive and false-negative results of the deep learning-based CAD system are shown in **Tables 5**, **6**. The subcategorization of asymptomatic and symptomatic breast lesions by the experienced radiologists is shown in **Table 7**. The ROC curves are shown in **Figures 5**, **6**.

Comparing the Performances of the Deep Learning-Based CAD System and Radiologists

For Asymptomatic Patients

The specificity (83.87 vs. 66.45%, $p < 0.001$), PPV (75.00 vs. 59.38%, $p = 0.013$) and accuracy (87.23 vs. 76.17%, $p = 0.002$), and AUC (0.89 to 0.81, $p = 0.0013$) of the CAD were all significantly higher than those of the radiologists. The sensitivity and NPV were not significantly different between the CAD and the radiologists ($p > 0.05$).

For Symptomatic Patients

The sensitivity (97.14 vs. 80.00%, $p < 0.001$), NPV (92.00 vs. 62.67%, $p < 0.001$) and accuracy (84.26 vs. 73.61%, $p = 0.002$), and AUC (0.79 to 0.71, $p = 0.040$) of the radiologists were all significantly higher than those of the CAD. The specificity and PPV were not significantly different between the radiologists and the CAD system ($p > 0.05$).

Comparing the Performances of the Deep Learning-Based CAD System for Asymptomatic Patients and for Symptomatic Patients

The sensitivity (93.75 vs. 80.00%), specificity (83.87 vs. 61.84%), and accuracy (87.23 vs. 73.61%), and AUC (0.89 vs. 0.71) of the CAD were higher for asymptomatic patients than for symptomatic patients.

For the Asymptomatic Screening Patients With Lesions <1 cm

In this study, a total of 87 lesions were <1 cm, of which, 61 were in asymptomatic patients. In the asymptomatic patients with lesions <1 cm, both the specificity (88.64 vs. 65.91%, $p = 0.002$) and accuracy (91.80 vs. 75.41%, $p = 0.014$) of the system were significantly higher than those of the experienced radiologists.

For the BI-RADS 4a Lesions of the Asymptomatic Patients

For the asymptomatic patients, 42 lesions were diagnosed as BI-RADS 4a by the radiologists. The pathologic results showed

TABLE 3 | The diagnostic performances of the deep learning-based CAD system and experienced radiologists for asymptomatic lesions and symptomatic lesions.

		SE(%) (95%CI)	SP(%) (95%CI)	PLR (95%CI)	NLR (95%CI)	PPV(%) (95%CI)	NPV(%) (95%CI)	Accuracy (%) (95%CI)	AUC (95%CI)
Asymptomatic lesions	CAD	93.75 (86.01–97.94)	83.87 (77.12–89.28)	5.81 (4.04–8.36)	0.07 (0.03–0.17)	75.00 (65.34–83.12)	96.30 (91.57–98.79)	87.23 (82.28–91.22)	0.89 (0.84–0.93)
	radiologists	95.00 (87.69–98.62)	66.45 (58.43–73.83)	2.83 (2.26–3.55)	0.08 (0.03–0.20)	59.38 (50.34–67.96)	96.26 (90.70–98.97)	76.17 (70.20–81.47)	0.81 (0.75–0.86)
Symptomatic lesions	CAD	80.00 (72.41–86.28)	61.84 (49.98–72.75)	2.10 (1.56–2.82)	0.32 (0.22–0.47)	79.43 (71.82–85.77)	62.67 (50.73–73.57)	73.61 (67.20–79.36)	0.71 (0.64–0.77)
	radiologists	97.14 (92.85–99.22)	60.53 (48.65–71.56)	2.46 (1.86–3.26)	0.05 (0.02–0.13)	81.93 (75.22–87.46)	92.00 (80.77–97.78)	84.26 (78.70–88.85)	0.79 (0.73–0.84)

SE, sensitivity; SP, specificity; PLR, positive likelihood ratio; NLR, negative likelihood ratio; PPV, positive predictive value; NPV, negative predictive value; AUC, area under the receiver operator characteristics curve; 95% CI, 95% confidence interval.

TABLE 4 | The diagnostic performances of the deep learning-based CAD system and experienced radiologists for lesions <1 cm.

		SE(%) (95%CI)	SP(%) (95%CI)	PLR (95%CI)	NLR (95%CI)	PPV(%) (95%CI)	NPV(%) (95%CI)	Accuracy (%) (95%CI)	AUC (95%CI)
Asymptomatic lesions	CAD	100.00 (80.49–100.00)	88.64 (75.44–96.21)	8.80 (3.86–20.09)	0.00	77.27 (54.63–92.18)	100.00 (90.97–100.00)	91.80 (81.90–97.28)	0.94 (0.85–0.99)
	Radiologists	100.00 (80.49–100.00)	65.91 (50.08–79.51)	2.93 (1.95–4.42)	0.00	53.13 (34.74–70.91)	100.00 (88.06–100.00)	75.41 (62.71–85.54)	0.83 (0.71–0.91)
Symptomatic lesions	CAD	60.00 (32.29–83.66)	72.73 (39.03–93.98)	2.20 (0.77–6.29)	0.55 (0.27–1.13)	75.00 (42.81–94.51)	57.14 (28.86–82.34)	65.38 (44.33–82.79)	0.66 (0.45–0.84)
	Radiologists	86.67 (59.54–98.34)	63.64 (30.79–89.07)	2.38 (1.06–5.34)	0.21 (0.05–0.82)	76.47 (50.10–93.19)	77.78 (39.99–97.19)	76.92 (56.35–91.03)	0.75 (0.54–0.90)

SE, sensitivity; SP, specificity; PLR, positive likelihood ratio; NLR, negative likelihood ratio; PPV, positive predictive value; NPV, negative predictive value; AUC, area under the receiver operator characteristics curve; 95% CI, 95% confidence interval.

TABLE 5 | False positive cases of deep learning-based CAD system.

False positive cases	Number of asymptomatic lesions (ratio%)	Number of symptomatic lesions (ratio%)	Total number (ratio%)
Fibroadenoma	6 (24)	8 (27.59)	14 (25.93)
Adenosis	9 (36)	7 (24.14)	16 (29.63)
Intraductal papilloma	6 (24)	8 (27.59)	14 (25.93)
Benign phyllodestumour	0 (0)	2 (6.90)	2 (3.70)
inflammation	3 (12)	4 (13.79)	7 (12.96)
Cyst	1 (4)	0 (0)	1 (1.85)
Total	25	29	54

TABLE 6 | False negative cases of deep learning-based CAD system.

False negative cases	Number of asymptomatic lesions (ratio%)	Number of symptomatic lesions (ratio%)	Total Number (ratio%)
Intraductal carcinoma	1 (20)	6 (21.42)	7 (21.21)
Invasive ductal carcinoma, not otherwise specified	2 (40)	10 (35.71)	12 (36.36)
Invasive lobular carcinoma	1 (20)	0 (0)	1 (3.03)
Papillary carcinoma	0 (0)	5 (17.86)	5 (15.15)
Mucinous carcinoma	1 (20)	3 (10.71)	4 (12.12)
Malignant phyllodestumours	0 (0)	4 (14.29)	4 (12.12)
Total	5	28	33

that 40 lesions (95.23%) were benign and 2 lesions (4.76%) were malignant. If the BI-RADS 4a lesions diagnosed by the radiologists in asymptomatic patients were downgraded to BI-RADS 3 according to the CAD system results, then 54.8% (23/42) of the lesions would avoid biopsy without missing the 2 malignant tumors.

DISCUSSION

As an important supplementary modality for mammography, ultrasound has the advantages of avoiding radiation and being simple and easy to use. Performing bilateral whole breast screening for Asian women with small breasts is

TABLE 7 | The subcategorization of asymptomatic and symptomatic breast lesions by the experienced radiologists.

	Radiologists diagnosis	Pathological result	
		Benign	**Malignant**
Asymptomatic lesions	BI-RADS 3	103	4
	BI-RADS 4a	40	2
	BI-RADS 4b	9	9
	BI-RADS 4c	3	27
	BI-RADS 5	0	38
Symptomatic lesions	BI-RADS 3	46	4
	BI-RADS 4a	21	8
	BI-RADS 4b	6	22
	BI-RADS 4c	3	39
	BI-RADS 5	0	67

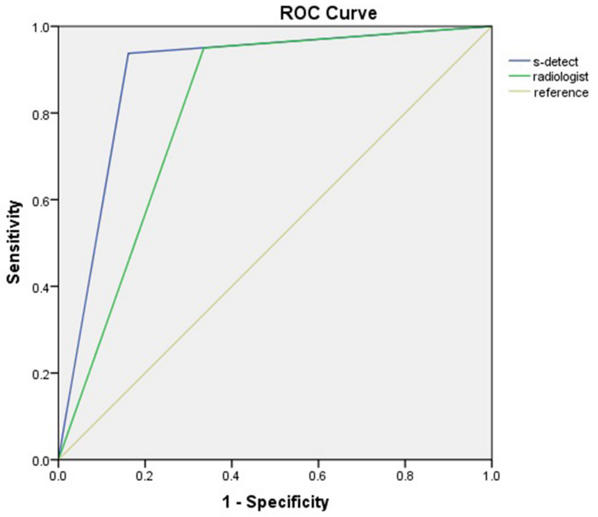


FIGURE 5 | ROC curves of asymptomatic patients.

easy, and the breast lesions can be observed in detail (12). However, ultrasound requires extensive experience since this modality is an operator-dependent examination with lower reproducibility, specificity and PPV than mammography (13). In recent years, CAD has been used to overcome this shortcoming and to increase diagnostic accuracy (14, 15), similar to elastography, which has been used as an adjunct tool to decrease the number of unnecessary biopsies while improving the specificity of ultrasound without losing sensitivity (16). Shibusawa et al. reported that CAD could significantly increase the AUC of the observers from 0.649 to 0.783 ($p = 0.0167$) (12). A recent study showed that adding CAD results to ultrasound significantly improved the specificity, accuracy, and PPV of radiologists without losing sensitivity and NPV (17).

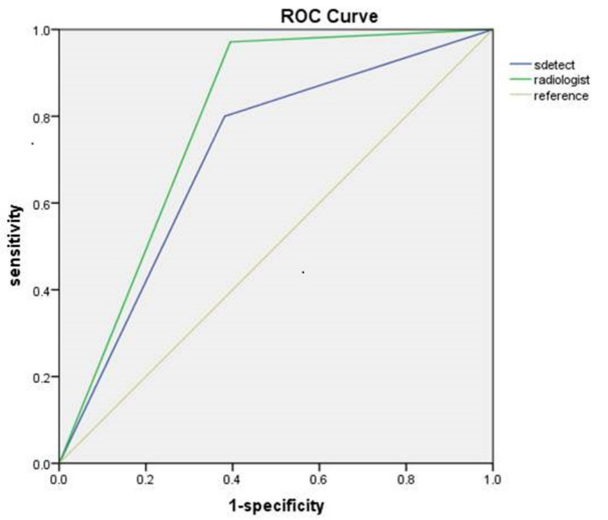


FIGURE 6 | ROC curves of symptomatic patients.

The Role of Deep Learning-Based CAD System in the Breast Lesions

The emergence of deep learning methods has profoundly influenced the medical field. Currently, deep learning techniques are considered the most advanced technology for image classification (18). Deep learning-based CAD systems are different from conventional CAD systems based on manual feature design. Deep learning-based CAD is superior to conventional CAD (19). The deep learning-based CAD system used in our study (Samsung corporation, Seoul Korea) is a newly developed CAD system for breast ultrasound based on deep learning of raw ultrasound signals through a convolutional neural network. After extensive learning and training on a large number of databases, the deep learning-based CAD system could extract high-order statistics and optimize the balance of input and output data through multiple hidden layers to provide an accurate diagnosis (9). The original unprocessed ultrasound signals were collected as the raw data and information for the deep learning-based CAD system to analyze through a complex hierarchical framework. Therefore, the deep learning-based CAD system did not have interference from artifacts or man-made interference, which leads to more realistic and reliable diagnoses. The analysis process of the deep learning-based CAD system is different from how by radiologists makes observations with their naked eyes, and more inherent information can be obtained by the CAD system. The analyses and descriptions of deep learning-based CAD include shape, echo and texture features using spatial gray-level dependence matrices, intensity in the tumor area, gradient magnitude in the tumor area, orientation, distance between the tumor shape and a best-fit ellipse, average gray value changes or histogram changes between the tissue and tumor area, comparison of the gray values of the tumor surroundings, the number of lobulation/protuberances/depressions, and the lobulation index (20). Moreover, deep learning-based CAD is economical, easy-to-operate, and capable of providing a rapid diagnosis; thus, this method can be easily incorporated in clinical practice (8). Segni et al. (21) reported that deep learning-based CAD had good performance. The sensitivity, specificity, PPV, NPV and AUC were 91.1, 70.8, 85.4, 81.0, and 0.81%, respectively. The AUC was consistent with that found in our study (0.81).

Ultrasound screening has a low specificity and PPV (4–6). Previous studies have shown that deep learning-based CAD could improve the specificity of ultrasound. Kim et al. (22) reported that the specificity (65.8 vs. 30.9%), PPV (58.3 vs. 46.2%), accuracy (70.8 vs. 56.2%) and AUC (0.725 vs. 0.653) of the deep learning-based CAD system were all significantly higher than those of the experienced radiologists ($p < 0.05$) when using BI-RADS 4a as the cutoff value. This finding indicated that deep learning-based CAD had good clinical value. Cho et al. (8) also showed that the sensitivity, specificity, PPV, NPV, accuracy and AUC of deep learning-based CAD were 72.2, 90.8, 86.7, 79.7, 82.4, and 0.815%, respectively. The specificity, PPV, and accuracy of the deep learning-based CAD system were all significantly higher than those of 2 experienced radiologists ($p < 0.05$). Thus, deep learning-based CAD could increase the specificity, PPV, and accuracy of ultrasound. For the asymptomatic patients in our study, the sensitivity, specificity, PPV, NPV, accuracy and AUC of the deep learning-based CAD system were 93.8, 83.9, 75.0, 96.3, 87.2, and 0.89%, respectively. The specificity (83.9 vs. 66.5%, $p < 0.001$), PPV (75.0 vs. 59.4%, $p = 0.013$), accuracy (87.2 vs. 76.2%, $p = 0.002$) and AUC (0.89 vs. 0.81, $p = 0.0013$) of the deep learning-based CAD system were all significantly higher than those of the radiologists. In our study, in the asymptomatic patients, the PLR (5.81 vs. 2.83) and PPV (75.00 vs. 59.38) of CAD were higher than those of radiologists. This means that, in the asymptomatic patients, the probability of a malignant diagnosis of CAD to be a true malignant lesion is higher than that of radiologists. In the symptomatic patients, the NLR (0.05 vs. 0.32) was lower of radiologist than that of CAD and the NPV (92.00 vs. 62.67) of radiologists was higher than that of CAD. This means that, in the symptomatic patients, the probability of a benign diagnosis of radiologist to be a true benign lesion is higher than that of CAD.

For Asymptomatic Patients

To the best of our knowledge, this is the first study to report the performance of a deep learning-based CAD system in the comparison of asymptomatic and symptomatic patients with breast lesions. Our study showed that the CAD system was more effective for asymptomatic patients than for symptomatic patients. Compared with those for the symptomatic patients, the sensitivity (93.8 vs. 80.0%), specificity (83.9 vs. 61.8%), accuracy (87.2 vs. 73.6%) and AUC (0.89 vs. 0.71) of the asymptomatic patients were all increased. These results indicate that the CAD system had a better performance in patients without clinical symptoms and medical or family histories. The CAD system is better than the human naked eye at extracting and analyzing inherent patterns from raw information data. Therefore, in the asymptomatic screening breast lesions, the diagnostic performance of radiologists could be improved by using a deep learning-based CAD approach.

For Symptomatic Patients

To diagnose breast lesions, many clinical factors are taken into account in addition to the images, such as the patient's age, symptoms, surgical histories, family histories, high-risk factors, clinical examination results, and other imaging findings, including those from mammography, MRI, color Doppler ultrasound, and elastography. The diagnosis is a comprehensive analysis and judgment. In our study, there were 5 malignant phyllodes tumors, 4 of which were postoperative recurrence. All 4 solid tumors had regular shapes and clear boundaries on the images. The radiologists correctly diagnosed these lesions as recurrent malignant phyllodes tumors, while the CAD misdiagnosed these lesions as benign tumors. In this study, one patient who previously underwent modified radical mastectomy for breast cancer 4 years ago had recurrence on the chest wall. The recurrent tumor manifested as a solid nodule with a regular shape, clear boundary, and rich internal blood flow. The radiologists correctly diagnosed this mass as a recurrent cancer, while the CAD also misdiagnosed this mass as a benign tumor. There were 15 inflammatory lesions in the present study, of which 7 were misdiagnosed as malignant by the CAD. These 7 lesions had irregular shapes and ill-defined borders; these lesions tended to be misdiagnosed as breast cancer without any medical histories or clinical symptoms. These observations indicated that the clinical diagnostic process and CAD techniques were significantly different. The clinical diagnostic process strongly depends on the medical history and clinical manifestations. In contrast, the CAD system only analyses imaging features without considering any non-imaging factors. Thus, the CAD has a better performance in the asymptomatic screening breast lesions. Adding clinical information into the CAD diagnostic process may be helpful in the future.

For the Asymptomatic Screening Patients With Lesions <1 cm

Small cancer with an invasive component <1 cm is considered unlikely to metastasize, and more than 90% of small cancers do not have axillary lymph node metastases, regardless of the histological grade (23). Therefore, detecting small cancers at the early stage is very important for the screening program. With the tumor size decreases, the characteristics of the cancer are also likely to decrease, such as desmoplastic changes and surrounding tissue changes to invasion (24). Therefore, correctly diagnosing small cancers is a true challenge for radiologists. In our study, the screening asymptomatic lesions were significantly smaller than the symptomatic lesions (1.44 vs. 2.42 cm, $p < 0.05$), which reveals the significance of breast screening for detecting small and early-stage breast cancer. In total, 87 lesions were smaller than 1 cm in our study, of which 61 lesions were from asymptomatic patients. Both the specificity (88.64 vs. 65.91%, $p = 0.002$) and accuracy (91.80 vs. 75.41%, $p = 0.014$) of the CAD were significantly higher than those of the experienced radiologists. These results suggest that for small breast cancers, the deep learning-based CAD system is more capable at extracting hidden information contain in the raw imaging data and recognizing the features of small cancers, which are indistinguishable to the radiologist's naked eye. The miniscule signs of malignant small breast cancer may be more easily identified by a deep learning-based CAD system than the naked human eye. Therefore, the diagnostic performance of radiologists for small cancer could be improved by a deep learning-based CAD system.

For the BI-RADS 4a Lesions of the Asymptomatic Patients

BI-RADS 4a lesions are worrisome lesions, most of which are benign. Correct diagnoses of BI-RADS 4a lesions can reduce unnecessary biopsies and decrease the false-positive rate, which has always been the goal of radiologists. In the asymptomatic patients of this study, 95.23% (40 of 42) of the BI-RADS 4a lesions were benign. If the diagnosis process for BI-RADS 4a lesions also involved the CAD results, then 54.76% (23 of 42) of the benign lesions could avoid being unnecessarily biopsied without missing any malignant tumors. Thus, deep learning-based CAD is helpful in distinguishing benign from worrisome lesions. Choi et al. (17) also found that deep learning-based CAD could improve the diagnostic performance of leading radiologists and enable radiologists to correctly diagnose lesions that are difficult to classify as BI-RADS 3 or 4a.

There were several limitations in this study. First, the proportion of ductal carcinoma *in situ* in this study was slightly low (30/220), which may be because ultrasound is not well-suited for detecting ductal carcinoma *in situ*, whose main feature is microcalcification. The CAD did not perform well for detecting ductal carcinoma *in situ* (21/30). Therefore, the results of this study may overestimate the diagnostic efficacy of the CAD. Second, the image acquisition for the CAD is also operator dependent. In the present study, the representative images analyzed by CAD were selected by two experienced radiologists with more than 12 years experience in breast ultrasound. The representative image might be better in this study, and the diagnostic performance of the CAD needs further verification. Third, the number of cases is limited and the sample size needs to be expanded in future studies or multicenter studies.

In conclusion, a deep learning-based CAD system has the advantages of convenient operation and accurate diagnosis of breast lesions, especially in the asymptomatic screening patients. For asymptomatic patients, we could rely more on the CAD results in the future. For patients with medical histories or symptoms, we should make comprehensive judgments based on the clinical histories and symptoms. The deep learning-based CAD approach also has good diagnostic performance for small breast cancer (<1 cm). Therefore, a deep learning-based CAD system has good screening value for asymptomatic breast cancer at an early stage.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the ethics committee of Peking Union Medical College hospital. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

MX and CZ: drafting the manuscript and organizing the database. JL and YJ: revising the work critically for important intellectual content. JZ: acquisition of data for the work. HL: interpretation of data for the work. MW, YO, and YZ: analysis and interpretation of data for the work. MX and QZ: acquisition, analysis, and interpretation of data for the work. QZ: substantial contributions to the conception or design of the work. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by Beijing Natural Science Foundation (7192177), Fundamental Research Funds for the Central Universities (2017320002), and CAMS Innovation Fund for Medical Sciences (2017-I2M-1–006).

ACKNOWLEDGMENTS

We would like to thank all of the patients for their participation in this study.

REFERENCES

1. Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J, Rosso S, Coebergh JW, Comber H, et al. Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. *Eur J Cancer.* (2013) 49:1374–403. doi: 10.1016/j.ejca.2012.12.027
2. Shen S, Zhou Y, Xu Y, Zhang B, Duan X, Huang R, et al. A multi-centre randomised trial comparing ultrasound vs mammography for screening breast cancer in high-risk Chinese women. *Br J Cancer.* (2015) 112:998–1004. doi: 10.1038/bjc.2015.33
3. Pan B, Yao R, Zhu QL, Wang CJ, You SS, Zhang J, et al. Clinicopathological characteristics and long-term prognosis of screening detected non-palpable breast cancer by ultrasound in hospital-based Chinese population (2001–2014). *Oncotarget.* (2016) 7:76840–51. doi: 10.18632/oncotarget.12319
4. Corsetti V, Houssami N, Ferrari A, Ghirardi M, Bellarosa S, Angelini O, et al. Breast screening with ultrasound in women with mammography-negative dense breasts: evidence on incremental cancer detection and false positives, and associated cost. *Eur J Cancer.* (2008) 44:539–44. doi: 10.1016/j.ejca.2008.01.009
5. Nothacker M, Duda V, Hahn M, Warm M, Degenhardt F, Madjar H, et al. Early detection of breast cancer: benefits and risks of supplemental breast ultrasound in asymptomatic women with mammographically dense breast tissue. a systematic review. *BMC Cancer.* (2009) 9:335. doi: 10.1186/1471–2407-9–335
6. Sprague BL, Stout NK, Schechter C, van Ravesteyn NT, Cevik M, Alagoz O, et al. Benefits, harms, and cost-effectiveness of supplemental ultrasonography screening for women with dense breasts. *Ann Intern Med.* (2015) 162:157–66. doi: 10.7326/M14–0692

7. Han S, Kang HK, Jeong JY, Park MH, Kim W, Bang WC, et al. A deep learning framework for supporting the classification of breast lesions in ultrasound images. *Phys Med Biol.* (2017) 62:7714–28. doi: 10.1088/1361–6560/aa82ec
8. Cho E, Kim EK, Song MK, Yoon JH. Application of computer-aided diagnosis on breast ultrasonography: evaluation of diagnostic performances and agreement of radiologists according to different levels of experience. *J Ultrasound Med.* (2018) 37:209–16. doi: 10.1002/jum.14332
9. Zhao C, Xiao M, Jiang Y, Liu H, Wang M, Wang H, et al. Feasibility of computer-assisted diagnosis for breast ultrasound: the results of the diagnostic performance of S-detect from a single center in China. *Cancer Manag Res.* (2019) 11:921–30. doi: 10.2147/CMAR.S190966
10. Xiao M, Zhao C, Zhu Q, Zhang J, Liu H, Li J, et al. An investigation of the classification accuracy of a deep learning framework-based computer-aided diagnosis system in different pathological types of breast lesions. *J Thorac Dis.* (2019) 11:5023–31. doi: 10.21037/jtd.2019.12.10
11. D'Orsi CJ, Sickles EA, Mendelson EB, Morris EA. *ACR BI-RADS Atlas, Breast Imaging Reporting and Data System*. 5th ed. Reston, VA: American College of Radiology (2013).
12. Shibusawa M, Nakayama R, Okanami Y, Kashikura Y, Imai N, Nakamura T, et al. The usefulness of a computer-aided diagnosis scheme for improving the performance of clinicians to diagnose non-mass lesions on breast ultrasonographic images. *J Med Ultrason.* (2016) 43:387–94. doi: 10.1007/s10396–016-0718–9
13. Abdullah N, Mesurolle B, El-Khoury M, Kao E. Breast imaging reporting and data system lexicon for US: interobserver agreement for assessment of breast masses. *Radiology.* (2009) 252:665–72. doi: 10.1148/radiol.2523080670
14. Jalalian A, Mashohor SB, Mahmud HR, Saripan MI, Ramli AR, Karasfi B, et al. Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: a review. *Clin Imaging.* (2013) 37:420–6. doi: 10.1016/j.clinimag.2012.09.024
15. Huang YL, Lin SH, Chen DR. Computer-aided diagnosis applied to 3-D US of solid breast nodules by using principal component analysis and image retrieval. In: *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference.* Shanghai: IEEE (2005). p. 1802-5. doi: 10.1109/IEMBS.2005.1616798
16. Lee SH, Cho N, Chang JM, Koo HR, Kim JY, Kim WH, et al. Two-view versus single-view shear-wave elastography: comparison of observer performance in differentiating benign from malignant breast masses. *Radiology.* (2014) 270:344–53. doi: 10.1148/radiol.13130561
17. Choi JS, Han BK, Ko ES, Bae JM, Ko EY, Song SH, et al. Effect of a deep learning framework-based computer-aided diagnosis system on the diagnostic performance of radiologists in differentiating between malignant and benign masses on breast ultrasonography. *Korean J Radiol.* (2019) 20:749–58. doi: 10.3348/kjr.2018.0530
18. Lee JG, Jun S, Cho YW, Lee H, Kim GB, Seo JB, et al. Deep learning in medical imaging: general overview. *Korean J Radiol.* (2017) 18:570–84. doi: 10.3348/kjr.2017.18.4.570
19. Cheng JZ, Ni D, Chou YH, Qin J, Tiu CM, Chang YC, et al. Computer-aided diagnosis with deep learning architecture: applications to breast lesions in US images and pulmonary nodules in CT scans. *Sci Rep.* (2016) 6:24454. doi: 10.1038/srep24454
20. Lee JH, Seong YK, Chang CH, Ko EY, Cho BH, Ku J, et al. Computer-aided lesion diagnosis in B-mode ultrasound by border irregularity and multiple sonographic features. In: *Proceedings SPIE 8670, Medical Imaging 2013: Computer-Aided Diagnosis.* Lake Buena Vista, FL (2013) 86701. doi: 10.1117/12.2007452
21. Di Segni M, de Soccio V, Cantisani V, Bonito G, Rubini A, Di Segni G, et al. Automated classification of focal breast lesions according to S-detect: validation and role as a clinical and teaching tool. *J Ultrasound.* (2018) 21:105–18. doi: 10.1007/s40477–018-0297–2
22. Kim K, Song MK, Kim EK, Yoon JH. Clinical application of S-Detect to breast masses on ultrasonography: a study evaluating the diagnostic performance and agreement with a dedicated breast radiologist. *Ultrasonography.* (2017) 36:3–9. doi: 10.14366/usg.16012
23. Tabar L, Duffy SW, Vitak B, Chen HH, Prevost TC. The natural history of breast carcinoma: what have we learned from screening? *Cancer.* (1999) 86:449-62. doi: 10.1002/(SICI)1097--0142(19990801)86:3<449::AID-CNCR13>3.0.CO;2-Q
24. Moon WK, Chen IL, Chang JM, Shin SU, Lo CM, Chang RF, et al. The adaptive computer-aided diagnosis system based on tumor sizes for the classification of breast tumors detected at screening ultrasound. *Ultrasonics.* (2017) 76:70–7. doi: 10.1016/j.ultras.2016.12.017

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling editor declared a shared committee affiliation, though no other collaboration, with one of the authors, QZ, at the time of review.

frontiers
in Oncology

REVIEW
published: 03 September 2020
doi: 10.3389/fonc.2020.01736

Study Processes and Applications of Ultrasomics in Precision Medicine

Rui Yin[1], Meng Jiang[2], Wen-Zhi Lv[3], Fan Jiang[4], Jun Li[5], Bing Hu[1]*, Xin-Wu Cui[2]* and Christoph F. Dietrich[6]

[1] Department of Ultrasound, Affiliated Renhe Hospital of China Three Gorges University, Yichang, China, [2] Sino-German Tongji-Caritas Research Center of Ultrasound in Medicine, Department of Medical Ultrasound, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [3] Department of Artificial Intelligence, Julei Technology, Wuhan, China, [4] Department of Ultrasound, The Second Affiliated Hospital of Anhui Medical University, Hefei, China, [5] Department of Ultrasound, The First Affiliated Hospital, School of Medicine, Shihezi University, Shihezi, China, [6] Department of Internal Medicine, Hirslanden Clinic, Bern, Switzerland

Ultrasomics is the science of transforming digitally encrypted medical ultrasound images that hold information related to tumor pathophysiology into mineable high-dimensional data. Ultrasomics data have the potential to uncover disease characteristics that are not found with the naked eye. The task of ultrasomics is to quantify the state of diseases using distinctive imaging algorithms and thereby provide valuable information for personalized medicine. Ultrasomics is a powerful tool in oncology but can also be applied to other medical problems for which a disease is imaged. To date there is no comprehensive review focusing on ultrasomics. Here, we describe how ultrasomics works and its capability in diagnosing disease in different organs, including breast, liver, and thyroid. Its pitfalls, challenges and opportunities are also discussed.

Keywords: ultrasomics, diagnosis, tumor, ultrasound, artificial intelligence, computer aided diagnosis

Edited by:
Richard Gary Barr,
Northeast Ohio Medical University,
United States

Reviewed by:
Laurence Gluch,
The Strathfield Breast Centre,
Australia
Wenjie Liang,
Zhejiang University, China

***Correspondence:**
Bing Hu
hubing32@163.com
Xin-Wu Cui
cuixinwu@live.cn

Specialty section:
This article was submitted to
Cancer Imaging and Image-directed
Interventions,
a section of the journal
Frontiers in Oncology

Received: *25 March 2020*
Accepted: *04 August 2020*
Published: *03 September 2020*

Citation:
Yin R, Jiang M, Lv W-Z, Jiang F,
Li J, Hu B, Cui X-W and Dietrich CF
(2020) Study Processes
and Applications of Ultrasomics
in Precision Medicine.
Front. Oncol. 10:1736.
doi: 10.3389/fonc.2020.01736

INTRODUCTION

Ultrasomics is the science of transforming digitally encrypted medical images that hold information related to tumor pathophysiology into mineable high-dimensional data (1, 2). The role of ultrasomics is to quantify the diseases using distinctive imaging algorithms and thereby provide valuable information for personalized medicine (3).

The Precision Medicine Initiative was launched in 2015 and studied the complex biological behaviors of tumors and their interactions. This initiative uses a holistic approach to explain the complexity of biological systems and starts with the recognition that the network that makes up an entire organism is not just the sum of its parts (4). In situations where traditional "one-on-one" diagnosis and treatment are unable to meet medical requirements, a multidisciplinary comprehensive diagnosis method is needed for both doctors and patients. This approach incorporates not only the relatively static genetic code but also the dynamic changes and heterogeneous nature of tumors (5). Radiomics plays a key role in precision medicine. Ultrasomics is a branch of radiomics that extracts vast arrays of quantitative features from ultrasound images and integrates them with the clinical data of patients. It can obtain the texture, shape, intensity, trends and wavelet features of a tumor, distinguish heterogeneity between tumors, and provide a comprehensive quantitative tumor phenotype for doctors (6). The aim of ultrasomics is to obtain the optimal efficacy and safety to ensure maximum quality of life and to avoid excessive and ineffective treatments (7). Ultrasomics does not aim to replace existing clinical decision-making tools but to provide a supplement to current measures by implementing a robust, low-cost, repeatable, and highly effective approach to current clinical practice (8).

BASIC TECHNIQUES OF ULTRASOMICS ANALYSIS

Ultrasomics is defined as quantitative mapping, that is, extracting medical imaging features related to predicted targets, analyzing the information contained, and finally establishing a model. The basic steps of ultrasomics include data acquisition, segmentation, feature calculation, and modeling (**Figure 1**).

Data Acquisition

Ultrasomics usually begins with a prediction target — the event a doctor wishes to predict. Building a successful model usually relies on access to a large number of medical images and clinical data to reveal correlations. Different data sources may have an unexpected impact on the results, therefore, it is important to use standardized imaging protocols to eliminate unnecessary confounding variability (9).

Data, including images, may be retrospectively collected or prospectively acquired, depending on the study design. The greatest obstacle to the reliability and stability of ultrasomics is the high variability of ultrasound images acquired by different operators. Single-center data are usually best obtained by a few radiologists with just a few machines, which guarantees better image consistency. However, multicenter study represents more authority and credibility as it requires a large number of representative teams to jointly obtain relevant data sets. This requires the participating hospitals to reach a clear agreements and establishment of standardized operating standards, including unifying the machine, frequency of the probe, gain of the image, focus, depth, resolution and gray value, and unifying whether to add blood flow, radiography, and patient posture (10). Retrospective image acquisition currently lacks standardization, and raw data are usually not available. Thus, reconstructed images must be used. Conversely, when images are acquired prospectively, an image acquisition standard suitable for ultrasomics should be selected for analysis. Standards in this situation are controllable and can maximize the information for subsequent work (11).

Segmentation

Defining the region of interest (ROI) can be undertaken by manual segmentation, semiautomatic segmentation and automatic segmentation (12). Regardless of which method is used, this step is time-consuming and challenging. On the one hand, there is no gold standard for ROI segmentation, and it is difficult to define the morphology, echogenicity and boundary of a variable lesion. On the other hand, a consensus for image standardization is difficult to reach for a diffuse disease or multiple lesions.

Ultrasomics segmentation includes outlining 2-dimensional (2D) ROIs and 3-dimensional (3D) volumes of interest (VOIs). In most studies, experts isolate the object of interest with a manual algorithm. Manual delineation has well-known limitations regarding inter- and intraobserver variability and should be performed by at least 2, and preferably more experts with predefined protocol consensus. This process is undoubtedly tedious, with inevitable variability. In recent years, deep learning has given researchers automatic or semiautomatic segmentation, which outperforms fixed thresholding, aiming to achieve higher accuracy. The common segmentation algorithms include region-growing, level setting, image cutting, active contour (snake) algorithms, semiautomatic segmentation, and livewire methods. However, since automatic segmentation techniques are in the exploratory stage and require much debugging and revision, their applications are still limited. Semiautomatic segmentation is a perfect combination of manual control and intelligence. Region-growing is one of the semiautomatic methods that was often used to the segmentation in computer-aided system (CAD) (13). The method is called "click and grow," that is, putting the seed points in the target area, then it will automatically grow around and automatically stop at the edge of the lesions. The seed point is generally selected in the center of the target tumor. This segmentation method saves both time and effort, but when the boundary of lesion appears to be unclear, the segmentation results may not be ideal and may need to be modified by a professional radiologist. Regardless of the segmentation method that researchers use, the ultimate aim is based on the reliability of the ROI result.

Feature Calculation

Ultrasomics features are automatically extracted by computer algorithms from the sketched ROI. The characteristics mined by different research institutes are different and have different content. In general, the features are divided into four parts: morphological features, first-order features, second-order features, and higher-order features.

Morphological features include lesion volume, shape (spherical, non-spherical, etc.), and boundary morphology (flat, round, clear, sharp, fuzzy, amorphous, and unclear, etc.). These features are acquired based purely on the experience of radiologists, but despite this, they still offer many possibilities for generating hypotheses.

The first-order features are the common statistical elements (gray signal-strength value) obtained from the image, which include the average pixel signal value, standard deviation, skewness, and kurtosis. These features are represented by a single value or frequency distribution histogram, which quantitatively summarizes the signal intensity of the target area. Although they have great appeal in ultrasomics due to their simplicity, these features do not include spatial information.

Second-order features are usually described as texture features, which include the gray intensity of adjacent pixels. Second-order features suggest indicative information for cancer judgment and are used to explain the internal heterogeneity and complexity of the spatial distribution of the tumors. Commonly used second-order statistical descriptors include

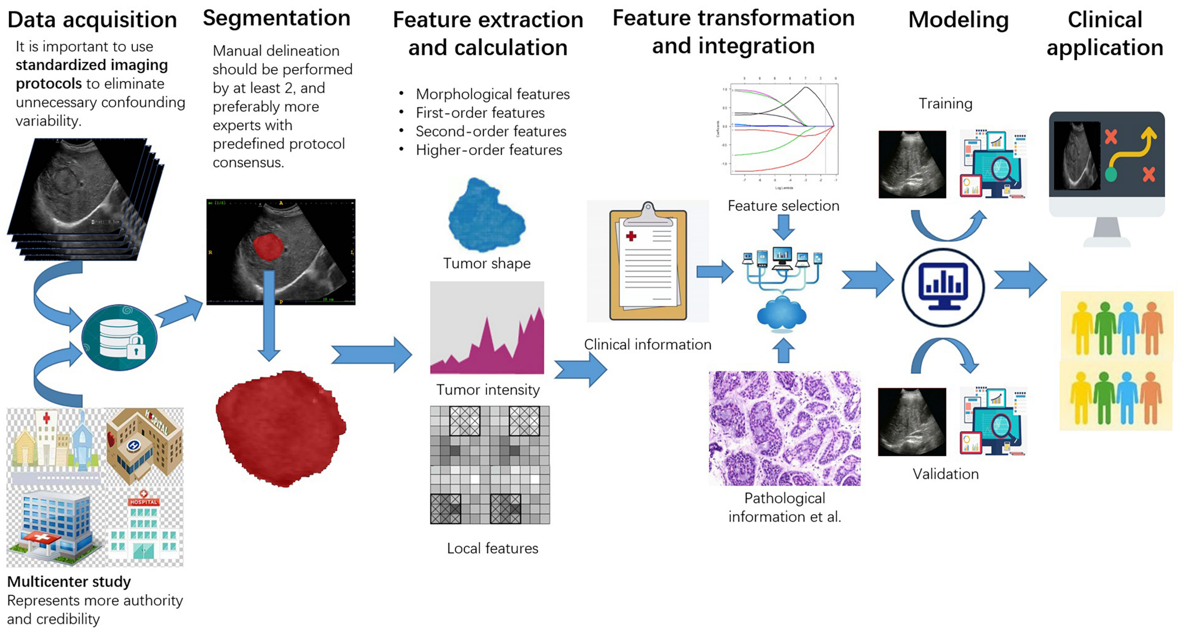


FIGURE 1 | The workflow of ultrasomics. The workflow includes data acquisition, segmentation, feature extraction and calculation, feature transformation and integration, modeling, and clinical application.

gray-level co-occurrence matrices (GLCMs) and gray-level run length matrices (GLRLMs). In GLCMs, the frequencies adjacent to (co-occurrence) pixels of the same signal strength are provided as a matrix to describe the density of the signal strength in a particular direction to reveal differences in regional heterogeneity (14). In GLRLMs, the heterogeneity of signal strength within the ROI can be determined by calculating the frequency at which the nearest-neighbor pixels match in intensity or the operating frequency with the same signal strength (15).

Higher-order features include filters and higher-order images to describe metrics. These features describe the local spatial organization of signal-strength values by applying and adjusting filters in multidimensional space. This represents a quantitative approach. Higher-order feature elements are usually obtained from gray values by the Fourier transform (FT), which converts spatial information to the frequency space and then reverses the conversion process back to the spatial domain (16). Typical techniques include the discrete orthonormal Stockwell transform (DOST), Gabor filter banks, the wavelet transform (WT), the Riesz transform, the Stockwell transform (ST), and the Laplacian of the Gaussian (17).

Modeling

Model building for ultrasomics includes three main steps: feature selection, modeling development, and validation. Selecting the required indicators from infinite features and avoiding overfitting should be the main focus. Ultrasomics and non-ultrasomics features should be combined with the prediction target to create a single dataset. This enables the investigation of relationships between features. Feature pruning is usually required, because computing a large number of features from several matrices can result in many redundant and/or highly correlated features, which greatly increases the complexity of the problem without adding useful information (18). Feature extraction is followed by pruning, methods for which usually include (1) the wrapper method, which scores features based on a performance classification to reflect the usefulness of each feature; (2) the filter method, which uses statistical methods to sort the features and select the highest-ranking feature to determine the intrinsic value of each feature; and (3) the embedded method, which is similar to the wrapper method in some aspects because characteristics are selected to optimize the performance of the learning algorithm. However, unlike the wrapper method, which uses the classification method as an external black box to sort the features, variable selection in the embedded method is an inherent part of the learning algorithm itself.

Model development methods are usually based on the skills and experience of the researcher, which has associated limitations. When training the model, the training samples with the corresponding clinical tags are paired with the training model. Through a predefined loss function, the relationships between the learning characteristics of the model and the clinical label are found, and finally, the model with good training results is selected for testing, this is called supervised learning. In unsupervised learning, the training model no longer needs clinical labeling, it divides similar samples into a set of final generated models according to the algorithm.

Verification is a tool to evaluate whether a model is useful. This research is considered successful only when both the internal and external verification results are satisfactory.

With the application of a classifier, it is necessary to use corresponding measures to evaluate the results as a way of verifying the stability of the generated model. Therefore, the measures below are considered for a confusion matrix of true positives (TPs), true negatives (TNs), false positives (FPs), and false negatives (FNs).

The performance of the model is evaluated by the following formulas:

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$F1\_score = 2 \times precision \times \frac{Sensitivity}{precision + Sensitivity}$$

Additionally, the area under the receiver operating characteristic curve (AUROC) is commonly used to describe the overall performance of a parameter. An AUROC value close to 1 represents an ideal value. A value less than 0.5 suggests that the parameter does not have any classification ability.

APPLICATIONS OF ULTRASOMICS IN PRECISION MEDICINE

The ultimate goal of ultrasomics is to assist radiologists in diagnosing diseases. Currently, studies on ultrasomics cover collecting imaging features, genetic features, and clinical features for data mining analysis and performing tumor screening, diagnosis, classification, and staging predictions. Ultrasomics can also analyze the molecular and biological characteristics of tumors, providing a scientific basis for targeted treatment programs. Regarding follow-up information, ultrasomics analysis on images before and after treatment can predict treatment effects and patient survival, thus assisting in the development of individualized and precise treatment plans. In this review, we briefly introduce the applications of ultrasomics in the breast, liver, and thyroid (**Table 1**).

Breast

Breast cancer is a major health problem in women. The early detection and identification of breast tumors is of great importance for improving quality of life. There are an increasing number of reports on the application of ultrasomics in breast diseases.

Screening, Diagnosis, Classification, and Staging

In traditional ultrasound diagnosis, images are purely used as pictures for human visual interpretation. This process relies heavily on the subjective scoring of images and the limited sensitivity of the naked eye. It can only extract completely macroscopic disease features and misses several pieces of important microbiological information (19). Ultrasomics extracts high-throughput information and performs quantitative analysis with a CAD, which can objectively describe and explain the features of tumors. At present, studies have combined conventional 2D ultrasound images, shear-wave elastography (SWE) images, strain elastography images, and contrast-enhanced ultrasound (CEUS) images with radiomics to detect and identify breast tumors (14, 20–22). These studies extracted high-throughput features to quantify tumor shape, hardness, and hardness heterogeneity to identify breast malignancies and benign tumors. Moreover, these studies found that quantitative ultrasound features were significantly associated with hormone receptor status, molecular subtype and histologic grade in breast invasive ductal carcinoma (IDC). Ultrasomics also makes it possible to evaluate biological parameters by non-invasive means (23–25). Moreover, in a study by Luo et al. 19 features selected by Least absolute shrinkage and selection operator (LASSO) were used to score the degree of malignancy of Breast Imaging Reporting and Data System (BI-RADS) 4 and 5 breast masses, and they obtained AUC values of 0.921 and 0.931 in the training and validation groups, respectively (26). This study showed the outstanding discriminative ability of ultrasomics in grading the possibility of malignancy. Ultrasomics reflects tissue structure and morphological features by quantitatively analyzing the gray value of medical images and then extracting the quantitative features with computer algorithms. This approach can effectively avoid the subjective description of radiologists and the large variability between observers (20, 27). Ultrasomics clarifies the correlation between the malignant potential of masses and image features and shows good prospects for tumor diagnosis.

Individualized Treatment and Survival Prediction

The core task of precision medicine is to identify patient phenotypes (disease, treatment response, adverse side effects, and survival prediction) to find individualized treatment options. Cancer cells exhibit a high degree of heterogeneity, even in different regions of the same tumor, different metastatic sites in the same patient, or the same type of tumor among different patients. This high degree of genetic variation explains the failure of targeted therapies and allows the emergence and proliferation of resistant clones (28). In this case, techniques for quantifying intra- and intertumor heterogeneity are critical because they may guide adaptive treatment (29). Lee et al. performed ultrasomics scoring on 901 lesions and ultimately obtained a model for distinguishing triple-negative breast cancer (TNBC) from breast fibroadenomas (30). Texture features were extracted using the GLCM and GLRLM in this study, they found that both tumor grade and receptor status had an impact on ultrasound performance. Tumors with ER^{+}, her 2^{-} are associated with irregular shapes, unbounded edges, or complex echo patterns, and rear shadowing. This may be caused by the relatively slow proliferation rate of cells, the long-term interaction between a tumor and host, and hyperplasia of the fibrous

TABLE 1 | Summary of ultrasomics studies in oncology.

Stydies	Study design	Cancer	No. of patients	Modality	Features	Feature classifier	Type of features	Statistical analysis	Endpoint	Result
Zhao et al. (37)	Retrospective Single center	Liver	177	BMUS/SWE /SWV	2560	SRT/SVM	GM/GEM/GEVM	Mann–Whitney *U* test	prognosis and diagnosis	AUC: 0.94 (benign/malignant) AUC: 0.97 (malignant subtyping) AUC: 0.97 (PD-1 prediction) AUC: 0.94 (Ki-67 prediction) AUC: 0.98 (MVI prediction)
Zhou et al. (20)	Retrospective Single center	Breast	205	SWE	4224	CNN	–	–	Diagnosis	Accuracy: 95.8% Sensitivity: 96.2% Specificity: 95.7%
Li et al. (21)	Retrospective Single center	Breast	178	BMUS/SWE/ CEUS	1226	SVM	Intensity/Texture/ Contourlet/Shape/Perfusion	Holdout test	Diagnosis	Accuracy: 84.12% Sensitivity: 92.86% Specificity:78.80% AUC: 0.919
Luo et al. (26)	Retrospective Single center	Breast	315	BMUS	1044	LASSO	Histogram/Texture/ RLM/Form factor	Multivariate regression analysis	Diagnosis	AUC: 0.928
Lee et al. (30)	Retrospective Single center	Breast	901	BMUS	730	LASSO	Intensity/Texture/ Wavelet	–	Diagnosis	AUC: 0.782
Zhang et al. (14)	Retrospective Single center	Breast	117	Sonoelastography	364	clusters derived	Shape/intensity/ GLCM/contourlet	Clusters derived/SVM	Diagnosis	AUC: 0.97 Accuracy: 88.0% Sensitivity: 85.7% Specificity: 89.3%
Qiu et al. (31)	Retrospective Single center	Lymph node	256	BMUS	843	LASSO and ridge regression	Shape/firstorder GLCM/gray-level size zone matrix/gray-level distance zone matrix/neighborhood gray-tone difference matrix/gray-level run length matrix	Elastic net logistic regression	Diagnosis	AUC: 0.816
Li et al. (33)	Prospective Single center	Liver	144	BMUS/CEMF	472	Spearman's correlation coefficient	Conventional radiomics/ORF/CEMF features	–	Diagnosis	Mean AUC: 0.78–0.85 (the multiparametric ultrasomics model)
Wang et al. (34)	Prospective Multicentre	Liver	654	SWE	–	CNN	–	Student's *t* test/Mann–Whitney *U* test	Prognosis	AUC: 0.97 (F4) AUC: 0.98 (F3) AUC: 0.85 (F2)
Hu et al. (38)	Retrospective Multicentre	Liver	482	CEUS	1044	LASSO	–	LASSO	Prognosis	AUC: 0.731 $p = 0.015$
Liang et al. (39)	Retrospective Multicentre	Thyroid	137	BMUS	1044	LASSO	–	Univariate logistic regression	Diagnosis	AUC: 0.921 (training cohort) AUC: 0.931 (validation cohort)
Liu et al. (40)	Retrospective Single center	Lymph node	1216	BMUS	614	combined feature selection strategy	Echo/posterior acoustic/calcification	–	Prognosis	AUC: 0.782
Park et al. (41)	Retrospective Single center	Thyroid	768	BMUS	730	LASSO	–	LASSO/Cox regression	Prognosis	C-index: 0.777; 95%[CI]: 0.735, 0.829
Liu et al. (42)	Retrospective Single center	Lymph node	75	BMUS/SE-US	684	SVM	–	Delong's test	Prognosis	AUC: 0.90 Accuracy: 0.85 Sensitivity: 0.77 Specificity: 0.88

The design of the studies, category of tumors, number of patients, number of features, type of features, mode build method, endpoint, diagnostic modality, and results of the studies were considered. The name of the first author and the reference number are indicated in the first column. BMUS, B-mode ultrasound; SWE, shear wave elastography; SWV, shear wave viscosity; CEUS, contrast-enhanced ultrasound; SE-US, strain ultrasound elastography; GM, the gray-scale modality; GEM, the gray-scale and elastography modality; GEVM, gray-scale, elastography and viscosity modality; SVM, support vector machine; LASSO, least absolute shrinkage and selection operator; RLM, gray level run-length matrix; CEMF, contrast-enhanced micro-flow; ORF, original radiofrequency; CEMF, contrast-enhanced micro-flow; and SRT, sparse representation theory.

tissue around the lesion, which results in uneven borders, burrs or leaves. Interstitial response and connective tissue hyperplasia cause different acoustic impedance differences, acoustic reflections, and echo attenuation behind the mass. However, TNBC tends to have oval or round shapes and circumscribed margins, reflecting a rapidly proliferating tumor prior to significant stromal reaction. It is also more likely to present with posterior acoustic enhancement since highly cellular circumscribed carcinomas tend to have enhanced through-transmission. This benign-looking might decrease the diagnostic efficacy of ultrasound and delay treatment. Radiomics based on texture analysis shows excellent diagnostic performance in the differential diagnosis of fibroadenoma and TNBC where it is indiscernible with the naked eye. Theoretically, ultrasound images may contain hidden information that can be difficult for radiologists to mine (29). Ultrasomics can find heterogeneities within a region from indistinguishable imaging data.

Sentinel lymph nodes are an important factor for the prognosis of breast cancer patients. The precise and non-invasive prediction of axillary lymph nodes before surgery is of great significance for staging, treatment and prognosis. Qiu X et al. combined ultrasomics with features of axillary lymph nodes on B-mode ultrasound images and found that a radiomics model with LASSO and ridge regression methods was able to predict axillary lymph node metastasis by using ultrasound features of primary breast tumors (31). This strategy might be an effective alternative to early screening for lymph node metastasis in clinically lymph node–negative breast cancer. It also showed the great potential to serve as an important decision support tool in clinical practice. It is expected to reduce the axillary lymph node dissection and sentinel lymph node biopsy and the corresponding postoperative complications accordingly.

Liver

Screening, Diagnosis, Classification, and Staging

Hepatitis B virus (HBV) infection is a serious problem around the world. Liver fibrosis, cirrhosis, and liver cancer are progressive diseases of chronic hepatitis B (CHB). An accurate assessment of liver status is essential for the prognosis, monitoring and management of CHB patients. D Souza et al. studied the B-mode ultrasound features of the liver in a rat model to assess liver fibrosis (32). The computer algorithm extracted quantitative parameters representing brightness (echo intensity and liver and kidney index) and variance (heterogeneity) to study the anisotropy of the liver. The echo intensity of DEN rats increased from 37.1 ± 7.8 to 53.5 ± 5.7∼57.5 ± 6.1, compared with an average of 34.5 ± 4.5 in the control group. Histological analysis revealed that fibrosis fractionation with METAVIR scores F2-F4 and specifically F0-F1 in DEN rats increased the imaging parameters. Wang et al. and Li et al. applied this technique in clinical practice. Li et al. acquired ultrasound radio frequency signals and dynamic perfusion information to construct an ultrasomics model, and they derived an optimal algorithm for assessing liver fibrosis in a small sample (33). Wang et al. suggested that deep learning radiomics of elastography (DLRE) could be successfully used to assess the liver fibrosis stage of patients with CHB and was comparable to the current grading criteria for cirrhosis and advanced fibrosis. The diagnostic accuracy of the model was higher than that of 2D-SWE for overcoming the influence of inflammation on cirrhosis assessments (34). Therefore, ultrasomics is a potential breakthrough in image diagnosis.

Individualized Treatment and Survival Prediction

Ultrasomics applies the identification, analysis, and integration of ultrasound images to reach a better solution for patients. The main factors in the recurrence of liver cancer are microvascular infiltration (MVI) and Ki-67 (35). MVI is a common predictor of the prognosis for patients with liver cancer. MVI is highly correlated with early recurrence and greatly influences treatment (liver resection or orthotopic liver transplant) (36). Yao et al. classified images by transforming them into high-throughput features, analyzed multiple parameters in the treatment area and used sparse representation theory (SRT), and support vector machine (SVM) methods to mine rich texture information. The authors found that malignant tumors had more complex textures and structural information than benign tumors. Their results indicated that predicting MVI (AUC = 0.98), Ki-67 (AUC = 0.94), and PD-1 (AUC = 0.97) with a non-invasive method based on radiomics is feasible (37). This finding showed that ultrasomics could improve the diagnostic efficiency of ultrasound and made it possible to diagnose FLL before operation. Additionally, ultrasomics can explain the biological behavior of tumors and improve the diagnostic efficacy and patient prognosis (38).

Thyroid

Screening, Diagnosis, Classification, and Staging

Thyroid disease has received widespread attention due to its high incidence. The Thyroid Imaging-Reporting and Data System (TI-RADS) is widely used to describe thyroid lesions and is unavoidably subjective. In a study, Liang et al. developed an ultrasomics model to diagnose malignant thyroid nodules. They performed LASSO to select features and found that ultrasomics could outperform ACR TI-RADS scoring, at least when performed by junior radiologists (39). In addition, the application of texture analysis and machine learning in thyroid nodule imaging can describe thyroid nodules better and more objectively. Liu et al. obtained predictive models though a SVM classifier from more than 50 traits of thyroid tumors, such as the volume, echo, margin, boundaries, posterior acoustic pattern, and calcification features. They obtained satisfactory results in predicting which thyroid nodules would develop lymph node metastasis (40). Lymph node metastasis is more likely to occur in patients with complex echoes in ultrasound images, uniform posterior regions, large calcifications or multiple calcifications (41). Clinically, the lymph nodes suspected malignant are re-examined by CT, fine needle aspiration cytology, or lymph node dissection (LND). LND has the risk of hyperparathyroidism and nerve injury. In addition, whether LND can improve the survival rate of PTC patients is still controversial, so full consideration must be given before it is used in patients. The radiomics evaluation has potential to predict LN status non-invasively based on preoperative ultrasound thyroid images. It made up for the shortcomings of traditional diagnosis. The lymph node status prediction model has the potential to promote early medical management for thyroid cancer patients and reduce overdiagnosis.

Individualized Treatment and Survival Prediction

Recurrence and metastasis are the key points during the treatment of cancer and are closely related to the survival time of patients. Long-term follow-up assessments are indispensable after tumor treatment. The most important part of this process is to determine whether there are any recurrences or residual lesions. Liu et al. combined the features extracted from B-mode ultrasound and strain ultrasound elastography (SE-US), and multimodal feature sets were obtained through image segmentation, quantitative feature extraction, feature selection and classification. This study used the sparse representation coefficient-based feature selection method with 10 bootstraps to reduce the dimensionality of the feature sets. A SVM with leave-one-out cross-validation was used to build the model to estimate LN status. The model had the best ability to diagnose lymph node metastases (42). Furthermore, ultrasomics could not only characterize the properties of thyroid nodules but also assess the disease-free survival of thyroid nodule patients. This is the first application of ultrasomics to predict the prognosis of thyroid cancer. The authors used relapsed or persistent disease-free survival as the study endpoint rather than mortality, demonstrating the great potential of ultrasomics (41).

CHALLENGES

First and foremost, the greatest problem with ultrasomics is the quality and quantity of the original data. A successful ultrasomics model needs a sufficient quantity of data to develop an effective knowledge system to support data integration, processing, and analysis, which is critical for research to be optimal. Currently, ultrasomics usually use a smaller population to extract more features, which may lead to overfitting and overoptimistic results. There are numerous methods for extracting useful biomarkers from separate or combined layers of ultrasomics and clinical data, but the results are still unsatisfactory. Second, ultrasound examinations are less reproducible than other imaging methods. Additionally, the device and experience of the radiologist have a great impact on the reliability of the diagnosis (43). Therefore, the inclusion and exclusion criteria for ultrasomics should be rigorously developed. Recently, the image biomarker standardization initiative (IBSI) was proposed to improve the reproducibility of high-throughput imaging analysis, which is a valuable step in improving radiological research. In addition, a radiomics quality score (RQS) was proposed to help evaluate radiomics studies (44). Moreover, most of the published ultrasomics studies are from a single center, with different patient numbers, different ultrasound equipment, and different study design methods. The differences in each step of the study design pose greater challenges to the repeatability of the study.

CONCLUSION AND PERSPECTIVES

Although many problems still need to be solved, the potential of ultrasomics is beyond doubt, and the field is evolving rapidly. The development of ultrasomics has occurred over only a few decades, and some impressive results have been achieved. This approach fills the gap in the clinical use of information and extracts and analyzes higher-dimensional and quantitative data to more accurately and more specifically describe and characterize tumors. The use of ultrasomics to improve disease diagnosis and care for patients shows great potential (45). In the future, we hope that ultrasomics will provide a more personalized, higher-quality, and more cost-effective care platform for patients. The advantages of ultrasomics, including its speed, low cost, reproducibility, and non-invasiveness, may make it a valuable clinical decision-making tool.

AUTHOR CONTRIBUTIONS

X-WC, BH, and CD established the design and conception of the manuscript. RY, MJ, W-ZL, FJ, and JL explored the literature data. RY provided the first draft of the manuscript, which was discussed and revised critically for intellectual content by RY, MJ, W-ZL, FJ, JL, BH, X-WC, and CD. All authors discussed the statement and conclusions and approved the final version to be published.

FUNDING

This manuscript was supported by the Health Commission of Hubei Province (WJ2019M077 and WJ2019H227), Shihezi Science and Technology Bureau (2019ZH11), Key project supported by the Xinjiang Construction Corps (2019DB012), and Research Fund for Excellent Dissertation of China Three Gorges University (2019SSPY109).

REFERENCES

1. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology.* (2016) 278:563–77. doi: 10.1148/radiol.2015151169
2. Avanzo M, Stancanello J, El Naqa I. Beyond imaging: the promise of radiomics. *Phys Med.* (2017) 38:122–39. doi: 10.1016/j.ejmp.2017.05.071
3. Li Q, Ye ZX. Radiomics: the process and applications in tumor research. *Chin J Oncol.* (2018) 40:801–4. doi: 10.3760/cma.j.issn.0253-3766.2018.11.001

4. Liu Z, Wang S, Dong D, Wei J, Fang C, Zhou X, et al. The applications of radiomics in precision diagnosis and treatment of oncology: opportunities and challenges. *Theranostics.* (2019) 9:1303–22. doi: 10.7150/thno.30309
5. Pinker K, Chin J, Melsaether AN, Morris EA, Moy L. Precision medicine and radiogenomics in breast cancer: new approaches toward diagnosis and treatment. *Radiology.* (2018) 287:732–47.
6. Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, et al. Radiomics: the process and the challenges. *Magn Reson Imaging.* (2012) 30:1234–48. doi: 10.1016/j.mri.2012.06.010
7. Holzinger A, Haibe-Kains B, Jurisica I. Why imaging data alone is not enough: AI-based integration of imaging, omics, and clinical data. *Eur J Nucl Med Mol Imaging.* (2019) 46:2722–30.
8. Keek SA, Leijenaar RT, Jochems A, Woodruff HC. A review on radiomics and the future of theranostics for patient selection in precision medicine. *Br J Radiol.* (2018) 91:20170926.
9. Sollini M, Antunovic L, Chiti A, Kirienko M. Towards clinical application of image mining: a systematic review on artificial intelligence and radiomics. *Eur J Nucl Med Mol Imaging.* (2019) 46:2656–72. doi: 10.1007/s00259-019-04372-x
10. Tran WT, Jerzak K, Lu FI, Klein J, Tabbarah S, Lagree A, et al. Personalized breast cancer treatments using artificial intelligence in radiomics and pathomics. *J Med Imaging Radiat Sci.* (2019) 50:S32–41. doi: 10.1016/j.jmir.2019.07.010
11. Hatt M, Le Rest CC, Tixier F, Badic B, Schick U, Visvikis D. Radiomics: data are also images. *J Nucl Med.* (2019) 60:38S–44S. doi: 10.2967/jnumed.118.220582
12. Ma M, Feng Z, Peng T, Yan H, Rong P, Jumbe MM. Radiomics and its advances in hepatocellular carcinoma. *Zhong Nan Da Xue Xue Bao Yi Xue Ban.* (2019) 44:225–32. doi: 10.11817/j.issn.1672-7347.2019.03.001
13. Elter M, Horsch A. CADx of mammographic masses and clustered microcalcifications: a review. *Med Phys.* (2009) 36:2052–68.
14. Zhang Q, Xiao Y, Suo J, Shi J, Yu J, Guo Y, et al. Sonoelastomics for breast tumor classification: a radiomics approach with clustering-based feature selection on sonoelastography. *Ultrasound Med Biol.* (2017) 43:1058–69. doi: 10.1016/j.ultrasmedbio.2016.12.016
15. Korn RL, Rahmanuddin S, Borazanci E. Use of precision imaging in the evaluation of pancreas cancer. *Cancer Treat Res.* (2019) 178:209–36. doi: 10.1007/978-3-030-16391-4_8
16. Drabycz S, Stockwell RG, Mitchell JR. Image texture characterization using the discrete orthonormal S-transform. *J Digit Imaging.* (2009) 22:696–708. doi: 10.1007/s10278-008-9138-8
17. Khan AM, El-Daly H, Simmons E, Rajpoot NM. HyMaP: a hybrid magnitude-phase approach to unsupervised segmentation of tumor areas in breast cancer histology images. *J Pathol Inform.* (2013) 4:S1. doi: 10.4103/2153-3539.109802
18. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol.* (2017) 14:749–62. doi: 10.1038/nrclinonc.2017.141
19. Valdora F, Houssami N, Rossi F, Calabrese M, Tagliafico AS. Rapid review: radiomics and breast cancer. *Breast Cancer Res Treat.* (2018) 169:217–29. doi: 10.1007/s10549-018-4675-4
20. Zhou Y, Xu J, Liu Q, Li C, Liu Z, Wang MA. Radiomics approach with CNN for shear-wave elastography breast tumor classification. *IEEE Trans Biomed Eng.* (2018) 65:1935–42.
21. Li Y, Liu Y, Zhang M, Zhang G, Wang Z, Luo J. Radiomics with attribute bagging for breast tumor classification using multimodal ultrasound images. *J Ultrasound Med.* (2020) 39:361–71. doi: 10.1002/jum.15115
22. Sollini M, Cozzi L, Chiti A, Kirienko M. Texture analysis and machine learning to characterize suspected thyroid nodules and differentiated thyroid cancer: where do we stand? *Eur J Radiol.* (2018) 99:1–8. doi: 10.1016/j.ejrad.2017.12.004
23. Bodalal Z, Trebeschi S, Beets-Tan R. Radiomics: a critical step towards integrated healthcare. *Insights Imaging.* (2018) 9:911–4. doi: 10.1007/s13244-018-0669-3
24. Ko ES, Lee BH, Kim HA, Noh WC, Kim MS, Lee SA. Triple-negative breast cancer: correlation between imaging and pathological findings. *Eur Radiol.* (2010) 20:1111–7.
25. Çelebi F, Pilancı KN, Ordu Ç, Ağacayak F, Alço G, İlgün S, et al. The role of ultrasonographic findings to predict molecular subtype, histologic grade, and hormone receptor status of breast cancer. *Diagn Interv Radiol.* (2015) 21:448–53. doi: 10.5152/dir.2015.14515
26. Luo WQ, Huang QX, Huang XW, Hu HT, Zeng FQ, Wang W, et al. Predicting breast cancer in breast imaging reporting and data system (BI-RADS) ultrasound category 4 or 5 lesions: a nomogram combining radiomics and BI-RADS. *Sci Rep.* (2019) 9:11921.
27. Gibbs P, Onishi N, Sadinski M, Gallagher KM, Hughes M, Martinez DF, et al. Characterization of sub-1 cm breast lesions using radiomics analysis. *J Magn Reson Imaging.* (2019) 50:1468–77. doi: 10.1002/jmri.26732
28. Rossi SH, Prezzi D, Kelly-Morland C, Goh V. Imaging for the diagnosis and response assessment of renal tumours. *World J Urol.* (2018) 36:1927–42. doi: 10.1007/s00345-018-2342-3
29. Nougaret S, Tibermacine H, Tardieu M, Sala E. Radiomics: an introductory guide to what it may foretell. *Curr Oncol Rep.* (2019) 21:70. doi: 10.1007/s11912-019-0815-1
30. Lee SE, Han K, Kwak JY, Lee E, Kim EK. Radiomics of US texture features in differential diagnosis between triple-negative breast cancer and fibroadenoma. *Sci Rep.* (2018) 8:13546.
31. Qiu X, Jiang Y, Zhao Q, Yan C, Huang M, Jiang T. Could ultrasound-based radiomics noninvasively predict axillary lymph node metastasis in breast cancer? *J Ultrasound Med.* (2020). doi: 10.1002/jum.15294 [Epub ahead of print].
32. D'Souza JC, Sultan LR, Hunt SJ, Schultz SM, Brice AK, Wood AKW, et al. B-mode ultrasound for the assessment of hepatic fibrosis: a quantitative multiparametric analysis for a radiomics approach. *Sci Rep.* (2019) 9:8708. doi: 10.1038/s41598-019-45043-z
33. Li W, Huang Y, Zhuang BW, Liu GJ, Hu HT, Li X, et al. Multiparametric ultrasomics of significant liver fibrosis: a machine learning-based analysis. *Eur Radiol.* (2019) 29:1496–506.
34. Wang K, Lu X, Zhou H, Gao Y, Zheng J, Tong M, et al. Deep learning Radiomics of shear wave elastography significantly improved diagnostic performance for assessing liver fibrosis in chronic hepatitis B: a prospective multicentre study. *Gut.* (2019) 68:729–41. doi: 10.1136/gutjnl-2018-316204
35. Ni M, Zhou X, Lv Q, Li Z, Gao Y, Tan Y, et al. Radiomics models for diagnosing microvascular invasion in hepatocellular carcinoma: which model is the best model? *Cancer Imaging.* (2019) 19:60.
36. Portolani N, Coniglio A, Ghidoni S, Giovanelli M, Benetti A, Tiberio GA, et al. Early and late recurrence after liver resection for hepatocellular carcinoma: prognostic and therapeutic implications. *Ann Surg.* (2006) 243:229–35. doi: 10.1097/01.sla.0000197706.21803.a1
37. Yao Z, Dong Y, Wu G, Zhang Q, Yang D, Yu JH, et al. Preoperative diagnosis and prediction of hepatocellular carcinoma: radiomics analysis based on multi-modal ultrasound images. *BMC Cancer.* (2018) 18:1089. doi: 10.1186/s12885-018-5003-4
38. Hu HT, Wang Z, Huang XW, Chen SL, Zheng X, Ruan SM, et al. Ultrasound-based radiomics score: a potential biomarker for the prediction of microvascular invasion in hepatocellular carcinoma. *Eur Radiol.* (2019) 29:2890–901. doi: 10.1007/s00330-018-5797-0
39. Liang J, Huang X, Hu H, Liu Y, Zhou Q, Cao Q, et al. Predicting malignancy in thyroid nodules: radiomics score versus 2017 american college of radiology thyroid imaging, reporting and data system. *Thyroid.* (2018) 28:1024–33. doi: 10.1089/thy.2017.0525
40. Liu T, Zhou S, Yu J, Guo Y, Wang Y, Zhou 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
41. Park VY, Han K, Lee E, Kim EK, Moon HJ, Yoon JH, et al. Association between radiomics signature and disease-free survival in conventional papillary thyroid carcinoma. *Sci Rep.* (2019) 9:4501. doi: 10.1038/s41598-018-37748-4
42. Liu T, Ge X, Yu J, Guo Y, Wang Y, Wang W, et al. Comparison of the application of B-mode and strain elastography ultrasound in the estimation of lymph node metastasis of papillary thyroid carcinoma

based on a radiomics approach. *Int J Comput Assist Radiol Surg.* (2018) 13:1617–27.

43. Rix A, Lederle W, Theek B, Lammers T, Moonen C, Schmitz G, et al. Advanced ultrasound technologies for diagnosis and therapy. *J Nucl Med.* (2018) 59:740–6.
44. Radiomics. (2020). Available online at: http://www.radiomics.world/ (accessed June 28, 2020).
45. Forghani R, Savadjiev P, Chatterjee A, Muthukrishnan N, Reinhold C, Forghani B. Radiomics and artificial intelligence for biomarker and prediction model development in oncology. *Comput Struct Biotechnol J.* (2019) 17:995–1008. doi: 10.1016/j.csbj.2019.07.001

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Oncology

ORIGINAL RESEARCH
published: 11 September 2020
doi: 10.3389/fonc.2020.557169

A Clinical Assessment of an Ultrasound Computer-Aided Diagnosis System in Differentiating Thyroid Nodules With Radiologists of Different Diagnostic Experience

Yichun Zhang[1,2†], *Qiong Wu*[1,2†], *Yutong Chen*[1,2] *and Yan Wang*[1,2*]

[1] *Department of Ultrasound in Medicine, Shanghai Sixth People's Hospital Affiliated to Shanghai Jiao Tong University, Shanghai, China,* [2] *Shanghai Institute of Ultrasound in Medicine, Shanghai, China*

Introduction: This study aimed to assess the diagnostic performance and the added value to radiologists of different levels of a computer-aided diagnosis (CAD) system for the detection of thyroid cancers.

Methods: 303 patients who underwent thyroidectomy from October 2018 to July 2019 were retrospectively reviewed. The diagnostic performance of the senior radiologist, the junior radiologist, and the CAD system were compared. The added value of the CAD system was assessed and subgroup analyses were performed according to the size of thyroid nodules.

Results: In total, 186 malignant thyroid nodules, and 179 benign thyroid nodules were included; 168 were papillary thyroid carcinoma (PTC), 7 were medullary thyroid carcinoma (MTC), 11 were follicular carcinoma (FTC), 127 were follicular adenoma (FA) and 52 were nodular goiters. The CAD system showed a comparable specificity as the senior radiologist (86.0% vs. 86.0%, $p > 0.99$), but a lower sensitivity and a lower area under the receiver operating characteristic (AUROC) curve (sensitivity: 71.5% vs. 95.2%, $p < 0.001$; AUROC: 0.788 vs. 0.906, $p < 0.001$). The CAD system improved the diagnostic sensitivities of both the senior and the junior radiologists (97.8% vs. 95.2%, $p = 0.063$; 88.2% vs. 75.3%, $p < 0.001$).

Conclusion: The use of the CAD system using artificial intelligence is a potential tool to distinguish malignant thyroid nodules and is preferable to serve as a second opinion for less experienced radiologists to improve their diagnosis performance.

Keywords: thyroid nodule, ultrasonography, computer-aided system, diagnosis, thyroid cancer

Edited by:
Katsutoshi Sugimoto,
Tokyo Medical University, Japan

Reviewed by:
Yoo Jin Lee,
Inje University Busan Paik Hospital, South Korea
Atsuo Kawamoto,
Tokyo Medical University Hospital, Japan

***Correspondence:**
Yan Wang
yannanfly@126.com

[†] *These authors have contributed equally to this work and share first authorship*

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *29 April 2020*
Accepted: *17 August 2020*
Published: *11 September 2020*

Citation:
Zhang Y, Wu Q, Chen Y and Wang Y (2020) A Clinical Assessment of an Ultrasound Computer-Aided Diagnosis System in Differentiating Thyroid Nodules With Radiologists of Different Diagnostic Experience. Front. Oncol. 10:557169. doi: 10.3389/fonc.2020.557169

INTRODUCTION

The incidence of thyroid nodules, up to 68% of the general population, continues to show increasing growth (1, 2). As one of the most extensively applied methods in the detection of thyroid nodules, the ultrasound has the advantages of accessibility, cost-effectiveness, and non-radiation. Although the particular ultrasound (US) features such as microcalcifications, hypoechogenicity, and irregular margins are commonly considered to relate to malignant thyroid disease, the presence of interobserver variation is inevitable (3, 4). Compared with seasoned radiologists, less experienced radiologists are at a greater risk of a misleading diagnosis of thyroid cancer.

Computer-aided diagnosis (CAD) has attracted great attention of researchers as a newly developed technique that has potential in enhancing radiologists' interpretation and overcoming subjective limitations. The CAD detection and diagnosis methods are based on machine learning approaches that extract features based on shape, texture, and statistical values, differentiating benign and malignant nodules (5–7).

Several studies have shown that CAD system has comparable performance to radiologists in terms of sensitivity (8–10). However, few studies have compared the distinction of diagnosis performance between the CAD system and radiologists with various levels of experience in the diagnosis of thyroid cancer and no detailed study has been conducted to focus on the influence of nodule size on CAD performance. Therefore, this retrospective study aimed to validate the clinical role of the CAD systems in thyroid cancer diagnosis and to evaluate their future developmental directions.

MATERIALS AND METHODS

Ethics and Consent

This prospective study was approved by our Institutional Review Board, and the requirement for informed consent was waived due to its retrospective nature.

Database

We retrospectively reviewed medical records of 303 patients who were treated at our center from October 2018 to July 2019. Patients who received an ultrasound examination prior to scheduled surgery with sufficient clinical information were enrolled. The histopathologic diagnosis of the nodules was established by surgery. Finally, there were 186 malignant nodules and 179 benign nodules included in our study.

Ultrasound Images Acquisition and Radiology Analysis

The US scans were operated with a 12–18 MHz linear probe (ACUSON S2000; Siemens Medical Solutions, Mountain View, CA, United States). The US images presented in a random fashion were assessed by a senior radiologist of more than 10 years' experience and a junior radiologist of 2 years' experience.

The CAD system used in the study was AI-SONIC for thyroid nodule (AI-SONIC; Demestics Medical Technology Co., Zhejiang, China), which can analyze the US images for real-time. A grayscale image of a transverse plane of each nodule was uploaded to the software and analyzed. The software is able to automatically mark the suspicious lesion with a square and rate the nodule on a scale of zero to one, with higher scores indicating the higher the risk of malignancy (**Figure 1**).

Statistical Analysis

The SPSS software (verion 20.0, IBM Corp, Armonk, NY, United States) and MedCalc software (version 15.2, Mariakerke, Belgium) were used to analyze the data. A statistically significant difference was considered as *p* value is less than 0.05. Figures were produced using GraphPad Prism (version 8.0, GraphPad Software, San Diego, CA, United States). The classification data were expressed as frequencies; Continuous variables were expressed as means and standard deviations. The diagnostic sensitivity, specificity, accuracy, positive predictive values (PPVs), negative predictive values (NPVs) of the CAD software were calculated by comparing the pathological results.

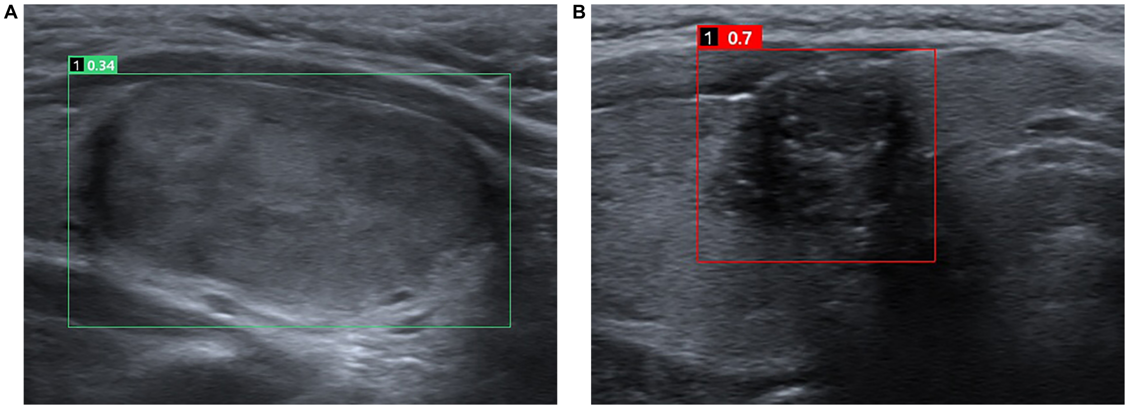


FIGURE 1 | Representative cases of benign **(A)** and malignant **(B)** thyroid nodules. For the benign nodule **(A)**, both the CAD system and the senior and the junior radiologists diagnosed it as a benign nodule. The CAD system rate the nodule of 0.34. For the malignant nodule **(B)**, both the CAD system and the senior and the junior radiologists diagnosed it as a malignant nodule. The CAD system rate the nodule of 0.7.

TABLE 1 | Characteristics of study subjects.

Parameter	Value
Mean age (years)	46.4 ± 14.6
Patient gender, *n* (%)	
Male	59 (19.5%)
Female	244 (80.5%)
No. of nodules, *n* (%)	
Benign nodules	179 (49%)
Malignant nodules	186 (51%)
Nodule sizes (mm)	
Total nodules	18.33 ± 13.5
Benign nodules	25.58 ± 13.7
Malignant nodules	11.36 ± 8.8

McNemar's test was used to compare the diagnostic sensitivity, specificity and accuracy of the CAD system and the senior and junior radiologists. The diagnostic performance of the radiologist assisted by the CAD system was defined as positive when the criteria meet one of the two categories: the radiologist and the CAD system. The diagnosis performance concerning nodule size was compared using the chi-squared (χ^2) test with Bonferroni adjustment, which result in a final *P*-value of 0.0167, based on three independent tests, considered statistically significant. The areas under the receiver operating characteristic (ROC) curve (AUC) were also analyzed to compare the diagnostic performance of different groups, and optimal cut-off value for CAD was defined by the Youden index J.

RESULTS

Patients Data

A total of 303 patients (Mage = 46.4 years; range 23–80 years) with 365 thyroid nodules were included in this study (**Table 1**). There were 186 cases (51.0%) that were malignant, including 168 cases of papillary carcinoma, 11 cases of follicular carcinoma, and 7 cases of medullary carcinoma. There were 179 cases (49.0%) that were benign, including 127 cases of follicular adenoma, and 52 cases of nodular goiters.

CAD Score Cut-Off Value for Predicting Malignant Thyroid Nodules

We determined the positive threshold of CAD scores using the Youden Index (= sensitivity + specificity-1) curve, which could suggest the best cut-off value by fitting optimal sensitivity and specificity (11). The maximum Youden Index pointed to 0.555 for the CAD system score with a sensitivity of 71.5% and a specificity of 86.0% (**Figure 2**).

Diagnostic Performance of the CAD System, Radiologists in the Different Groups and CAD-Assisted Radiologists

The diagnostic performances of the CAD system, radiologists in the different groups, and CAD-assisted radiologists for detecting thyroid cancer were summarized in **Table 2** and **Figure 3**. The CAD system exhibited no statistically significant difference in terms of specificity compared with the senior radiologist (86.0% vs. 86.0%, $p > 0.99$), while the sensitivity and accuracy were markedly lower in the CAD system than those in the senior radiologist (71.5% vs.95.2%, $p < 0.001$; 78.6% vs.90.7%, $p < 0.001$, respectively). When compared with the junior radiologist, the CAD system resulted in increased specificity and similar sensitivity and accuracy in the classification of thyroid cancer (86.0% vs.78.8%, $p = 0.024$; 71.5% vs.75.3%, $p = 0.419$; 78.6% vs.77.0%, $p = 0.552$, respectively). When the CAD system was used to assist the senior and junior radiologists, the diagnostic sensitivity improved (97.8% vs. 95.2%, $p = 0.063$; 88.2% vs. 75.3%, $p < 0.001$, respectively), while the specificity declined (76.0% vs. 86.0%, $p < 0.001$; 79.9% vs. 84.4%, $p = 0.008$, respectively). A ROC analysis comparing the diagnostic values of the CAD system, radiologists, and CAD-assisted radiologists is illustrated in **Figure 3** and **Table 2**. The AUCs were 0.788

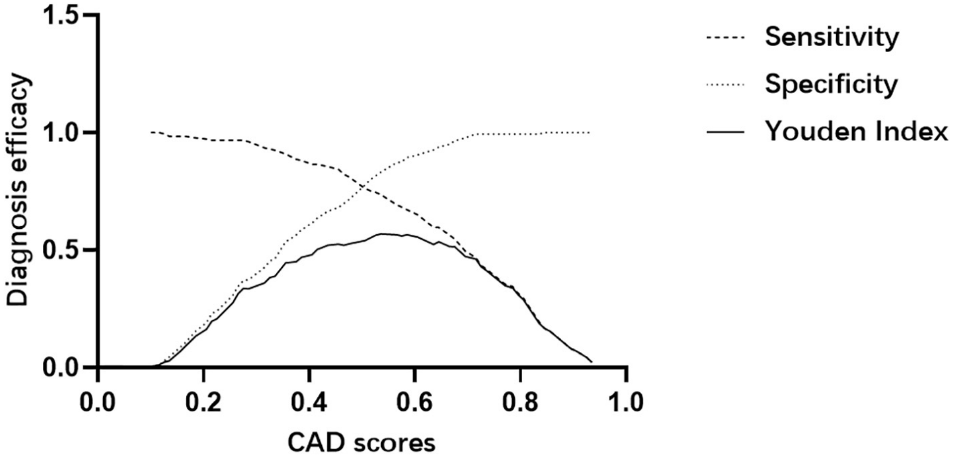


FIGURE 2 | Determination of the positive threshold for the CAD system score through the Youden index.

TABLE 2 | Diagnostic performance of CAD System, radiologists and CAD-assisted radiologists.

Diagnostic measures (%)	Sensitivity	Specificity	PPV	NPV	Accuracy	Area under the ROC curve
CAD system	71.5 (133/186)	86.0 (154/179)	84.2 (133/158)	74.4 (154/207)	78.6 (287/365)	0.788
Senior radiologist	95.2 (177/186)	86.0 (154/179)	87.6 (177/202)	94.5 (154/163)	90.7 (331/365)	0.906
CAD + senior radiologist	97.8 (182/186)	76.0 (136/179)	80.9 (182/225)	97.1 (136/140)	87.1 (318/365)	0.869
Junior radiologist	75.3 (140/186)	78.8 (141/179)	78.7 (140/178)	75.4 (141/187)	77.0 (281/365)	0.770
CAD + Junior radiologist	88.2 (164/186)	74.3 (133/179)	78.1 (164/210)	85.8 (133/155)	81.4 (297/365)	0.812
P-value*	<0.001	>0.99			<0.001	<0.001
P-value**	0.063	<0.001			0.011	0.031
P-value‡	0.419	0.024			0.586	0.552
P-value†	<0.001	0.008			0.121	0.022
P-value†	0.015	0.003			<0.001	<0.001

P-value is that of the CAD system vs. the senior radiologist; P-value** is that of the CAD-assisted senior radiologist vs. the senior radiologist; P-value*‡ *is that of the CAD system vs. the junior radiologist; P-value*† *is that of the CAD-assisted junior radiologist vs. the junior radiologist; P-value*† *is that of the CAD-assisted junior radiologist vs. the senior radiologist; NPV, negative predictive value; PPV, positive predictive value.*

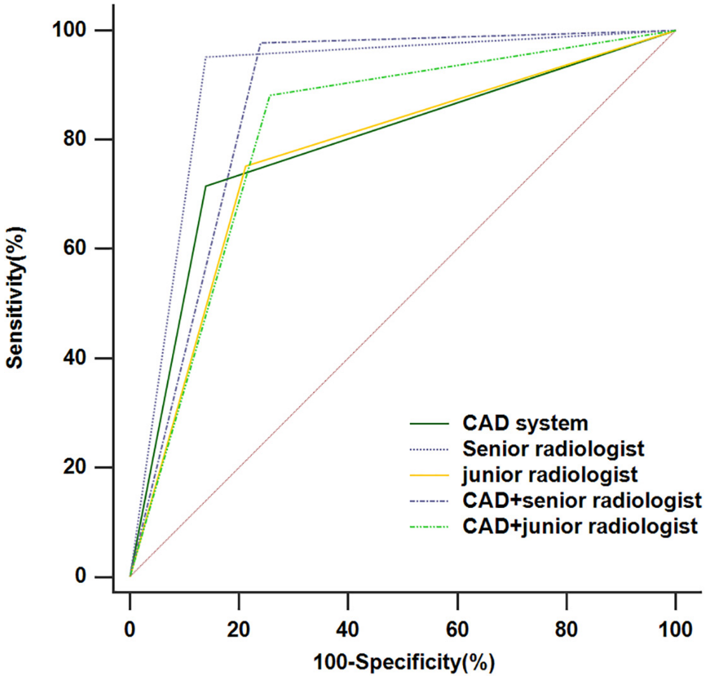


FIGURE 3 | The receiver operating characteristic (ROC) curves for the performance of the computer-aided diagnosis (CAD) system, the senior radiologist, the junior radiologist, and CAD-assisted radiologists.

(0.742, 0.829) for the CAD system, 0.906 (0.871, 0.934) for the senior radiologist, 0.869 (0.830, 0.902) for the CAD-assisted senior radiologist, 0.770 (0.724, 0.812) for the junior radiologist, and 0.812 (0.768, 0.851) for the CAD-assisted junior radiologist.

Comparison of the Diagnostic Performance of the CAD System for the Diagnosis of Thyroid Nodules of Different Sizes

The sensitivity, specificity, NPV, PPV, and accuracy of the CAD system for diagnosing thyroid nodules of different sizes were summarized in **Table 3**. The sensitivity, specificity, and accuracy of the CAD system for the diagnosis of small thyroid nodules diagnosis did not differ from those of medium-sized thyroid nodules after applying Bonferroni correction ($P = 041$; $P = 0.025$; $P = 0.818$, respectively). The sensitivity of the CAD system for the diagnosis of large thyroid nodules was significantly less than for small thyroid nodules ($P < 0.0167$), whereas the specificity and the accuracy of the CAD system in the diagnosis of large thyroid nodules were significantly higher than that of small thyroid nodules ($P < 0.001$; $P < 0.001$, respectively). The sensitivity and specificity of the CAD system for the diagnosis of large thyroid nodules diagnosis did not differ from those of medium-sized thyroid nodules ($P = 0.486$; $P = 0.062$, respectively).

TABLE 3 | Comparison of the diagnostic performance of the CAD system in differentiating the thyroid nodules of different sizes.

	Sensitivity (%)	Specificity (%)	Accuracy (%)
$d \leq 15$ mm	76.3 (116/152)	65.5 (36/55)	73.4 (152/207)
$15 < d \leq 25$ mm	55.0 (11/20)	87.5 (28/32)	75 (39/52)
$d > 25$ mm	42.9 (6/14)	97.8 (90/92)	90.5 (96/106)
P-value			
$d \leq 15$ mm vs. $15 < d \leq 25$ mm	0.041	0.025	0.818
$d \leq 15$ mm vs. $d > 25$ mm	0.016*	<0.001*	<0.001*
$15 < d \leq 25$ mm vs. $d > 25$ mm	0.486	0.062	0.009*

**P values considered significant after applying Bonferroni correction.*

DISCUSSION

Ultrasonography is playing a crucial role in the greatly increasing detection rate of thyroid nodules (12, 13). However, the usefulness of ultrasound may be limited for the diagnostic performance of it is various from person to person, which depends on the experience of a radiologist to a large extent (14). The original CAD system was used to diagnose the breast tumor in the 1960s (15). The CAD system which based on artificial intelligence has been developed to assist radiologists in analyzing images, shortening the time cost of the diagnostic process, and reducing interobserver variability.

In this study, a clinical assessment was performed to evaluate the value of an ultrasound CAD system in the ultrasound diagnosis of thyroid cancer. This retrospective study showed that the CAD system generally performed comparably to qualitative assessments by the senior radiologist in terms of specificity, but had a lower sensitivity and accuracy. In addition, the specificity of the CAD system was greatly higher than that of the junior radiologist and the CAD system demonstrated similar sensitivity and accuracy to the junior radiologist.

Since the diagnostic performance for thyroid lesions of the CAD system initially reported (16), several studies have already revealed that CAD approaches improved the diagnostic manifestations of thyroid ultrasound (8, 9, 17–19). More recently, Chung et al. compared the diagnostic performance of a real-time CAD system with that of a 7-year experienced radiologist, CAD system had comparable sensitivity but lower specificity than the experienced radiologist (20). However, Gitto et al. reported that the CAD system had a significantly lower sensitivity than the experienced radiologist and there was no statistical difference in specificity (21).

The added value of the CAD system was also evaluated in this study. With the assistant of the CAD system, the junior radiologist showed a significant increase in sensitivity from 75.3 to 88.2%. Also, the AUC was greatly improved from 0.770 to 0.812 ($P = 0.022$). The improved sensitivity, NPV, and AUC indicated that the CAD system might function as a supplementary opinion to avoid the missed diagnosis, especially for less-experienced radiologists. As was shown in the study, the CAD system had a comparable specificity to that of a senior radiologist, which implied that the CAD system could play a constructive role in avoiding overdiagnosis and help to reduce unnecessary biopsies for the thyroid nodule diagnosis.

In this study, we further analyzed whether the efficiency of diagnosis of the CAD systems were affected by nodule sizes. It was shown that the diagnostic performance of the CAD system was not consistent in each group depending on the size of the lesion. The sensitivities of the CAD system in identifying small were significantly higher than those of large nodules. These results may be attributed to that large thyroid nodules tend to occupy most of the thyroid gland in the US image, which makes it hard for the CAD system to distinguish between the nodules and the normal thyroid gland. This should be considered when the CAD system is used in clinical practice.

The study contributes to several clinical implications. First, the CAD system in this study can automatically recognize and analyze the thyroid nodules of US images, which demonstrates an opportunity for the combination between clinician and machine in future clinical practice. Second, the CAD system exhibited no statistically significant difference in terms of specificity compared with the senior radiologist, although the sensitivity was lower. This finding implied that the CAD system could cut down unnecessary biopsies and also help to lighten the load of physicians. Besides, the use of the CAD system significantly improved the diagnostic sensitivity and AUC of the junior radiologist, which suggested the possibility that it could serve as a second opinion for less experienced radiologists to minimize missed diagnosis. Lastly, the diagnostic efficiency of the CAD system for thyroid nodules of different sizes was evaluated, which was able to reflect the clinical value of the CAD system further.

This study also has some limitations. First of all, the sample capacity was relatively small and selection bias was inevitable due to the retrospective study nature. Second, the diagnostic criteria for the CAD system-assisted radiologist diagnosis are artificially defined. The actual help of the CAD system for the radiologists in clinical needs to be substantiated in the future. Further, although this study enrolled five pathological types of thyroid nodules, most of the malignant nodules were PTCs. However, the follicular thyroid carcinoma appears with different sonographic characteristics from PTC and tend to show more benign US features (22, 23), which make it difficult for CAD systems to distinguish FTC from thyroid nodules. Large-scale multicenter studies are needed to overcome these drawbacks and generalize the findings.

In conclusion, the CAD system assessed in this study shows comparable specificity to that of the senior radiologist and helps to improve the diagnostic sensitivity and AUC of the junior radiologist significantly. The nodule size of thyroid nodules are potential influencers of CAD diagnostic performance. Further efforts are required to improve its diagnostic performance and future researches are necessary to evaluate the clinical role of CAD in thyroid nodule diagnosis.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The study was reviewed and approved by Ethics Committee of Shanghai Sixth People's Hospital. All procedures performed in the study involving human participants were in accordance with the ethical standards of the institutional research committee and the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Formal consent is not required for this type of study.

AUTHOR CONTRIBUTIONS

YZ and QW were major contributors in writing the manuscript. YZ, QW, and YW conceived and designed the experiments. YZ did the literature research. QW analyzed the data. YC provided basic information on all cases. All authors contributed to the article and approved the submitted version.

FUNDING

This work was funded by National Natural Science Foundation of China (Nos. 81671700 and 81701706), Shanghai Key Clinical Disciplines Fund (Grant No. shslczdzk03203), Shanghai Key Discipline of Medical Imaging (No. 2017ZZ02005). These four parties have provided funding financial support for this research.

REFERENCES

1. Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, 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) 26:1–133. doi: 10.1089/thy.2015.0020
2. Miller KD, Siegel RL, Lin CC, Mariotto AB, Kramer JL, Rowland JH, et al. Cancer treatment and survivorship statistics, 2016. *CA Cancer J Clin.* (2016) 66:271–89.
3. Hun KS, Suk PC, Lyung JS, Joo KB, Young KJ, Jung CJ, et al. Observer variability and the performance between faculties and residents: US criteria for benign and malignant thyroid nodules. *Korean J Radiol.* (2010) 11:149–55. doi: 10.3348/kjr.2010.11.2.149
4. Park CS, Kim SH, Jung SL, Kang BJ, Kim JY, Choi JJ, et al. Observer variability in the sonographic evaluation of thyroid nodules. *J Clin Ultras.* (2010) 38:287–93.
5. Faust O, Acharya UR, Tamura T. Formal design methods for reliable computer-aided diagnosis: a review. *IEEE Rev Biomed Eng.* (2012) 5:15–28. doi: 10.1109/rbme.2012.2184750
6. Lam J, Ying M, Cheung SY, Yeung KH, Yu PH, Cheng HC, et al. A comparison of the diagnostic accuracy and reliability of subjective grading and computer-aided assessment of intranodal vascularity in differentiating metastatic and reactive cervical lymphadenopathy. *Ultraschall Med.* (2016) 37:63–7. doi: 10.1055/s-0034-1384939
7. Chang T-C. The role of computer-aided detection and diagnosis system in the differential diagnosis of thyroid lesions in ultrasonography. *J Med Ultras.* (2015) 23:177–84. doi: 10.1016/j.jmu.2015.10.002
8. Jin YY, Ha Eun J, Cho Yoon J, Kim Hye L, Han M, Kang Y. So, computer-aided diagnosis of thyroid nodules via ultrasonography: initial clinical experience. *Korean J Radiol.* (2018) 19:665–72. doi: 10.3348/kjr.2018.19.4.665
9. Choi YJ, Baek JH, Park HS, Shim WH, Kim TY, Shong YK, et al. System using artificial intelligence for the diagnosis and characterization of thyroid nodules on ultrasound: initial clinical assessment. *Thyroid.* (2017) 27:546–52. doi: 10.1089/thy.2016.0372
10. Wang L, Yang S, Yang S, Zhao C, Tian G, Gao Y, et al. Automatic thyroid nodule recognition and diagnosis in ultrasound imaging with the YOLOv2 neural network. *World J Surg Oncol.* (2019) 17:12. doi: 10.1186/s12957-019-1558-z
11. Grmec S, Gasparovic V. Comparison of APACHE II, MEES and glasgow coma scale in patients with nontraumatic coma for prediction of mortality. Acute physiology and chronic health evaluation. mainz emergency evaluation system. *Critical Care (Lond Engl).* (2001) 5:19–23. doi: 10.1186/cc973
12. Burman KD, Wartofsky L. Clinical practice. Thyroid nodules. *New Engl J Med.* (2015) 373:2347–56. doi: 10.1056/NEJMcp1415786
13. Fisher SB, Perrier ND. The incidental thyroid nodule. *CA Cancer J Clin.* (2018) 68:97–105. doi: 10.3322/caac.21447
14. Acharya UR, Swapna G, Sree SV, Molinari F, Suri JS, Review A. on Ultrasound-based thyroid cancer tissue characterization and automated classification. *Technol Cancer Res Treat.* (2013) 13:289–301. doi: 10.7785/tcrt.2012.500381
15. Takahashi R, Kajikawa Y. Computer-aided diagnosis: a survey with bibliometric analysis. *Int J Med Inform.* (2017) 101:58–67. doi: 10.1016/j.ijmedinf.2017.02.004
16. Lim KJ, Choi CS, Yoon DY, Chang SK, Kim KK, Han H, et al. Computer-aided diagnosis for the differentiation of malignant from benign thyroid nodules on ultrasonography. *Acad Radiol.* (2008) 15:853–8. doi: 10.1016/j.acra.2007.12.022
17. Gao L, Liu R, Jiang Y, Song W, Wang Y, Liu J, et al. Computer-aided system for diagnosing thyroid nodules on ultrasound: a comparison with radiologist-based clinical assessments. *Head Neck.* (2018) 40:778–83. doi: 10.1002/hed.25049
18. Jeong EY, Kim HL, Ha EJ, Park SY, Cho YJ, Han M. Computer-aided diagnosis system for thyroid nodules on ultrasonography: diagnostic performance and reproducibility based on the experience level of operators. *Eur Radiol.* (2019) 29:1978–85. doi: 10.1007/s00330-018-5772-9
19. Reverter JL, Vazquez F, Puig-Domingo M. Diagnostic performance evaluation of a computer-assisted imaging analysis system for ultrasound risk stratification of thyroid nodules. *Am J Roentgenol.* (2019) 19:1–6. doi: 10.2214/ajr.18.20740
20. Chung SR, Baek JH, Lee MK, Ahn Y, Choi YJ, Sung TY, et al. Computer-aided diagnosis system for the evaluation of thyroid nodules on ultrasonography: prospective non-inferiority study according to the experience level of radiologists. *Korean J Radiol.* (2020) 21:369–76. doi: 10.3348/kjr.2019.0581
21. Gitto S, Grassi G, De Angelis C, Monaco CG, Sdao S, Sardanelli F, et al. A computer-aided diagnosis system for the assessment and characterization of low-to-high suspicion thyroid nodules on ultrasound. *La Radiol Med.* (2019) 124:118–25. doi: 10.1007/s11547-018-0942-z
22. Cordes M, Kondrat P, Uder M, Kuwert T, Sasiadek M. Differential diagnostic ultrasound criteria of papillary and follicular carcinomas: a multivariate analysis. *Rofo.* (2014) 186:489–95. doi: 10.1055/s-0034-1366282
23. Park JW, Kim DW, Kim D, Baek JW, Lee YJ, Baek HJ. Korean thyroid imaging reporting and data system features of follicular thyroid adenoma and carcinoma: a single-center study. *Ultrasonography.* (2017) 36:349–54. doi: 10.14366/usg.17020

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Oncology

ORIGINAL RESEARCH
published: 11 September 2020
doi: 10.3389/fonc.2020.01719

Ultrasound Radiomics Effective for Preoperative Identification of True and Pseudo Gallbladder Polyps Based on Spatial and Morphological Features

Hai-xia Yuan [1,2†], *Qi-hui Yu* [3†], *Yan-qun Zhang* [1,2], *Qing Yu* [4], *Qi Zhang* [3,5*] and *Wen-ping Wang* [1,2,4*]

[1] Department of Ultrasound, Zhongshan Hospital of Fudan University, Shanghai, China, [2] Department of Ultrasound, Xiamen Branch, Zhongshan Hospital of Fudan University, Xiamen, China, [3] The SMART (Smart Medicine and AI-based Radiology Technology) Lab, School of Communication and Information Engineering, Shanghai University, Shanghai, China, [4] Shanghai Institute of Medical Imaging, Shanghai, China, [5] Hangzhou YITU Healthcare Technology, Hangzhou, China

Edited by:
Wei Wang,
First Affiliated Hospital of Sun Yat-sen University, China

Reviewed by:
Cai Chang,
Fudan University, China
Baiying Lei,
Shenzhen University, China

***Correspondence:**
Wen-ping Wang
puguang61@126.com
Qi Zhang
zhangq@t.shu.edu

[†] These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *29 February 2020*
Accepted: *31 July 2020*
Published: *11 September 2020*

Citation:
Yuan H-x, Yu Q-h, Zhang Y-q, Yu Q, Zhang Q and Wang W-p (2020) Ultrasound Radiomics Effective for Preoperative Identification of True and Pseudo Gallbladder Polyps Based on Spatial and Morphological Features. Front. Oncol. 10:1719. doi: 10.3389/fonc.2020.01719

Purpose: To explore the value of ultrasound radiomics in the preoperative identification of true and pseudo gallbladder polyps and to evaluate the associated diagnostic accuracy.

Methods: Totally, 99 pathologically proven gallbladder polyps in 96 patients were enrolled, including 58 cholesterol polyps (55 patients) and 41 gallbladder tubular adenomas (41 patients). Features on preoperative ultrasound images, including spatial and morphological features, were acquired for each lesion. Following this, two-stage feature selection was adopted using Fisher's inter-intraclass variance ratios and Z-scores for the selection of intrinsic features important for differential diagnosis achievement with support vector machine use.

Results: Eighty radiomic features were extracted from each polyp. Eight intrinsic features were identified after two-stage selection. The contrast 14 (Cont14) and entropy 6 (Entr6) values in the cholesterol polyp group were significantly higher than those in the gallbladder adenoma group (4.063 ± 1.682 vs. 2.715 ± 1.867, $p < 0.001$ for Cont14; 4.712 ± 0.427 vs. 4.380 ± 0.720, $p = 0.003$ for Entr6); however, the homogeneity 13 (Homo13) and energy 8 (Ener8) values in the cholesterol polyp group were significantly lower (0.500 ± 0.069 vs. 0.572 ± 0.057, $p < 0.001$ for Homo13; 0.050 ± 0.023 vs. 0.068 ± 0.038, $p = 0.002$ for Ener8). These results indicate that the pixel distribution of cholesterol polyps was more uneven than that of gallbladder tubular adenomas. The dispersion degree was also significantly lower in the cholesterol polyp group than the gallbladder adenoma group (0.579 ± 0.054 vs. 0.608 ± 0.041, $p = 0.005$), indicating a lower dispersion of high-intensity areas in the cholesterol polyps. The long axis length of the fitting ellipse (Maj.Len), diameter of a circle equal to the lesion area (Eq.Dia) and perimeter (Per) values in the cholesterol polyp group were significantly lower than those in the gallbladder adenoma group (0.971 ± 0.485 vs. 1.738 ± 0.912, $p < 0.001$ for Maj.Len; 0.818 ± 0.393 vs. 1.438 ± 0.650, $p < 0.001$ for Eq.Dia; 2.637 ± 1.281 vs. 5.033 ± 2.353, $p < 0.001$ for Per), demonstrating that the cholesterol polyps were smaller and more regular in terms of morphology. The classification accuracy, sensitivity, specificity, and area under the curve values were 0.875, 0.885, 0.857, and 0.898, respectively.

Conclusions: Ultrasound radiomic analysis based on the spatial and morphological features extracted from ultrasound images effectively contributed to the preoperative diagnosis of true and pseudo gallbladder polyps and may be valuable in their clinical management.

Keywords: gallbladder true-polyps, gallbladder pseudo-polyps, ultrasound radiomics, gallbladder cholesterol polyp, gallbladder adenoma, preoperative identification

INTRODUCTION

With the development of high-resolution ultrasound equipment and increased frequency of periodic health examinations, numerous gallbladder polyps are now diagnosed at an early phase. Although the reported incidence rate in adults is∼0.3–12.3%, only about 5% of polyps are true polyps (1, 2). Postoperative pathological gallbladder polyp types include cholesterol polyps, inflammatory polyps, adenomyomas, adenomas, and early gallbladder cancer. Gallbladder cholesterol polyps and gallbladder adenoma polyps are the two most commonly observed types and are associated with different clinical procedures. Gallbladder cholesterol polyps are a type of pseudo polyps and are usually caused by the accumulation of cholesterol crystals in the inner wall of the gallbladder that are swallowed by macrophages. This subsequently promotes the formation of foam cells at the surface of the gallbladder mucosa; most cholesterol polyps tend to remain in a benign state (3, 4). Inversely, gallbladder adenomas are true polyps and usually coexist with atypical hyperplasia; they tend to progress to gallbladder cancer (5, 6). Therefore, the preoperative identification of gallbladder true polyps is vital.

At present, the accurate identification of the aforementioned polyps before cholecystomy using the existing imaging techniques is extremely challenging. Ultrasonography is the preferred imaging method owing to its characteristics that include radiation absence, clear imaging and scanning section flexibility. However, few studies have focused specifically on how gallbladder cholesterol polyps and adenomas can be distinguished from each other. Park et al. (7) found different types of gallbladder adenomas and cholesterol polyps in progression by the application of endoscopic ultrasound, the use of which is limited in clinical practice due to its invasiveness. With the use of contrast-enhanced ultrasound, our previous study (8) revealed that gallbladder adenomas exhibit uniformly eccentric enhanced characteristics and slower regression compared to gallbladder cancer. However, as some hospitals do not use contrast-enhanced ultrasound, the distinction of gallbladder adenomas from cholesterol polys is a tremendous challenge for radiologists. Further reliable and objective methods are needed for a larger number of imaging features to be obtained for differential diagnosis.

Nowadays, surgical guidelines recommend that gallbladder polyps of size >1 cm be surgically resected as gallbladder adenomas and carcinomas are larger than benign polyps (9). However, of 1,541 cases of gallbladder polyps investigated in our hospital from January 2011 to November 2018, only ∼30% of gallbladder polyps were pathologically proven as being gallbladder adenomas, adenomas with severe atypia, or cancerous adenomas, indicating that the remaining 70% were pseudo gallbladder polyps, including cholesterol polyps, adenoma-like hyperplasia, and inflammatory polyps. Therefore, there is an urgent need for clinical surgery aimed at the identification of a novel imaging method with higher diagnostic accuracy that may allow for the avoidance of unnecessary cholecystectomy, reduce the wastage of medical resources, and relieve patient suffering.

The field of radiomic technology based on artificial intelligence (AI) has been developing rapidly in recent years, with computers processing massive datasets through layered mathematical models that can detect patterns not otherwise decipherable using biostatistics (10). Many researchers have made progress in the field of radiomics. Wang et al. (11) showed that the newly developed deep learning radiomics of elastography (DLRE) was valuable in liver fibrosis stage prediction. Liu et al. (12) developed a radiomics model that incorporated radiomics signatures and independent clinicopathological risk factors, that allowed for the performance of the individualized, non-invasive prediction of pathologic complete response to neoadjuvant chemoradiotherapy in patients with locally advanced rectal cancer. Song et al. (13) demonstrated the individualized prediction of progression-free survival probability associated with epidermal growth factor receptor tyrosine kinase inhibitor therapy in non-small cell lung cancer on the basis of computed tomography features. In general, AI is widely used in the field of medical radiomics analysis, with computers capturing changes in the protein genes on macroscopic images using information of a higher dimension. This is expected to provide accurate and reliable diagnostic recommendations for doctors' clinical decisions (14–17).

Abbreviations: GLCM, Gray-level co-occurrence matrix; IMean, Mean of the pixels within the lesion; IMedian, Median of the pixels within the lesion; HE, Histogram entropy; RImedian, Corresponding ratio of the median of the pixels; RImean, Corresponding ratio of the mean of the pixels; Ener, Energy; Cont, Contrast; Entr, Entropy; Homo, Homogeneity; AR, Area ratio; CDD, Center deviation degree; DD, Dispersion degree; Area, Area of the lesion; C.area, Area of minimum convex polygon; Maj.Len, Long axis length of the fitting ellipse; Min.Len, Short axis length of the fitting ellipse; Per, Perimeter; Ori, Orientation; Eq.Dia, Diameter of a circle equal to the lesion area; Sol, Solidity; AUC, Area under the curve; SVM, Support vector machine; AI, Artificial intelligence.

In this study, we aimed to retrospectively analyse the preoperative two-dimensional ultrasound images of patients with gallbladder adenomas and gallbladder cholesterol polyps. Multiple groups of imaging features were extracted automatically for the detection of early imaging differences between the two diseases so as to provide accurate diagnosis.

MATERIALS AND METHODS

Patients

Approval was obtained from the Institutional Ethics Committee for the retrospective review of images and patients' medical records(Y2020-188).

The exclusion criteria were as follows: (1) diagnosis of gallbladder carcinoma on previous imaging; (2) insufficient liver, kidney or heart function; (3) presence of a thickened gallbladder wall lesion; and (4) imaging scanning demonstrated liver metastasis.

From July 2018 to December 2019, 263 patients with gallbladder polyps (size >7 mm) were referred to our hospital for surgical treatment (**Figure 1**), and all of them underwent ultrasound. After a discussion with their surgeons, 152 cases chose clinical follow-up, and 111 underwent cholecystectomy. The polyps were pathologically proven as being cholesterol polyps ($n = 58$) in 55 patients and gallbladder tubular adenomas ($n = 41$) in 41 patients. Gallbladder polypoid adenocarcinomas ($n = 8$), inflammation polyps ($n = 4$), and adenomyomas ($n = 3$) were also observed. Patients with polypoid adenocarcinomas, inflammation polyps and adenomyomas were excluded from this study owing to the small sample size. Finally, we enrolled 99 gallbladder polyps (cholesterol polyps and adenomas) in 96 patients (40 men and 56 women; mean age 36.5 years, age range 27–71 years).

Ultrasound Scanning and Instruments

All patients fasted for at least 8 h before undergoing ultrasound examination. Gray-scale and color Doppler ultrasound were performed. The target area was magnified to ensure the ideal plane for the display of the whole gallbladder and adjacent liver parenchyma. Ultrasound was performed by two experienced technologists using one of the following ultrasonographic systems: Aplio 500 (Canon Healthcare, Japan; PVT-375BT, 1.9–6 MHz), Ascendus (Hitachi Medical Systems, Japan; EUP-C715, 1–5 MHz), Resona 7s (Mindray Medical Systems, China; SC5-1U, 1–5 MHz), and Mylab Twice (Esaote Medical Systems, Italy, CA431, 1–5 MHz). The maximum diameter of the polyp was measured, and the original ultrasound images of the lesion were captured for further analysis.

Ultrasound Radiomics Analysis Procedure

Overall Design

The radiomic analysis based on ultrasound images comprised seven steps, as shown in **Figure 2**.

Image Processing

In this retrospective study, the ultrasound images of cholesterol polyps and gallbladder adenomas were acquired, and the edge of each lesion was circled with a red curve by the drawing software (**Figure 3A**). Then, the images were binarised with the thresholding method to obtain mask images on which the outline was filled with white inside and the rest set to black. The area (orange rectangle) showing the gallbladder polyp was zoomed partially (**Figure 3B**), and the mask of the gallbladder polyp lesion was shown as in **Figure 3C**.

Spatial Feature Extraction

The imaging features of the lesion were extracted based on the original ultrasound image and the corresponding mask image. Some spatial features were extracted based on the ultrasonic gray-scale image, which included first-order statistic features and gray-level co-occurrence matrix (GLCM) texture features. Additionally, binary texture spatial features were extracted based on the ultrasonic binary mask image for the reflection of pixel distribution inside the lesion.

The first-order statistic features included the mean (IMean), median (IMedian), standard deviation, coefficient of variation, histogram entropy, skewness, and kurtosis of the pixels within the lesion. The corresponding ratio of the median (mean) of the pixels was calculated, which was within the lesion and within the reference area (the rectangular area expanding outwards from the lesion), and the ratio was defined as RImedian (RImean).

The GLCM is an important technique for texture analysis (18), which represents the characteristics of the intensity distribution and respective distance of the intensity levels in the original image. In this study, the GLCM texture features were of four types: energy (Ener), contrast (Cont), entropy (Entr), and homogeneity (Homo), and each type of GLCM feature was constructed for different values of offset d. Here, d was an integer between 1 and 15 pixels. Therefore, each type of GLCM feature included 15 texture features; a total of 60 GLCM texture features was extracted for each lesion.

The binary texture features included the following: the area ratio (AR), which denotes the ratio of the high-intensity area to the whole lesion area; center deviation degree, which characterizes the normalized distance between each pixel point in the high-intensity area of the lesion and the center point of the lesion; and dispersion degree (DD), which characterizes the mean of the normalized Euclidean distance between each pixel point in the high-intensity area of the lesion and the center point of the high-intensity area (19).

Morphological Feature Extraction

As shown in **Figure 4**, the morphological features of the lesion were extracted, including the area of the lesion (Area), area of the minimum convex polygon corresponding to the lesion (C.area), long axis length (Maj.Len), and short axis length (Min.Len) of the fitting ellipse with the same standard second order center distance as the lesion, number of contour pixel points of the lesion (perimeter, Per), angle between the long axis of the fitting ellipse and X-axis (orientation, Ori), diameter of a circle equal to the lesion area (equivalent diameter, Eq.Dia), and ratio of the lesion area to convex area (solidity, Sol).

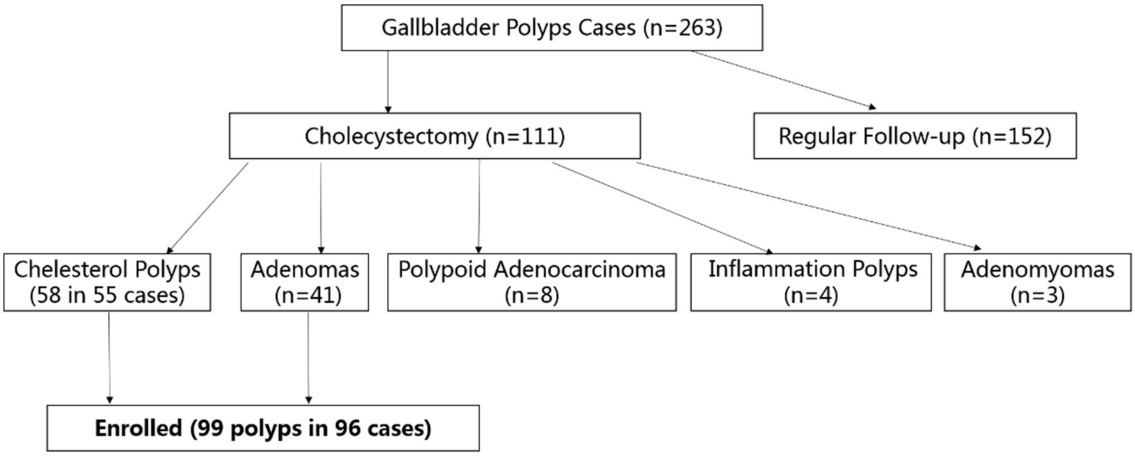


FIGURE 1 | Flow chart of participant enrolment. Totally, 263 patients with gallbladder polyps (size >7 mm) were referred to our hospital: 152 cases chose clinical follow-up, 111 underwent cholecystectomy for pathological proven cholesterol polyps ($n = 58$), gallbladder tubular adenomas ($n = 41$), gallbladder polypoid adenocarcinomas ($n = 8$), inflammation polyps ($n = 4$), and adenomyomas ($n = 3$). Finally, 99 gallbladder polyps (cholesterol polyps and adenomas) in 96 patients were enrolled.

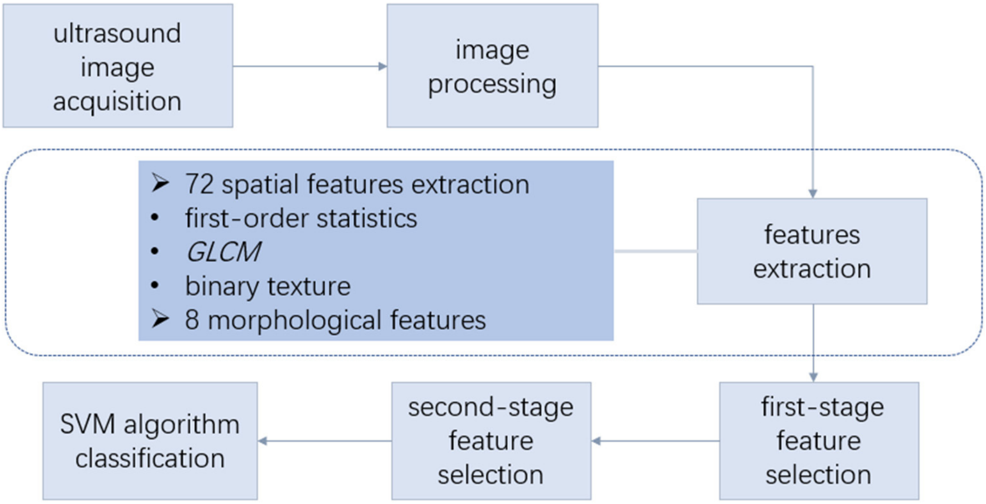


FIGURE 2 | Overall design of radiomic analysis based on ultrasound images. The radiomic analysis comprised the following steps: (1) acquisition of the ultrasound images and lesion contour delineation; (2) image processing to obtain the mask image; (3) spatial feature extraction based on the ultrasonic gray-scale and binary image; (4) morphological feature extraction; (5) first-stage feature selection; (6) second-stage feature selection; (7) support vector machine (SVM) classification using the selected features. GLCM, gray-level co-occurrence matrix.

First-Stage Feature Selection

Assuming the features were normally distributed, the non-paired t-test was used to analyse the features of the gallbladder cholesterol polyps and gallbladder tubular adenomas. Otherwise, the Kruskal–Wallis test was used to analyse the features. Here, p-values lower than 0.05 indicated statistical significance.

In order to further enhance the reliability of the features and select intrinsic features from among all the significant features, we adopted two indicators—Fv and F values. Here, the Fv-value was Fisher's inter-intraclass variance ratio and F-value was defined using Z-scores (20):

$$F_v = \frac{|\bar{x}_0 - \bar{x}_1|}{\sqrt{(\sigma_0^2 + \sigma_1^2)}} \tag{1}$$

$$\begin{aligned} f &= \left| mean_i(z\text{-}score_{oi}) - mean_i(z\text{-}score_{1i}) \right| \\ &= \left| mean_i\left(\frac{x_{0i} - \bar{x}_0}{\sigma_0}\right) - mean_i\left(\frac{x_{1i} - \bar{x}_1}{\sigma_1}\right) \right| \end{aligned} \tag{2}$$

where the subscripts 0 and 1 represented the gallbladder cholesterol polyps and gallbladder tubular adenomas, respectively, $\bar{x}$ and σ denoted the mean and standard

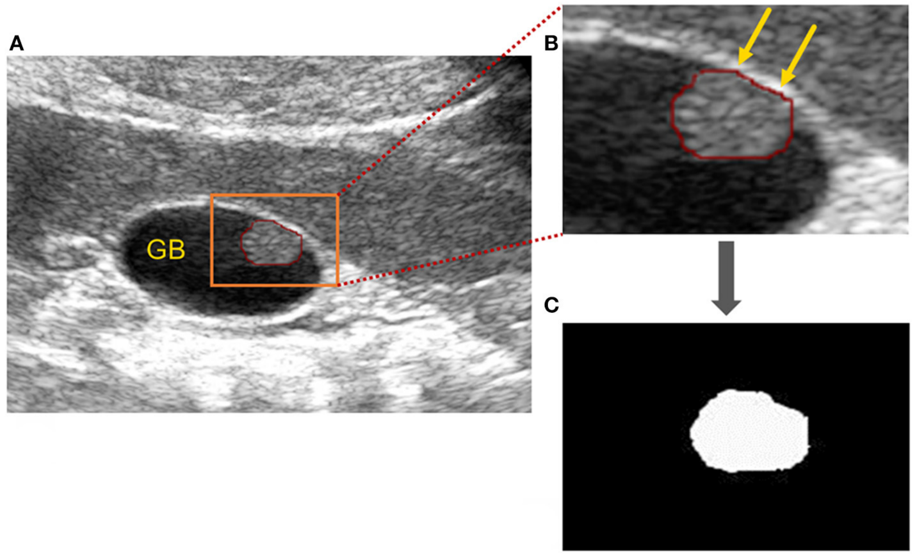


FIGURE 3 | Image processing of gallbladder cholesterol polyps. **(A)** Original ultrasound image of a gallbladder polyp: the polyp is circled (red line) and the area (orange rectangle) showing the gallbladder polyp) has been zoomed in partially to **(B)**. **(B)** Partially enlarged gallbladder polyp (orange arrows). **(C)** Mask of gallbladder polyp after binarisation processing.

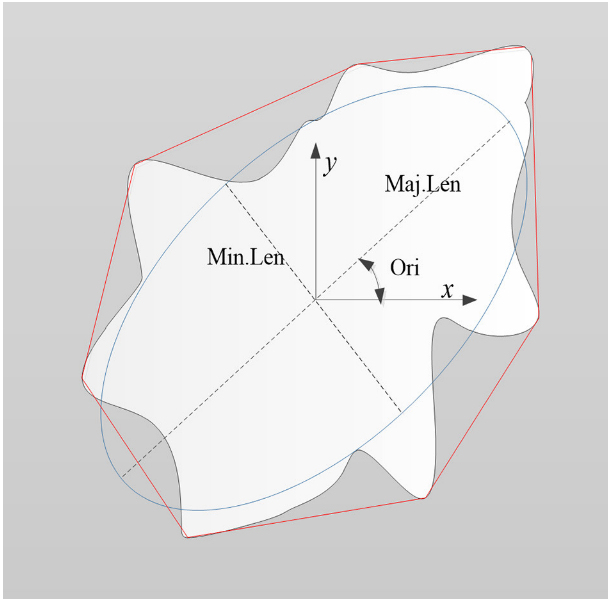


FIGURE 4 | Schematic diagram illustrating the morphological features of a lesion (denoted as the white region). Features included the area of the lesion (Area), area of minimum convex polygon encompassing the lesion (C.area; denoted as the area inside the red line), long axis length (Maj.Len), and short axis length (Min.Len) of the fitting ellipse (blue line), perimeter (Per), orientation (Ori), equivalent diameter (Eq.Dia), and solidity (Sol).

deviation of a feature. X_0i represented the i-th data of a feature in class 0, and X_1i represented the i-th data of a feature in class 1.

Considering the presence of a large number of GLCM features (60) and the likelihood of internal redundancy, we first retained the feature with the largest *Fv* value in each type of GLCM feature as the representative feature of GLCM. In addition to the GLCM features, other statistically significant features were retained.

Second-Stage Feature Selection

Next, we selected a few more important features after first-stage feature selection. In terms of spatial features, we retained the features that satisfied both the following criteria: (1) *Fv* value was greater than the median *Fv* value of the alternative spatial domain features. (2) F value was greater than the median corresponding F value of the alternative spatial domain features. Similarly, in terms of morphological features, we retained the features that satisfied both the following criteria: (1) *Fv* value was greater than the median *Fv* value of the alternative morphological features. (2) F value was greater than the median corresponding F value of the alternative morphological features.

Classification

For the classification of features, we used the supervised support vector machine (SVM) algorithm. The SVM is used for the identification of a decision boundary to maximize the margin between two classes and is a very popular classification method (21). First, we divided the data set into the training set and test set in a ratio of 6:4. In the training set, we used 5-fold cross validation for the identification of the optimal model of the features, which was then used for the test set classification. Finally, we acquired the classification performance of the test set, including the classification accuracy, classification sensitivity, specificity, Youden index, and area under the curve (AUC).

RESULTS

Features After First-Stage Feature Selection

The enrolled cases were confirmed by surgical pathology, and included 58 cases of cholesterol polyps in 55 patients and 41 cases of gallbladder tubular adenomas in 41 patients. Each case corresponded to 72 spatial features and eight morphological features. Finally, 69 significant features were obtained from among all the features of the two diseases, including 52 GLCM features. The spatial and morphological features obtained after first-stage feature selection are shown in **Table 1**, in which, if a feature was normally distributed, its mean and standard deviation are shown, otherwise its median and interquartile range are given.

Features After Second-Stage Feature Selection

In the second-stage feature selection, we retained the spatial features that satisfied $Fv>=0.396$ and $F>=0.564$. Similarly, we retained the morphological features that satisfied $Fv>=0.740$ and $F>=0.964$. Finally, a total of eight features was selected, as shown in **Table 2**.

As **Table 2** indicates, in terms of spatial features, the Cont14 and Entr6 values in the cholesterol polyp group were significantly higher than those in the gallbladder adenoma group, but the Homo13 and Ener8 values in the cholesterol polyp group were significantly lower than those in the gallbladder adenoma group. These results indicate that the pixel distribution of the cholesterol polyp lesions was more uneven than that of the gallbladder tubular adenomas. The DD was also significantly lower in the cholesterol polyps than gallbladder adenomas, indicating a lower degree of dispersion of the highlight area in the cholesterol polyps. In addition, in terms of morphological characteristics, the Maj.Len, Eq.Dia and Per values in the cholesterol polyp group were significantly lower than those in the gallbladder adenoma group, demonstrating that the cholesterol polyps were smaller and more regular in appearance than the gallbladder tubular adenomas.

As shown in **Figure 5**, it is very hard to manually and visually distinguish gallbladder adenomas (**Figures 5C,D**) from cholesterol polyps (**Figures 5A,B**) based on their ultrasound images. Using radiomic analysis, the Cont14 value was found to be significantly higher (3.850 and 2.387) than that of the gallbladder adenomas (1.460 and 1.898). These results indicate that the pixel distribution of the cholesterol polyp lesions was more uneven than that of the gallbladder tubular adenomas, and that ultrasound radiomics based on spatial and morphological features may be valuable for the differential diagnosis of these two diseases.

Classification Results of SVM

Finally, we used the SVM to obtain the optimal models of five spatial features and three morphological features. The classification performance in the test set (**Table 3**) indicated that the accuracy of the spatial feature model was higher than that of the morphological feature model, but the sensitivity and specificity the spatial feature model were more unbalanced than those of the morphological feature model. When we applied the SVM to all eight features for the classification performance of the test set, the accuracy, sensitivity, and specificity of the model including all features increased to 0.875, 0.885, and 0.857, respectively. Additionally, while comparing the AUC values between the SVM models including three morphological features, five spatial features and all eight features, the AUC of the all features model (0.898) was the highest, while that of the spatial feature model (0.886) was higher than the AUC of the morphological feature model (0.862) (**Figure 6**).

DISCUSSION

In this study, we demonstrated that ultrasound radiomics analysis, based on the spatial and morphological features extracted from ultrasound images, effectively contributed to the preoperative diagnosis of true and pseudo gallbladder polyps, and may be valuable in the clinical management of gallbladder polyps.

For true gallbladder polyps, cholecystectomy is indeed required for the prevention of malignancy development. The 5-year survival rate associated with gallbladder cancer is 2∼80%, which is closely correlated to the stage of gallbladder cancer in surgery. The 5-year survival rate of gallbladder carcinoma *in situ* is as high as 80%, while it decreases to 8% in cases with lymph node metastasis, and even drops to values as low as 2% in stage 4b gallbladder cancer (22). Therefore, it is of significance to improve the diagnostic accuracy of gallbladder cancer or precancerous lesions at an early stage.

Recently, several imaging methods have been applied in the examination of gallbladder tumors, such as transabdominal ultrasound, high-frequency ultrasound, contrast-enhanced ultrasound, endoscopic ultrasound, enhanced computed tomography, and enhanced magnetic resonance imaging. As the preferred imaging method for gallbladder lesion examination, conventional trans-abdominal ultrasound is widely used in different levels of hospitals for gallbladder polyp screening and follow-up. However, it is unreliable to distinguish true and pseudo polyps only based on the results of lesion echo, morphology, and blood flow obtained by conventional trans-abdominal ultrasound. Compared to traditional low-frequency ultrasound scans, high-frequency ultrasound scans greatly heighten the accuracy of the determination of the preoperative stage of gallbladder cancer as well as differentiating benign and malignant lesions (23, 24). However, an obvious limitation of high-frequency ultrasound is that it is not effective when the polyps are located deep within the gallbladder body or neck. Moreover, due to the low resolution, contrast-enhanced computed tomography, and enhanced magnetic resonance imaging too do not provide satisfying results in terms of true gallbladder polyp diagnosis.

Owing to the significantly high potential of malignancy development in larger polyps, clinical surgery guidelines highly recommend the performance of cholecystectomy in cases

TABLE 1 | Features showing statistical significance.

	Features	Gallbladder cholesterol polyps (Class 0)	Gallbladder tubular adenomas (Class 1)	*P*	*Fv*	*F*
Spatial features	CDD	0.614 ± 0.052	0.639 ± 0.031	0.007	0.413	0.545
	DD	0.579 ± 0.054	0.608 ± 0.041	0.005	0.421	0.564
	AR	0.536 ± 0.119	0.580 ± 0.090	0.046	0.298	0.408
	Imedian	88.474 ± 24.727	104.256 ± 31.886	0.007	0.391	0.550
	Imean	88.237 ± 24.154	102.897 ± 30.604	0.009	0.376	0.530
	CoV	0.307 ± 0.105*	0.263 ± 0.084	0.001	0.322	0.674
	Kurtosis	2.858 ± 0.858*	3.279 ± 1.051*	0.012	0.310	0.235
	Cont14	4.063 ± 1.682*	2.715 ± 1.867	<0.001	0.536	0.883
	Ener8	0.050 ± 0.023*	0.068 ± 0.038*	0.002	0.414	0.583
	Homo13	0.500 ± 0.069	0.572 ± 0.057	<0.001	0.796	0.980
	Entr6	4.712 ± 0.427*	4.380 ± 0.720	0.003	0.396	0.564
Morphological features	Area	0.525 ± 0.489*	1.623 ± 1.405*	<0.001	0.738	0.883
	Maj.Len	0.971 ± 0.485*	1.738 ± 0.912*	<0.001	0.742	1.045
	Min.Len	0.651 ± 0.298*	1.135 ± 0.602*	<0.001	0.720	1.117
	C.Area	0.537 ± 0.511*	1.707 ± 1.423*	<0.001	0.774	0.882
	Eq.Dia	0.818 ± 0.393	1.438 ± 0.650*	<0.001	0.816	1.131
	Ori	15.261 ± 41.013*	33.606 ± 44.117	0.002	0.305	0.562
	Per	2.637 ± 1.281*	5.033 ± 2.353*	<0.001	0.894	1.124
	Sol	0.980 ± 0.030*	0.963 ± 0.053*	0.005	0.273	0.283

**Parameters with non-normal distribution.*

CDD, center deviation degree; DD, dispersion degree; CoV, coefficient of variance; C.area, area of minimum convex polygon; Maj.Len, long axis length of the fitting ellipse; Min.Len, short axis length of the fitting ellipse; Per, perimeter; Ori, orientation; Eq.Dia, diameter of a circle equal to the lesion area; Sol, solidity; AR, area ratio; IMean, mean of the pixels within the lesion; Imedian, median of the pixels within the lesion; Cont, Contrast; Ener, energy; Homo, homogeneity; Entr, entropy.

TABLE 2 | Features filtered using the *Fv* value and *F* value.

	Features	Gallbladder cholesterol polyps	Gallbladder tubular adenomas	*P*	*Fv*	*F*
Spatial features	Cont14	4.063 ± 1.682*	2.715 ± 1.867	<0.001	0.536	0.883
	Ener8	0.050 ± 0.023*	0.068 ± 0.038*	0.002	0.414	0.583
	Homo13	0.500 ± 0.069	0.572 ± 0.057	<0.001	0.796	0.980
	Entr6	4.712 ± 0.427*	4.380 ± 0.720	0.003	0.396	0.564
	DD	0.579 ± 0.054	0.608 ± 0.041	0.005	0.421	0.564
Morphological features	Maj.Len	0.971 ± 0.485*	1.738 ± 0.912*	<0.001	0.742	1.045
	Eq.Dia	0.818 ± 0.393	1.438 ± 0.650*	<0.001	0.816	1.131
	Per	2.637 ± 1.281*	5.033 ± 2.353*	<0.001	0.894	1.124

DD, dispersion degree; Maj.Len, long axis length of the fitting ellipse; Per, perimeter; Eq.Dia, diameter of a circle equal to the lesion area; Cont, Contrast; Ener, energy; Homo, homogeneity; Entr, entropy.

with a gallbladder polyp diameter >1 cm (25). However, this recommendation is being questioned by a growing number of scholars and clinical doctors, with their concerns predominantly centring on the fact that many pseudo non-cancerous gallbladder polyps have a diameter larger than 1 cm and that cholecystectomy performance in such cases may lead to injury and huge wastage of the health system resources. Meanwhile, it has been deemed unreasonable to "watch" the growth of malignant polyps with atypical hyperplasia that have diameters smaller than 1 cm (i.e., 6∼10 mm) by ultrasound in the early phase (26–28). Therefore, there is a need for a larger number of studies focusing on the development of new imaging methods to distinguish such true gallbladder polyps for the performance of cholecystectomy as early as possible, as well as efficiently increase the 5-year survival rate of patients and reduce public health resource wastage.

As a medical research hot spot, AI technology is now being applied in medical imaging. In particular, the use of AI in magnetic resonance imaging has proven successful in terms of pathological slide reading (29–32). With the use of computer-based big data analysis, hundreds of unbiased data of image features from existing images can be obtained in a reasonable

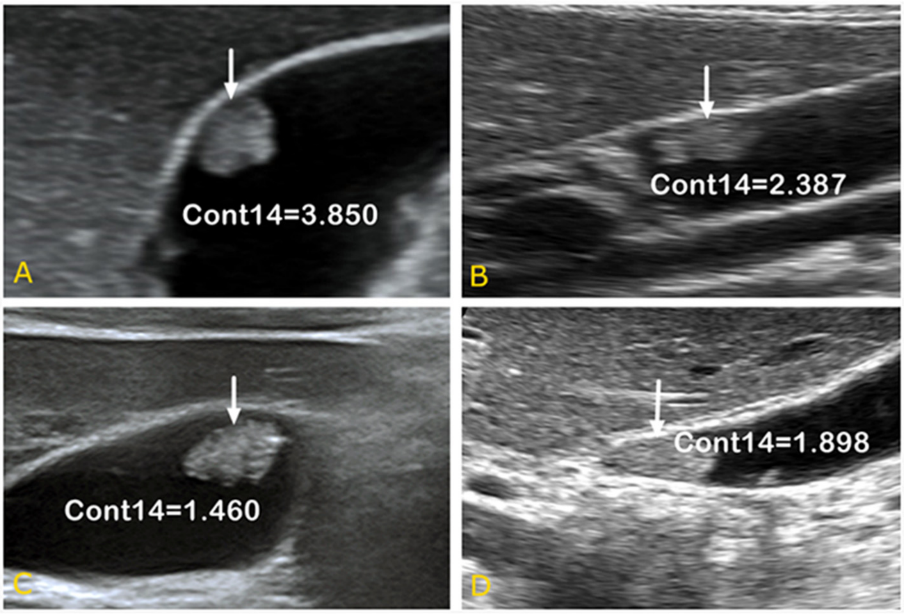


FIGURE 5 | Ultrasound images of gallbladder cholesterol polyps **(A,B)** and gallbladder tubular adenomas **(C,D)**. When combined with spatial parameter analysis, the Cont14 values were significantly higher (3.850 and 2.387) than those of the gallbladder adenomas (1.460 and 1.898), indicating the pixel distribution of the cholesterol polyp lesions was more uneven than that of the gallbladder tubular adenomas. Cont, contrast.

TABLE 3 | Classification performance on the test set using SVM.

Features	Acc	Sen	Spc	Yi	AUC
All features	0.875	0.885	0.857	0.742	0.898
Morphological features	0.825	0.826	0.824	0.650	0.862
Spatial features	0.850	0.864	0.833	0.697	0.886

SVM, support vector machine; Acc, accuracy; Sen, sensitivity; Spc, specificity; Yi, Youden index; AUC, area under the curve.

span of time. With a resolution that far exceeds that of the human eyes, the characteristics for the differentiation of benign and malignant polyps can be obtained by computers from the analysis results of a large number of cases, which can further be used to train computers for deep learning.

In our current study, computer aided high-throughput imaging analysis was applied for the analysis of the medical images of the 99 gallbladder polyps. According to the existing literature, gallbladder cholesterol polyps and adenomas display different patterns of echoes, as obtained by endoscopic ultrasonography (7, 33, 34). Impressively, we found that compared to gallbladder adenomas, cholesterol polyps exhibit a greater degree of unevenness in terms of the pixel distribution of the lesion area and higher aggregability of the highlight area. Meanwhile, our results revealed that the cholesteric polyps exhibited smaller lesion area perimeters and showed greater regularity than the gallbladder tubular adenomas. Particularly, these imaging features of gallbladder cholesterol polyps are closely correlated to their pathophysiological characteristics. Due to the cholesterol crystals in foam cells (3, 6, 35), the images of cholesterol polyps by conventional ultrasound usually show point-like strong echoes or high echoes. For small-size polyps,

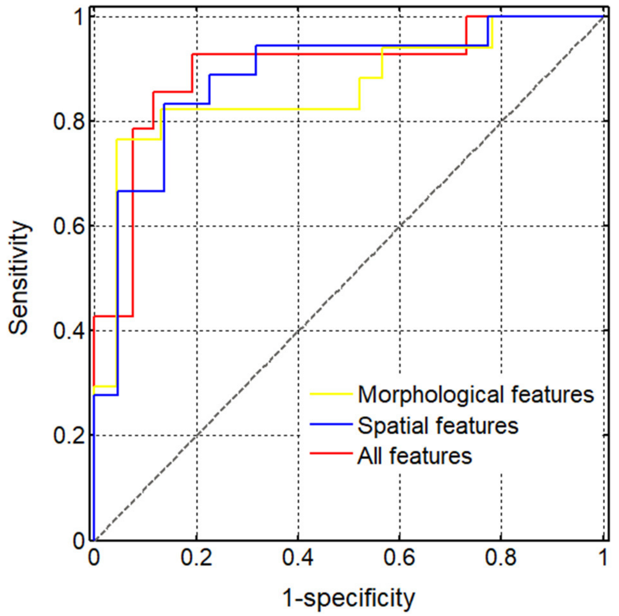


FIGURE 6 | Diagnostic performance of different support vector machine (SVM) models. The (areas under the curve) AUCs obtained using the SVM models of various feature sets were compared, including three morphological features (0.862), five spatial features (0.886), and all eight selected features (0.898).

these echoes are too weak for their detection by the human eye, but may be well-obtained by computers, which have a greater sensitivity. In contrast to the echoes of cholesterol polyps, those of gallbladder adenomas are more uniform in nature as a result of the smaller surface area and similar acoustic impendence inside the adenoma that comprises proliferating glandular epithelial cells and mesenchymal cells. With real-time harmonic contrast ultrasound use, small focal areas of non-enhancement within the peaks could be detected in gallbladder cholesterol polyps, while the enhancements within the peaks usually showed more uniformity in adenomas (8). Strikingly, these newly revealed features are consistent with the pathological characteristics of the lesion. Additionally, these features have potential classification ability. The present study also demonstrated that compared to cholesterol polyps, gallbladder adenomas have a relatively larger girth and volume and show greater shape-related irregularity, consistent with previous reports. Although statistically, the diameter of adenoma is significantly larger than that of cholesterol polyps, for individual cases, we cannot accurately determine true or false polyps by the size of the lesions. In cases with a lesion size of 1 cm with similar echo appearance, it was extremely difficult for the radiologist to provide a pathological diagnosis using conventional ultrasound. However, when combined with AI analysis, including potential morphological and spatial features, a higher diagnostic accuracy in distinguishing true and pseudo gallbladder polyps could be achieved.

Our study also have some limitations. As an initial attempt aimed at the application of up-to-date radiomics technology to distinguishing true and pseudo polyps in the gallbladder, we did not collect a large number of cases. In our following studies, the sample size will be expanded, and deep learning will further be performed on various ultrasound instruments, to provide more promising and reliable parameters for clinical diagnosis. Moreover, we will also attempt to introduce radiomics to the study susing multi-modal ultrasound to obtain more novel indicators. Moreover, combining the automated radiomics technique with the traditional 2D image descriptors, assessed visually by radiologists, could integrate more useful information, which may contribute to more accurate differential diagnosis and deserves further study.

CONCLUSION

Ultrasound radiomic analysis based on the spatial and morphology features of original ultrasound images could effectively improve the preoperative diagnostic ability of true and pseudo gallbladder polyps, which may inform gallbladder polyp procedure-related decision-making. Compared to gallbladder adenomas, gallbladder cholesterol polyps showed a greater degree of unevenness and the highlight area showed a higher degree of clustering; these characteristics can be useful in the performance of differential diagnosis in such settings.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/supplementary material.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Institutional Ethics Committee of Fudan University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

Each author of the manuscript participated in the study and approved the manuscript for submission. WW and QZ designed the study and guided the data analysis. HY and QY performed the ultrasound scanning and image analysis. YZ and QY performed the data collection and analysis. HY and Q-hY contributed to the manuscript drafting as well as critical revision and editing. All authors approved the final version.

FUNDING

This study was supported by the National Natural Science Foundation of China (Nos. 61671281 and 61911530249 to QZ), Shanghai Municipal Key Clinical Specialty (shslczdzk03501 to WW), Fujian Province for Health and Science Research Project (2019-ZQNB-39 to HY), and a program from Xiamen Science and Technology Plan (3502Z20184002 to HY).

REFERENCES

1. Mccain RS, Diamond A, Jones C, Coleman HG. Current practices and future prospects for the management of gallbladder polyps: a topical review. *World J Gastroenterol*. (2018) 24:2844–52. doi: 10.3748/wjg.v24.i26.2844
2. Sahiner IT, Dolapci M. When should gallbladder polyps be treated surgically? *Adv Clin Exp Med*. (2018) 27:1697–700. doi: 10.17219/acem/75678
3. Oestmann A. [Polyps of the gallbladder]. *Praxis*. (2012) 101:581–4. doi: 10.1024/1661-8157/a000919
4. Limaiem F, Sassi A, Talbi G, Bouraoui S, Mzabi S. Routine histopathological study of cholecystectomy specimens. Useful? A retrospective study of 1960 cases. *Acta Gastroenterol Belg*. (2017) 80:365–70.
5. Guettier C. [Pathology of gallbladder and extrahepatic bile ducts. Case 7. Biliary-type tubulopapillary adenoma without dysplasia]. *Ann Pathol*. (2014) 34:315–23. doi: 10.1016/j.annpat.2014.06.001
6. Kai K, Aishima S, Miyazaki K. Gallbladder cancer: clinical and pathological approach. *World J Clin Cases*. (2014) 2:515–21. doi: 10.12998/wjcc.v2.i10.515
7. Park CH, Chung MJ, Oh TG, Park JY, Bang S, Park SW, et al. Differential diagnosis between gallbladder adenomas and cholesterol polyps on contrast-enhanced harmonic endoscopic ultrasonography. *Surg Endosc*. (2013) 27:1414–21. doi: 10.1007/s00464-012-2620-x
8. Yuan HX, Cao JY, Kong WT, Xia HS, Wang X, Wang WP. Contrast-enhanced ultrasound in diagnosis of gallbladder adenoma. *Hepatobiliary Pancreat Dis Int*. (2015) 14:201–7. doi: 10.1016/S1499-3872(15)60351-4
9. Wennmacker SZ, Van Dijk AH, Raessens JHJ, Van Laarhoven C, Drenth JPH, De Reuver PR, et al. Polyp size of 1 cm is insufficient to discriminate neoplastic

and non-neoplastic gallbladder polyps. *Surg Endosc*. (2019) 33:1564–71. doi: 10.1007/s00464-018-6444-1

10. Miller DD, Brown EW. Artificial intelligence in medical practice: the question to the answer? *Am J Med*. (2018) 131:129–33. doi: 10.1016/j.amjmed.2017.10.035
11. Wang K, Lu X, Zhou H, Gao Y, Zheng J, Tong M, et al. Deep learning radiomics of shear wave elastography significantly improved diagnostic performance for assessing liver fibrosis in chronic hepatitis B: a prospective multicentre study. *Gut*. (2019) 68:729–41. doi: 10.1136/gutjnl-2018-316204
12. Liu Z, Zhang XY, Shi YJ, Wang L, Zhu HT, Tang Z, et al. Radiomics analysis for evaluation of pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer. *Clin Cancer Res*. (2017) 23:7253–62. doi: 10.1158/1078-0432.CCR-17-1038
13. Song J, Shi J, Dong D, Fang M, Zhong W, Wang K, et al. A new approach to predict progression-free survival in stage IV EGFR-mutant NSCLC patients with EGFR-TKI therapy. *Clin Cancer Res*. (2018) 24:3583–92. doi: 10.1158/1078-0432.CCR-17-2507
14. Haralick RM, Shanmugam K, Dinstein IH. Textural features for image classification. *IEEE Trans Syst Man Cybern*. (1973) 3:610–21. doi: 10.1109/TSMC.1973.4309314
15. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, De Jong EEC, Van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol*. (2017) 14:749–62. doi: 10.1038/nrclinonc.2017.141
16. Limkin EJ, Sun R, Dercle L, Zacharaki EI, Robert C, Reuze S, et al. Promises and challenges for the implementation of computational medical imaging (radiomics) in oncology. *Ann Oncol*. (2017) 28:1191–206. doi: 10.1093/annonc/mdx034
17. Verma V, Simone CB III, Krishnan S, Lin SH, Yang J, Hahn SM. The rise of radiomics and implications for oncologic management. *J Natl Cancer Inst*. (2017) 109:djx055. doi: 10.1093/jnci/djx055
18. Tahir MA. Pattern analysis of protein images from fluorescence microscopy using gray level co-occurrence matrix. *J King Saud Univ Sci*. (2018) 30:29–40. doi: 10.1016/j.jksus.2016.12.004
19. Zhang Q, Li C, Han H, Yang L, Wang Y, Wang W. Computer-aided quantification of contrast agent spatial distribution within atherosclerotic plaque in contrast-enhanced ultrasound image sequences. *Biomed Signal Process Control*. (2014) 13:50–61. doi: 10.1016/j.bspc.2014.03.005
20. Zhang Q, Xiao Y, Suo J, Shi J, Yu J, Guo Y, et al. Sonoelastomics for breast tumor classification: a radiomics approach with clustering-based feature selection on sonoelastography. *Ultrasound Med Biol*. (2017) 43:1058–69. doi: 10.1016/j.ultrasmedbio.2016.12.016
21. Hearst MA, Dumais ST, Osman E, Platt J, Scholkopf B. Support vector machines. *IEEE Intell Syst Their Appl*. (1998) 13:18–28. doi: 10.1109/5254.708428
22. Amin MB, Greene FL, Edge SB, Compton CC, Gershenwald JE, Brookland RK, et al. The Eighth Edition AJCC cancer staging manual: continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. *CA Cancer J Clin*. (2017) 67:93–9. doi: 10.3322/caac.21388
23. Kim JH, Lee JY, Baek JH, Eun HW, Kim YJ, Han JK, et al. High-resolution sonography for distinguishing neoplastic gallbladder polyps and staging gallbladder cancer. *AJR Am J Roentgenol*. (2015) 204:W150–9. doi: 10.2214/AJR.13.11992
24. Bonatti M, Vezzali N, Lombardo F, Ferro F, Zamboni G, Tauber M, et al. Gallbladder adenomyomatosis: imaging findings, tricks and pitfalls. *Insights Imaging*. (2017) 8:243–53. doi: 10.1007/s13244-017-0544-7
25. Babu BI, Dennison AR, Garcea G. Management and diagnosis of gallbladder polyps: a systematic review. *Langenbecks Arch Surg*. (2015) 400:455–62. doi: 10.1007/s00423-015-1302-2
26. Wiles R, Varadpande M, Muly S, Webb J. Growth rate and malignant potential of small gallbladder polyps–systematic review of evidence. *Surgeon*. (2014) 12:221–6. doi: 10.1016/j.surge.2014.01.003
27. Lu D, Radin R, Yung E, Tchelepi H. Malignant transformation of a 5-mm gallbladder polyp over 2 years: a case report and review of current literature. *Ultrasound Q*. (2015) 31:66–8. doi: 10.1097/RUQ.0000000000000094
28. Chae HD, Lee JY, Jang JY, Chang JH, Kang J, Kang MJ, et al. Photoacoustic imaging for differential diagnosis of benign polyps versus malignant polyps of the gallbladder: a preliminary study. *Korean J Radiol*. (2017) 18:821–7. doi: 10.3348/kjr.2017.18.5.821
29. Colling R, Pitman H, Oien K, Rajpoot N, Macklin P, CM-Path AI in Histopathology Working Group, et al. Artificial intelligence in digital pathology: a roadmap to routine use in clinical practice. *J Pathol*. (2019) 249:143–50. doi: 10.1002/path.5310
30. Dalmis MU, Gubern-Merida A, Vreemann S, Bult P, Karssemeijer N, Mann R, et al. Artificial intelligence-based classification of breast lesions imaged with a multiparametric breast MRI protocol with ultrafast DCE-MRI, T2, and DWI. *Invest Radiol*. (2019) 54:325–32. doi: 10.1097/RLI.0000000000000544
31. O'sullivan S, Heinsen H, Grinberg LT, Chimelli L, Amaro E Jr, Do Nascimento S, et al. The role of artificial intelligence and machine learning in harmonization of high-resolution post-mortem MRI (virtopsy) with respect to brain microstructure. *Brain Inform*. (2019) 6:3. doi: 10.1186/s40708-019-0096-3
32. Weisberg EM, Chu LC, Park S, Yuille AL, Kinzler KW, Vogelstein B, et al. Deep lessons learned: radiology, oncology, pathology, and computer science experts unite around artificial intelligence to strive for earlier pancreatic cancer diagnosis. *Diagn Interv Imaging*. (2020) 101:111–5. doi: 10.1016/j.diii.2019.09.002
33. Sugiyama M, Xie XY, Atomi Y, Saito M. Differential diagnosis of small polypoid lesions of the gallbladder: the value of endoscopic ultrasonography. *Ann Surg*. (1999) 229:498–504. doi: 10.1097/00000658-199904000-00008
34. Sugiyama M, Atomi Y, Yamato T. Endoscopic ultrasonography for differential diagnosis of polypoid gall bladder lesions: analysis in surgical and follow up series. *Gut*. (2000) 46:250–4. doi: 10.1136/gut.46.2.250
35. Taira A, Tanaka K. [Cholesterol polyp]. *Ryoikibetsu Shokogun Shirizu*. (1996). 280–82.

Conflict of Interest: QZ was a consultant of Hangzhou YITU Healthcare Technology.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 24 September 2020
doi: 10.3389/fonc.2020.01646

Ultrasound-Based Radiomics Analysis for Preoperatively Predicting Different Histopathological Subtypes of Primary Liver Cancer

Yuting Peng[1†], Peng Lin[1†], Linyong Wu[1], Da Wan[1], Yujia Zhao[1], Li Liang[1], Xiaoyu Ma[1], Hui Qin[1], Yichen Liu[1], Xin Li[2], Xinrong Wang[2], Yun He[1] and Hong Yang[1*]*

[1] Department of Medical Ultrasonics, The First Affiliated Hospital of Guangxi Medical University, Nanning, China, [2] GE Healthcare, Shanghai, China

Edited by:
Wei Wang,
The First Affiliated Hospital of Sun Yat-sen University, China

Reviewed by:
JIanhua Zhou,
Sun Yat-sen University Cancer Center (SYSUCC), China
Guang-Jian Liu,
The Sixth Affiliated Hospital of Sun Yat-sen University, China
Isaac Shiri,
Geneva University Hospitals (HUG), Switzerland

***Correspondence:**
Yun He
228388072@qq.com
Hong Yang
yanghong@gxmu.edu.cn

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *14 March 2020*
Accepted: *27 July 2020*
Published: *24 September 2020*

Citation:
Peng Y, Lin P, Wu L, Wan D, Zhao Y, Liang L, Ma X, Qin H, Liu Y, Li X, Wang X, He Y and Yang H (2020) Ultrasound-Based Radiomics Analysis for Preoperatively Predicting Different Histopathological Subtypes of Primary Liver Cancer. Front. Oncol. 10:1646. doi: 10.3389/fonc.2020.01646

Background: Preoperative identification of hepatocellular carcinoma (HCC), combined hepatocellular–cholangiocarcinoma (cHCC-ICC), and intrahepatic cholangiocarcinoma (ICC) is essential for treatment decision making. We aimed to use ultrasound-based radiomics analysis to non-invasively distinguish histopathological subtypes of primary liver cancer (PLC) before surgery.

Methods: We retrospectively analyzed ultrasound images of 668 PLC patients, comprising 531 HCC patients, 48 cHCC-ICC patients, and 89 ICC patients. The boundary of a tumor was manually determined on the largest imaging slice of the ultrasound medicine image by ITK-SNAP software (version 3.8.0), and then, the high-throughput radiomics features were extracted from the obtained region of interest (ROI) of the tumor. The combination of different dimension-reduction technologies and machine learning approaches was used to identify important features and develop the moderate radiomics model. The comprehensive ability of the radiomics model can be evaluated by the area under the receiver operating characteristic curve (AUC).

Results: After digitally processing tumor ultrasound images, 5,234 high-throughput radiomics features were obtained. We used the Spearman + least absolute shrinkage and selection operator (LASSO) regression method for feature selection and logistics regression for modeling to develop the HCC-vs-non-HCC radiomics model (composed of 16 features). The Spearman + statistical test + random forest methods were used for feature selection, and logistics regression was applied for modeling to develop the ICC-vs-cHCC-ICC radiomics model (composed of 19 features). The overall performance of the radiomics model in identifying different histopathological types of PLC was moderate, with AUC values of 0.854 (training cohort) and 0.775 (test cohort) in the

HCC-vs-non-HCC radiomics model and 0.920 (training cohort) and 0.728 (test cohort) in the ICC-vs-cHCC-ICC radiomics model.

Conclusion: Ultrasound-based radiomics models can help distinguish histopathological subtypes of PLC and provide effective clinical decision making for the accurate diagnosis and treatment of PLC.

Keywords: primary liver cancer, histopathological subtype, radiomics, ultrasound, identification

INTRODUCTION

Primary liver cancer (PLC) is one of the most lethal and prevailing tumors, which is estimated to rank the fifth in cancer mortality among men and the seventh among women. In recent years, the incidence of PLC has continued to increase, rising faster than that of other cancers (1, 2). In the same solid malignant neoplasm, PLC can be classified according to histological sources. A tumor that contains only cancerous hepatocytes is defined as hepatocellular carcinoma (HCC), only cancerous bile duct cells are defined as intrahepatic cholangiocarcinoma (ICC), and a mixture of HCC and ICC is defined as combined hepatocellular–cholangiocarcinoma (cHCC-ICC) (3, 4).

cHCC-ICC is a relatively rare subtype of PLC with a variably reported incidence between 0.4 and 14.2%, and its overall prognosis is worse than that of either HCC or ICC alone (5, 6). Studies have revealed that in patients with PLC undergoing liver resection surgery, the survival outcome of cHCC-ICC is worse than that of HCC and that it is similar to or worse than that of ICC patients (7). HCC patients who meet the Milan criteria are indicated for liver transplantation, and their transplantation effect is excellent (8). However, increasing evidence indicates that the prognosis for cHCC-ICC patients undergoing liver transplantation is worse than that of patients with HCC alone and that cHCC-ICC is regarded as a relative contraindication for liver transplantation (9–11). Considering the scarcity of liver sources available for transplantation and the poor prognosis for cHCC-ICC, the correct identification of different PLC subtypes before surgery is a necessary condition for the reasonable selection of surgical candidates for liver transplantation and liver resection surgery, and it can improve overall survival outcomes (12, 13). PLC is often diagnosed as advanced, and many patients do not qualify for a curable treatment; systemic treatments that are effective for either HCC or ICC alone appears to be ineffective for cHCC-ICC (5). Therefore, precise and proper preoperative diagnosis is important for patient management to distinguish cHCC-ICC from HCC and ICC since different PLC subtypes may determine different treatment decisions.

Due to the high heterogeneity in the proportion and existing forms of the two tumor components, the imaging manifestations of cHCC-ICC have lacked specificity. At present, most cases of cHCC-ICC are misdiagnosed as simple HCC or ICC. Theodora et al. showed that the liver imaging reporting and data system (LI-RADS) as a common method for qualitative diagnosis of liver tumors applied in liver-contrast-enhanced ultrasound (CEUS) diagnosis may misdiagnose 54.1% of cHCC-ICC lesions as HCC (14). In contrast-enhanced imaging, cHCC-ICC has overlapping imaging modes with HCC and ICC. The main tissue in the tumor largely determines the main imaging features, making it difficult to distinguish cHCC-ICC from HCC and ICC (15). Moreover, most tumors can be diagnosed with core needle biopsy before surgery, but due to the different proportions of ICC and HCC in cHCC-ICC and sampling error, even histological biopsy may lead to preoperative diagnosis error and misdiagnosis of cHCC-ICC as HCC or ICC (16). Therefore, although accurate preoperative diagnosis of the three subtypes of PLC is important, it is still difficult.

Radiomics, a newly emerging concept in recent years, uses computers to extract a large amount of non-visual quantitative image information to realize the extraction of tumor features and model establishment, and it further excavates and analyzes image data information to assist doctors in diagnosis (17). Through the radiomics approach, the features that can be identified by human eyes and extracted by computers build a complementary relationship; in addition, radiomics combined with currently effective clinical evaluation indicators can improve the accuracy of medical diagnosis (18, 19). Tumor features vary from different tumor morphologies and biological behaviors. Radiomics as a method of deep mining high-dimensional image features can capture the characteristics of tumors more comprehensively, providing a feasible new method for identifying different tumors. Rafael et al. extracted 2D texture features and 3D texture features from T1-weighed MR images of 67 brain metastases and established a radiomics model using a random forest method. This model was helpful in distinguishing the primary tumors from brain metastases (breast cancer, lung cancer, and melanoma) (20). In the research by Yin et al., the radiomics model based on MR images can effectively identify different sacral tumors for preoperative identification of chordoma, giant cell tumor, and metastatic tumor (21).

Currently, the diagnosis of cHCC-ICC is usually based on postoperative pathology. Radiomics studies based on ultrasound evaluation of three different PLC subtypes are still lacking, and relevant reports have not been reported. In different imaging examinations, ultrasound technology has the advantages of no radiation, real-time observation, and simplicity with regard to liver disease examinations. An ultrasound-based radiomics approach may be better than other approaches in identifying three types of PLC to provide additional information. In this study, an ultrasound-based machine learning method was used to extract radiomics features and develop radiomics models to identify different pathological types of PLC.

MATERIALS AND METHODS

Study Population

This study was approved by the Ethics Committee of the First Affiliated Hospital of Guangxi Medical University. A comprehensive retrospective research was implemented on the medical records of patients diagnosed with PLC after surgery in the First Affiliated Hospital of Guangxi Medical University from January 2017 to September 2019.

The following inclusion and exclusion criteria were implemented in this study. Inclusion criteria included the following: (1) the lesions were primary liver tumors; (2) the target nodule was confirmed by surgery pathology; (3) liver ultrasound examination was performed within 14 days before resection; and (4) the target lesions were displayed clearly on the ultrasound images. Exclusion criteria included the following: (1) anticancer treatment before surgery; (2) poor image quality; and (3) uncompleted clinical data.

Finally, 668 eligible patients (544 male/124 female; mean age, 50.5 ± 11.4 years; age range, 22–79 years) were enrolled (**Figure 1**). The pathological tissue of the lesions was obtained by surgical hepatic resection for pathological diagnosis to determine the histological classification of PLC, of which there were 531 HCC patients, 89 ICC patients, and 48 cHCC-ICC patients.

Patient Clinical Pathological Parameters

Basic patient information was collected including data on gender, age, tumor size, cirrhosis, hepatitis, and serum tumor markers. Serological data included carbohydrate antigen 19-9 (CA19-9), alpha fetoprotein (AFP), and carcinoembryonic antigen (CEA) levels. These data were measured within 2 weeks before surgery.

We also collected patient pathological information, including tumor differentiation, microvascular invasion (MVI), TNM stage, and immunohistochemical information on Ki67, p53, and vascular endothelial growth factor (VEGF). MVI referred to the observation of a nest of cancer cells in a blood vessel lining the endothelial cells by microscopy. In this study, the TNM staging of PLC patients was analyzed according to the American Joint Cancer Commission (AJCC) eighth edition staging system (22, 23).

Radiomics Analysis

The research of radiomics mainly includes the following steps: tumor segmentation, data preprocessing and feature selection, modeling, and evaluation (**Figure 2**). In the training cohort, we performed a combination of different dimension-reduction technologies and machine learning approaches to establish radiomics models. Finally, the test cohort was taken to evaluate the generalization performance of the model.

Ultrasound Imaging and Tumor Segmentation

GE Logiq E9 ultrasound diagnostic instruments (GE Healthcare, United States, C5-1 abdominal probe, 2.8–5.0 MHz), Philips EPIQ 5 ultrasound diagnostic instruments (Philips Medical Systems, United States, C5-1 abdominal probe, 1–5 MHz), and Aloka EZU-MT28-S1 ultrasound diagnostic instruments (Aloka, Japan, abdominal probe, 2–6 MHz) were used to collect images. We conducted a retrospective review of the image data and selected two-dimensional ultrasound images in digital imaging and communications in medicine (DICOM) format that clearly showed the largest cross section of each lesion. We imported the images into the ITK-SNAP software (version 3.8.0)[1] to manually draw the tumor boundary and determine the tumor region of interest (ROI) (**Figure 3**). Under the supervision of a radiologist with over 20 years of ultrasound diagnosis experience, another radiologist with 15 years of ultrasound diagnosis experience completed the ROIs for all tumors.

Feature Extraction and Data Preprocessing

Intelligence Foundry software (GE Healthcare, version 1.3) was used for radiomics analysis. Since the images were collected by different ultrasound equipment and the feature vectors had a wide range, we preprocessed the data before modeling analysis to improve the accuracy of the calculation, including ultrasound system supplier data alignment, median value replacement of missing value processing, and data normalization processing.

We used 256 as the bin size to discretize the gray value of the images and used the ComBat method to standardize the radiomics features. The ComBat method was previously used in radiomics studies of different PTE or MRI protocols (24, 25). The wavelet features were based on the original gray value image for wavelet transformation (including HLH, LLL, and HHL, with eight local matrices); the energy, skewness, and other series of parameters were extracted from the obtained wavelet transform matrix. In the same way, the shearlet change and the gabor operator transformation were also carried out, and different step lengths were used in the change to obtain multiple sets of transformation intermediate value matrices. Based on the above transformations, the radiomics parameters were extracted, and finally, we obtained 5,234 high-throughput features. The types of features included the following: first-order features (energy, mean, skewness, kurtosis, etc.), shape features (minor axis length, major axis length, elongation, etc.), wavelet features and textural features [gray level co-occurrence matrix (GLCM) features, grey level run length matrix (GLRLM) features, etc] (**Supplementary Part A**). The feature parameters extracted by the Intelligence Foundry software (GE Healthcare, version 1.3) were algorithms provided using the pyradiomics package, which calculated the radiomics features in accordance with the feature definition described in the 2016 version of the image biomarker standardization initiative (IBSI) (26, 27). The median was used to fill in missing extracted feature values and substitute abnormal value. Z-score normalization was used to convert different data to the same order of magnitude, and the calculation formula was as follows:

$$y = (x - \mu)/\sigma$$

where μ is the mean and σ is the standard deviation.

---

[1] http://www.itksnap.org

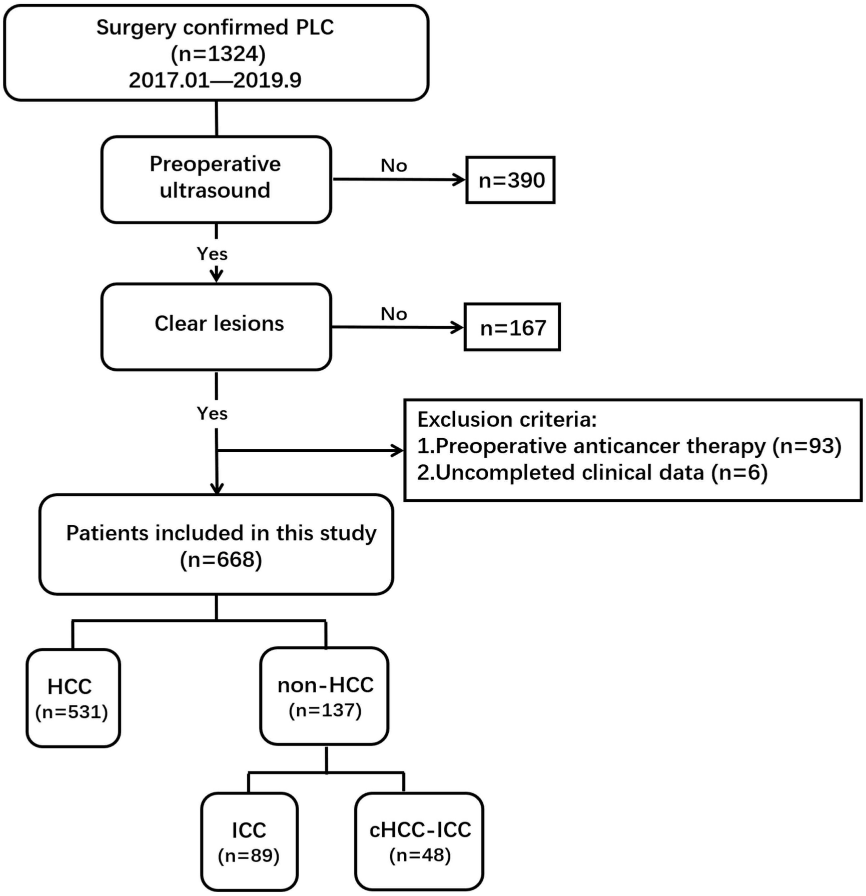


FIGURE 1 | Flow chart of study population screening.

The PLC patients were labeled according to different histological types into different layers. In the HCC-vs-non-HCC model, the non-HCC label was "0," and the HCC label was "1." In the ICC-vs-cHCC-ICC model, the cHCC-ICC label was "0," and the ICC label was "1." Then, PLC patients with different histological types were grouped based on a 7:3 ratio (training cohort : test cohort) in each layer using the method of stratified sampling. The training cohort was used to build the model, and the test cohort was an independent external verification to evaluate the model established by the training cohort.

Feature Selection

We obtained 5,234 high-throughput radiomics features and normalized the quantitative expression values of the radiomics features using the *Z*-score method. Considering that some highly correlated and redundant features in the data may affect the classification effect of the model, we calculated the Spearman correlation coefficient. A correlation coefficient between the two variables close to 1 indicated that the linear relationship between them was strong and that one of the two variables could be used instead of the other. In this study, the high-correlation features were removed with a threshold of 0.95 (HCC vs. non-HCC) and 0.75 (ICC vs. cHCC-ICC). Then, we used the statistical test method to screen for features that had differences.

Finally, we used four dimension-reduction technologies to further deal with the features that were processed above. Dimension-reduction technologies included random forest, max-relevance and min-redundancy (mRMR), logistic regression, and support vector machine recursive feature elimination (SVM-RFE) (**Supplementary Part B**).

Modeling and Evaluation

The final selected radiomics features were imported into the classifier to build a model for evaluating three different histopathological types of PLC. Ten machine learning approaches were used in this study, which were decision tree, naïve Bayes, k-nearest neighbor (KNN), logistics regression, support vector machine (SVM), bagging, random forest, extremely randomized trees, AdaBoost, and gradient boosting tree (**Supplementary Part B**).

We extracted 5,234 features from the ultrasound images. We quantify the discriminative ability of the radiomics model by calculating the receiver operating characteristic curve (AUC). We constructed the model by separately combining the above four dimension-reduction technologies and the above 10 machine

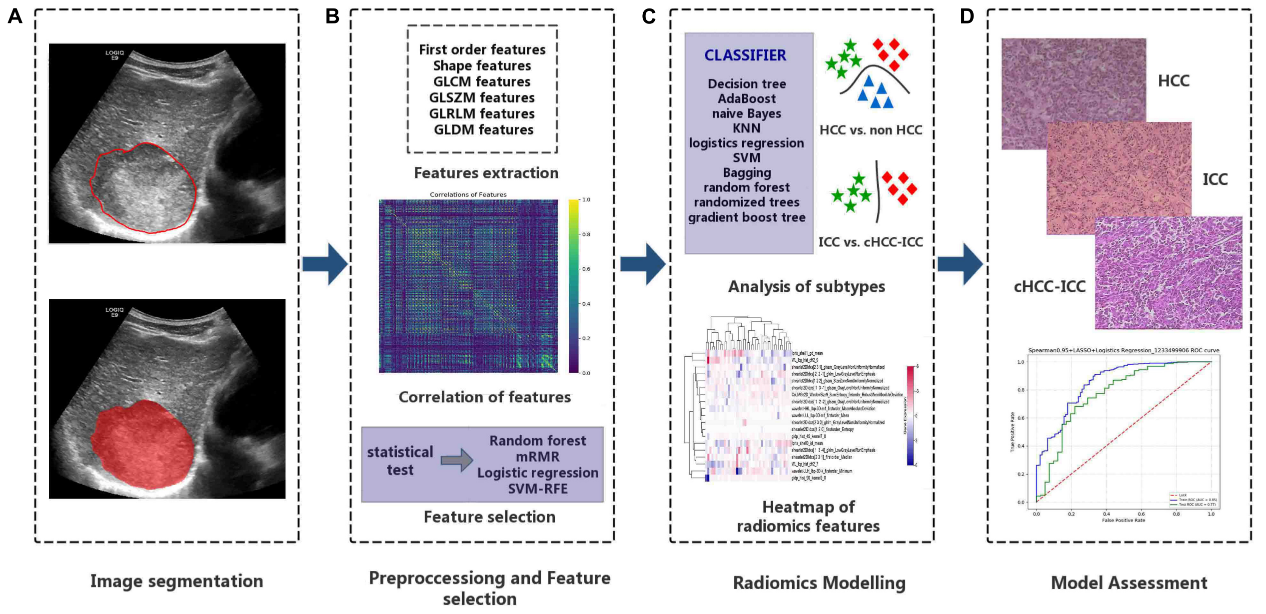


FIGURE 2 | Important steps for the radiomics research. **(A)** Tumor regions-of-interest identification. **(B)** After the tumor images are digitalized, a total of 5,234 quantitative features were obtained, and data were standardized for preprocessing. **(C)** The combination of dimension reduction and classifier was performed to develop radiomics models to identify primary liver cancers of different histopathological types. **(D)** Evaluation of the classification effects of the radiomics model in identifying different histopathological types of primary liver cancer.

learning approaches and chose the combination with the highest AUC to build the optimal radiomics model. In the training cohort, to avoid overfitting the classifier, we used a 10-fold cross-validation method.

We performed a receiver operating characteristic (ROC) curve analysis and calculated the accuracy and precision. We also used the confusion matrix as a measure of the quality of the machine learning approaches to verify whether the prediction results were consistent with the actual results. The confusion matrix is a useful tool for evaluating the classification ability of radiomics models (28).

In the HCC-vs-non-HCC and ICC-vs-cHCC-ICC radiomics models, we performed univariate and multivariate logistic regression analyses to analyze the relevant factors of different pathological types of PLC. Univariate analysis factors with *P*-values less than 0.1 were further analyzed by multivariate logistic regression analysis. In multivariate analysis, a *P*-value less than 0.05 was considered significant.

Statistical Analysis

R software (version 3.6.0) and SPSS software (version 22.0) were applied for statistical analysis. In the quantitative data with a normal distribution, the completely random design *t*-test was performed for the two-samples contrast, the analysis of variance was used to contrast several independent samples, and variables were summarized as the mean ± standard deviation (SD). For quantitative data with a skewed distribution, the Mann–Whitney *U* test was performed to compare two independent samples, the Kruskal–Wallis *H* test was used to compare several independent samples, and variables were summarized as the median (q1–q3). Qualitative data were compared using chi-square tests, with variables described as percentages. *P*-values below 0.05 was considered to be statistically significant differences. In the R software (version 3.6.3), the "pheatmap" package was used to draw heat maps of features.

RESULTS

Clinicopathological Data of PLC Patients

A total of 668 PLC patients were adopted in this research (**Figure 1**). The clinicopathological parameters of the training and test cohorts were shown in **Table 1**. There were no significant differences in the distribution of clinicopathological features between the two cohorts, including gender, age, tumor size, hepatitis, cirrhosis, serum tumor markers, pathological subtype, immunohistochemistry, or tumor stage. These results showed the rationality of our training and test cohort partitions.

In the HCC-vs-non-HCC group, the study sample included 467 people in the training cohort (379 male/88 female, mean age, 50.5 ± 11.4 years), 371 cases of HCC, and 96 cases of non-HCC. There were 201 patients in the test cohort (165 male/36 female, mean age, 50.6 ± 11.3 years), 160 cases of HCC, and 41 cases of non-HCC. In the ICC-vs-cHCC-ICC group, the study sample included 95 people in the training cohort (65 male/30 female, mean age, 49.4 ± 11.6 years), 33 cases of cHCC-ICC and 62 cases of ICC. There were 42 patients in the test cohort (27 male/15 female, mean age, 51.8 ± 10.3 years), 15 cases of cHCC-ICC, and 27 cases of ICC.

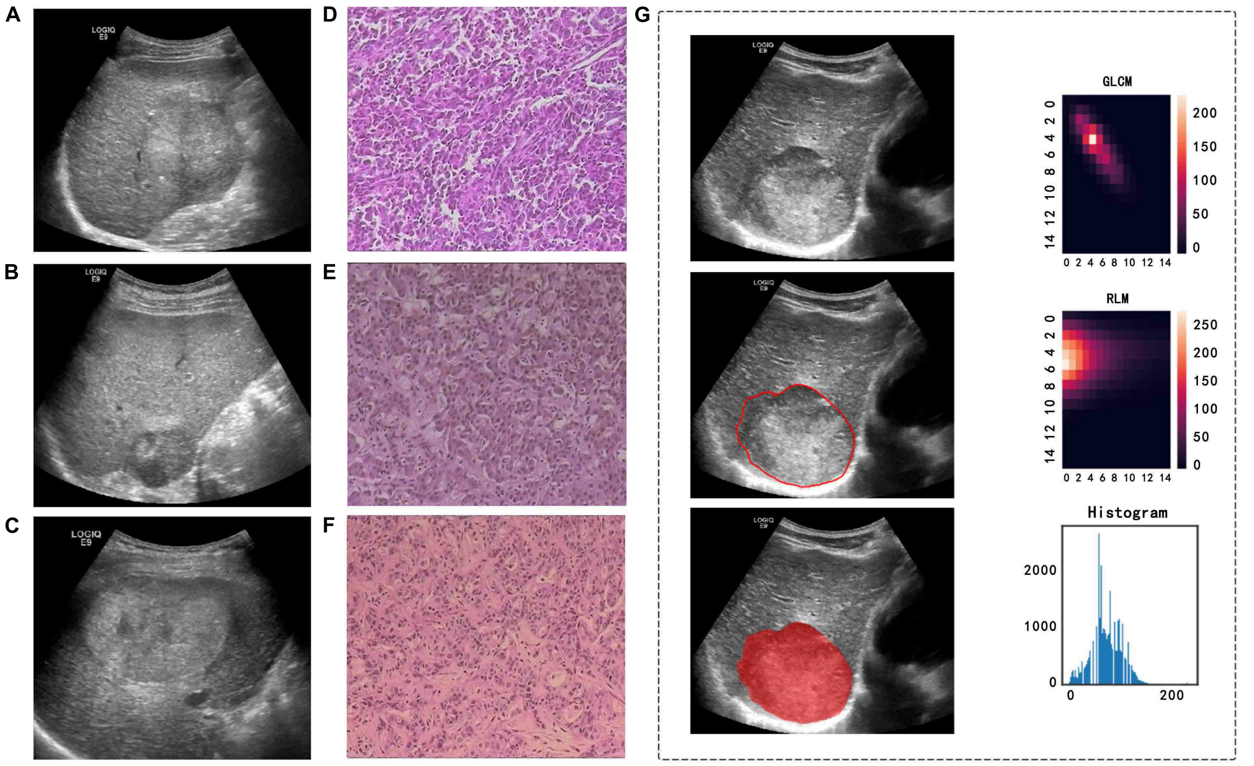


FIGURE 3 | Ultrasound and pathological images, tumor segmentation, and feature extraction of three different pathological types of PLC. **(A,D)** A 57-year-old man with a pathological diagnosis of cHCC-ICC. **(B,E)** A 64-year-old woman with a pathological diagnosis of HCC. **(C,F)** A 44-year-old man with a pathological diagnosis of ICC. **(G)** An example of manually sketching the region of interest (ROI) of a tumor on an ultrasound image and gray level co-occurrence matrix (GLCM) features, run length matrix (RLM) features, and grayscale histogram feature extraction from the grayscale ultrasound image.

Identification of the Radiomics Signature

In the HCC-vs-non-HCC group, we used the LASSO regression method for dimension reduction and modeling with the logistics regression method. In the ICC-vs-cHCC-ICC group, we used the random forest method for dimension reduction, feature selection with a threshold value of 1.25 times the mean value, and modeling with the logistics regression method. Finally, we respectively identified 16 and 19 optimal radiomics features for HCC-vs-non-HCC model and ICC-vs-cHCC-ICC model predictions (**Table 2**). **Figure 4** showed the heat map of 16 features (HCC-vs-non-HCC model) and 19 features (ICC-vs-cHCC-ICC model) of the final radiomics models.

Radiomics Model Assessment

The results showed that the radiomics models we built had a high overall classification performance for identifying three subtypes of PLC. The AUC values in the training cohort and test cohort were 0.854 and 0.775 (HCC vs. non-HCC) and 0.920 and 0.728 (ICC vs. cHCC-ICC), respectively (**Figures 5A,B**). The confusion matrix was shown in **Figures 5C,D**. In the HCC-vs-non-HCC model, the predicted results showed that of the 160 actual HCC patients, 155 were correctly predicted to be HCC. In the ICC-vs-cHCC-ICC model, the 15 patients with actual cHCC-ICC, 6 were predicted to be cHCC-ICC, and among the 27 actual ICC patients, 22 were correctly predicted to be ICC. These results indicated that the radiomics models can moderately distinguish three different histological types of PLC and performed best at HCC identification.

Tables 3, **4** showed the results of univariate and multivariate logistic regression analyses of HCC-vs-non-HCC and ICC-vs-cHCC-ICC radiomics models. In the HCC-vs-non-HCC radiomics model, gender, hepatitis, AFP, CA19-9, CEA, stage, and radiomics score were independent factors related to HCC ($P < 0.05$). In the ICC-vs-cHCC-ICC radiomics model, AFP and radiomics score were independent factors related to cHCC-ICC ($P < 0.05$).

DISCUSSION

In this research, as far as we know, we are the first to identify an ultrasound-based radiomics models that can be used to predict HCC, ICC, and cHCC-ICC. The radiomics models achieved good diagnostic efficiency in both the training cohort and the test cohort, which is expected to help doctors improve the accuracy of presurgical diagnosis and guide the further treatment of PLC patients.

TABLE 1 | Clinicopathological profiles of two radiomics models in the training cohort and test cohort.

	HCC vs. non-HCC Model				ICC vs. cHCC-ICC Model		
Variables	**Training cohort (*n* = 467)**	**Test cohort (*n* = 201)**	***P*-value**	**Variables**	**Training cohort (*n* = 95)**	**Test cohort (*n* = 42)**	***P*-value**
Gender				**Gender**			
Male	379 (81.2)	165 (41.8)	0.78	Male	65 (68.4)	27 (64.3)	0.63
Female	88 (18.8)	36 (58.2)		Female	30 (31.6)	15 (35.7)	
Age (years)				**Age (years)**			
<40	88 (18.8)	35 (174)	0.89	<40	21 (22.1)	5 (11.9)	0.27
40–60	280 (60.0)	124 (61.7)		40–60	59 (62.1)	27 (64.3)	
>60	99 (21.2)	42 (20.9)		>60	15 (15.8)	10 (23.8)	
Tumor size (cm)				**Tumor size (cm)**			
≤5	246 (52.7)	115 (57.2)	0.28	≤5	40 (42.1)	15 (35.7)	0.48
>5	221 (47.3)	86 (42.8)		>5	55 (57.9)	27 (64.3)	
Hepatitis				**Hepatitis**			
Yes	364 (80.2)	159 (79.9)	0.15	Yes	58 (61.1)	23 (54.8)	0.09
No	103 (19.8)	59 (20.1)		No	37 (38.9)	19 (45.2)	
Cirrhosis				**Cirrhosis**			
Yes	218 (46.7)	108 (53.7)	0.09	Yes	36 (37.9)	14 (33.3)	0.61
No	249 (53.3)	93 (46.3)		No	59 (62.1)	28 (66.7)	
AFP (μ g/ml)				**AFP (μ g/ml)**			
≤400	344 (73.7)	139 (69.2)	0.23	≤400	76 (80.0)	38 (90.5)	0.13
>400	123 (26.3)	62 (30.8)		>400	19 (20.0)	4 (9.5)	
CA19-9 (U/ml)				**CA19-9 (U/ml)**			
≤37	389 (83.3)	158 (78.6)	0.15	≤37	62 (65.3)	23 (54.8)	0.24
>37	78 (16.7)	43 (21.4)		>37	33 (34.7)	19 (45.2)	
CEA (μ g/ml)				**CEA (μ g/ml)**			
≤5	411 (88.0)	179 (89.1)	0.70	≤5	75 (78.9)	33 (78.6)	0.96
>5	56 (12.0)	22 (10.9)		>5	20 (21.1)	9 (21.4)	
Histological type				**Histological type**			
HCC	371 (79.4)	160 (79.6)	0.96	cHCC-ICC	33 (34.7)	15 (35.7)	0.91
Non-HCC	96 (20.6)	41 (20.4)		ICC	62 (65.3)	27 (64.3)	
Differentiation				**Differentiation**			
Well	21 (4.5)	5 (2.5)	0.61	Well	1 (1.0)	0 (0)	0.87
Moderate	331 (70.9)	149 (74.1)		Moderate	60 (63.2)	25 (59.5)	
Poor	83 (17.8)	33 (16.4)		Poor	23 (24.2)	11 (26.2)	
No data	32 (6.8)	14 (7.0)		No data	11 (11.6)	6 (14.3)	
Immunohistochemistry, Positive/Negative				**Immunohistochemistry, Positive/Negative**			
Ki67, >10%/≤10%	286/181 (61.2/38.8)	129/72 (64.2/35.8)	0.47	Ki67, >10%/≤10%	46/14 (48.4/51.6)	32/10 (76.2/23.8)	0.96
P53	254/213 (54.4/45.6)	122/79 (60.7/39.3)	0.13	P53	63/32 (66.3/33.7)	28/14 (66.7/33.3)	0.96
VEGF	219/248 (46.9/53.1)	102/99 (50.7/49.3)	0.36	VEGF	40/55 (42.1/57.9)	20/22 (47.6/52.3)	0.55
Microvascular invasion	132/335 (28.3/71.7)	62/139 (30.8/69.2)	0.50	Microvascular invasion	33/62 (34.7/65.3)	11/31 (26.2/73.8)	0.32

(Continued)

TABLE 1 | Continued

	HCC vs. non-HCC Model				ICC vs. cHCC-ICC Model		
Variables	**Training cohort (*n* = 467)**	**Test cohort (*n* = 201)**	***P*-value**	**Variables**	**Training cohort (*n* = 95)**	**Test cohort (*n* = 42)**	***P*-value**
Depth of invasion				**Depth of invasion**			
T1	255 (54.6)	111 (55.2)	0.53	T1	35 (36.8)	24 (57.2)	0.15
T2	111 (23.8)	55 (27.4)		T2	30 (31.6)	8 (19.0)	
T3	6 (1.3)	3 (1.5)		T3	1 (1.1)	0 (0)	
T4	95 (20.3)	32 (15.9)		T4	29 (30.5)	10 (23.8)	
Lymph node metastasis				**Lymph node metastasis**			
N0	450 (96.4)	194 (96.5)	0.92	N0	81 (85.3)	38 (90.5)	0.40
N1	17 (3.6)	7 (3.5)		N1	14 (14.7)	4 (9.5)	
Distant metastasis				**Distant metastasis**			
M0	451 (96.6)	198 (98.5)	0.17	M0	87 (91.6)	40 (95.2)	0.45
M1	16 (3.4)	3 (1.5)		M1	8 (8.4)	2 (4.8)	
Stage				**Stage**			
I	251 (53.7)	107 (53.2)	0.38	I	30 (31.6)	23 (54.8)	0.08
II	107 (22.9)	54 (26.9)		II	27 (28.4)	7 (16.7)	
III	89 (19.1)	36 (17.9)		III	30 (31.6)	10 (23.8)	
IV	20 (4.3)	4 (2.0)		IV	8 (8.4)	2 (4.7)	
Radiomics score	1.58 (0.97–2.04)	1.58 (0.95–2.10)	0.74	**Radiomics score**	0.91 (−0.31−2.39)	0.85 (−0.40−1.76)	0.29

Values were shown as the number of patients (percentage) unless otherwise explained. Radiomics score data were shown as median (Q1 – Q3). AFP, alpha fetoprotein; CA19-9, carbohydrate antigen 19-9; CEA, carcino-embryonic antigen; VEGF, vascular endothelial growth factor.

Another highlight of this study is that we constructed the optimal model through a variety of combinations of dimension-reduction technologies and classifiers. Shiri et al. found that the performance of machine learning models depends on the type of data or application and that there was no general algorithm or single model (29). Different combinations of feature selection methods and classifiers can provide different results (30–32). In the current study, we performed different dimension-reducing technologies and machine learning approaches to find the optimal models to predict HCC vs. non-HCC and ICC vs. cHCC-ICC. Therefore, the models that we obtained comprehensively captured the potential of radiomics-based differential diagnosis of PLC in the current clinical medical environment.

In the current clinical practice, physicians preoperatively rely on clinical symptoms, tumor serum markers, and imaging tests to determine the type of PLC patient, but these data can sometimes lead to false diagnoses because they may overlap. In addition, due to high heterogeneity in the proportion and existing forms of the two tumor components, the imaging findings of mixed HCC currently lack performance, and most cases are misdiagnosed as simple HCC or ICC. Preoperative differentiation of PLC subtypes has important clinical significance, as different types are associated with different treatment options and prognosis. Improving the accuracy of initial diagnosis can provide more optimized and active treatment for cHCC-ICC patients (16). In addition, clinical medicine is currently moving toward a trend of precision and personalized medicine. In the precise medical environment, medical imaging as an important diagnostic tool is also rapidly evolving and gradually playing an important role (33). Radiomics, which provides a non-invasive method to assess lesions and performs well in the diagnosis and prediction of tumors, is widely considered to be a step in the evolution of imaging toward a concept of personalized cancer management (34, 35).

So far, only a few studies have attempted to identify three different tissue types of PLC by imaging methods, and most previous studies have been based on CT and MR images. Wang et al. previously attempted to use preoperative CT and MR imaging to identify cHCC-ICC with HCC and ICC. The study found that compared with ICC and cHCC-ICC, the incidence of HCC pseudocapsule was significantly higher. Compared with their occurrence in HCC and cHCC-ICC, rim enhancement, abnormal perfusion, capsular retraction, and biliary dilatation were more common in ICC. However, in that study, the number of features obtained from images was small, and imaging features, such as tumor size, were all visible to the naked eye; the approach failed to identify and analyze microscopic image features with potential value for clinical diagnosis (36). Lewis et al. used MR images of 65 liver cancer patients. The tumor characteristics and LI-RADS classification were evaluated by two independent observers. Among the two independent observers, the combined AUC of sex and LI-RADS and apparent diffusion coefficient

TABLE 2 | Features and corresponding coefficients of HCC vs. non-HCC radiomics model and ICC vs. cHCC-ICC radiomics model.

HCC vs. non-HCC Model		**ICC vs. cHCC-ICC Model**	
Radiomics features	**Coefficient**	**Radiomics features**	**Coefficient**
Roughness index of boundary	−0.034	Ipris_shell0_id_mean	0.019
Textural_phenotype_level_20–30%	−0.423	Ipris_shell1_gd_mean	0.026
Wavelet-LHL_lbp-3D-m1_firstorder_InterquartileRange	−0.068	CoLIAGe2D_WindowSize9_Sum Entropy_firstorder_RobustMean Absolute Deviation	0.021
Shearlet2didxs[1 2 -2]_glszm_Small Area Emphasis	−0.020	Wavelet-LLH_lbp-3D-k_firstorder_Minimum	0.019
Shearlet2didxs[1 2 -2]_glszm_Small Area High GrayLevel Emphasis	−0.006	Wavelet-HHL_lbp-3D-m1_firstorder_MeanAbsoluteDeviation	0.026
Shearlet2didxs[1 2 -1]_glszm_Small Area High GrayLevel Emphasis	−0.053	Wavelet-LLL_lbp-3D-m1_firstorder_Mean	0.029
shearlet2DIdxs[1 3 4]_glszm_GrayLevel Non-Uniformity	−0.058	Shearlet2didxs[1 2 -2]_glszm_GrayLevel Non-Uniformity Normalized	0.028
Shearlet2didxs[2 3 -3]_firstorder_Maximum	−0.028	Shearlet2didxs[1 2 0]_firstorder_Entropy	0.023
Shearlet2didxs[2 3 -2]_firstorder_Minimum	0.017	Shearlet2didxs[1 2 2]_glszm_Size Zone NonUniformity Normalized	0.028
Shearlet2didxs[2 3 0]_firstorder_Skewness	−0.038	shearlet2DIdxs[1 3 −4] _glrlm_Low GrayLevel Run Emphasis	0.021
Shearlet2didxs[2 3 2]_firstorder_Minimum	0.023	Shearlet2didxs[1 3 −1]_glszm_GrayLevel Non-Uniformity Normalized	0.025
Shearlet2didxs[2 3 3]_firstorder_Maximum	−0.165	Shearlet2didxs[2 2 −1]_glrlm_Low GrayLevel Run Emphasis	0.024
glbp_hist_kernel1_2	−0.323	Shearlet2didxs[2 3 0]_glrlm_GrayLevel Non-Uniformity Normalized	0.023
glbp_hist_kernel4_3	−0.008	Shearlet2didxs[2 3 1]_firstorder_Median	0.022
gLTCoPs1_hist_kernel6_1	0.056	Shearlet2didxs[2 3 1]_glszm_GrayLevel Non-Uniformity Normalized	0.022
gLTCoPs1_hist_kernel6_2	0.061	gldp_hist_45_kernel7_0	0.019
		gldp_hist_90_kernel9_0	0.031
		WL_lbp_hist_cH2_7	0.022
		WL_lbp_hist_cH2_9	0.021

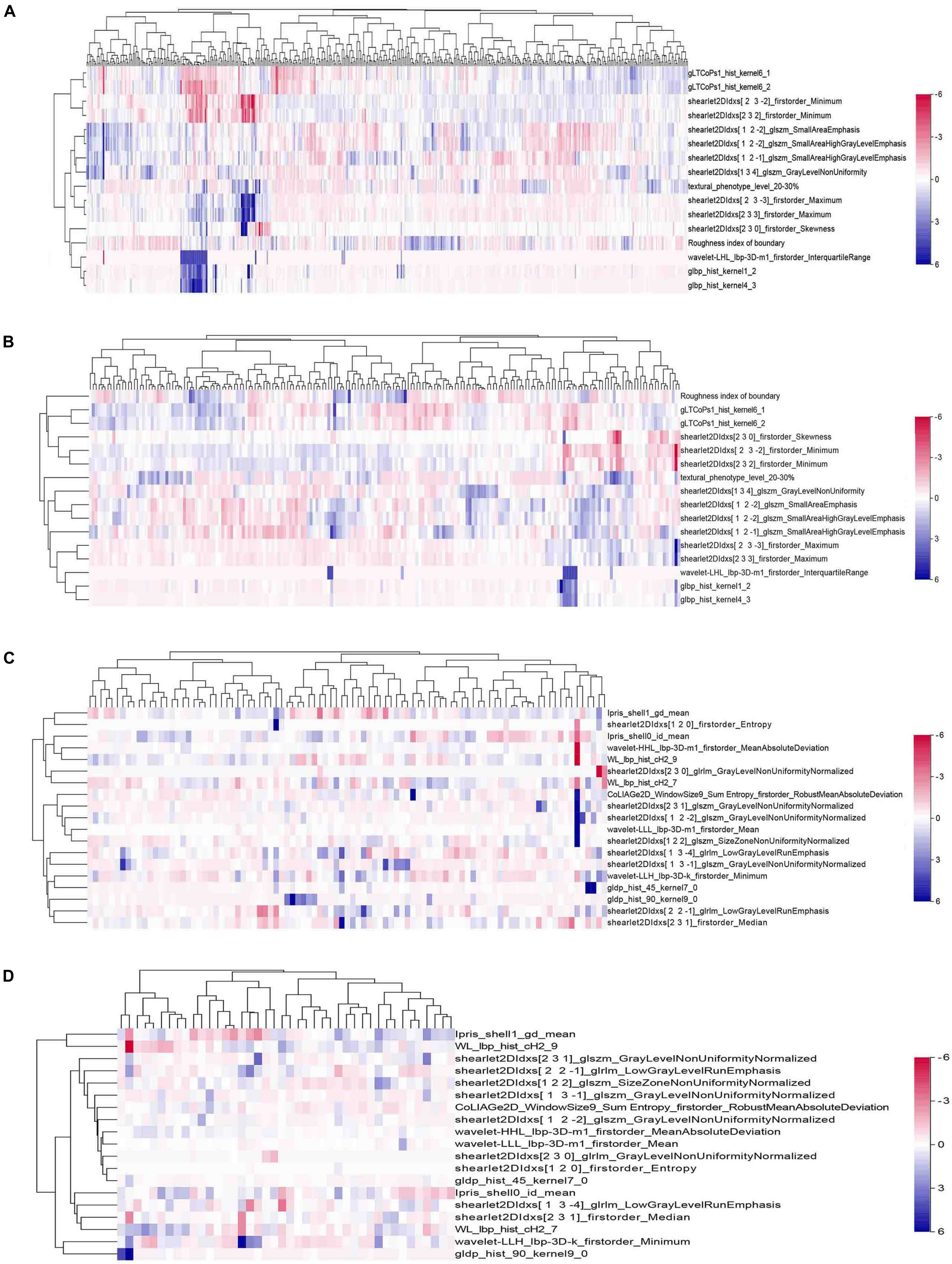


FIGURE 4 | Heat maps of the final features of radiomics models. A total of 16 features were used to build the HCC-vs-non-HCC model, and 19 features were used to build the ICC-vs-cHCC-ICC model. The radiomics features were normalized by *Z*-score. **(A)** Training cohort in the HCC-vs-non-HCC model. **(B)** Test cohort in the HCC-vs-non-HCC model. **(C)** Training cohort in the ICC-vs-cHCC-ICC model. **(D)** Test cohort in the ICC-vs-cHCC-ICC model.

(ADC) at the fifth percentile for the diagnosis of liver cancer were 0.90 and 0.89, respectively. This result showed that HCC can be better distinguished from ICC and cHCC-ICC by combining the ADC histogram parameters and LI-RADS categorization. However, the number of samples included in that study and the number of extracted features were small, and the study did

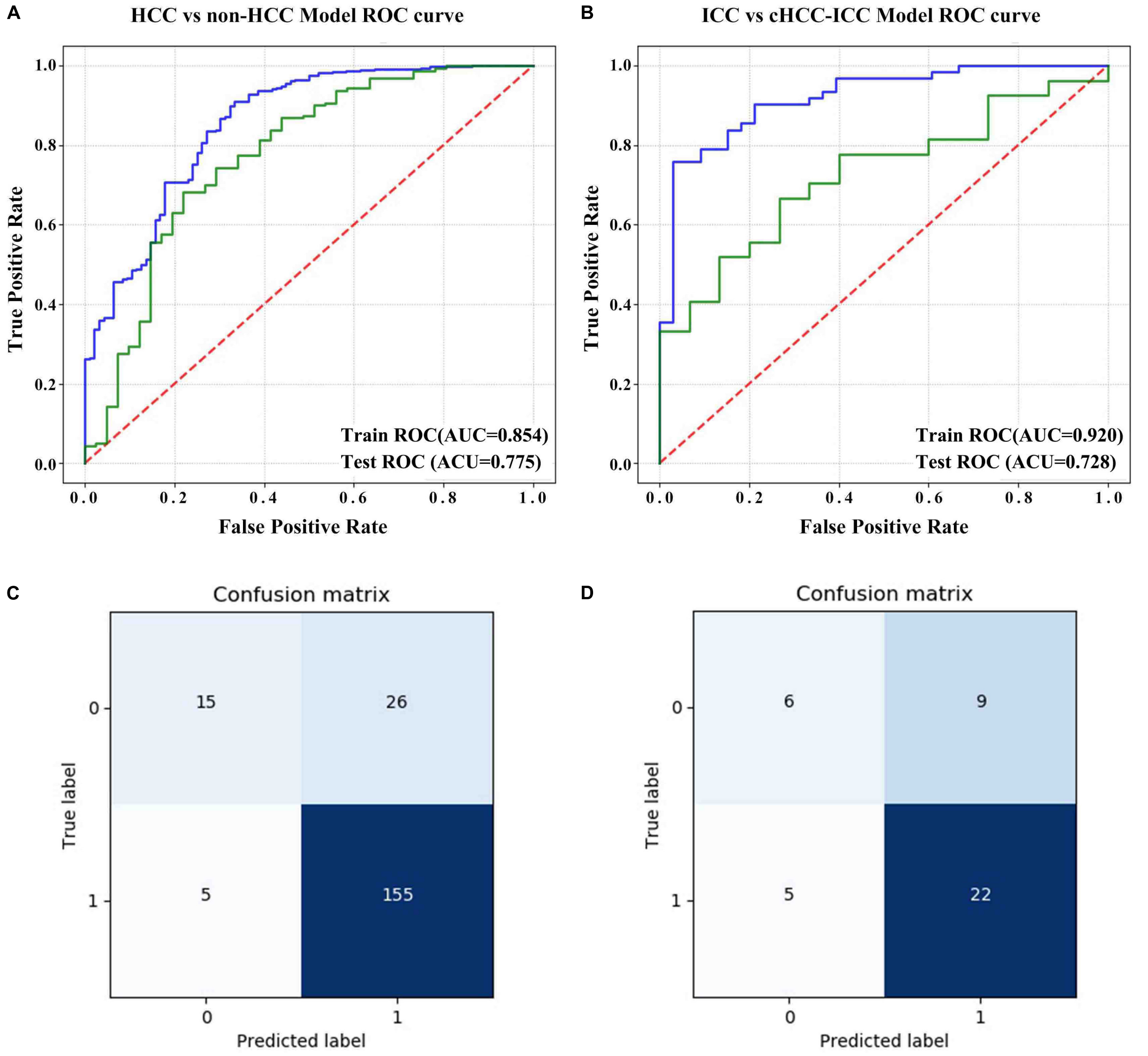


FIGURE 5 | Evaluation of the predictive performance of the radiomics models. **(A)** ROC curve of the HCC-vs-non-HCC model in the training cohort and test cohort. **(B)** ROC curve of the ICC-vs-cHCC-ICC model in the training cohort and test cohort. **(C)** Confusion matrix of the HCC-vs-non-HCC model in the test cohort. The non-HCC label was "0," and the HCC label was "1." **(D)** Confusion matrix of the ICC-vs-cHCC-ICC model in the test cohort. The cHCC-ICC label was "0," and the ICC label was "1." The abscissa represents the predicted label, and the ordinate represents the actual label.

not distinguish between ICC and cHCC-ICC (37). Compared with CT/MRI, ultrasound examination has the advantages of simplicity and real-time observation, and it plays a vital role in the diagnosis and treatment of liver tumors. However, no radiomics study has sought to identify HCC, cHCC-ICC, and ICC. In view of this knowledge gap, we used ultrasound images to establish radiomics models to distinguish three different pathological classifications of PLC, and we obtained promising results.

Our results showed that the radiomics models we built have a good overall AUC and could well to accurately predict pure HCC, while obtaining lower accuracy in cHCC-ICC. Our findings are roughly consistent with the results of some previous studies that suggest that identifying cHCC-ICC from PLC remains challenging, possibly due to the greater histological heterogeneity of cHCC-ICC. Wang et al. studied the CT and MR images of 136 patients with PLC and found that the features of capsular retraction, abnormal perfusion, and rim enhancement showed better performance in the identification of HCC and ICC, while the ability to distinguish cHCC-ICC from the other two types of PLC was not significant (36). Many image features such as shape, size, edge, position, and enhancement mode in cHCC-ICC mostly behave like ICC or HCC, creating some difficulties in its diagnosis (38).

We finally used the LASSO and random forest methods for feature selection. LASSO regression is also called L1 regularization of linear regression, which is a popular method

TABLE 3 | Results of the univariate and multivariate analyses in HCC-vs-non-HCC Model.

Factors (reference)	**Univariate analysis**		**Multivariate analysis**	
	OR (95% CI)	***p* value**	**OR (95% CI)**	***p* value**
Gender (female)	0.357 (0.233–0.549)	0.000*	0.379 (0.190–0.758)	0.006
Age (years)				
<40	Reference		–	–
40–60	0.804 (0.436–1.482)	0.485	–	–
>60	0.797 (0.487–1.305)	0.367	–	–
Tumor size (>5 cm)	0.508 (0.347–0.745)	0.001	1.618 (0.855–3.061)	0.139
Hepatitis (yes)	3.433 (2.279–5.172)	0.000*	2.642 (1.360–5.133)	0.004
Cirrhosis (yes)	1.883 (1.279–2.774)	0.001	1.436 (0.775–2.661)	0.250
AFP (μg/ml)				
≤400	Reference		Reference	
>400	2.176 (1.340–3.533)	0.002	3.533 (1.702–7.335)	0.001
CA19-9 (U/ml)				
≤37	Reference		Reference	
>37	0.244 (0.159–0.374)	0.000*	0.232 (0.118–0.456)	0.000*
CEA (μg/ml)				
≤5	Reference		Reference	
>5	0.379 (0.229–0.627)	0.000*	0.427 (0.189–0.965)	0.041
Differentiation				
Well	Reference		Reference	
Moderate	10.366 (1.350–79.590)	0.025	4.266 (0.361–50.333)	0.249
Poor	1.927 (1.212–3.063)	0.006	1.681 (0.867–3.258)	0.124
Immunohistochemistry, Negative/Positive				
Ki67, ≤10%/>10%	0.407 (0.263–0.629)	0.000*	0.632 (0.306–1.303)	0.214
P53	0.586 (0.395–0.868)	0.008	0.531 (0.275–1.025)	0.059
VEGF	1.241 (0.850–1.810)	0.264	–	–
Microvascular invasion	0.832 (0.555–1.248)	0.374	–	–
Stage				
I	Reference		Reference	
II	4.111 (1.735–9.736)	0.001	4.077 (1.152–14.425)	0.029
III	2.668 (1.090–6.532)	0.032	5.245 (1.410–19.504)	0.013
IV	1.518 (0.621–3.712)	0.360	2.267 (0.616–8.342)	0.218
Radiomics score	3.555 (2.789–4.532)	0.000*	4.295 (3.098–5.953)	0.000*

*In univariate analysis, variables with $P < 0.1$ were included in multivariate logistic regression analysis. In multivariate analysis, $P < 0.05$ was considered significant. AFP, alpha fetoprotein; CA19-9, carbohydrate antigen 19-9; CEA, carcino-embryonic antigen; VEGF, vascular endothelial growth factor. * represents $P < 0.0001$.*

used in radiomics researches. The basic idea of LASSO is to minimize the residual sum of squares under the constraint that the sum of the absolute values of the regression coefficients is less than a constant, so as to produce some regression coefficients strictly equal to 0 to get an interpretable model. Essentially, it is a process of seeking a sparse expression of the model (39, 40). Random forest is an ensemble learning algorithm based on decision tree analysis and has a good performance in classification and regression. Random forest can also be used as a feature selection technology, and it has been widely used in machine learning, determining the importance of features during model training (28, 41, 42).

The texture features showed high importance in our prediction model. Image texture is a visual feature that reflects homogeneous phenomena in the image, and it reflects the surface structure organization and arrangement properties of the object with slow or periodic changes. The texture can be layered by the statistical order of the information encoded in the image, which can be divided into first-order texture features, second-order texture features, and high-order texture features (43). Texture features are widely recognized as quantitative biomarkers of tumor heterogeneity (44, 45).

The large sample size of our study helped to improve the generality and stability of our results. However, our research also has certain limitations. First, all ultrasound imaging data were from a unitary center, and the study was retrospective in nature. The grayscale ultrasound images used in our study were collected by different commercial ultrasound systems. Although the data extracted from the images were preprocessed, the imaging of different instruments may still have some influence on the results of feature extraction, so whether the model can play a prospective role remains an open question. Therefore, it is necessary to conduct a multicenter prospective study with a rigorous control of ultrasound machines to

TABLE 4 | Results of the univariate and multivariate analyses in ICC-vs-cHCC-ICC Model.

Factors (reference)	Univariate analysis		Multivariate analysis	
	OR (95% CI)	*p* value	OR (95% CI)	*p* value
Gender (female)	2.943 (1.272–6.814)	0.012	1.924 (0.638–5.806)	0.245
Age (years)				
<40	Reference		–	–
40–60	0.431 (0.129–1.434)	0.170	–	–
>60	0.560 (0.202–1.551)	0.265	–	–
Tumor size (>5 cm)	1.781 (0.874–3.632)	0.112	–	–
Hepatitis (yes)	0.246 (0.109–0.554)	0.001	0.572 (0.178–1.832)	0.347
Cirrhosis (yes)	0.413 (0.199–0.854)	0.017	0.700 (0.232–2.112)	0.527
AFP (µg/ml)				
≤400	Reference		Reference	
>400	0.198 (0.077–0.507)	0.001	0.205 (0.057–0.735)	0.015
CA19-9 (U/ml)				
≤37	Reference		Reference	
>37	2.449 (1.128–5.318)	0.024	1.222 (0.400–3.740)	0.725
CEA (µg/ml)				
≤5	Reference		Reference	
>5	6.190 (1.765–21.711)	0.004	4.554 (0.919–22.571)	0.063
Differentiation				
Well/Moderate	Reference		–	–
Poor	2.174 (0.849–5.569)	0.106	–	–
Immunohistochemistry, Negative/Positive				
Ki67, ≤ 10%/> 10%	0.703 (0.294–1.678)	0.427	–	–
P53	0.733 (0.344–1.565)	0.423	–	–
VEGF	0.523 (0.257–1.065)	0.074	0.570 (0.211–1.540)	0.267
Microvascular invasion	0.596 (0.284–1.250)	0.171	–	–
Stage				
I	Reference		–	–
II	0.302 (0.059–1.559)	0.153	–	–
III	0.281 (0.052–1.523)	0.141	–	–
IV	1.417 (0.240–8.367)	0.701	–	–
Radiomics score	2.292 (1.662–3.160)	0.000*	2.395 (1.636–3.506)	0.000*

*In univariate analysis, variables with $P < 0.1$ were included in multivariate logistic regression analysis. In multivariate analysis, $P < 0.05$ was considered significant. AFP, alpha fetoprotein; CA19-9, carbohydrate antigen 19-9; CEA, carcino-embryonic antigen; VEGF, vascular endothelial growth factor. * represents $P < 0.0001$.*

further explore the diagnostic potential of radiomics-based modeling. Second, our study included only PLC and did not include benign and metastatic tumors of the liver. The identification of more types of tumors is more challenging. We will add data for other types of liver tumors in future studies to optimize the universality and clinical value of the model. Third, we took into account the characteristics of general clinical applications of ultrasound, and this is a retrospective study, so we finally adopted two-dimensional ultrasound images. However, the quantitative features extracted based on two-dimensional ultrasound images cannot stand for the overall lesion, and a more precise radiomics analysis depends on the acquisition of 3D images. Further research on three-dimensional ultrasound radiomics is necessary in the future. Fourth, our study focused on the relationship between high-throughput imaging features extracted from tumor ROI and pathological typing. In order to quantify the heterogeneity of tumors more comprehensively, it is necessary to pay more attention to the peritumoral information and combine more clinicopathological information to establish a more accurate individualized disease assessment model. Therefore, in the future, we need to optimize our model based on the above limitations and carry out prospective studies, which may be helpful to improve the discrimination performance of radiomics model for PLC.

In summary, we developed and validated the ultrasound-based radiomics models to distinguish different histopathological types of PLC, thus providing a new approach for doctors to non-invasively identify HCC, cHCC-ICC, and ICC.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

ETHICS STATEMENT

This study was approved by the Ethics Committee of the First Affiliated Hospital of Guangxi Medical University.

AUTHOR CONTRIBUTIONS

HY and YH: guarantor of the article. HY, YH, YP, and PL: conception and design. YP, PL, LW, YZ, LL, XM, DW, YL, and HQ: collection and assembly of data. YP, PL, LW, XL, and XW: data analysis and interpretation. All authors: manuscript writing and final approval of manuscript.

FUNDING

The study was supported by funds from the Guangxi Science and Technology Program (Grant no. GuiKeAB17195020) and the National Natural Science Foundation of China (Grant nos. NSFC81860319 and NSFC81960329).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2020.01646/full#supplementary-material

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin*. (2019) 69:7–34. doi: 10.3322/caac.21551
2. Sia D, Villanueva A, Friedman SL, Llovet JM. Liver cancer cell of origin, molecular class, and effects on patient prognosis. *Gastroenterology*. (2017) 152:745–61. doi: 10.1053/j.gastro.2016.11.048
3. Connell LC, Harding JJ, Shia J, Abou-Alfa GK. Combined intrahepatic cholangiocarcinoma and hepatocellular carcinoma. *Chin Clin Oncol*. (2016) 5:66. doi: 10.21037/cco.2016.10.02
4. Akiba J, Nakashima O, Hattori S, Tanikawa K, Takenaka M, Nakayama M, et al. Clinicopathologic analysis of combined hepatocellular-cholangiocarcinoma according to the latest WHO classification. *Am J Surg Pathol*. (2013) 37:496–505. doi: 10.1097/PAS.0b013e31827332b0
5. Stavraka C, Rush H, Ross P. Combined hepatocellular cholangiocarcinoma (cHCC-CC): an update of genetics, molecular biology, and therapeutic interventions. *J Hepatocell Carcinoma*. (2019) 6:11–21. doi: 10.2147/JHC.S159805
6. O'Connor K, Walsh JC, Schaeffer DF. Combined hepatocellular-cholangiocarcinoma (cHCC-CC): a distinct entity. *Ann Hepatol*. (2014) 13:317–22.
7. Lee JH, Chung GE, Yu SJ, Hwang SY, Kim JS, Kim HY, et al. Long-term prognosis of combined hepatocellular and cholangiocarcinoma after curative resection comparison with hepatocellular carcinoma and cholangiocarcinoma. *J Clin Gastroenterol*. (2011) 45:69–75. doi: 10.1097/MCG.0b013e3181ce5dfa
8. Mazzaferro V, Regalia E, Doci R, Andreola S, Pulvirenti A, Bozzetti F, et al. Liver transplantation for the treatment of small hepatocellular carcinomas in patients with cirrhosis. *N Engl J Med*. (1996) 334:693–9. doi: 10.1056/NEJM199603143341104
9. Park YH, Hwang S, Ahn CS, Kim KH, Moon DB, Ha TY, et al. Long-term outcome of liver transplantation for combined hepatocellular carcinoma and cholangiocarcinoma. *Transplant Proc*. (2013) 45:3038–40. doi: 10.1016/j.transproceed.2013.08.056
10. Chang CC, Chen YJ, Huang TH, Chen CH, Kuo FY, Eng HL, et al. Living donor liver transplantation for combined hepatocellular carcinoma and cholangiocarcinoma: experience of a single center. *Ann Transplant*. (2017) 22:115–20. doi: 10.12659/aot.900779
11. Lee DD, Croome KP, Musto KR, Melendez J, Tranesh G, Nakhleh R, et al. Liver transplantation for intrahepatic cholangiocarcinoma. *Liver Transplant*. (2018) 24:634–44. doi: 10.1002/lt.25052
12. Pinna AD, Yang T, Mazzaferro V, De Carlis L, Zhou J, Roayaie S, et al. Liver transplantation and hepatic resection can achieve cure for hepatocellular carcinoma. *Ann Surg*. (2018) 268:868–75. doi: 10.1097/SLA.0000000000002889
13. Meirelles Junior RF, Salvalaggio P, Rezende MB, Evangelista AS, Guardia BD, Matielo CE, et al. Liver transplantation: history, outcomes and perspectives. *Einstein*. (2015) 13:149–52. doi: 10.1590/S1679-45082015RW3164
14. Potretzke TA, Tan BR, Doyle MB, Brunt EM, Heiken JP, Fowler KJ. Imaging features of biphenotypic primary liver carcinoma (Hepatocholangiocarcinoma) and the potential to mimic hepatocellular carcinoma: LI-RADS analysis of CT and MRI features in 61 cases. *AJR Am J Roentgenol*. (2016) 207:25–31. doi: 10.2214/AJR.15.14997
15. Li F, Han J, Han F, Wang JW, Luo RZ, Li AH, et al. Combined hepatocellular cholangiocarcinoma (Biphenotypic) tumors: potential role of contrast-enhanced ultrasound in diagnosis. *AJR Am J Roentgenol*. (2017) 209:767–74. doi: 10.2214/AJR.16.17513
16. Gera S, Ettel M, Acosta-Gonzalez G, Xu R. Clinical features, histology, and histogenesis of combined hepatocellular-cholangiocarcinoma. *World J Hepatol*. (2017) 9:300–9. doi: 10.4254/wjh.v9.i6.300
17. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology*. (2016) 278:563–77. doi: 10.1148/radiol.2015151169
18. Gu D, Hu Y, Ding H, Wei J, Chen K, Liu H, et al. CT radiomics may predict the grade of pancreatic neuroendocrine tumors: a multicenter study. *Eur Radiol*. (2019) 29:6880–90. doi: 10.1007/s00330-019-06176-x
19. Hu HT, Wang Z, Huang XW, Chen SL, Zheng X, Ruan SM, et al. Ultrasound-based radiomics score: a potential biomarker for the prediction of microvascular invasion in hepatocellular carcinoma. *Eur Radiol*. (2019) 29:2890–901. doi: 10.1007/s00330-018-5797-0
20. Ortiz-Ramon R, Larroza A, Ruiz-Espana S, Arana E, Moratal D. Classifying brain metastases by their primary site of origin using a radiomics approach based on texture analysis: a feasibility study. *Eur Radiol*. (2018) 28:4514–23. doi: 10.1007/s00330-018-5463-6
21. Yin P, Mao N, Zhao C, Wu J, Chen L, Hong N. A triple-classification radiomics model for the differentiation of primary chordoma, giant cell tumor, and metastatic tumor of sacrum based on T2-weighted and contrast-enhanced T1-weighted MRI. *J Magn Reson Imaging*. (2019) 49:752–9. doi: 10.1002/jmri.26238
22. Kamarajah SK, Frankel TL, Sonnenday C, Cho CS, Nathan H. Critical evaluation of the American joint commission on cancer (AJCC) 8th edition staging system for patients with hepatocellular carcinoma (HCC): a surveillance, epidemiology, end results (SEER) analysis. *J Surg Oncol*. (2018) 117:644–50. doi: 10.1002/jso.24908
23. Lee AJ, Chun YS. Intrahepatic cholangiocarcinoma: the AJCC/UICC 8th edition updates. *Chin Clin Oncol*. (2018) 7:52. doi: 10.21037/cco.2018.07.03
24. Lucia F, Visvikis D, Vallieres M, Desseroit MC, Miranda O, Robin P, et al. External validation of a combined PET and MRI radiomics model for prediction of recurrence in cervical cancer patients treated with chemoradiotherapy. *Eur J. Nuclear Med Mol Imaging*. (2019) 46:864–77. doi: 10.1007/s00259-018-4231-9
25. Orlhac F, Frouin F, Nioche C, Ayache N, Buvat I. Validation of a method to compensate multicenter effects affecting CT radiomics. *Radiology*. (2019) 291:53–9. doi: 10.1148/radiol.2019182023
26. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res*. (2017) 77:e104–7. doi: 10.1158/0008-5472.CAN-17-0339
27. Zwanenburg A, Leger S, Vallières M, Löck S. Image biomarker standardisation initiative. *arXiv* [preprint]. (2016):doi: 10.1148/radiol.2020191145
28. Yu L, Tao G, Zhu L, Wang G, Li Z, Ye J, et al. Prediction of pathologic stage in non-small cell lung cancer using machine learning algorithm based on CT

image feature analysis. *BMC Cancer*. (2019) 19:464. doi: 10.1186/s12885-019-5646-9

29. Shiri I, Maleki H, Hajianfar G, Abdollahi H, Ashrafinia S, Hatt M, et al. Next-generation radiogenomics sequencing for prediction of EGFR and KRAS mutation status in NSCLC patients using multimodal imaging and machine learning algorithms. *Mol Imaging Biol*. (2020) 22:1132–48. doi: 10.1007/s11307-020-01487-8
30. Parmar C, Grossmann P, Rietveld D, Rietbergen MM, Lambin P, Aerts HJ. Radiomic machine-learning classifiers for prognostic biomarkers of head and neck cancer. *Front Oncol*. (2015) 5:272. doi: 10.3389/fonc.2015.00272
31. Hajianfar G, Shiri I, Maleki H, Oveisi N, Haghparast A, Abdollahi H, et al. Noninvasive O6 methylguanine-DNA methyltransferase status prediction in glioblastoma multiforme cancer using magnetic resonance imaging radiomics features: univariate and multivariate radiogenomics analysis. *World Neurosurg*. (2019) 132:e140–61. doi: 10.1016/j.wneu.2019.08.232
32. Leger S, Zwanenburg A, Pilz K, Lohaus F, Linge A, Zophel K, et al. A comparative study of machine learning methods for time-to-event survival data for radiomics risk modelling. *Sci Rep*. (2017) 7:13206. doi: 10.1038/s41598-017-13448-3
33. Saini A, Breen I, Pershad Y, Naidu S, Knuttinen MG, Alzubaidi S, et al. Radiogenomics and radiomics in liver cancers. *Diagnostics (Basel)*. (2018) 9:4. doi: 10.3390/diagnostics9010004
34. Youk JH, Kwak JY, Lee E, Son EJ, Kim JA. Grayscale ultrasound radiomic features and shear-wave elastography radiomic features in benign and malignant breast masses. *Ultraschall Med*. (2019) 41:390–6. doi: 10.1055/a-0917-6825
35. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol*. (2017) 14:749–62. doi: 10.1038/nrclinonc.2017.141
36. Wang Y, Yang Q, Li S, Luo R, Mao S, Shen J. Imaging features of combined hepatocellular and cholangiocarcinoma compared with those of hepatocellular carcinoma and intrahepatic cholangiocellular carcinoma in a Chinese population. *Clin Radiol*. (2019) 74:407.e1–407.10. doi: 10.1016/j.crad.2019.01.016
37. Lewis S, Peti S, Hectors SJ, King M, Rosen A, Kamath A, et al. Volumetric quantitative histogram analysis using diffusion-weighted magnetic resonance imaging to differentiate HCC from other primary liver cancers. *Abdom Radiol*. (2019) 44:912–22. doi: 10.1007/s00261-019-01906-7
38. Panjala C, Senecal DL, Bridges MD, Kim GP, Nakhleh RE, Nguyen JH, et al. The diagnostic conundrum and liver transplantation outcome for combined hepatocellular-cholangiocarcinoma. *Am J Transplant*. (2010) 10:1263–7. doi: 10.1111/j.1600-6143.2010.03062.x
39. Ji GW, Zhu FP, Zhang YD, Liu XS, Wu FY, Wang K, et al. A radiomics approach to predict lymph node metastasis and clinical outcome of intrahepatic cholangiocarcinoma. *Eur Radiol*. (2019) 29:3725–35. doi: 10.1007/s00330-019-06142-7
40. Wu S, Zheng J, Li Y, Yu H, Shi S, Xie W, et al. A radiomics nomogram for the preoperative prediction of lymph node metastasis in bladder cancer. *Clin Cancer Res*. (2017) 23:6904–11. doi: 10.1158/1078-0432.CCR-17-1510
41. Kaissis G, Ziegelmayer S, Lohofer F, Algul H, Eiber M, Weichert W, et al. A machine learning model for the prediction of survival and tumor subtype in pancreatic ductal adenocarcinoma from preoperative diffusion-weighted imaging. *Eur Radiol Exp*. (2019) 3:41. doi: 10.1186/s41747-019-0119-0
42. Chen T, Li M, Gu Y, Zhang Y, Yang S, Wei C, et al. Prostate cancer differentiation and aggressiveness: assessment with a radiomic-based model vs. PI-RADS v2. *J. Magn Reson Imaging*. (2019) 49:875–84. doi: 10.1002/jmri.26243
43. Chitalia RD, Kontos D. Role of texture analysis in breast MRI as a cancer biomarker: a review. *J Magn Reson Imaging*. (2019) 49:927–38. doi: 10.1002/jmri.26556
44. Yu H, Scalera J, Khalid M, Touret AS, Bloch N, Li B, et al. Texture analysis as a radiomic marker for differentiating renal tumors. *Abdom Radiol*. (2017) 42:2470–8. doi: 10.1007/s00261-017-1144-1
45. Kirienko M, Cozzi L, Rossi A, Voulaz E, Antunovic L, Fogliata A, et al. Ability of FDG PET and CT radiomics features to differentiate between primary and metastatic lung lesions. *Eur J Nuclear Med Mol Imaging*. (2018) 45:1649–60. doi: 10.1007/s00259-018-3987-2

Conflict of Interest: XL and XW were employed by GE Healthcare.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling editor declared a past co-authorship with one of the authors XL.

frontiers in Oncology

ORIGINAL RESEARCH
published: 27 October 2020
doi: 10.3389/fonc.2020.581321

Predicting Axillary Lymph Node Status With a Nomogram Based on Breast Lesion Ultrasound Features: Performance in N1 Breast Cancer Patients

Yanwen Luo†, Chenyang Zhao†, Yuanjing Gao, Mengsu Xiao, Wenbo Li, Jing Zhang, Li Ma, Jing Qin, Yuxin Jiang and Qingli Zhu**

Department of Ultrasound, Chinese Academy of Medical Sciences and Peking Union Medical College Hospital, Beijing, China

Edited by:
Katsutoshi Sugimoto,
Tokyo Medical University, Japan

Reviewed by:
Hirohito Takeuchi,
Tokyo Medical University, Japan
Mutsumi Nishida,
Hokkaido University Hospital, Japan

***Correspondence:**
Yuxin Jiang
jiangyuxinxh@163.com
Qingli Zhu
zqlpumch@126.com

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: 08 July 2020
Accepted: 29 September 2020
Published: 27 October 2020

Citation:
Luo Y, Zhao C, Gao Y, Xiao M, Li W, Zhang J, Ma L, Qin J, Jiang Y and Zhu Q (2020) Predicting Axillary Lymph Node Status With a Nomogram Based on Breast Lesion Ultrasound Features: Performance in N1 Breast Cancer Patients. Front. Oncol. 10:581321. doi: 10.3389/fonc.2020.581321

Objective: To develop a nomogram for predicting axillary lymph node (ALN) metastases using the breast imaging reporting and data system (BI-RADS) ultrasound lexicon.

Methods: A total of 703 patients from July 2015 to January 2018 were included in this study as a primary cohort for model construction. Moreover, 109 patients including 51 pathologically confirmed N1 patients (TNM staging) and 58 non-metastatic patients were recruited as an external validation cohort from March 2018 to August 2019. Ultrasound images and clinical information of these patients were retrospectively reviewed. The ultrasonic features based on the BI-RADS lexicon were extracted by two radiologists. The features extracted from the primary cohort were used to develop a nomogram using multivariate analysis. Internal and external validations were performed to evaluate the predictive efficacy of the nomogram.

Results: The nomogram was based on two features (size, lesion boundary) and showed an area under the curve of 0.75 (95% confidence interval [CI], 0.70–0.79) in the primary cohort and 0.91 (95% CI, 0.84–0.97) in the external validation cohort; it achieved an 88% sensitivity in N1 patients.

Conclusion: The nomogram based on BI-RADS ultrasonic features can predict breast cancer ALN status with relatively high accuracy. It has potential clinical value in improving the sensitivity and accuracy of the preoperative diagnosis of ALN metastases, especially for N1 patients.

Keywords: nomogram, breast cancer, axillary lymph node metastasis, ultrasound, prediction model

INTRODUCTION

Breast cancer, posing a serious threat to women's health and social economy, has drawn great attention from researchers for years (1). Axillary lymph node (ALN) status plays an essential role in treatment planning for breast cancer (2), being the most significant prognostic indicator for early stage patients (3). Preoperative staging of ALN status can make a way for optimized clinical decision making. While, currently recognized method for identifying ALN status is sentinel lymph node biopsy (SLNB), which is performed during surgery and requires pathological diagnosis. The SLNB-negative patients would be diagnosed as pN0 in TNM staging (4, 5).

In current clinical practice, axillary ultrasound (US) is commonly recommended for all patients with breast cancer to evaluate ALN status preoperatively (6, 7). However, the SLN cannot be identified by grayscale US, and metastases of isolated tumor cells or micro-metastases are not visible on US. As a consequence, it is difficult for conventional US to achieve high accuracy in identifying axillary nodal metastases. It was reported that US has a sensitivity of 45% to 87% in diagnosing ALN metastases and specificity of 55% to 97% (8). Zhang et al. proved that among N1-3 patients, axillary US had the highest false-negative rate in pathologic N1 patients (9). Hence, it is crucial to improve the preoperative diagnostic accuracy of US in identifying ALN metastases, especially for patients with a minimal number of abnormal nodes.

Previous studies have demonstrated that some ultrasonic features of breast lesions, such as tumor size, margin, and location might be associated with breast cancer nodal metastases and thus can help predict ALN status (10–13). However, in those studies, US findings and tumor clinicopathologic characteristics were simultaneously incorporated to predict ALN metastases (11–13), or a risk model was developed for predicting ALN metastases in a subgroup of patients with invasive ductal carcinoma (10, 11, 13). Considering that the clinicopathologic characteristics, such as histological type, histological grade, and molecular subtype, might directly be related to the probability of ALN metastases, it is necessary to explore the independent contributions of breast lesion US features in determining the likelihood of positive lymph nodes in a preoperative patient population. Therefore, we aimed to construct a predictive model for ALN metastases based on breast lesion US features, to investigate the feasibility of using only US features in identifying nodal metastases preoperatively.

In this study, we summarize the ultrasonic features of the malignant lesions using the breast imaging reporting and data system (BI-RADS) lexicon, the widely accepted standard for defining ultrasonic feature of breast lesions (14). We analyzed the correlations of these ultrasonic features with nodal metastases, developed an ALN metastases predictive model based on these features, and presented it as a nomogram. Such a tool is expected to improve preoperative diagnostic efficacy, especially for N1 patients.

MATERIALS AND METHODS

This study is retrospective and was approved by the Institutional Review Board of Perking Union Medical College Hospital.

Patient Recruitment

A total of 1,024 female patients with breast cancer were enrolled consecutively for model construction and internal validation from July 2015 to January 2018. The clinical data, US images, and pathological results were reviewed. The inclusion and exclusion criteria for establishing the primary and internal validation cohorts were as follows.

Inclusion criteria:

(1) patients pathologically diagnosed as having breast cancer;
(2) ALN status clearly illustrated by pathology after SLNB or ALN dissection (ALND);
(3) breast US scanning performed within one month before surgery;
(4) only a single lesion pathologically identified in each patient, with a diameter less than 5 cm (T1 and T2 stage).

Exclusion criteria:

(1) neoadjuvant chemotherapy or biopsy performed before US scanning;
(2) multiple malignant lesions;
(3) target neoplasms that could not be visualized on US;
(4) incomplete clinical and pathological information.

Finally, a total of 703 consecutive patients were included in this study for model construction and internal validation from July 2015 to January 2018. Then, to validate the efficacy of the prediction model in early breast cancer patients, based on the inclusion and exclusion criteria described above, another 109 patients with pN1/pN0 were recruited at 1:1 ratio as the external validation cohort after primary cohort (From March 2018 to August 2019). Including 51 patients classified as having N1 according to the TNM classification (with one to three metastatic ALN nodes) by postoperative pathology and 58 patients with no ALN metastases (15).

Clinical and Pathological Information Collection

The clinical and pathological features of the patients, including age, pathological results, and ALN status (LN-positive or LN-negative), were extracted from the medical records.

Ultrasound Scanning and Imaging Acquisition

All the included patients underwent US scanning before surgery in our Department. Our study did not specify US equipment. The high-quality US images are acquired by four different commercial US devices, which are RS85A (Samsung), IU22 (Philips), Logic 9 (GE) and RS85A (Samsung) with Linear probes (3–12 MHz, centered at 10 MHz). And do not affect the handcrafted extraction of BI-RADS features. The recorded imaging data of the patients were carefully reviewed and selected for further analysis by one experienced radiologist (QZ, 23-year experience in breast US), blinded to the clinical and pathological results. The grayscale and color-Doppler ultrasonic images of both longitudinal section and cross-section were acquired for feature extraction. The largest diameter of each lesion was measured on the grayscale US images.

BI-RADS-Based US Feature Extraction

Referring to the BI-RADS lexicon and previous researches (16–18), a total of eight ultrasonic features were selected in this study as evaluation indices (**Table 1**). Image reading and feature extraction were conducted by the two radiologists (CZ, 4-year experience in breast US, and YL, 2-year experience in breast US), who were also blinded to the patient's clinical and pathological information. As discrepancies occurred, the agreement would be reached through discussion. Before participating in the study, the two radiologists received systematic training on the BI-RADS lexicon. Inter-observer reliability was assessed by comparing the results of the 2 radiologists in 100 randomly chosen lesions. CZ performed the second feature extraction from 100 randomly selected lesions after 1 week with the same procedure. Then by comparing the results of CZ at two different time points evaluated intra-observer reliability. Finally the inter-observer and intra-observer agreement were measured by kappa statistics.

Model Construction and Validation

The prediction model was built based on multivariate logistic regression analysis. Before construction, multicollinearity analysis was performed by calculating the variance inflation factor (VIF) among the features; a VIF value > 10 was considered to indicate multicollinearity, and the corresponding variables were excluded from the model. All the US features were modeled as categorical data with a dummy variable, adding age as continuous variables, to construct models. In multivariate models, a backward stepwise variable selection procedure was used for model selection based on the Akaike information criterion (AIC). The final model thus built was tested for predictive power using both internal and external validation. Internal validation was performed with the bootstrap resampling method by randomly drawing 500 samples from the primary dataset to avoid overoptimism. The developed model underlying the nomogram was used to predict ALN status of the patients in the external validation cohort. The diagnostic performance of the model in the primary and validation cohorts was evaluated by calculating sensitivity, specificity, positive likelihood ratio, negative likelihood ratio, positive predictive value, and negative predictive value. Receiver operating curves (ROC) and the corresponding area under the curve (AUC) values were used to assess the discriminating ability of the nomogram.

Statistical Analysis

Statistical analysis was performed using R (http://www.R-project.org) and EmpowerStats software (X&Y Solutions). The variables were compared using Student's t-test (continuous data) and the Pearson chi-squared test (categorical data). Continuous variables are expressed as the mean ± SD, categorical variables as percentages (%), and p values < 0.05 were considered statistically significant. The degree of intra-observer and inter-observer agreement between the two readers was measured using the κ value, which was interpreted as follows: $\kappa < 0$, poor agreement; $0 < \kappa < 0.20$, slight agreement; $0.20 < \kappa < 0.40$, fair agreement; $0.40 < \kappa < 0.60$, moderate agreement; $0.60 < \kappa < 0.80$, substantial agreement; and $0.80 < \kappa < 1$, perfect agreement. The "glm" function was used for the univariate and multivariate logistic regression analyses. The "Hmisc" package was used to plot the nomogram. The "pROC" package was used to plot the ROC curves and measure the AUCs. The "calibration curve" function was used to plot the calibration curves.

TABLE 1 | Extracted US features.

Feature		Number	Description
Shape	regular	1	A mass that is oval (egg-shaped or elliptical) or round (spherical, ball-shaped).
	irregular	2	Neither oval nor round.
Orientation	horizontal	1	The long axis of the lesion is parallel to the skin line ("wider-than-tall").
	vertical	2	The anterior-posterior or vertical dimension is greater than the transverse or horizontal dimension ("taller-than-wide").
Margin	circumscribed	1	The demarcation is well defined and clear, with abrupt transition between the lesion and the surrounding tissue.
	not circumscribed	2	The boundary is poorly defined, and can be characterized as indistinct, angular, microlobulated, or spiculated.
Lesion boundary	abrupt interface	1	The demarcation between the lesion and the surrounding tissue is imperceptible or is a distinct well-defined echogenic rim without any thickness.
	echogenic halo	2	A band bridged by an echogenic transition zone can be perceived.
Echo pattern	hypoechoic	1	The mass has decreased echogenicity compared with fat.
	complex	2	A complex mass containing both anechoic (cystic) and echogenic (solid) components.
Posterior acoustic features	no	1	No shadowing or enhancement is present deep in the mass; the echogenicity of the area immediately behind the mass is not different from that of the adjacent tissue at the same depth.
	enhancement	2	Sound transmission is unimpeded in its passage through the mass. Enhancement appears as a more echogenic (whiter) column deep into the mass. Enhancement is a criterion for cyst diagnosis.
	shadowing	3	Shadowing, i.e., posterior attenuation of acoustic transmission. Sonographically, the area posterior to the mass appears darker.
Calcification	no	1	No calcification.
	macrocalcification	2	Macrocalcifications: coarse calcifications 0.5 mm or greater in size are depicted.
	microcalcification	3	Microcalcifications embedded in the mass are well depicted. The punctate, hyperechoic foci appear conspicuous in a hypoechoic mass.
Vascularity	no	1	Little or No vascularity.
	adjacent	2	present immediately adjacent to lesion
	diffusely increased	3	Diffusely increased vascularity surrounding lesion.

RESULTS

Clinical Characteristics and Ultrasonic Features of the Primary and External Validation Cohorts

Table 2 shows the baseline clinical characteristics and ultrasonic features of the 703 patients in the primary cohort and 109 patients in the external validation cohort. A total of 167 (23.9%) patients with ALN metastases were included in the primary cohort and 51 patients (46.8%) with ALN metastases in the external validation cohort.

The inter-operator agreements for the ultrasonic features ranged from 0.77 to 0.92 (shape: 0.87; orientation: 0.92; margin: 0.91; lesion boundary: 0.77; echo pattern: 0.92; posterior acoustic features: 0.90; calcification: 0.78; vascularity: 0.81). The intra-operator agreements for the ultrasonic features ranged from 0.79 to 0.96 (shape: 0.91; orientation: 0.94; margin: 0.89; lesion boundary: 0.82; echo pattern: 0.96; posterior acoustic features: 0.91; calcification: 0.85; vascularity: 0.90).

Diagnostic Performance of the Nomogram

Using multivariate logistic regression analysis, several multivariate models were generated. And after stepwise model selection, two features showed independent correlation with the

TABLE 2 | Baseline characteristics in the primary and external validation cohorts.

Variable	Primary cohort			External validation cohort		
	Negative for LN metastasis (n = 536)	**Positive for LN metastasis(n = 167)**	**P-value**	**Negative for LN metastasis (n = 58)**	**Positive for LN metastasis (n = 51)**	**P-value**
Age	51.3 ± 11.6	50.6 ± 11.4	0.517	51.6 ± 11.5	55.4 ± 11.6	0.088
Size	2.1 ± 0.9	2.7 ± 1.0	<0.001	1.7 ± 0.8	2.5 ± 1.1	<0.001
Shape			<0.001			<0.001
regular	206 (38.4%)	35 (21.0%)		46 (79.3%)	4 (7.8%)	
irregular	330 (61.6%)	132 (79.0%)		12 (20.7%)	47 (92.2%)	
Orientation			0.342			0.002
horizontal	327 (61.0%)	95 (56.9%)		43 (74.1%)	23 (45.1%)	
vertical	209 (39.0%)	72 (43.1%)		15 (25.9%)	28 (54.9%)	
Margin			0.004			<0.001
circumscribed	82 (15.3%)	11 (6.6%)		58 (100.0%)	1 (2.0%)	
not circumscribed	454 (84.7%)	156 (93.4%)		0 (0.0%)	50 (98.0%)	
Lesion boundary			<0.001			<0.001
abrupt interface	327 (61.0%)	40 (24.0%)		55 (94.8%)	12 (23.5%)	
echogenic halo	209 (39.0%)	127 (76.0%)		3 (5.2%)	39 (76.5%)	
Echo pattern			0.333			0.056
hypoechoic	524 (97.8%)	161 (96.4%)		54 (93.1%)	51 (100.0%)	
complex	12 (2.2%)	6 (3.6%)		4 (6.9%)	0 (0.0%)	
Posterior acoustic features			0.054			<0.001
no	396 (73.9%)	109 (65.3%)		39 (67.2%)	24 (47.1%)	
enhance	69 (12.9%)	24 (14.4%)		18 (31.0%)	11 (21.6%)	
decrease	71 (13.2%)	34 (20.4%)		1 (1.7%)	16 (31.4%)	
Calcification			0.531			0.556
no	375 (70.0%)	110 (65.9%)		47 (81.0%)	37 (72.5%)	
macro	6 (1.1%)	3 (1.8%)		2 (3.4%)	2 (3.9%)	
micro	6 (1.1%)	54 (32.3%)		9 (15.5%)	12 (23.5%)	
Vascularity			0.050			0.068
no	210 (39.2%)	48 (28.7%)		21 (36.2%)	13 (25.5%)	
minimal	202 (37.7%)	73 (43.7%)		18 (31.0%)	27 (52.9%)	
abundant	124 (23.1%)	46 (27.5%)		19 (32.8%)	11 (21.6%)	
Histological type			0.163			0.471
invasive ductal carcinoma	396(73.9%)	135(80.8%)		42 (72.4%)	41 (80.4%)	
invasive lobular carcinoma	23(4.3%)	7(4.2%)		9 (15.5%)	5(9.8%)	
ductal carcinoma in situ	101(18.8%)	19(11.4%)		4 (6.9%)	2 (3.9%)	
Others	16(3.0%)	6(3.6%)		3 (5.2%)	3 (5.9%)	
pN status			<0.001			<0.001
pN0	536(100%)	0(0.0%)		58(100%)	(0.0%)	
pN1	0(0.0%)	98(58.7%)		(0.0%)	51(100%)	
pN2	0(0.0%)	29(17.4%)		(0.0%)	(0.0%)	
pN3	0(0.0%)	40(23.9%)		(0.0%)	(0.0%)	

risk of ALN metastases (**Table 3**) and thus were incorporated into the final nomogram, namely, size and lesion boundary. The nomogram is presented in **Figure 1**.

TABLE 3 | Results of Univariate and Multivariate logistic regression analysis in the primary cohort.

Exposure	Univariate analyses	Final multivariate model
Age	1.0 (1.0, 1.0) 0.734	
Size	1.7 (1.4, 2.1) < 0.001	1.7 (1.4, 2.0) < 0.001
Shape		
regular	1.0	
irregular	2.1 (1.3, 3.4) 0.002	
Orientation		
horizontal	1.0	
vertical	1.2 (0.8, 1.7) 0.501	
Margin		
circumscribed	1.0	
not circumscribed	1.6 (0.8, 3.2) 0.169	
Lesion boundary		
abrupt interface	1.0	1.0
echogenic halo	3.7 (2.4, 5.8) < 0.001	4.5 (3.0, 6.7) < 0.001
Echo pattern		
hypoechoic	1.0	
complex	1.5 (0.5, 4.5) 0.434	
Posterior acoustic features		
no	1.0	
enhance	1.1 (0.6, 2.1) 0.651	
decrease	1.4 (0.8, 2.4) 0.249	
Calcification		
no	1.0	
macro	2.2 (0.5, 9.5) 0.285	
micro	1.3 (0.8, 2.0) 0.250	
Vascularity		
no	1.0	
minimal	1.4 (0.9, 2.3) 0.141	
abundant	1.6 (0.9, 2.7) 0.105	

The diagnostic performance of the nomogram in the primary dataset is shown in **Table 4**. The ROC curve of the nomogram showed good predictive power, with an AUC of 0.75 [95% confidence interval (CI), 0.70–0.79] (**Figure 2**).

Good calibration was observed for the probability of ALN metastases in the primary cohort (**Figure 3**).

Nomogram Validation in N1 Patients

An external validation cohort of 109 patients was enrolled using the same criteria used to select the primary cohort and included 51 patients (46.8%) with ALN metastases (the mean number of metastatic ALN nodes was 1.57). The nomogram demonstrated good predictive power (**Table 4**) with an AUC of 0.91 (95% CI: 0.84–0.97) in these N1 patients (**Figure 4**).

DISCUSSION

Axillary imaging plays an essential role in evaluating ALN status. Axillary US is the primary method for evaluation of axillary

TABLE 4 | Diagnostic performance of the nomogram.

Diagnostic performance	P set	EV Set
AUC	0.7468 (0.7038–0.7898)	0.9065 (0.8424–0.9707)
Specificity	0.6124	0.8966
Sensitivity	0.7711	0.8824
Accuracy	0.6500	0.8899
Positive likelihood ratio	1.9892	8.5294
Negative likelihood ratio	0.3738	0.1312
Positive predictive value	0.3821	0.8824
Negative predictive value	0.8959	0.8966

P, Primary dataset; EV, External validation dataset.

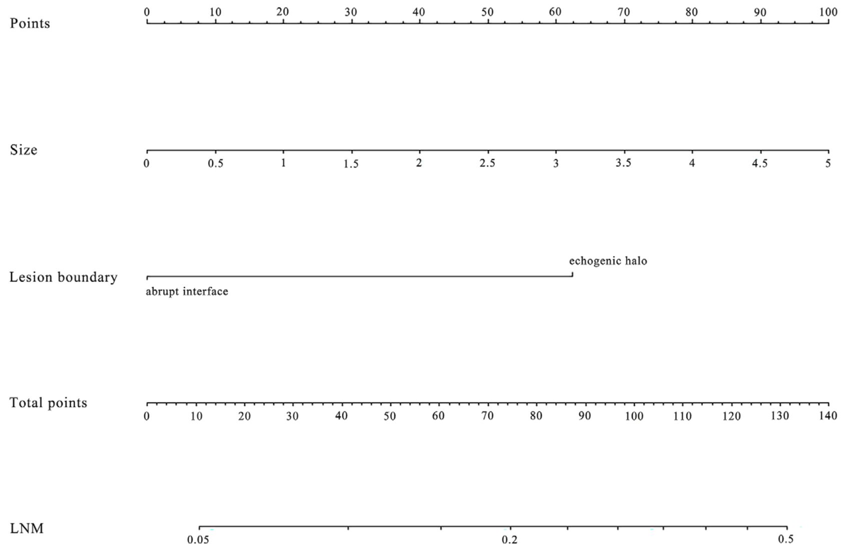


FIGURE 1 | The nomogram was developed in the primary dataset. it included two factors (size, lesion boundary). The nomogram plot provides a visual way to predict the risk of LN metastases for breast cancer patients.

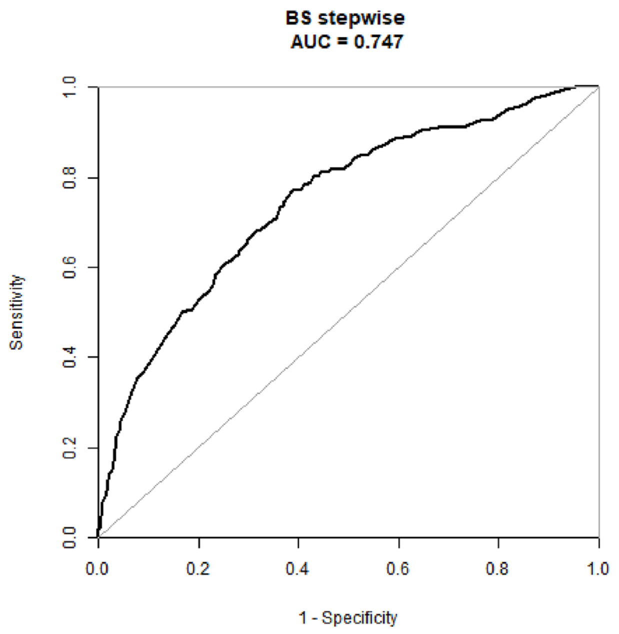


FIGURE 2 | The ROC curves of the prediction model in the primary dataset.

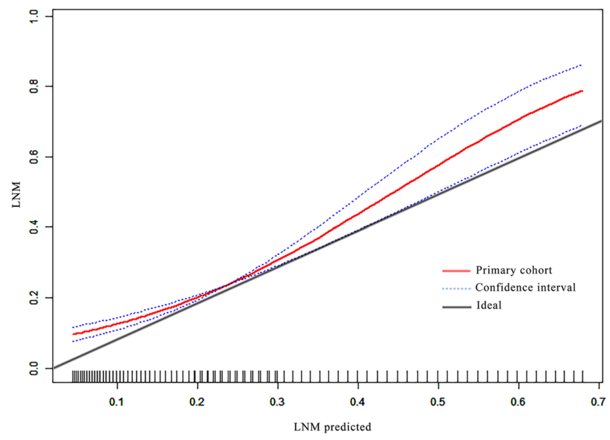


FIGURE 3 | The calibration curves of the nomogram in the primary cohort.

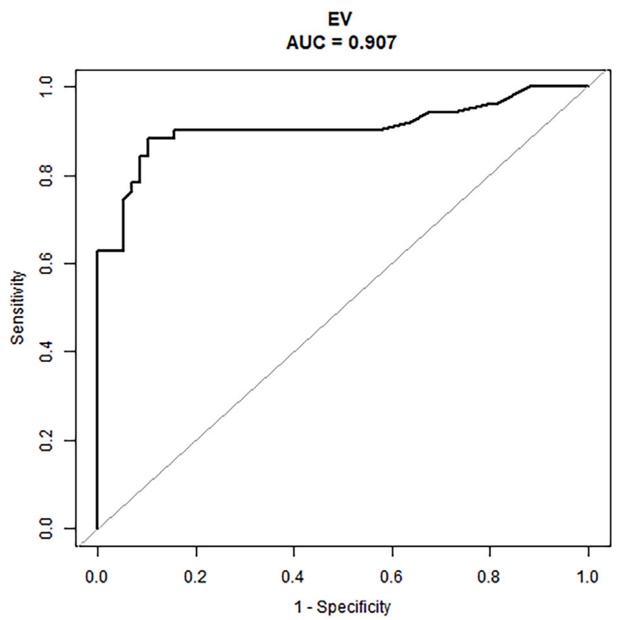


FIGURE 4 | The ROC curve of the prediction model in the N1 patients of the external validation cohort.

nodes, especially in the evaluation of early ALN metastasis. Breast MRI can better demonstrate lymph node metastasis on higher stations (19). However, the use of axillary US in evaluating ALN has been limited by its moderate accuracy and considerable discrepancy among the studies. Some studies have shown that malignant lymph nodes detected by US had a higher node burden than those detected by SLNB, implying a disparity between "ultrasound positive" and "SLNB positive" (20, 21). Moreover, according to previous studies, axillary US tends to perform poorly in identifying metastases in pathologic N1 patients, characterized by one to three abnormal nodes (9). Therefore, to improve the US diagnostic performance for ALN metastases, it is important to improve its accuracy and lower its false-negative rate in N1 patients.

In our study, we developed a prediction model based on BI-RADS ultrasonic features to predict the risk of LN metastases, achieving an accuracy of 65.0% in the primary cohort,and 89.0% in the external validation cohort. A nomogram, incorporating two factors among the lesion US features, showed significant discriminating ability in the primary cohort, and also showed high predictive power in an external validation cohort of early-stage breast cancer patients.

Recent studies have investigated the potential value of ultrasonic images of breast lesions in predicting nodal metastases, with reported AUCs ranging from 0.731 to 0.848 (22–25). Some of these studies showed that US features of breast lesion and axillary lymph nodes are correlated with ALN status (22), and in some studies, high-throughput features of ultrasonic images were proved useful for the prediction of ALN metastases (24, 25). Taken together, these results demonstrate that ultrasonic images of breast lesions can potentially be useful in the preoperative diagnosis of ALN metastases. Considering the nonspecific ultrasonic presentations of metastatic ALNs and the disparity in positive rates between US and SLNB, the images of breast lesions are worth exploring, as they might contain helpful information for the prediction of nodal metastases.

In 2003, a standard protocol for breast US was established in the BI-RADS lexicon and received worldwide recognition (18). The definition and description of the ultrasonic features, the lesion classification, and the reporting system were all clearly defined and illustrated in the lexicon, allowing reliable feature identification. Previous studies have validated clinical-pathological factors and US BI-RADS features of masses could predict breast cancer LN metastasis. Zong et al. (26) suggest that US features of breast mass, like margin, microcalcification, and blood flow signals are significantly correlated with ALN metastasis in early breast cancer. Besides, Guo et al. (12) have proven that irregular shape and high color Doppler flow imaging grades are independent impact factors of ALN metastasis. However, both of them incorporated some clinical-pathological factors simultaneously, like immunohistochemical analysis (ER, PR, Ki-67, and so on) and the histologic grade, which are also highly associated with ALN status. To figure out the independent contributions of breast lesion US features in determining the likelihood of ALN metastasis preoperatively, and to develop a simple and practical nomogram based on US features, we adopted the ultrasonic features defined by the BI-RADS lexicon in 2013 to construct our models (17). A total of eight features were included for modeling, which has been commonly used in differentiating benign and malignant breast lesions. Our results show that some features are also related to ALN status. As shown by the nomogram, tumor size and lesion boundary had more significant impacts on total scores than other features. The prediction model displayed a remarkable ability to predict ALN status, especially in N1 patients, yielding an AUC of 0.901. More importantly, it achieved 88% sensitivity for N1 patients, compared with that in previous studies, which presented false-negative rates as high as 46.2% (9). These results indicate the potential value of our model in increasing sensitivity in the identification of abnormal lymph nodes, as well as in decreasing the rate of preoperatively missed diagnoses, thus bringing benefits to early-stage breast cancer patients.

To note, US readers can predict the probability of ALN metastases associated with the lesion using this nomogram, after routinely extracting the standardized features from the breast lesion ultrasonic images. Apart from its high accuracy, compared with some complex models using additional image processing software, the prediction process used by this model is simple and time-saving. We hope that this model will be widely used in clinical practice as a supplementary to conventional breast US, allowing improved accuracy of preoperative diagnosis of nodal metastases.

Our predictive model has several limitations. First, the sample size of the external cohort was relatively small, and increasing the sample size would be necessary to obtain more convincing results. Moreover, the single-center design of the study might lead to an un recognized bias in patient recruitment, imaging acquisition, and image analysis. Adding data from other medical centers would be helpful in further improving the clinical efficacy of the model.

In this study, a nomogram based on ultrasonic features of breast lesions was developed to predict the risk of ALN metastases in breast cancer patients. The model demonstrated clinical potential in providing a non-invasive, effective, and easy-to-use approach to identify ALN metastases preoperatively, which might aid in clinical decision making.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

This retrospective study was approved by the Institutional Review Board of Peking Union Medical College Hospital.

AUTHOR CONTRIBUTIONS

YJ and QZ conceived and designed the study. WL, JZ, LM, and MX collected the clinical and image data. YL and JQ performed image pre-processing. CZ and YG analyzed the image data and performed the statistical analysis. YL and CZ wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by CAMS Innovation Fund for Medical Sciences (2017-I2M-1-006).

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: Cancer J Clin* (2018) 68(6):394–424. doi: 10.3322/caac.21492
2. Rao R, Euhus D, Mayo HG, Balch C. Axillary node interventions in breast cancer: a systematic review. *JAMA* (2013) 310(13):1385–94. doi: 10.1001/jama.2013.277804
3. Cianfrocca M, Goldstein LJ. Prognostic and predictive factors in early-stage breast cancer. *Oncol* (2004) 9(6):606–16. doi: 10.1634/theoncologist.9-6-606
4. Janni W, Kuhn T, Schwentner L, Kreienberg R, Fehm T, Wockel A. Sentinel node biopsy and axillary dissection in breast cancer: the evidence and its limits. *Deutsches Arzteblatt Int* (2014) 111(14):244–9. doi: 10.3238/arztebl.2014.0244
5. Lyman GH, Giuliano AE, Somerfield MR, Benson 3rd AB, Bodurka DC, Burstein HJ, et al. American Society of Clinical Oncology guideline recommendations for sentinel lymph node biopsy in early-stage breast cancer. *J Clin Oncol* (2005) 23(30):7703–20. doi: 10.1200/jco.2005.08.001
6. Surgery tAoB. *Consensus statement on management of the malignant axilla in early breast cancer*. (2015).
7. Ecanow JS, Abe H, Newstead GM, Ecanow DB, Jeske JM. Axillary staging of breast cancer: what the radiologist should know. *Radiographics* (2013) 33 (6):1589–612. doi: 10.1148/rg.336125060
8. Alvarez S, Anorbe E, Alcorta P, Lopez F, Alonso I, Cortes J. Role of sonography in the diagnosis of axillary lymph node metastases in breast cancer: a systematic review. *AJR Am J Roentgenol* (2006) 186(5):1342–8. doi: 10.2214/ajr.05.0936
9. Zhang YN, Wang CJ, Xu Y, Zhu QL, Zhou YD, Zhang J, et al. Sensitivity, Specificity and Accuracy of Ultrasound in Diagnosis of Breast Cancer

Metastasis to the Axillary Lymph Nodes in Chinese Patients. *Ultrasound Med Biol* (2015) 41(7):1835–41. doi: 10.1016/j.ultrasmedbio.2015.03.024

10. Li XL, Xu HX, Li DD, He YP, Yue WW, Xu JM, et al. A Risk Model based on Ultrasound, Ultrasound Elastography, and Histologic Parameters for Predicting Axillary Lymph Node Metastasis in Breast Invasive Ductal Carcinoma. *Sci Rep* (2017) 7(1):3029. doi: 10.1038/s41598-017-03582-3
11. Yun SJ, Sohn YM, Seo M. Risk Stratification For Axillary Lymph Node Metastases in Breast Cancer Patients: What Clinicopathological and Radiological Factors of Primary Breast Cancer Can Predict Preoperatively Axillary Lymph Node Metastases? *Ultrasound Q* (2017) 33(1):15–22. doi: 10.1097/ruq.0000000000000249
12. Guo Q, Dong Z, Zhang L, Ning C, Li Z, Wang D, et al. Ultrasound Features of Breast Cancer for Predicting Axillary Lymph Node Metastasis. *J Ultrasound Med* (2018) 37(6):1354–3. doi: 10.1002/jum.14469
13. Tran HT, Pack D, Mylander C, Martino L, Rosman M, Tafra L, et al. Ultrasound-Based Nomogram Identifies Breast Cancer Patients Unlikely to Harbor Axillary Metastasis: Towards Selective Omission of Sentinel Lymph Node Biopsy. *Ann Surg Oncol* (2020) 27(8):2679–86. doi: 10.1245/s10434-019-08164-3
14. Bromham N, Schmidt-Hansen M, Astin M, Hasler E, Reed MW. Axillary treatment for operable primary breast cancer. *Cochrane Database Syst Rev* (2017) 1:Cd004561. doi: 10.1002/14651858.CD004561.pub3
15. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* (2015) 162(1):W1–73. doi: 10.7326/m14-0698
16. Yu X, Hao X, Wan J, Wang Y, Yu L, Liu B. Correlation between Ultrasound Appearance of Small Breast Cancer and Axillary Lymph Node Metastasis. *Ultrasound Med Biol* (2018) 44(2):342–9. doi: 10.1016/j.ultrasmedbio.2017.09.020
17. Radiology ACo. *Breast Imaging Reporting and Data System® (BI-RADS®) 5*. Reston: American College of Radiology (2013).
18. Radiology ACo. *Breast Imaging Reporting and Data System® (BI-RADS®) 4*. Reston: American College of Radiology (2003).
19. Chang JM, Leung JWT, Moy L, Ha SM, Moon WK. Axillary Nodal Evaluation in Breast Cancer: State of the Art. *Radiology* (2020) 295(3):500–15. doi: 10.1148/radiol.2020192534
20. Caudle AS, Kuerer HM, Le-Petross HT, Yang W, Yi M, Bedrosian I, et al. Predicting the extent of nodal disease in early-stage breast cancer. *Ann Surg Oncol* (2014) 21(11):3440–7. doi: 10.1245/s10434-014-3813-4
21. Verheuvel NC, van den Hoven I, Ooms HW, Voogd AC, Roumen RM. The role of ultrasound-guided lymph node biopsy in axillary staging of invasive breast cancer in the post-ACOSOG Z0011 trial era. *Ann Surg Oncol* (2015) 22 (2):409–15. doi: 10.1245/s10434-014-4071-1
22. Akissue de Camargo Teixeira P, Chala LF, Shimizu C, Filassi JR, Maesaka JY, de Barros N. Axillary Lymph Node Sonographic Features and Breast Tumor Characteristics as Predictors of Malignancy: A Nomogram to Predict Risk. *Ultrasound Med Biol* (2017) 43(9):1837–45. doi: 10.1016/j.ultrasmedbio.2017.05.003
23. Meretoja TJ, Heikkila PS, Mansfield AS, Cserni G, Ambrozay E, Boross G, et al. A predictive tool to estimate the risk of axillary metastases in breast cancer patients with negative axillary ultrasound. *Ann Surg Oncol* (2014) 21 (7):2229–36. doi: 10.1245/s10434-014-3617-6
24. Moon WK, Chen IL, Yi A, Bae MS, Shin SU, Chang RF. Computer-aided prediction model for axillary lymph node metastasis in breast cancer using tumor morphological and textural features on ultrasound. *Comput Methods Programs Biomed* (2018) 162:129–37. doi: 10.1016/j.cmpb.2018.05.011
25. Moon WK, Lee YW, Huang YS, Lee SH, Bae MS, Yi A, et al. Computer-aided prediction of axillary lymph node status in breast cancer using tumor surrounding tissue features in ultrasound images. *Comput Methods Programs Biomed* (2017) 146:143–50. doi: 10.1016/j.cmpb.2017.06.001
26. Zong Q, Deng J, Ge W, Chen J, Xu D. Establishment of Simple Nomograms for Predicting Axillary Lymph Node Involvement in Early Breast Cancer. *Cancer Manag Res* (2020) 12:2025–35. doi: 10.2147/CMAR.S241641

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Oncology

MINI REVIEW
published: 21 December 2020
doi: 10.3389/fonc.2020.594580

Artificial Intelligence in Medical Imaging and Its Application in Sonography for the Management of Liver Tumor

*Naoshi Nishida** *and Masatoshi Kudo*

Department of Gastroenterology and Hepatology, Kindai University Faculty of Medicine, Osaka-Sayama, Japan

Recent advancement in artificial intelligence (AI) facilitate the development of AI-powered medical imaging including ultrasonography (US). However, overlooking or misdiagnosis of malignant lesions may result in serious consequences; the introduction of AI to the imaging modalities may be an ideal solution to prevent human error. For the development of AI for medical imaging, it is necessary to understand the characteristics of modalities on the context of task setting, required data sets, suitable AI algorism, and expected performance with clinical impact. Regarding the AI-aided US diagnosis, several attempts have been made to construct an image database and develop an AI-aided diagnosis system in the field of oncology. Regarding the diagnosis of liver tumors using US images, 4- or 5-class classifications, including the discrimination of hepatocellular carcinoma (HCC), metastatic tumors, hemangiomas, liver cysts, and focal nodular hyperplasia, have been reported using AI. Combination of radiomic approach with AI is also becoming a powerful tool for predicting the outcome in patients with HCC after treatment, indicating the potential of AI for applying personalized medical care. However, US images show high heterogeneity because of differences in conditions during the examination, and a variety of imaging parameters may affect the quality of images; such conditions may hamper the development of US-based AI. In this review, we summarized the development of AI in medical images with challenges to task setting, data curation, and focus on the application of AI for the managements of liver tumor, especially for US diagnosis.

Keywords: artificial intelligence, ultrasound, imaging, liver cancer, neural network, diagnosis

Edited by:
Hui-Xiong Xu,
Tongji University, China

Reviewed by:
Christoph Dietrich,
Hirslanden Private Hospital Group,
Switzerland
Ming-de Lu,
Sun Yat-sen University, China

***Correspondence:**
Naoshi Nishida
naoshi@med.kindai.ac.jp

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *13 August 2020*
Accepted: *16 November 2020*
Published: *21 December 2020*

Citation:
Nishida N and Kudo M (2020) Artificial Intelligence in Medical Imaging and Its Application in Sonography for the Management of Liver Tumor. Front. Oncol. 10:594580. doi: 10.3389/fonc.2020.594580

INTRODUCTION

Artificial intelligence (AI) is generally considered as the intelligence performed by compactional statistics, where machine learning is a subset of AI. Recently, AI is emerging as a major constituent in the field of medicine and healthcare. In particular, AI can be easily applied to imaging data because these data are electronically organized, and AI excels at recognizing unique and complex features of images and facilitates quantitative assessments in an automated fashion. This characteristic of AI is ideal in the constrained clinical setting wherein medical staff must interpret large image datasets based on their visual perception with uncertainty, in which human errors are inevitable. For example, AI is a powerful tool in radiomics where extracting a large number of features form medical images is required. Based on this advantage, AI have been applied for classification of lesions, such as liver tumors, and prediction of the prognosis using image data from computed tomography and magnetic resonance imaging (MRI) (1). In addition, AI-based image processing techniques have also introduced in the field of ultrasonography (US). This review shows the recent progress in AI for medical imaging, especially for an AI-aided diagnosis for the detection, characterization, subsequent monitoring, and prediction of outcomes in patients with liver cancer, especially in the field of US diagnosis.

HISTORY AND RECENT PROGRESS OF AI IN MEDICAL IMAGING

The application of pattern recognition in medical issues has been proposed in the early 1960s. In the 1980s, the prevalence of computers induced the development of medical AI in radiology using a quantitatively computable domain. After the emergence of deep neural network, the rate at which AI is evolving radiology is rapidly growing that is proportional to the growth of data volume in medical image and computational power (2).

For the image analysis, a convolutional neural network (CNN) is commonly applied, which is a class of deep neural networks using pixel value and assembling complex patterns to smaller and simpler patterns (2). The algorism contains multiped hidden layer with multiple convolutional and pooling layers. A trained CNN-based AI model using ≥120,000 retinal fundus images has been demonstrated to show high performance comparable to that of an experienced ophthalmologist for detecting referable diabetic retinopathy, which is expected to effectively assist ophthalmologists in the clinical workflow (3). Assessment of AI models for detecting lymph node metastasis of breast cancer based on whole microscopic slide images showed the superior performance of AI for detecting cancer cells in specimens to that of pathologists (4). A pre-trained CNN-based AI model for the diagnosis of skin cancer achieves performance on par with that by expert dermatologists in terms of the discrimination of skin cancers from corresponding benign lesions on dermography (5). AI models for the detection of pediatric pneumonia on chest radiography images and for the discrimination of diabetic macular edema from age-related macular degeneration on optical coherence tomography images are also reported with high performance, comparable to that of human experts (6). An AI-based colonoscopy system has been shown to accurately differentiate neoplastic lesions from non-neoplastic lesions on stained endocytoscopic images and endocytoscopic narrow-band images in endoscopic evaluation of small colon polyps (7). The application of AI for US-based diagnosis has been mainly reported for the diagnosis of malignant tumors, such as mammary and thyroid cancers (8–11). Le et al. reported an AI model for the diagnosis of thyroid cancer pre-trained with 312,399 B-mode US images of cancer and healthy controls (12). The model's diagnostic performance was validated in three test datasets with AUCs of 0.908–0.947. The AI model showed higher specificity in identifying thyroid cancer and comparable sensitivity to those corresponding to experienced radiologists. Another report described a real-time detection system of thyroid tumors based on real-time images using the "You Only Look Once" (YOLO) algorithm. This model achieved a similar sensitivity, positive predictive value, negative predictive value, and accuracy for the diagnosis of malignant thyroid tumors with higher specificity compared to those corresponding to experienced radiologists (12, 13). For the detection of breast cancer, Kumar et al. reported a real-time segmentation model of breast tumors using a CNN (14). This system can reportedly segment tumor images in real-time, suggesting its potential for clinical applications. Collectively, diagnostic accuracy of well-trained AI model for medical image is, at least, on par with human experts with much quicker output, suggesting the higher efficiency for diagnosis in clinical setting.

On the other hand, recently, Skrede et al. reported the use of AI for the prediction of outcomes after colorectal cancer resection using a pre-trained CNN-based model with pathological images (15). They discriminated the cases of poor prognosis from those of good prognosis, indicating the potential of medical AI for the management of cancer, such as the identification of patients who would benefit from adjuvant treatment after resection.

PROCESS FOR DEVELOPING AI MODELS FOR IMAGING DIAGNOSIS

Setting Tasks for AI in Medical Imaging

For the development of AI in medical imaging, it is important to select tasks that reflect important needs at clinical sites. For example, large-volume screening of medical images requires extensive effort, which is time consuming and invites human errors. In this setting, AI should be a powerful tool for clinicians because of its advantageous for precise detection of subtle features of lesions, segmentation, and quick output. AI models that can estimate the risk of disease may contribute to avoiding invasive examinations, representing an attractive task (16).

Data Sets for Developing AI Models for Medical Images

Generally, three independent datasets are required for developing medical AI (17). A training set is required for the training of AI models, which contains many images to update model parameters. A tuning set is for the selection of a model's hyperparameters that are necessary for the best expected output. A test set is for the final assessment of the performance of AI models. The splitting of curated data must be clean, and each dataset should be completely independent without any overlap with respect to lesions to avoid overfitting the output.

For disease classification, such as that corresponding to diagnosis, the data volume in each subclass should be similar because imbalances in data volumes among subclasses may lead to overfitting of the output, which may limit the performance of an AI model. For the image of rare diseases, the AI-based image created through generative adversarial networks might also be applicable.

AI Algorithm

During training, AI models automatically detect specific features of images through the fitting of model parameters, which improves the performance. CNNs are commonly applied for AI algorithm of imaging data (2). However, US examinations require real-time output, and an algorithm that requires many mathematical operations might not be appropriate for analyzing US images. The YOLO-based algorithm is suitable for the real-time detection and classification of lesions with high-speed processing. The process of selecting a model's architecture and training essentially involves a balance between model underfitting and overfitting (17). Underfitting occurs when a low-capacity model is used relative to the problem complexity and data size. Overfitting indicates that the evaluation overestimates the model's performance on previously unencountered data, in which case low performance on the test set is observed. Because there is a large diversity among US images in terms of the conditions of the examination and image parameter settings, larger volumes of data are required compared to those required for the development of other medical imaging AI.

Evaluation of Performance and Potential Impact

One of the major categories of evaluation of AI-aided imaging diagnosis is the ability to discriminate the lesions, such as benign or malignant. The area under the receiver operating characteristic curve (AUC) is commonly used as a threshold-free discriminative metric. Evaluation may also be based on other metrics, such as sensitivity (recall), specificity, and precision (positive predictive value); these are threshold-dependent. On the other hand, calibration, which evaluates how effectively the predicted probability matches the actual diagnosis should also be estimated (17). In addition, variability in the probability in the same lesion may also need to be analyzed because there can be variations among US images even within the same lesion, which is attributed to differences in parameter settings. Validation for accuracy is a critical process in the transitional process of medical AI. The performance of AI models must be evaluated using independent test cohorts and be compared with an experienced human control in real-world scenarios.

CURRENT AI MODELS FOR MEDICAL IMAGING OF LIVER LESIONS

AI Using Medical Image for the Management of Liver Tumors

Recently, many reports have described the development of AI models for the detection and diagnosis of liver tumors; some studies have aimed to predict outcomes after treatments, which may be applicable for the personalized management of patients (18, 19).

Hamm et al. reported the classification of 6 types of liver tumors by a pre-trained CNN using MRI data of 494 lesions from 334 cases (20). After data augmentation of the images for training, the established AI model demonstrated 90% sensitivity and 98% specificity for the test cohort. The average sensitivity and specificity for the radiologist were 82.5 and 96.5%, respectively. For the diagnosis of hepatocellular carcinoma (HCC), the sensitivities were 90% for the AI model and 60–70% for the radiologists. Considering the short processing time (only 6.6 ms) for output, the pre-trained AI model showed superior performance compared to that of the human radiologists.

On the other hand, AI is also useful for the detection of specific radiological features that may reflect histopathological characteristics associated with the biological behavior of a tumor. From this point of view, the development of AI for the prediction of outcomes after treatment, including tumor recurrence after surgery, may be possible. If pathological diagnosis is applied for constructing an AI model for medical imaging, it may be a non-invasive substitute for biopsy, which may significantly impact the management of cancer. Fent et al. reported a preoperative prediction model for microvascular invasion in patients with resectable HCC who do not show macroscopic vascular invasion through training using gadolinium-ethoxybenzyl (EOB)-diethylenetriamine-enhanced MRI data (21). The AI model selected ten specific features of EOB-enhanced MRI data to predict microvascular invasion. The performance of the AI model showed an AUC of 0.83 with 90.0, 75.0, and 84.0% sensitivity, specificity, and accuracy, respectively, which were much better than those of human radiologists. Kim et al. reported an AI model for the prediction of early and late recurrence of tumors after surgery using EOB-MRI data from solitary HCC cases (22). They established their AI model using a random survival forest to predict disease-free survival and found that peritumoral image features 3 mm outside the tumor border are important for the prediction of early recurrence after curative surgery.

AI Using Histopathological Images for Diagnosis and Management of Liver Cancers

It has also been reported that an AI model pre-trained with histopathological images of liver cancer using transfer learning can distinguish cancerous tissue from healthy liver tissue (23). Saillard et al. showed that a deep-learning model of histopathological images predicts survival after resection of HCC (24). They developed two kinds of AI models pre-trained with supervised image data, which was annotated based on the tumor portion in the slide images by pathologists, and non-supervised data without human annotations. The concordance indices for survival prediction were 0.78 and 0.75 for the pre-trained AI models with supervised and non-supervised data, respectively. Reportedly, these histopathological AI models showed a higher discriminatory power than that derived from a combination of known clinical risk factors. Some pathological findings, including vascular space, a macrotrabecular pattern of tumor cell architecture, a high degree of cytological atypia, and nuclear hyperchromasia, effectively predicted poor survival, and immune infiltrates and fibrosis in tumor and non-tumor tissues were associated with a low risk of short survival. These studies indicate that histopathological images yield useful training data for the prediction of prognosis in HCC cases (25).

AI-Aided Diagnosis for Liver Tumors in Ultrasonography

Generally, US images are heterogeneous because of the multiple image parameters and conditions of examination compared to other kind of medical images. Such heterogeneity of image data makes it difficult to develop the AI for US diagnosis, especially for liver tumors (18).

AI models are trained using cropped images of regions of interest that specifically focus on tumors for applying neural network and can be evaluated using cross-validation methods for small sample cohorts. The studies regarding the application of B-mode US images on machine learning for the diagnosis of liver tumor are summarized in **Table 1**. Virmani et al. reported the machine learning for discriminating HCC and metastatic liver tumor using support vector machine (SMV), where overall accuracy was 91.6 %; sensitivity of 90% for HCC and 93.3% for metastatic tumor were achieved (26). Hwang et al. tried to extract textural features of liver tumors including cysts, hemangiomas, and malignant lesions for the diagnosis; they examined the accuracy of two-class discriminations for cyst vs. hemangioma, cyst vs. malignant tumor, and hemangioma vs. malignant tumor, demonstrating the accuracy of more than 95% for each comparison (28). On the other hand, the study using artificial neural network (ANN) show 4-class discrimination for normal liver, cyst, hemangioma, and HCC: accuracy of almost 90%, and similar levels of sensitivity, and specificity are reported (29). Generally, these early studies failed to show the superiority of neural network for the diagnostic accuracy of liver tumors compared to the conventional machine learning because of the small size of learning cohort. Schmauch et al. reported the performance of an AI model for the diagnosis of liver tumors from B-mode US images (30). They reported an AI model for lesion detection and diagnosis from whole-liver US images using a 50-layer residual network. Despite the relatively small volume of training data, the performance for tumor detection and 5-class discrimination (HCC, metastatic tumors, hemangiomas, cysts, and focal nodular hyperplasia) achieved considerable AUCs (0.953 and 0.916) for tumor detection and discrimination, respectively, by cross validation. To reduce the heterogeneity of the US images, they cropped the images maximally to remove the black borders and standardize the aspect ratio. They also performed rescaling of the image intensity for normalization based on the intensity of the abdominal wall.

In addition to the gray scale B-mode US, doppler US, contrast-enhanced US (CEUS), shear wave elastography (SWE) and three-dimensional US images are also applicable for the training of AI models. Still image of contrast-enhanced US (CEUS) was applied for the learning data for more accurate discrimination of liver tumors. Streba et al. applied ANN for 4-class discrimination of liver tumor with 94.5, 94.2, and 89.7% for accuracy, sensitivity and specificity, respectively, for the discriminaton (31). Gatos et al. and Kondo et al. reported the 4-class classification of benign tumors, hepatocellular carcinoma, and metastatic tumors using SMV pretrained with CEUS images (32, 33). A contrast agent, Sonazoid, was used and, reportedly, sensitivity, specificity, and accuracy that discriminate malignant lesions from benign were 94.0, 87.1, and 91.8%, respectively (33). Another report applied a pretrained SMV using CEUS images and achieved the accuracy, sensitivity, and specificity of 90.4, 93.6, and 86.9%, respectively, for the different diagnosis of benign and malignant liver tumors (34). Discrimination of benign and malignant lesions is a critical task for the management of patients with liver tumors, and CEUS images yield attractive data for the development of AI models to detect malignant tumors.

On the other hand, because of the development of new treatments in HCC, management of this type of cancer is becoming complex (39). Recently, in addition to detection and diagnosis, AI model regarding the management of HCC, such as prediction of microvascular invasion, pathological grading, and treatment outcomes have been reported. Hu et al. proposed US-based radiomics score consisted of six selected features was an independent predictor of microvascular invasion in HCC (35). On the other hand, model for predicting pathological grading of HCC before surgery was also reported using ultrasomics of CEUS images (36). Liu et al. developed an AI model for the prediction of responses to transarterial chemoembolization in patients with HCC through training with B-mode US and CEUS images (37). They reported AUCs of 0.93 and 0.81 for the AI based on CEUS and B-mode US images, respectively, indicating a higher performance of the model pre-trained with CEUS images than that with B-mode US images. They also reported AI models for predicting outcomes in patients with HCC after two types of treatment—radiofrequency ablation (RFA) and liver resection—from radiomics information based on CEUS images (38). For the prediction of two-year progression-free survival (PFS), both models provided high prediction accuracy. Interestingly, the models showed that some patients who underwent RFA and surgery should swap their treatments, so that a higher probability of increased 2-year PFS would be achieved. In addition, another report showed radiomic signature from grayscale US images of gross-tumoral region had potential for prediction of microvascular invasion of HCC before surgery, suggesting the potential of radiomic approach for the prediction of outcome (40). Such AI prediction models using radiomic signature may be applicable for personalized medicine in HCC treatment.

The grading of liver fibrosis and steatosis is also an important task for the management of liver disease because these backgrounds may confer a risk of liver cancer. Several reports have described the classification of fibrosis and steatosis based on disease progression using AI models trained with B-mode US and SWE images (18, 41). Deep-learning models show hyper-performance in terms of detection and risk stratification of fatty

TABLE 1 | Performance for diagnosis of liver tumor based on the machine learning using image of ultrasonography.

Algorism	Liver lesions:number of the cases	Performance	References
Pre-trained AI using B-mode US images			
SVM	normal liver: 15 cirrhotic liver: 16 HCC: 25	accuracy: 88.8%	(26, 27)
SVM	HCC: 27 metastatic tumor: 27	overall accuracy: 91.6% sensitivity: 90% for HCC 93.3% for metastatic carcinoma	(26)
ANN	cyst; 29 hemangioma: 37 malignant tumor: 33	cyst vs. hemangioma accuracy: 99.7% cyst vs. malignant accuracy: 98.7% hemangioma vs. malignant accuracy: 96.1%	(28)
ANN (sparse autoencoder)	normal liver: 16 cyst: 44 hemangioma: 18 HCC: 30	accuracy: 90.5% sensitivity: 91.6% specificity: 88.5%	(29)
CNN	non-tumorus liver: 258 hemangioma: 17 metastatic tumor: 48 HCC: 6 cyst: 30 focal nodular hyperplasia: 8	AUC for tumor detection: 0.935 AUC for tumor discrimination: 0.916 (mean)	(30)
Pre-trained AI using CEUS images			
ANN	hemangioma: 16 focal fatty liver: 23 HCC: 41 metastatic tumor: 32 hypervascular: 20 hypovascular: 12	accuracy: 94.5% sensitivity: 93.2% specificity: 89.7%	(31)
SVM	benign tumor: 30 malignant tumor: 22	accuracy: 90.3% sensitivity: 93.1% specificity: 86.9%	(32)
SVM	benign tumor, HCC, or metastatic tumor 98	benign vs. malignant accuracy: 91.8% sensitivity: 94.0% specificity: 87.1% benign vs. HCC vs. metastatic carcinoma accuracy: 85.7% sensitivity: 84.4% specificity: 87.7%	(33)
SVM (multiple kernel learning)	benign tumor: 46 malignant tumor: 47	accuracy: 90.4% sensitivity: 93.6% specificity: 86.9%	(34)
Logistic regression analyses	Solitary HCC without macrovascular invasion: 468	Prediction of microvascular invasion using ultrasomics feature: AUC = 0.731	(35)
SVM	HCC: 235 High-grade: 65 Low-grade: 170	Discrimination of HCC pathological grades using ultrasomics and clinical factors: AUC = 0.785	(36)
CNN	HCC before TACE: 130	Prediction of response to TACE: AUC = 0.93	(37)
CNN	HCC: 419 patients who underwent RFA; 214 patients who underwent resection; 215	prediction of RFS for 2 years after curative treatment C-index 0.726 for RFA C-index 0.726 for resection	(38)

SVM, support vector machine; ANN, artificial neural network; CNN, convolutional neural network; CEUS, contrast-enhanced ultrasonography; HCC, hepatocellular carcinoma; AUC, area under the receiver operating characteristic curve; TACE, transarterial chemoembolization; RFA, radiofrequency ablation; C-index, concordance index.

liver disease compared to that corresponding to conventional machine-learning models (42). AI models pre-trained with color images of US-SWE can also discriminate chronic liver disease from healthy cases (43). Reportedly, the combination of B-mode US images, raw radiofrequency data, and dynamic contrast-enhanced microflow is a useful dataset for developing AI models that classify the stage of liver fibrosis (44), where datasets involving raw radiofrequency data provide better predictive value than those from conventional US image only. Therefore, it should be possible that AI using multiparametric ultrasomics can help improve the performance of the model. For the development of AI that determine the stage of liver fibrosis more accurately, Gatos et al. reported a detection algorithm that excludes unreliable regions on SWE images, which contributes to a reduction in interobserver variability (45). Applying these AI models may be an alternative to invasive liver biopsy for predicting the progression of liver disease, which may be associated with a risk of liver cancer.

CONCLUSION

Among the imaging modalities, US is the most commonly used in clinical practice for detection of liver tumors because of its low-cost, non-ionizing, and portable point-of-care characteristics providing real-time images. From this point of view, the AI-powered US carries more advantage in routine clinical applications compared to that in CT and MRI (46). Although, US images involve operator-, patient-, and scanner-dependent variations, AI-aided US diagnosis is becoming mature that is attributed to the recent advancement in the US equipment and increase in computing power to identify the complex imaging features. In addition to the B-mode image, images from CEUS and US elastography is becoming promising data applicable in AI-based diagnosis in the field of liver tumor according to the prevalence of high-end US equipment (46, 47). These could also be a safeguard for misdiagnosis in the actual workflow. The development of AI-aided technologies for the detection and diagnosis of malignant tumors may carry sufficient potential to reduce cancer-related mortality in the near future.

AUTHOR CONTRIBUTIONS

Conceptualization: NN. Writing—original draft preparation: NN. Writing—review and editing: NN. Supervision: MK. Funding acquisition: NN and MK. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by AMED (Japan Agency for Medical Research and Development) under Grant No. JP19lk1010035. (NN and MK) and ROIS NII Open Collaborative Research 2020 under Grant No. 20S0601 (NN).

REFERENCES

1. Haj-Mirzaian A, Kadivar A, Kamel IR, Zaheer A. Updates on Imaging of Liver Tumors. *Curr Oncol Rep* (2020) 22:46. doi: 10.1007/s11912-020-00907-w
2. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts H. Artificial intelligence in radiology. *Nat Rev Cancer* (2018) 18:500–10. doi: 10.1038/s41568-018-0016-5
3. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA* (2016) 316:2402–10. doi: 10.1001/jama.2016.17216
4. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, van Ginneken B, Karssemeijer N, Litjens G, et al. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. *JAMA* (2017) 318:2199–210. doi: 10.1001/jama.2017.14580
5. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* (2017) 542:115–8. doi: 10.1038/nature21056
6. Kermany DS, Goldbaum M, Cai W, Valentim CCS, Liang H, Baxter SL, et al. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. *Cell* (2018) 172:1122–31.e1129. doi: 10.1016/j.cell.2018.02.010
7. Kudo SE, Misawa M, Mori Y, Hotta K, Ohtsuka K, Ikematsu H, et al. Artificial Intelligence-assisted System Improves Endoscopic Identification of Colorectal Neoplasms. *Clin Gastroenterol Hepatol* (2020) 18:1874–81.e1872. doi: 10.1016/j.cgh.2019.09.009
8. Zhang Z, Zhang X, Lin X, Dong L, Zhang S, Zhang X, et al. Ultrasonic Diagnosis of Breast Nodules Using Modified Faster R-CNN. *Ultrason Imaging* (2019) 41:353–67. doi: 10.1177/0161734619882683
9. Fujioka T, Kubota K, Mori M, Kikuchi Y, Katsuta L, Kasahara M, et al. Distinction between benign and malignant breast masses at breast ultrasound using deep learning method with convolutional neural network. *Jpn J Radiol* (2019) 37:466–72. doi: 10.1007/s11604-019-00831-5
10. Zhao WJ, Fu LR, Huang ZM, Zhu JQ, Ma BY. Effectiveness evaluation of computer-aided diagnosis system for the diagnosis of thyroid nodules on ultrasound: A systematic review and meta-analysis. *Med (Baltimore)* (2019) 98:e16379. doi: 10.1097/MD.0000000000016379
11. Ciritsis A, Rossi C, Eberhard M, Marcon M, Becker AS, Boss A. Automatic classification of ultrasound breast lesions using a deep convolutional neural network mimicking human decision-making. *Eur Radiol* (2019) 29:5458–68. doi: 10.1007/s00330-019-06118-7
12. Li X, Zhang S, Zhang Q, Wei X, Pan Y, Zhao J, et al. Diagnosis of thyroid cancer using deep convolutional neural network models applied to sonographic images: a retrospective, multicohort, diagnostic study. *Lancet Oncol* (2019) 20:193–201. doi: 10.1016/S1470-2045(18)30762-9
13. Verburg F, Reiners C. Sonographic diagnosis of thyroid cancer with support of AI. *Nat Rev Endocrinol* (2019) 15:319–21. doi: 10.1038/s41574-019-0204-8
14. Kumar V, Webb JM, Gregory A, Denis M, Meixner DD, Bayat M, et al. Automated and real-time segmentation of suspicious breast masses using convolutional neural network. *PloS One* (2018) 13:e0195816. doi: 10.1371/journal.pone.0195816
15. Skrede OJ, De Raedt S, Kleppe A, Hveem TS, Liestol K, Maddison J, et al. Deep learning for prediction of colorectal cancer outcome: a discovery and validation study. *Lancet* (2020) 395:350–60. doi: 10.1016/S0140-6736(19)32998-8
16. Poplin R, Varadarajan AV, Blumer K, Liu Y, McConnell MV, Corrado GS, et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat BioMed Eng* (2018) 2:158–64. doi: 10.1038/s41551-018-0195-0
17. Chen PC, Liu Y, Peng L. How to develop machine learning models for healthcare. *Nat Mater* (2019) 18:410–4. doi: 10.1038/s41563-019-0345-0
18. Nishida N, Yamakawa M, Shiina T, Kudo M. Current status and perspectives for computer-aided ultrasonic diagnosis of liver lesions using deep learning technology. *Hepatol Int* (2019) 13:416–21. doi: 10.1007/s12072-019-09937-4
19. Spann A, Yasodhara A, Kang J, Watt K, Wang B, Goldenberg A, et al. Applying Machine Learning in Liver Disease and Transplantation: A Comprehensive Review. *Hepatology* (2020) 71:1093–105. doi: 10.1002/hep.31103
20. Hamm CA, Wang CJ, Savic LJ, Ferrante M, Schobert I, Schlachter T, et al. Deep learning for liver tumor diagnosis part I: development of a convolutional neural network classifier for multi-phasic MRI. *Eur Radiol* (2019) 29:3338–47. doi: 10.1007/s00330-019-06205-9
21. Feng ST, Jia Y, Liao B, Huang B, Zhou Q, Li X, et al. Preoperative prediction of microvascular invasion in hepatocellular cancer: a radiomics model using Gd-EOB-DTPA-enhanced MRI. *Eur Radiol* (2019) 29:4648–59. doi: 10.1007/s00330-018-5935-8
22. Kim S, Shin J, Kim DY, Choi GH, Kim MJ, Choi JY. Radiomics on Gadoxetic Acid-Enhanced Magnetic Resonance Imaging for Prediction of Postoperative Early and Late Recurrence of Single Hepatocellular Carcinoma. *Clin Cancer Res* (2019) 25:3847–55. doi: 10.1158/1078-0432.CCR-18-2861
23. Sun C, Xu A, Liu D, Xiong Z, Zhao F, Ding W. Deep Learning-Based Classification of Liver Cancer Histopathology Images Using Only Global

Labels. *IEEE J BioMed Health Inform* (2020) 24:1643–51. doi: 10.1109/JBHI.2019.2949837

24. Saillard C, Schmauch B, Laifa O, Moarii M, Toldo S, Zaslavskiy M, et al. Predicting survival after hepatocellular carcinoma resection using deep-learning on histological slides. *Hepatology* (2020). doi: 10.1002/hep.31207
25. Chaudhary K, Poirion OB, Lu L, Garmire LX. Deep Learning-Based Multi-Omics Integration Robustly Predicts Survival in Liver Cancer. *Clin Cancer Res* (2018) 24:1248–59. doi: 10.1158/1078-0432.CCR-17-0853
26. Virmani J, Kumar V, Kalra N, Khandelwal N. Characterization of primary and secondary malignant liver lesions from B-mode ultrasound. *J Digit Imaging* (2013) 26:1058–70. doi: 10.1007/s10278-013-9578-7
27. Virmani J, Kumar V, Kalra N, Khandelwal N. SVM-based characterization of liver ultrasound images using wavelet packet texture descriptors. *J Digit Imaging* (2013) 26:530–43. doi: 10.1007/s10278-012-9537-8
28. Hwang YN, Lee JH, Kim GY, Jiang YY, Kim SM. Classification of focal liver lesions on ultrasound images by extracting hybrid textural features and using an artificial neural network. *BioMed Mater Eng* (2015) 26 Suppl 1:S1599–1611. doi: 10.3233/BME-151459
29. Tarek M, Hassan M, El-Sayed S. Diagnosis of focal liver diseases based on deep learning technique for ultrasound images. *Arab J Sci Eng* (2017) 42:3127–40. doi: 10.1007/s13369-016-2387-9
30. Schmauch B, Herent P, Jehanno P, Dehaene O, Saillard C, Aube C, et al. Diagnosis of focal liver lesions from ultrasound using deep learning. *Diagn Interv Imaging* (2019) 100:227–33. doi: 10.1016/j.diii.2019.02.009
31. Streba CT, Ionescu M, Gheonea DI, Sandulescu L, Ciurea T, Saftoiu A, et al. Contrast-enhanced ultrasonography parameters in neural network diagnosis of liver tumors. *World J Gastroenterol* (2012) 18:4427–34. doi: 10.3748/wjg.v18.i32.4427
32. Gatos I, Tsantis S, Spiliopoulos S, Skouroliakou A, Theotokas I, Zoumpoulis P, et al. A new automated quantification algorithm for the detection and evaluation of focal liver lesions with contrast-enhanced ultrasound. *Med Phys* (2015) 42:3948–59. doi: 10.1118/1.4921753
33. Kondo S, Takagi K, Nishida M, Iwai T, Kudo Y, Ogawa K, et al. Computer-Aided Diagnosis of Focal Liver Lesions Using Contrast-Enhanced Ultrasonography With Perflubutane Microbubbles. *IEEE Trans Med Imaging* (2017) 36:1427–37. doi: 10.1109/TMI.2017.2659734
34. Guo LH, Wang D, Qian YY, Zheng X, Zhao CK, Li XL, et al. A two-stage multi-view learning framework based computer-aided diagnosis of liver tumors with contrast enhanced ultrasound images. *Clin Hemorheol Microcirc* (2018) 69:343–54. doi: 10.3233/CH-170275
35. Hu HT, Wang Z, Huang XW, Chen SL, Zheng X, Ruan SM, et al. Ultrasound-based radiomics score: a potential biomarker for the prediction of microvascular invasion in hepatocellular carcinoma. *Eur Radiol* (2019) 29:2890–901. doi: 10.1007/s00330-018-5797-0
36. Wang W, Wu SS, Zhang JC, Xian MF, Huang H, Li W, et al. Preoperative Pathological Grading of Hepatocellular Carcinoma Using Ultrasomics of Contrast-Enhanced Ultrasound. *Acad Radiol* (2020). doi: 10.1016/j.acra.2020.05.033
37. Liu D, Liu F, Xie X, Su L, Liu M, Xie X, et al. Accurate prediction of responses to transarterial chemoembolization for patients with hepatocellular carcinoma by using artificial intelligence in contrast-enhanced ultrasound. *Eur Radiol* (2020) 30:2365–76. doi: 10.1007/s00330-019-06553-6
38. Liu F, Liu D, Wang K, Xie X, Ming L. Deep learning radiomics based on contrast-enhanced ultrasound might optimize curative treatments for very early or early stage hepatocellular carcinoma patients. *Liver Cancer* (2020) 9 (4):397–413. doi: 10.1159/000505694
39. Nishida N, Kudo M. Immune checkpoint blockade for the treatment of human hepatocellular carcinoma. *Hepatol Res* (2018) 48:622–34. doi: 10.1111/hepr.13191
40. Dong Y, Zhou L, Xia W, Zhao XY, Zhang Q, Jian JM, et al. Preoperative Prediction of Microvascular Invasion in Hepatocellular Carcinoma: Initial Application of a Radiomic Algorithm Based on Grayscale Ultrasound Images. *Front Oncol* (2020) 10:353. doi: 10.3389/fonc.2020.00353
41. Wang K, Lu X, Zhou H, Gao Y, Zheng J, Tong M, et al. Deep learning Radiomics of shear wave elastography significantly improved diagnostic performance for assessing liver fibrosis in chronic hepatitis B: a prospective multicentre study. *Gut* (2019) 68:729–41. doi: 10.1136/gutjnl-2018-316204
42. Biswas M, Kuppili V, Edla DR, Suri HS, Saba L, Marinhoe RT, et al. Symtosis: A liver ultrasound tissue characterization and risk stratification in optimized deep learning paradigm. *Comput Methods Programs BioMed* (2018) 155:165–77. doi: 10.1016/j.cmpb.2017.12.016
43. Gatos I, Tsantis S, Spiliopoulos S, Karnabatidis D, Theotokas I, Zoumpoulis P, et al. A Machine-Learning Algorithm Toward Color Analysis for Chronic Liver Disease Classification, Employing Ultrasound Shear Wave Elastography. *Ultrasound Med Biol* (2017) 43:1797–810. doi: 10.1016/j.ultrasmedbio.2017.05.002
44. Li W, Huang Y, Zhuang BW, Liu GJ, Hu HT, Li X, et al. Multiparametric ultrasomics of significant liver fibrosis: A machine learning-based analysis. *Eur Radiol* (2019) 29:1496–506. doi: 10.1007/s00330-018-5680-z
45. Gatos I, Tsantis S, Spiliopoulos S, Karnabatidis D, Theotokas I, Zoumpoulis P, et al. Temporal stability assessment in shear wave elasticity images validated by deep learning neural network for chronic liver disease fibrosis stage assessment. *Med Phys* (2019) 46:2298–309. doi: 10.1002/mp.13521
46. Akkus Z, Cai J, Boonrod A, Zeinoddini A, Weston AD, Philbrick KA. Erickson BJ. A Survey of Deep-Learning Applications in Ultrasound: Artificial Intelligence-Powered Ultrasound for Improving Clinical Workflow. *J Am Coll Radiol* (2019) 16:1318–28. doi: 10.1016/j.jacr.2019.06.004
47. Gillies RJ, Schabath MB. Radiomics Improves Cancer Screening and Early Detection. *Cancer Epidemiol Biomarkers Prev* (2020) 29(12):2556–67. doi: 10.1158/1055-9965.EPI-20-0075

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Oncology

ORIGINAL RESEARCH
published: 18 February 2021
doi: 10.3389/fonc.2020.614201

Multiple U-Net-Based Automatic Segmentations and Radiomics Feature Stability on Ultrasound Images for Patients With Ovarian Cancer

Juebin Jin [1†], ***Haiyan Zhu*** [2,3†], ***Jindi Zhang*** [3], ***Yao Ai*** [4†], ***Ji Zhang*** [4], ***Yinyan Teng*** [5], ***Congying Xie*** [4,6*] **and** ***Xiance Jin*** [4*]

[1] Department of Medical Engineering, Wenzhou Medical University First Affiliated Hospital, Wenzhou, China, [2] Department of Gynecology, Shanghai First Maternal and Infant Hospital, Tongji University School of Medicine, Shanghai, China, [3] Department of Gynecology, Wenzhou Medical University First Affiliated Hospital, Wenzhou, China, [4] Department of Radiotherapy Center, Wenzhou Medical University First Affiliated Hospital, Wenzhou, China, [5] Department of Ultrasound Imaging, Wenzhou Medical University First Affiliated Hospital, Wenzhou, China, [6] Department of Radiation and Medical Oncology, Wenzhou Medical University Second Affiliated Hospital, Wenzhou, China

Few studies have reported the reproducibility and stability of ultrasound (US) images based radiomics features obtained from automatic segmentation in oncology. The purpose of this study is to study the accuracy of automatic segmentation algorithms based on multiple U-net models and their effects on radiomics features from US images for patients with ovarian cancer. A total of 469 US images from 127 patients were collected and randomly divided into three groups: training sets (353 images), validation sets (23 images), and test sets (93 images) for automatic segmentation models building. Manual segmentation of target volumes was delineated as ground truth. Automatic segmentations were conducted with U-net, U-net++, U-net with Resnet as the backbone (U-net with Resnet), and CE-Net. A python 3.7.0 and package Pyradiomics 2.2.0 were used to extract radiomic features from the segmented target volumes. The accuracy of automatic segmentations was evaluated by Jaccard similarity coefficient (JSC), dice similarity coefficient (DSC), and average surface distance (ASD). The reliability of radiomics features were evaluated by Pearson correlation and intraclass correlation coefficients (ICC). CE-Net and U-net with Resnet outperformed U-net and U-net++ in accuracy performance by achieving a DSC, JSC, and ASD of 0.87, 0.79, 8.54, and 0.86, 0.78, 10.00, respectively. A total of 97 features were extracted from the delineated target volumes. The average Pearson correlation was 0.86 (95% CI, 0.83–0.89), 0.87 (95% CI, 0.84–0.90), 0.88 (95% CI, 0.86–0.91), and 0.90 (95% CI, 0.88–0.92) for U-net++, U-net, U-net with Resnet, and CE-Net, respectively. The average ICC was 0.84 (95% CI, 0.81–0.87), 0.85 (95% CI, 0.82–0.88), 0.88 (95% CI, 0.85–0.90), and 0.89 (95% CI, 0.86–0.91) for U-net++, U-net, U-net with Resnet, and CE-Net, respectively. CE-Net based segmentation achieved the best radiomics reliability. In conclusion, U-net based

OPEN ACCESS

Edited by:
Wei Wang,
The First Affiliated Hospital of Sun Yat-Sen University, China

Reviewed by:
Jinghao Duan,
Shandong University, China
Qingtao Qiu,
Shandong Cancer Hospital, China

***Correspondence:**
Congying Xie
wzxiecongying@163.com
Xiance Jin
jinxc1979@hotmail.com

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *05 October 2020*
Accepted: *29 December 2020*
Published: *18 February 2021*

Citation:
Jin J, Zhu H, Zhang J, Ai Y, Zhang J, Teng Y, Xie C and Jin X (2021) Multiple U-Net-Based Automatic Segmentations and Radiomics Feature Stability on Ultrasound Images for Patients With Ovarian Cancer. Front. Oncol. 10:614201. doi: 10.3389/fonc.2020.614201

automatic segmentation was accurate enough to delineate the target volumes on US images for patients with ovarian cancer. Radiomics features extracted from automatic segmented targets showed good reproducibility and for reliability further radiomics investigations.

Keywords: automatic segmentation, U-net, ultrasound images, radiomics, ovarian cancer

INTRODUCTION

Ovarian cancer remains the second most common gynecological malignancy and the leading cause of death in women with gynecological cancer (1). Several imaging modalities, such as computed tomography (CT), ultrasonography (US), positron emission tomography (PET), and magnetic resonance imaging (MRI) have been used as diagnostic and treatment assessment tools for gynecological cancer all over the world (2, 3). US is a well recognized and most common applied image modality for diagnosis and assessment of ovarian cancer due to its advantage characteristics of non-invasive, no radiation, cheap and affordable (4, 5). Recently, the emerging radiomics to find association between clinical characteristics and qualitative and quantitative information extracted from US images, has further expanded the application and importance of US images for gynecological cancer (6).

By converting medical images into quantitative information, which was then analyzed subsequently using conventional biostatistics, machine learning techniques, and artificial intelligence (7), radiomics has been developed rapidly for clinical application to promote precision diagnostics and cancer treatment (8, 9). Multiple processes, such as imaging acquisition, region of interests (ROIs) segmentation, image feature extraction, and modeling, were involved in the radiomics analysis, in which ROI segmentation is the most critical, challenging, and contentious step (7).

Segmentation is the step of extracting or distinguishing a ROI from its background. It is a common and crucial stage in the quantitative and qualitative analysis of medical images, and usually it is one of the most important and earliest steps of image processing (10). Due to the low contrast, speckle noise, low signal noise ratio and artifacts inherently associated with ultrasound images, it presents unique challenges for the analysis on US images, especially for accurate segmentation of different structures and tumor volumes compared with other image modalities, e.g., CT, MRI (11, 12). The image quality of US has a high intra- and inter-observer variability across different institutes and manufactures. It also highly depends on the abundance and experience of operators or diagnosticians. All these render manual segmentation more variable and significantly impact the quantitative (e.g., radiomics) and geometric analyses with US images (13, 14).

The US segmentation problems have been the hot research topics and rapidly evolved over the past few years (11). Currently, no golden standard for tumor segmentation had been established and manual segmentation is usually applied (15). However, except for the inter and intra varieties mentioned above, the manual segmentation is also quite time consuming and boring. More recently, automatic segmentation techniques based on deep learning have become a main stream and show significant improvement in image classification predictions and recognition tasks (16). A well-known U-net architecture for biomedical imaging segmentation (17), which built uponfully convolutional network (18), has been successfully adapted to segment US images of breast (19), arterial walls (20), and gynecological cancer (21). Studies reported that the reproducibility and reliability of radiomics features could be deeply affected by the segmentation methods for CT (22), MR (23), and PET images (24). However, few studies have reported the reproducibility and stability of US based radiomics features obtained in oncology.

Previously, the feasibility of radiomics based on US images to predict the lymph node status for patients with gynecological cancer had been investigated (6). The purpose of this study is to investigate the accuracy of automatic segmentation algorithms based on multiple U-net models and their effects on radiomics features from US images for patients with ovarian cancer.

MATERIALS AND METHODS

Patients and Images

Patients with ovarian cancer underwent radical hysterectomy and transvaginal US diagnosis at authors' hospital from January 2002 to December 2016 were retrospectively reviewed in this study. The US images were acquired with a transvaginal ultrasonography using Voluson-E8 (GE Healthcare, Wilmington, USA) at 5–9 MHz, Philips (ATL HDI 5000, Netherland) at 4–8 MHz, and Esaote (MyLab classC) at 3–9 MHz or Hitachi (HI Vison Preirus) (Hitachi Ltd, Tokyo, Japan) at 4–8 MHz. All the images were reviewed with a Picture Archiving and Communication Systems (PACS).

Manual segmentation of target volumes was contoured by a radiologist with 7 years of experience in gynecological imaging and was further confirmed by a senior radiologist with > 15 years of experience in gynecological imaging. This retrospective study was approved by the Ethics Committee in Clinical Research (ECCR) of authors' hospital (ECCR#2019059). ECCR waived the need of written informed consent for this retrospective study. Patient data confidentiality was confirmed.

Automatic Segmentation Models

In this work, the classical U-net scheme and its multiple variations were used for the automatic segmentation task. Generally, the U-net is a symmetrical U-shaped model consisting of an encoder-decorder architecture (17). The left side encoder is a down-sampling used to get feature map, similar to a compression operation, while the right side decoder is an up-sampling used to restore the encoded features to the original image size and to output the results. Skip-connection was added to encoder-decoder networks in order to concatenate the features of high- and low-level together (17). When Resnet is used as a fixed feature encoder to deepen the layers of the network and solve the vanishing gradient, the U-net structure is changed to U-net with Resnet as the backbone (U-net with Resnet) (25). Resnet34 was preferred in this study.

A so-called context encoder network (CE-Net) was also employed in this study, which consists of three major parts: a feature encoder module, a feature decoder module and a context extractor. In CE-Net, Resnet block is used as a fixed feature extractor; a residual multi-kernel pooling (RMP) block and a dense atrous convolution (DAC) block consist of the context extractor module (26). U-net++ is a modified U-net with deeply-supervised encoder-decoder network, in which a series of nested, dense skip pathways are applied to connect the encoder and decoder sub-networks (27). A typical U-net structure was shown in **Figure 1**.

Image Preprocessing

Image clipping was performed on each image set in order to satisfy the size requirement of U-net and to shift the center of clipping box so as to make the training model robust (28). The tumor center minus the offset (a number from 360 to 0 at -60 intervals) was selected as the starting point for a 480 * 512 clipping box. The clipping box should not exceed the image edge. A typical image preprocessing was shown in **Figure 2**.

Radiomics Feature Extraction

After manual and automatic segmentations, the arbitrary gray intensity values on US images were transformed into a standardized intensity range by intensity normalizing. A python 3.7.0 and package Pyradiomics 2.2.0 were used to extract radiomic features from the segmented target volumes. According to different matrices capturing the spatial intensity distributions with four different scales, 79 texture features and 18 first-order histogram statistics were extracted from neighborhood gray-level different matrix (NGLDM), gray level co-occurrence matrix (GLCM), grey-level zone length matrix (GLZLM), and gray-level run length matrix (GLRLM).

Evaluation and Statistical Analysis

The automatic segmentation models were built with the image dataset randomly divided into training sets, validation sets and test sets. The results of automatic segmentation models were evaluated by comparing them with manually segmented targets. Jaccard similarity coefficient (JSC), dice similarity coefficient (DSC), and average surface distance (ASD) were applied during the evaluation of delineation using the four U-net-related models with test data sets (29). The effects of segmentation on the radiomics features were evaluated with Pearson correlation coefficient and intraclass correlation coefficients (ICC), in which the agreement of a certain radiomic feature (e.g., shape features, texture features) between automatic and manual segmentation was evaluated by ICC (30). General statistical analyses were performed in SPSS Statistics (version 20.0.0). Statical significance was considered as a $p < 0.05$.

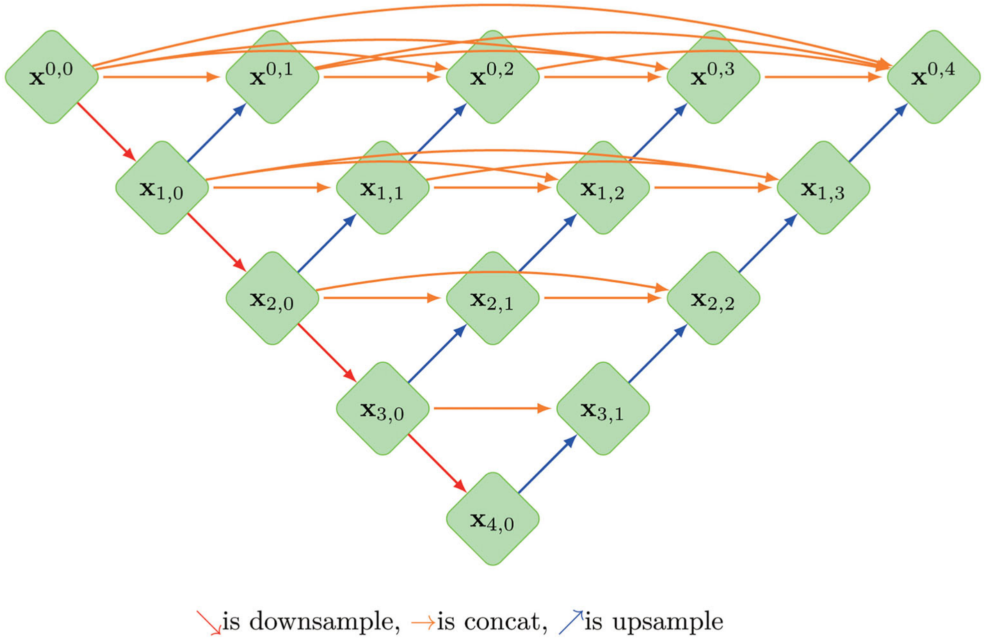


FIGURE 1 | The architecture of a typical U-net model, where $X^{i,\ j}$ is the operation of convolution block; Every $X^{i,\ j}$(j>0)'s input is concatenated from the up-sampling of $X^{i+1,\ j-1}$ from the previous convolution layer of the same dense block and all of $X^{i,\ k}$(k<j) from same pyramid level.

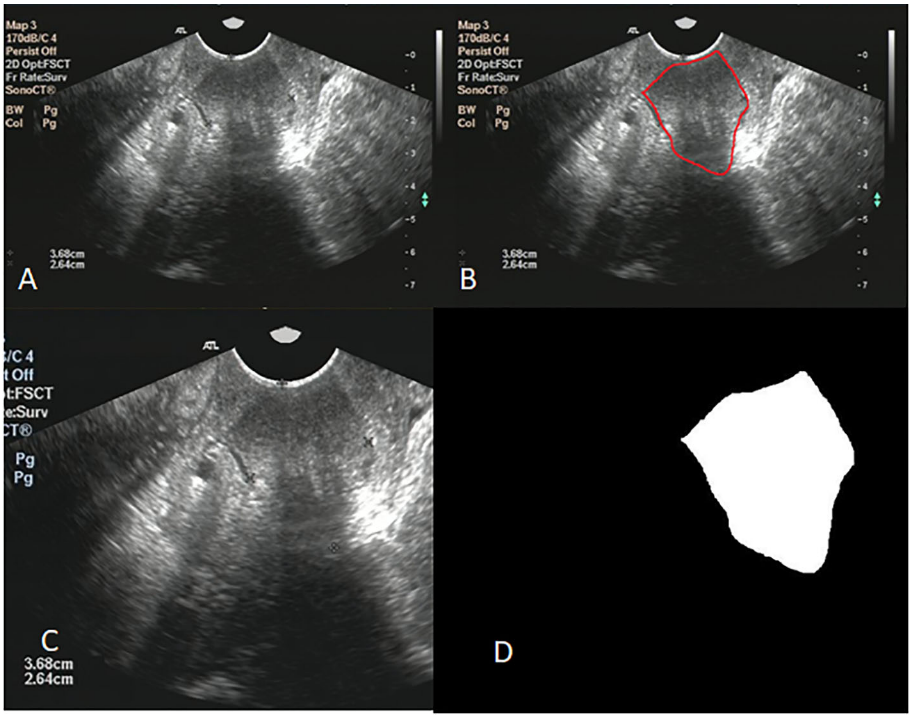


FIGURE 2 | **(A)** shows the original ultrasound image; **(B)** shows ovarian tumor segmented by radiologist; **(C)** shows the image after clipping; **(D)** shows the mask of ovarian.

RESULTS

There were 127 patients with ovarian cancer and with transvaginal US images included in the study. The median age of these patients was 56 years old (from 23 to 80 years). A total of 469 US images were analyzed and randomly divided into three groups: training sets (353 images), validation sets (23 images), and test sets (93 images) for the building of automatic segmentation models. Detailed characteristics of patients and images were presented in **Table 1**. No significant difference among the training, validation, testing sets in terms of age, histological type, and tumor stages was observed.

U-net, CE-Net, U-net++, and U-net with Resnet were applied to delineate automatically the target volumes of ovarian cancer

TABLE 1 | Clinical characteristics of enrolled patients and images.

Category		**Patients characteristics**	**Images**			
			Training sets	**Validation sets**	**Testing sets**	**p**
Total number		127	353	23	93	
Age (years)						0.344
	Mean	54.65	53.96	56.22	55.53	
	Median	56	54	59	56	
	Range	23~80	23~80	32~73	23~80	
	SD	11.85	11.38	11.01	9.79	
Histological types						0.679
	Epithelial	105	308	21	83	
	None epithelial	16	38	2	7	
	N.A.	2	12	1	5	
Tumor stages						0.691
	I	36	87	4	24	
	II	15	34	5	11	
	III	68	213	13	54	
	IV	3	19	1	4	
	N.A.	1	5	1	2	

Note. p value is calculated from the univariate association test between sub-groups, one-factor ANOVA for continues variables, Fisher's exact test for categorized variables.

on US images. **Figure 3** presents typical contours achieved by these automatic segmentation models and their comparison with manual contours. Detailed results of segmentation accuracy metrics were presented in **Table 2**. CE-Net and U-net with Resnet achieved a DSC and JSC of 0.87, 0.79, and 0.86, 0.78, respectively. The ASD of CE-Net and U-net with Resnet were 8.54 and 10.00, respectively.

There were 97 features extracted from the delineated target volumes. **Figure 4** shows the hot maps of Pearson correlation and ICC for the comparison between features extracted from automatic segmentations and manual contours. The average Pearson correlation was 0.86 (95% CI, 0.83–0.89), 0.87 (95% CI, 0.84–0.90), 0.88 (95% CI, 0.86–0.91), and 0.90 (95% CI, 0.88–0.92) for U-net++, U-net, U-net with Resnet, and CE-Net, respectively. The average ICC was 0.84 (95% CI, 0.81–0.87), 0.85 (95% CI, 0.82–0.88), 0.88 (95% CI, 0.85–0.90), and 0.89 (95% CI, 0.86–0.91) for U-net++, U-net, U-net with Resnet, and CE-Net, respectively.

High correlations were observed for most of the features except for some features of shape GLZLM. Detailed results of Pearson correlation and ICC for all the 97 features were presented in **Supplementary Tables 1** and **2**. Further analysis on the shape GLZLM features was shown in **Figure 5**. Sphericity and PerimeterSurfaceRatio were the two shape features that showed weak correlation between automatic and manual segmentations. Excluding these two shape features, the Pearson coefficient and ICC between features extracted by CE-Net and manual segmentation ranged from 0.71–0.98, and 0.70–0.97, respectively.

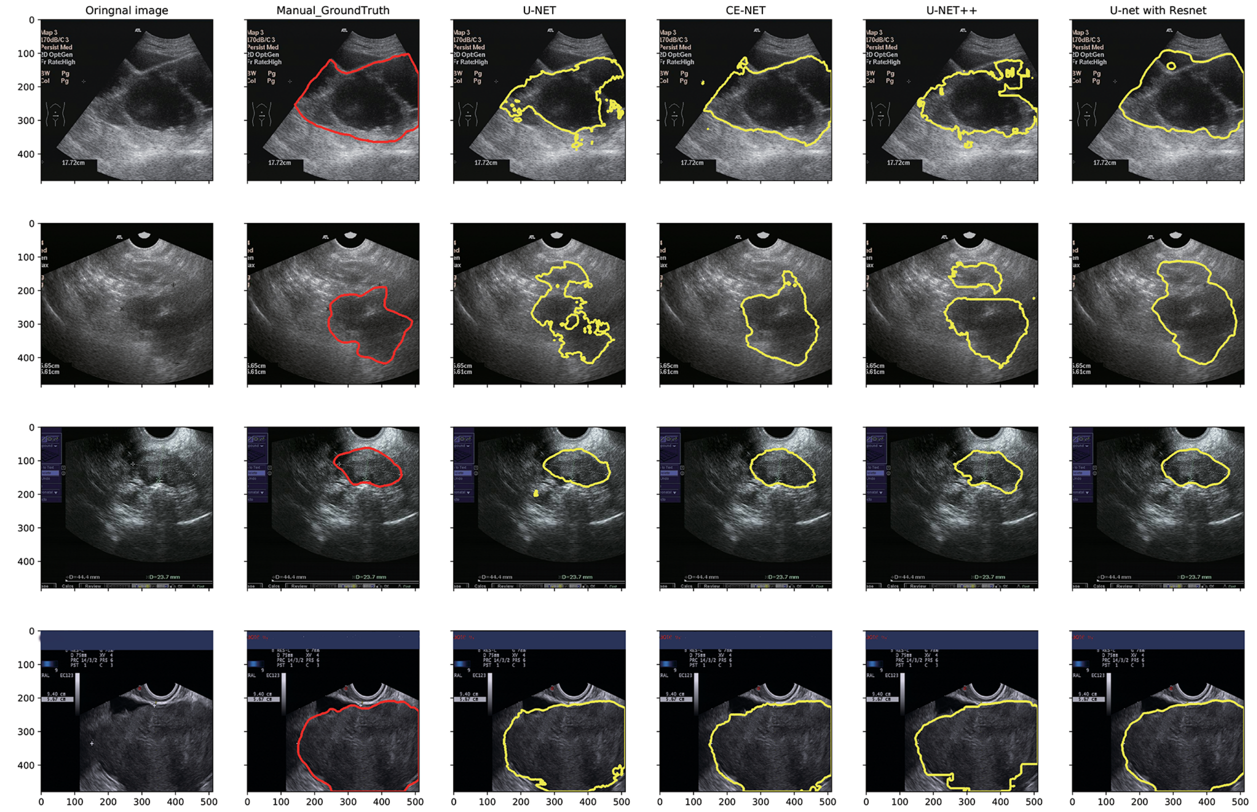


FIGURE 3 | Typical segmentation results from manual delineation, U-Net, CE-NET, U-net++, and U-net with Resnet models.

TABLE 2 | Automatic segmentation accuracy metrics for U-net-related models.

Models	*Evaluation metrics*					
	JSC (95%CI)		**DSC (95%CI)**		**ASD (95%CI)**	
U-net	0.71	0.68~0.75	0.81	0.79~0.85	10.57	9.25~11.89
CE-Net	0.79	0.76~0.82	0.87	0.85~0.90	8.54	7.21~9.86
U-net++	0.72	0.68~0.75	0.82	0.79~0.85	10.15	8.90~11.40
U-net with Resnet	0.78	0.75~0.82	0.86	0.83~0.89	10.00	8.03~11.97

Note. JSC, Jaccard similarity coefficient; DSC, Dice similarity coefficient; ASD, Average surface distance.

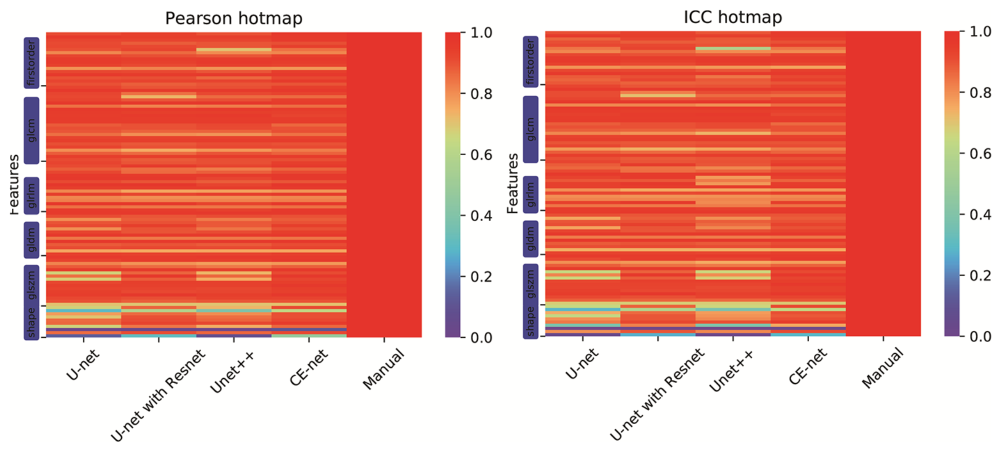


FIGURE 4 | Hot maps of Pearson correlation and intraclass correlation coefficients for radiomics features extracted from manual segmentation and U-net models based automatic segmentations.

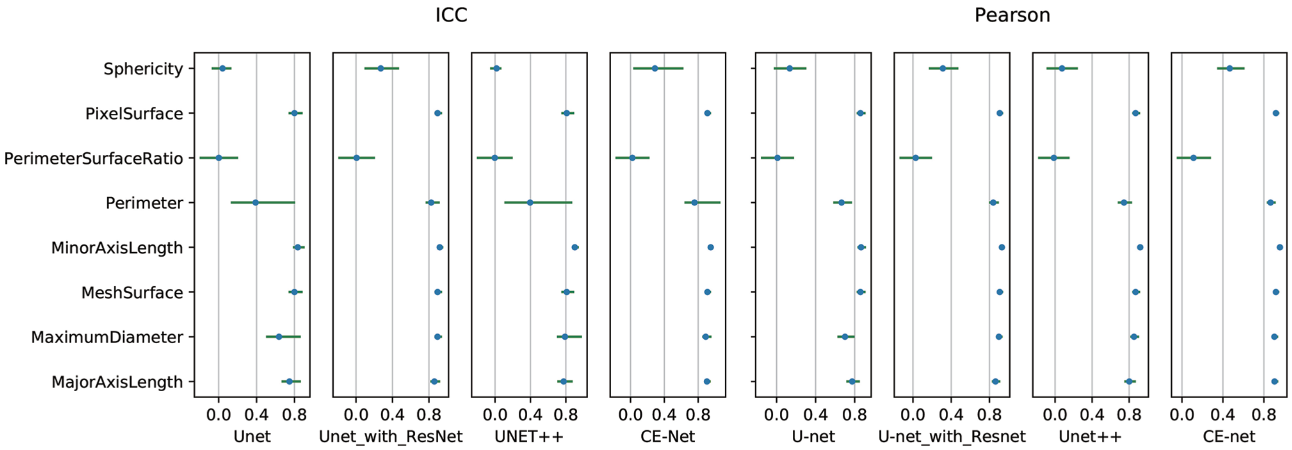


FIGURE 5 | Pearson correlation and intraclass correlation coefficients for shape features extracted from different U-net automatic segmentations.

DISCUSSION

Automatic segmentation of target volumes for ovarian cancer on US images were generated with multiple U-net models. The segmentation accuracy and its effects on radiomics features were evaluated in this study. CE-Net and U-net with Resnet models achieved a relatively higher accuracy on target delineation. Except for some shape features, most features extracted with automatic segmentation algorithms achieved high Pearson correlation and ICC in correlation with features extracted from manual contours.

US is a standard imaging modality for lots of diagnostic and monitoring purposes, had has been significantly investigated with deep learning based automatic segmentation (31). Yang et al. (32) used a fine-grained recurrent neural network to segment prostate US images automatically and achieved a high DSC around 0.92. Ghavami et al. (33) also proposed convolutional neural networks (CNNs) to automatically segment transrectal US images of prostate and got a mean DSC of 0.91 ± 0.12. Automatic segmentations on cardiac and carotid artery US images were proposed by Chen et al. (34) and Mechon-Lara et al. (35) using deep learning methods. Amiri et al. (36) fine-tuned the U-Net on breast US images and got a mean DSC of 0.80 ± 0.03. Similarly, a DSC of 0.83 to 0.90 was achieved on US images of ovarian cancer using different U-net models in this study.

U-net is a structure for medical image segmentation with superior skip connections design for different stages of the network, which had inspired the development of many variations (37). Marques et al. (21) explored different U-Net architectures with various hyperparameters in their automatic segmentations on the transvaginal US images of ovary and ovarian follicles, and indicated that architecture takes into account the spatial context of ROI is important for a better performance (21). In this study, Unet++, U-net, CE-Net, and

U-net with Resnet were applied for the automatic segmentations. As shown in **Table 2**, CE-Net and U-net with Resnet exhibited higher mean DSC and JSF, and lower mean ASD compared with U-net++ and U-net, where CE-Net achieved the best performance.

In radiomics analysis, usually the ROI contoured is the region analyzed. The reproducibility and reliability of radiomics features were highly impacted by the segmentation methods. Parmar et al. (38) demonstrated that semi-automatic segmentation (ICC: 0.85 ± 0.15) provided a better feature extraction reproducibility than manual segmentation (ICC, 0.77 ± 0.17) in CT images for 20 non-small cell lung cancer patients. Heye et al. (23) achieved an ICC of 0.99 with a semiautomatic segmentation on dynamic contrast material-enhanced MR images. In this study, a highest Pearson correlation and ICC of 0.90 (95% CI, 0.88–0.92) and 0.89 (95% CI, 0.86–0.91) were achieved with CE-Net automatic segmentation. Similarly, Lin et al. (39) achieved an ICC of 0.70–0.99 on first-order apparent diffusion coefficient radiomics parameters using U-net automatic segmentation for cervical cancer.

However, a few of shape textures showed worse correlation, as shown in **Figures 4** and **5**. This may be caused by artifacts resulted from less optimal automatic segmentation algorithms as shown in **Figure 3**, which could be improved by manual correction during clinical practice. This also indicated that automatic segmentation for US images needs further investigation to improve the reliability and reproducibility of delineated volumes and radiomics features. Future evaluation of the reliability and reproducibility may be focused on prediction modeling level instead of at the level of radiomics features.

CONCLUSIONS

U-net based automatic segmentation was accurate enough to delineate the target volumes on US images for patients with ovarian cancer. Radiomics features extracted from automatic segmented ROI showed high reliability and reproducibility for further radiomics investigations.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding authors.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee in Clinical Research (ECCR) of the First Affiliated Hospital of Wenzhou Medical University. The ethics committee waived the requirement of written informed consent for participation.

AUTHOR CONTRIBUTIONS

CX and XJ conceptualized and designed the study. XJ and CX provided administrative support. CX and HZ provided the study materials or patients. JJ and HZ collected and assembled the data. YA, JZ, and YT analyzed and interpreted the data. JJ and XJ wrote the manuscript. JJ, HZ, JZ, YA, JDZ, YT, CX, and XJ gave the final approval of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was partially funded by the Wenzhou Municipal Science and Technology Bureau (2018ZY016, 2019) and National Natural Science Foundation of China (No.11675122, 2017).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2020.614201/full#supplementary-material

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2017. *CA Cancer J Clin* (2017) 67:7–30. doi: 10.3322/caac.21387
2. Williams AD, Cousins C, Soutter WP, Mubashar M, Peters AM, Dina R, et al. Detection of Pelvic Lymph Node Metastases in Gynecologic Malignancy. *Am J Roentgenol* (2001) 177:343–8. doi: 10.2214/ajr.177.2.1770343
3. Alcázar JL, Arribas S, Mínguez JA, Jurado M. The Role of Ultrasound in the Assessment of Uterine Cervical Cancer. *J Obstet Gynecol India* (2014) 64:311–6. doi: 10.1007/s13224-014-0622-4
4. Meys EMJ, Kaijser J, Kruitwagen RFPM, Slangen BFM, Van Calster B, Aertgeerts B, et al. Subjective assessment versus ultrasound models to diagnose ovarian cancer: A systematic review and meta-analysis. *Eur J Cancer* (2016) 58:17–29. doi: 10.1016/j.ejca.2016.01.007
5. Kyriazi S, Kaye SB, Desouza NM. Imaging ovarian cancer and peritoneal metastases - current and emerging techniques. *Nat Rev Clin Oncol* (2010) 7:381–93. doi: 10.1038/nrclinonc.2010.47
6. Jin X, Ai Y, Zhang J, Zhu H, Jin J, Teng Y, et al. Noninvasive prediction of lymph node status for patients with early-stage cervical cancer based on radiomics features from ultrasound images. *Eur Radiol* (2020) 30:4117–24. doi: 10.1007/s00330-020-06692-1
7. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images are more than pictures, they are data. *Radiology* (2016) 278:563–77. doi: 10.1148/radiol.2015151169
8. Huang YQ, Liang CH, He L, Tian J, Liang CS, Chen X, et al. Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. *J Clin Oncol* (2016) 34:2157–64. doi: 10.1200/JCO.2015.65.9128
9. Liu Z, Zhang XY, Shi YJ, Wang L, Zhu HT, Tang Z, et al. Radiomics analysis for evaluation of pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer. *Clin Cancer Res* (2017) 23:7253–62. doi: 10.1158/1078-0432.CCR-17-1038
10. Shi C, Cheng Y, Wang J, Wang Y, Mori K, Tamura S. Low-rank and sparse decomposition based shape model and probabilistic atlas for automatic pathological organ segmentation. *Med Image Anal* (2017) 38:30–49. doi: 10.1016/j.media.2017.02.008
11. Noble JA, Boukerroui D. Ultrasound image segmentation: A survey. *IEEE Trans Med Imaging* (2006) 25:987–1010. doi: 10.1109/TMI.2006.877092

12. Shi C, Cheng Y, Liu F, Wang Y, Bai J, Tamura S. A hierarchical local region-based sparse shape composition for liver segmentation in CT scans. *Pattern Recognit* (2016) 50:88–106. doi: 10.1016/j.patcog.2015.09.001
13. Zhao J, Zheng W, Zhang L, Tian H. Segmentation of ultrasound images of thyroid nodule for assisting fine needle aspiration cytology. *Heal Inf Sci Syst* (2013) 1:5. doi: 10.1186/2047-2501-1-5
14. Yazbek J, Ameye L, Testa AC, Valentin L, Timmerman D, Holland TK, et al. Confidence of expert ultrasound operators in making a diagnosis of adnexal tumor: effect on diagnostic accuracy and interobserver agreement. *Ultrasound Obstet Gynecol* (2010) 35:89–93. doi: 10.1002/uog.7335
15. Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, et al. Radiomics: The process and the challenges. *Magn Reson Imaging* (2012) 30:1234–48. doi: 10.1016/j.mri.2012.06.010
16. Nair AA, Tran TD, Reiter A, Lediju Bell MA. A Deep Learning Based Alternative to Beamforming Ultrasound Images. *ICASSP IEEE Int Conf Acoust Speech Signal Process - Proc* (2018) 2018-April:3359–63. doi: 10.1109/ICASSP.2018.8461575
17. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. *Lect Notes Comput Sci (Incl Subser Lect Notes Artif Intell Lect Notes Bioinforma)* (2015) 9351:234–41. doi: 10.1007/978-3-319-24574-4_28
18. Shelhamer E, Long J, Darrell T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Trans Pattern Anal Mach Intell* (2017) 39:640–51. doi: 10.1109/TPAMI.2016.2572683
19. Yap MH, Goyal M, Osman FM, Martí R, Denton E, Juette A, et al. Breast ultrasound lesions recognition: end-to-end deep learning approaches. *J Med Imaging (Bellingham Wash)* (2019) 6:11007. doi: 10.1117/1.JMI.6.1.011007
20. Yang J, Faraji M, Basu A. Robust segmentation of arterial walls in intravascular ultrasound images using Dual Path U-Net. *Ultrasonics* (2019) 96:24–33. doi: 10.1016/j.ultras.2019.03.014
21. Marques S, Carvalho C, Peixoto C, Pignatelli D, Beires J, Silva J, et al. Segmentation of gynaecological ultrasound images using different U-Net based approaches. *IEEE Int Ultrason Symp IUS* (2019) 2019-Octob:1485–8. doi: 10.1109/ULTSYM.2019.8925948
22. Armato SG, McLennan G, Bidaut L, McNitt-Gray MF, Meyer CR, Reeves AP, et al. The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A completed reference database of lung nodules on CT scans. *Med Phys* (2011) 38:915–31. doi: 10.1118/1.3528204
23. Heye T, Merkle EM, Reiner CS, Davenport MS, Horvath JJ, Feuerlein S, et al. Reproducibility of dynamic contrast-enhanced MR imaging part II. Comparison of intra- and interobserver variability with manual region of interest placement versus semiautomatic lesion segmentation and histogram analysis. *Radiology* (2013) 266:812–21. doi: 10.1148/radiol.12120255
24. Hatt M, Lee JA, Schmidtlein CR, El Naqa I, Caldwell C, De Bernardi E, et al. Classification and evaluation strategies of auto-segmentation approaches for PET: Report of AAPM task group No. 211. *Med Phys* (2017) 44:e1–e42. doi: 10.1002/mp.12124
25. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV (2016). p. 770–8. doi: 10.1109/CVPR.2016.90
26. Gu Z, Cheng J, Fu H, Zhou K, Hao H, Zhao Y, et al. CE-Net: Context Encoder Network for 2D Medical Image Segmentation. *IEEE Trans Med Imaging* (2019) 38:2281–92. doi: 10.1109/TMI.2019.2903562
27. Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, Liang J. Unet++: A nested u-net architecture for medical image segmentation. *Lect Notes Comput Sci (Incl Subser Lect Notes Artif Intell Lect Notes Bioinforma)* (2018) 11045 LNCS:3–11. doi: 10.1007/978-3-030-00889-5_1
28. Aithal S, Krishna PK. Two Dimensional Clipping Based Segmentation Algorithm for Grayscale Fingerprint Images. *Soc Sci Electron Publ* (2017) 1:51–6. doi: 10.5281/zenodo.1037627
29. Heimann T, Van Ginneken B, Styner MA, Arzhaeva Y, Aurich V, Bauer C, et al. Comparison and evaluation of methods for liver segmentation from CT datasets. *IEEE Trans Med Imaging* (2009) 28:1251–65. doi: 10.1109/TMI.2009.2013851
30. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med* (2016) 15:155–63. doi: 10.1016/j.jcm.2016.02.012
31. Yang X, Yu L, Wu L, Wang Y, Ni D, Qin J, Heng PA, et al. Fine-grained recurrent neural networks for automatic prostate segmentation in ultrasound images. In: *31st AAAI Conference on Artificial Intelligence, AAAI*. San Francisco, CA (2017).
32. Yang X, Yu L, Li S, Wang X, Wang N, Qin J, Heng PA. Towards automatic semantic segmentation. In: *Medical Image Computing and Computer Assisted Intervention (MICCAI)*. Springer, Cham (2017). doi: 10.1007/978-3-319-66182-7_81
33. Ghavami N, Hu Y, Bonmati E, Rodell R, Gibson E, Moore C, et al. Integration of spatial information in convolutional neural networks for automatic segmentation of intraoperative transrectal ultrasound images. *J Med Imaging* (2018) 6:011003. doi: 10.1117/1.jmi.6.1.011003
34. Chen H, Zheng Y, Park JH, Heng PA, Zhou SK. Iterative multi-domain regularized deep learning for anatomical structure detection and segmentation from ultrasound images. In: *Medical Image Computing and Computer Assisted Intervention (MICCAI)*. Springer, Cham (2016). doi: 10.1007/978-3-319-46723-8_56
35. Menchón-Lara RM, Sancho-Gómez JL. Fully automatic segmentation of ultrasound common carotid artery images based on machine learning. *Neurocomputing* (2015) 45:215–28. doi: 10.1016/j.neucom.2014.09.066
36. Amiri M, Brooks R, Rivaz H. Fine-Tuning U-Net for Ultrasound Image Segmentation: Different Layers, Different Outcomes. IEEE Trans Ultrason Ferroelectr Freq Control (2020) 67:2510–18. doi: 10.1109/TUFFC.2020.3015081
37. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. "3D U-net: Learning dense volumetric segmentation from sparse annotation", In: *Medical Image Computing and Computer Assisted Intervention (MICCAI)*. Springer, Cham (2016). doi: 10.1007/978-3-319-46723-8_49
38. Parmar C, Velazquez ER, Leijenaar R, Jermoumi M, Carvalho S, Mak RH, et al. Robust radiomics feature quantification using semiautomatic volumetric segmentation. *PLoS One* (2014) 9:e102107. doi: 10.1371/journal.pone.0102107
39. Lin YC, Lin CH, Lu HY, Chiang HJ, Wang HK, Huang YT, et al. Deep learning for fully automated tumor segmentation and extraction of magnetic resonance radiomics features in cervical cancer. *Eur Radiol* (2019). doi: 10.1007/s00330-019-06467-3

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 04 March 2021
doi: 10.3389/fonc.2021.625646

A Radiomic Nomogram for the Ultrasound-Based Evaluation of Extrathyroidal Extension in Papillary Thyroid Carcinoma

Xian Wang[1†], Enock Adjei Agyekum[2†], Yongzhen Ren[2†], Jin Zhang[1], Qing Zhang[1], Hui Sun[3], Guoliang Zhang[4], Feiju Xu[1], Xiangshu Bo[1], Wenzhi Lv[5], Shudong Hu[6] and Xiaoqin Qian[1*]*

[1] Department of Ultrasound, Affiliated People's Hospital of Jiangsu University, Zhenjiang, China, [2] School of Medicine, Jiangsu University, Zhenjiang, China, [3] Department of Pathology, Affiliated People's Hospital of Jiangsu University, Zhenjiang, China, [4] Department of General Surgery, Affiliated People's Hospital of Jiangsu University, Zhenjiang, China, [5] Department of Artificial Intelligence, Julei Technology Company, Wuhan, China, [6] Department of Radiology, The Affiliated Hospital, Jiangnan University, Wuxi, China

Edited by:
Hui-Xiong Xu,
Tongji University, China

Reviewed by:
Hang-Tong Hu,
First Affiliated Hospital of Sun Yat-Sen University, China
Jinhua Yu,
Fudan University, China

Correspondence:
Xiaoqin Qian
yz_tyz1030@126.com
Shudong Hu
hsd2001054@163.com

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *03 November 2020*
Accepted: *11 January 2021*
Published: *04 March 2021*

Citation:
Wang X, Agyekum EA, Ren Y, Zhang J, Zhang Q, Sun H, Zhang G, Xu F, Bo X, Lv W, Hu S and Qian X (2021) A Radiomic Nomogram for the Ultrasound-Based Evaluation of Extrathyroidal Extension in Papillary Thyroid Carcinoma. Front. Oncol. 11:625646. doi: 10.3389/fonc.2021.625646

Purpose: To construct a sequence diagram based on radiological and clinical factors for the evaluation of extrathyroidal extension (ETE) in patients with papillary thyroid carcinoma (PTC).

Materials and Methods: Between January 2016 and January 2020, 161 patients with PTC who underwent preoperative ultrasound examination in the Affiliated People's Hospital of Jiangsu University were enrolled in this retrospective study. According to the pathology results, the enrolled patients were divided into a non-ETE group and an ETE group. All patients were randomly divided into a training cohort (n = 97) and a validation cohort (n = 64). A total of 479 image features of lesion areas in ultrasonic images were extracted. The radiomic signature was developed using least absolute shrinkage and selection operator algorithms after feature selection using the minimum redundancy maximum relevance method. The radiomic nomogram model was established by multivariable logistic regression analysis based on the radiomic signature and clinical risk factors. The discrimination, calibration, and clinical usefulness of the nomogram model were evaluated in the training and validation cohorts.

Results: The radiomic signature consisted of six radiomic features determined in ultrasound images. The radiomic nomogram included the parameters tumor location, radiological ETE diagnosis, and the radiomic signature. Area under the curve (AUC) values confirmed good discrimination of this nomogram in the training cohort [AUC, 0.837; 95% confidence interval (CI), 0.756–0.919] and the validation cohort (AUC, 0.824; 95% CI, 0.723–0.925). The decision curve analysis showed that the radiomic nomogram has good clinical application value.

Conclusion: The newly developed radiomic nomogram model is a noninvasive and reliable tool with high accuracy to predict ETE in patients with PTC.

Keywords: nomogram, ultrasound radiomics, papillary thyroid carcinoma, extrathyroidal extension, thyroid neoplasms, ultrasonography

INTRODUCTION

Papillary thyroid carcinoma (PTC) occurs in 90% of patients with thyroid carcinoma (1–3). PTCs are inert, differentiated cancers with relatively low recurrence and incidence rates. However, some histologic PTC subtypes (high cell count, diffuse sclerosing type, infiltrative type) show aggressive behavior and recurrence with extrathyroidal extension (ETE), vascular invasion, and distant metastasis.

According to the TNM classification of differentiated thyroid cancer by the Eighth Edition of the American Joint Committee on Cancer (AJCC) (2–5), minimal ETE (i.e., T3) refers to a primary tumor of more than 4 cm that is limited to the thyroid gland or has invaded the strap muscles surrounding the thyroid, whereas extensive ETE (i.e., T4) describes the primary tumor invasion of the subcutaneous soft tissue, larynx, trachea, oesophagus, recurrent laryngeal nerve, prevertebral fascia, carotid artery, or mediastinal vessels. Several studies demonstrated that ETE is an independent risk factor for high recurrence and mortality in PTC patients (3, 4). The detection of ETE is also clinically significant regarding the selection of the optimal treatment. PTC surgery mainly involves either total/subtotal thyroidectomy or resection of the affected thyroid lobe and isthmus, but both surgical procedures have no significant effect on postoperative distant metastasis and mortality. Besides, surgical procedures targeting only one thyroid lobe or the isthmus do not only retain some functionality of the thyroid gland but protect also parathyroid functions and prevent injuries of the contralateral laryngeal recurrent nerve. Total/subtotal thyroidectomy is usually suggested for PTC patients with ETE, whereas PTC patients without ETE are treated with resection of the affected thyroid lobe and isthmus. Thus, there is a need for a noninvasive method to evaluate ETE, thereby avoiding total/subtotal thyroidectomy in patients without ETE.

Ultrasound is the most common imaging modality for preoperative PTC diagnosis (6–9), but it is subjective and relies on the experience level of the operator. Computed tomography (CT) has a certain advantage in evaluating whether PTC has invaded neighboring tissue, but CT requires ionizing radiation. Magnetic resonance imaging can improve soft tissue resolution; however, it is expensive and not widely used in the examination of thyroid tumors. Ultrasound radiomics (USR) is a new tool that can extract hundreds of quantitative features from medical images and combine the key features into a radiomic signature, an image-based biomarker, that can be used for cancer diagnosis (10, 11). Some studies demonstrated that ultrasound imaging has great value in the diagnosis of various diseases, as well as the assessment of their prognosis (12–14). It has been reported that the texture features of ultrasound images have good predictive value for cervical lymph node metastasis in PTC (12–15). However, only a few studies analyzed ETE in PTC using USR. Therefore, we developed and validated in this study a USR model for the noninvasive prediction of the preoperative ETE status in PTC.

MATERIALS AND METHODS

Patients

This retrospective study was approved by the local ethics committee at the Affiliated People's Hospital of Jiangsu University, and the requirement for informed consent was waived. Between January 2016 and January 2020, 161 patients of our hospital were retrospectively selected. **Figure 1** shows the enrolment procedure. The inclusion criteria were defined as follows: (1) preoperative thyroid ultrasound examination, providing relevant ultrasound image data, and a PTC diagnosis; (2) the postoperative pathology confirmed the PTC diagnosis; (3) a single unilateral lesion; and (4) no prior thyroidectomy. Exclusion criteria were: (1) the primary tumor was not unequivocally identifiable on the ultrasound image and (2) the maximum diameter of the primary tumor was <5 mm. The clinical and pathological information of the enrolled patients included age, sex, tumor size, tumor position, and tumor location. The AJCC deleted in the eighth edition of the TNM classification the definition of minimal ETE because it was not considered an independent risk factor related to PTC prognosis. However, this concept remains controversial, and some researchers suggest that minimal ETE increases the risk of PTC recurrence (3, 4). Therefore, in this study, minimal ETE and extensive ETE were uniformly classified as ETE.

Ultrasound Examination

Preoperatively, all patients underwent a routine ultrasound examination, performed by well-trained technicians using a Philips Q5, Philips iU22 (both Healthcare, Eindhoven, the Netherlands) or a GE LOGIQ s8, LOGIQ E20, LOGIQ E9 (GE Medical Systems, American General) ultrasound system with a 5–12 MHz linear array transducer. The patient was placed in a supine position with the pillow removed to lower and slightly recline the head. This exposed the neck region as much as possible to carry out the ultrasound examination of the thyroid and cervical area using longitudinal, horizontal continuous scanning. This allowed the observation of the thyroid tumor size (maximum long axis of the nodule), tumor position (left lobe, right lobe, or isthmus), tumor location (upper, middle, or lower pole), internal echo pattern (uniform, less uniform, or nonuniform), tumor border (clear, less clear, or fuzzy), tumor vascularization (without, rare, or abundant), elastic properties (0 points: tumor tissue color-coded between red and green; 1 point: uniform green tumor; 2 points: tumor mainly between green and blue-green; 3 points: tumor mainly between blue-green and blue; 4 points: uniform blue tumor), and the ETE diagnosis. Preoperative imaging criteria for US diagnosis ETE were as follows: the lesions contact the thyroid capsule >25% or protrude the thyroid capsule to invade the sternum thyroid

---

Abbreviations: AJCC, American Joint Committee on Cancer; AUC, area under the curve; CI, confidence interval; CT, computed tomography; ETE, extrathyroidal extension; GLCM, gray-level co-occurrence matrix; GLDM, gray-level dependence matrix; GLRLM, gray-level run length matrix; GLSZM, grey-level size zone matrix; ICC, interclass correlation coefficient; LASSO, least absolute shrinkage and selection operator; mRMR, maximum relevance minimum redundancy; NGTDM, neighbourhood grey-tone dependency matrix; PTC, papillary thyroid carcinoma; ROI, region of interest; USR, ultrasound radiomics.

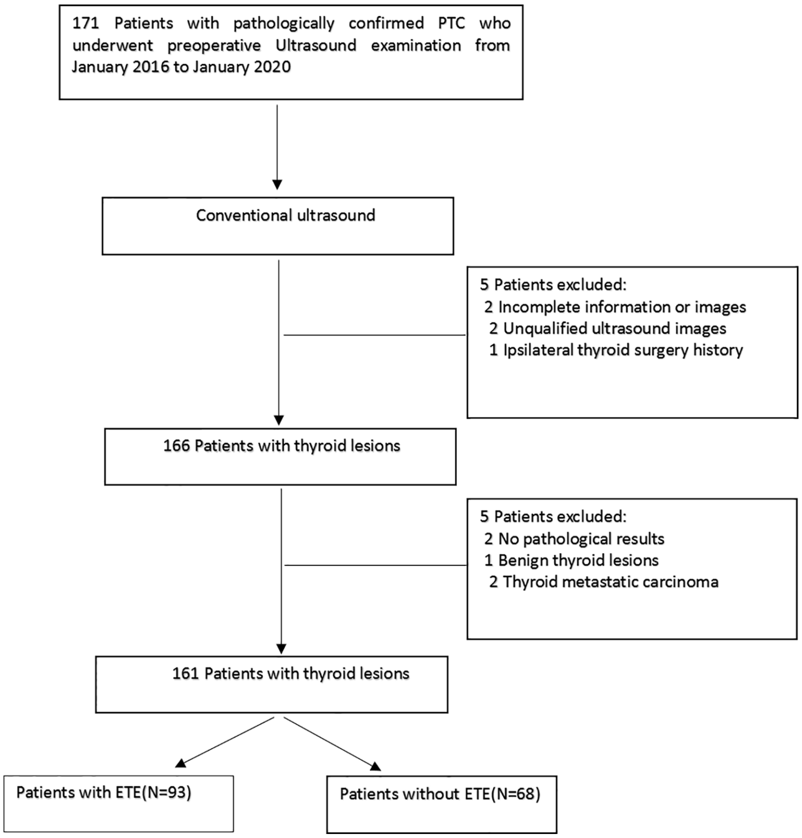


FIGURE 1 | Schematic diagram of the patient selection. PTC, papillary thyroid carcinoma.

muscle and the soft tissue around the thyroid gland, the fat space between the trachea, esophagus, trachea and esophagus sulcus, cervical sheath vessels, and the thyroid gland disappears.

Diagnostic Criteria for Image Analysis

Two radiologists blinded to the clinical and pathological information assessed the ultrasound images, and any disagreement was resolved by agreement. According to the AJCC guidelines (3–7), ETE can be diagnosed when one of the following two criteria exists: (1) >25% of the circumference of the lesion is in contact with the thyroid capsule or the envelope line of the contact between the lesion and the thyroid gland disappears; (2) a tumor of any size exceeds the thyroid capsule and invades the subcutaneous soft tissue, larynx, trachea, oesophagus, recurrent laryngeal nerve, carotid artery, or mediastinal vessels.

Region of Interest (ROI) Segmentation and Radiomic Feature Extraction

To indicate focal areas within the thyroid gland, ROIs were manually drawn on ultrasound images by one radiologist with 15 years of experience in the diagnosis of thyroid diseases using the software ITK-SNAP (version 3.8.0, http://www.itksnap.org). The ROI was placed on the solid component of the tumor, avoiding necrotic, hemorrhagic, and cystic areas.

To assess the consistency of the ROI placements, 30 patients were randomly selected, and a second physician with 8 years of experience in thyroid ultrasound diagnosis independently placed ROIs on the relevant structures.

From these ROIs on ultrasound images, 479 image features were extracted using PyRadiomics (version 2.2.0, https://github.com/Radiomics/pyradiomics). These features included 18 first-order features, 14 shape features, 16 grey-level run length matrix (GLRLM) features, 16 grey-level size zone matrix (GLSZM) features, 14 grey-level dependence matrix (GLDM) features, 5 neighbourhood grey-tone dependency matrix (NGTDM) features, 24 grey-level co-occurrence matrix (GLCM) features, and 372 features derived from first-order GLCM, GLRLM, GLSZM, GLDM, and NGTDM features using wavelet filter images.

Feature Selection and Radiomic Signature Construction

The consistency of the extracted ROI characteristics was evaluated using the interclass correlation coefficient (ICC). The analysis revealed an ICC of >0.8, demonstrating a good consistency of these characteristics. Next, we used the independent sample t-test or Mann-Whitney U test in the two groups to eliminate nonsignificant features with P-values of >0.05. The minimum redundancy maximum relevance (mRMR) algorithm was employed to assess the relevance and redundancy of the remaining features, and the top 10 features with high relevance and low redundancy were selected for the following analyses. The least absolute shrinkage and selection operator (LASSO) logistic regression model with 10-fold cross-validation was adopted for further feature selection and radiomic signature construction in the training cohort. The radiomic signature was generated by LASSO regression using a linear combination of the selected features with nonzero coefficient weight. Finally, the potential association of the radiomic signature with ETE was evaluated in the training and validation cohorts. **Figure 2** shows the workflow of this study.

Development of the Ultrasound Radiomic Nomogram

Based on multivariate logistic regression analysis, the clinical model was generated using clinical factors with P-values of <0.05. In this model, clinical risk factors such as age, sex, tumor size, tumor position, tumor location, internal echo pattern, tumor border, tumor vascularization, elastic properties, and radiological ETE diagnosis were included. A radiomic nomogram incorporating the radiomic signature and clinical risk factors was developed and used to intelligently predict ETE based on the multivariate analysis in the training cohort (**Figure 2**). For comparison, a clinical model was developed using the independent clinical risk factors alone.

Performance and Clinical Utility of the Radiomic Nomogram

The radiomic nomogram was evaluated using a calibration curve and the Hosmer-Lemeshow test (a nonsignificant test denotes that the model calibrates perfectly). The nomogram-predicted probability of each patient was calculated according to the nomogram algorithm. The discrimination performance of the radiomic nomogram-predicted probability was evaluated based on the receiver operating characteristic curve, sensitivity, and specificity. Then, the performances of the radiomic nomogram-predicted probabilities were tested in the training and validation cohorts. The decision curve analysis was applied in determining the clinical usefulness of the radiomic nomogram by calculating the net benefits at different threshold values in the combined training and validation cohort (16–18).

Histopathological Examination

PTC specimens of paraffin embedding slice, after HE dyeing, by two attending pathologists according to the American Thyroid Association (ATA) published guidelines for the diagnosis and treatment of thyroid cancer malignant degree classification standard, classification under the lens. The tumor is considered to be invasive if one of the following pathological manifestations is present: (1) the vascular or enveloped thyroid gland is invaded by the tumor; or (2) tumor invasion beyond the thyroid, tumor regional metastasis, and distant metastasis.

Statistical Analysis

Statistical analyses were processed using the R software (version 3.6.1, https://www.r-project.org). Pearson's chi-square or Fisher's exact test was used to compare differences for categorical characteristics. The independent sample t-test was performed for continuous factors with normal distribution, whereas the Mann-Whitney U test was used for continuous factors without normal distribution. A two-sided $P < 0.05$ denoted statistically significant differences.

RESULTS

Clinical Characteristics

A total of 161 patients with PTC were enrolled with an average age of 46.09 ± 11.79 years and a male-to-female ratio of 38:123. The pathology excluded ETE in 68 patients and confirmed it in 93 patients. Using stratified sampling, all patients were randomly divided into a training group ($n = 97$) and a validation group ($n = 64$). The clinical data of the training group and the validation group are shown in **Table 1**. There was no significant difference between the two groups in pathology and ultrasound image characteristics (all $P > 0.05$).

Radiomic Signature Construction and Diagnostic Validation

A total of 479 imaging features were extracted from each greyscale ultrasound image. Of those, 87 image features with ICC values ≤0.8 were deleted, and 256 radiomic features with no statistical significance in the training cohort according to the

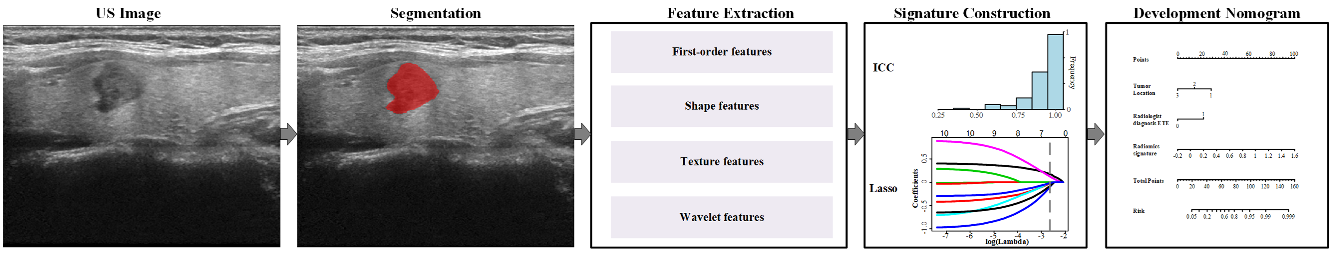


FIGURE 2 | Radiomics workflow in this study.

TABLE 1 | Patient characteristics of the training and validation cohorts.

Characteristic	Training cohort (n = 97)	Validation cohort (n = 64)	P
Age, mean ± SD, years	45.57 ± 11.87	46.89 ± 11.73	0.972
Sex, n			
Female	72	51	0.425
Male	25	13	
Tumor size in ultrasound	10.25 ± 7.99	11.58 ± 8.75	0.089
Tumor location			
Left lobe	45	30	
Right lobe	5	3	0.991
Isthmus	47	31	
Tumor position			
Upper pole	50	31	
Middle pole	32	14	0.066
Inferior pole	15	19	
Internal echo pattern			
Uniform	12	11	
Owe uniform	44	27	0.691
Nonuniform	41	26	
Tumor border			
Clear	35	33	0.254
Less clear	39	21	
Fuzzy	19	9	
Tumor vascularization			
Without	33	19	
Rare	28	33	0.831
Abundant	16	12	
Elastic properties classification			
2	11	7	
3	70	48	0.907
4	16	9	
Radiological ETE diagnosis			
Without ETE	20	22	0.052
With ETE	77	42	

ETE, extrathyroidal extension; SD, standard deviation.

t-test or Mann-Whitney U test were eliminated. Then, 10 image feature subsets with the best ETE discrimination were screened using the mRMR method. Finally, using LASSO regression and 10-fold cross-validation, six features with nonzero coefficients were selected in the training cohort (**Figure 3**). According to the result of the LASSO regression analysis, the mathematical expression of the radiomic signature was:

$$
\begin{aligned}
&Radiomic\ \ signature \\
&\quad = 0.335 + 0.176 \times LL_{glszm} \\
&\qquad + 0.122 \times ngtdm - 0.100 \times HH_{glcm} \\
&\qquad - 0.080 \times LH_{glcm\ \ Corr} - 0.024 \times LH_{glcm\ \ Clus} \\
&\qquad - 0.016 \times LH_{glszm}
\end{aligned}
$$

th LL_{glszm} being Zone Entropy of the wavelet low frequency filtered image GLSZM, (wavelet.LL_glszm_ZoneEntropy), ngtdm being Busyness of the original image NGTDM (original_ngtdm_Busyness), HH_{glcm} being Cluster Shade of the wavelet high frequency filtered image GLCM (wavelet.HH_glcm_ClusterShade), $LH_{glcmCorr}$ being wavelet low frequency filtered image has correlation of GLCM (wavelet.LH_glcm_Correlation), $LH_{glcmClus}$ being Cluster Shade of GLCM (wavelet.LH_glcm_ClusterShade), and LH_{glszm} being grey-level nonuniformity normalized of GLSZM (wavelet.LH_glszm_GLNUN).

Development and Validation of the Radiomic Nomogram

The clinical model was built using the variables age, sex, tumor size, tumor position, tumor location, internal echo pattern, tumor border, tumor vascularization, elastic properties, and radiological ETE diagnosis. The univariate logistic regression analysis selected three statistically significant clinical factors, namely tumor size (P = 0.039), tumor position (P < 0.001), and radiological ETE diagnosis (P = 0.002). Multivariate logistic

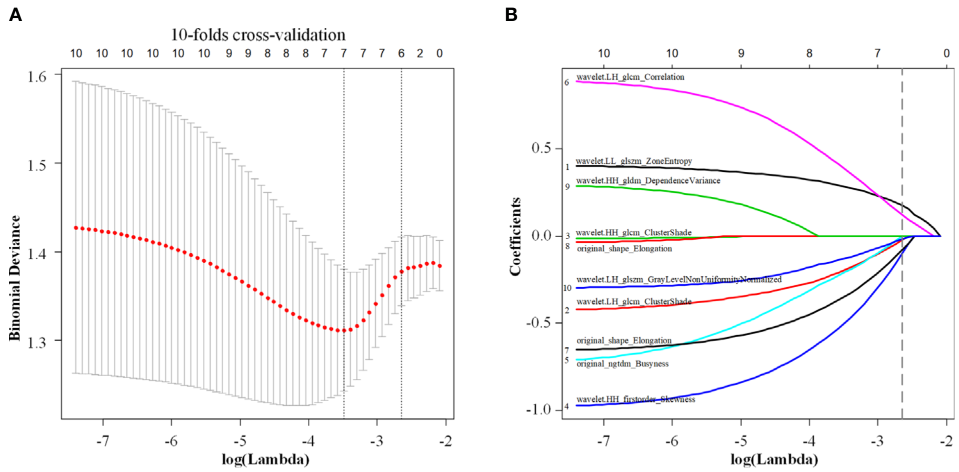


FIGURE 3 | Least absolute shrinkage and selection operator (LASSO) regression with 10-fold cross-validation **(A)** was used to reduce the dimension of the grouping characteristics **(B)**. Six features corresponded to the minimum error.

TABLE 2 | Multivariate logistic regression analysis.

Variable	Clinical model			Radiomic nomogram model		
	β	OR (95% CI)	P	B	OR (95% CI)	P
Cut off	0.671			–0.363		
Tumor location	–1.070	0.343 (0.180–0.653)	0.001	1.065	0.345 (0.172–0.690)	0.003
Radiological ETE diagnosis	1.786	5.964 (1.827–19.46)	0.003	1.645	5.183 (1.470–18.28)	0.011
Radiomic signature	NA	NA	NA	4.130	62.167 (4.760–811.9)	0.002

β, beta coefficient; CI, confidence interval; ETE, extrathyroidal extension; NA, not applicable; OR, odds ratio.

regression analysis was then used to further analyze the influence of these parameters. Because the parameter tumor size was not statistically significant (P = 0.081), this variable was excluded from the model. Thus, tumor position and radiological ETE diagnosis constituted the clinical model for ETE diagnosis in PTC patients (**Table 2**).

Next, a radiomic nomogram with the parameters tumor location, radiological ETE diagnosis, and radiomic signature was developed. Multivariate logistic regression analysis was used to assess this radiomic nomogram, and all three predictors were statistically significant (**Table 2**, **Figure 4A**). **Table 3** shows the results of the clinical model and the radiomic nomogram model in distinguishing ETE in PTC patients.

In the training cohort, the radiomic nomogram model showed the best discrimination (**Table 3**) with an area under the curve (AUC) value of 0.837 (95% confidence interval [CI]: 0.756–0.919). This AUC value of the radiomic nomogram was higher than that of the clinical model alone (AUC: 0.768, 95% CI: 0.676–0.860; DeLong test, P = 0.031) and the radiomic signature alone (AUC: 0.736, 95% CI: 0.633–0.838; DeLong test, P = 0.024). The radiomic nomogram model also presented the best discrimination (**Table 3**) in the validation cohort with an AUC of 0.824 (95% CI: 0.723–0.925), which was higher than that of the clinical model alone (AUC: 0.741, 95% CI: 0.619–0.863; DeLong test, P = 0.012). The calibration curve and the Hosmer-Lemeshow test showed a good calibration in the training set (**Figure 4B**, P = 0.077) and the validation set (**Figure 4C**, P = 0.531). Thus, the nomogram model performed well in both training and validation sets.

The decision curve analysis demonstrated that the radiomic nomogram provided a high overall net benefit and was more beneficial than either the treat-all or the treat-none strategy (**Figure 5**).

DISCUSSION

PTC patients with ETE have higher recurrence and mortality rates than those without ETE (1–4). PTC patients with ETE require total/subtotal thyroidectomy; postoperatively, these patients will inevitably be affected by chronic hypothyroidism. Moreover, this procedure is more demanding for the surgeon requiring higher levels of skills, may impair the postoperative parathyroid function, and increases the probability of laryngeal recurrent nerve injury (13–15). Therefore, the accurate diagnosis of ETE before the operation can help the surgeon determine the most suitable surgical plan and reduce the risk of reoperation.

Previous studies showed (6–8) that of the 176 included PTC patients with pathology-confirmed ETE, only 84 patients were accurately diagnosed by ultrasound. Therefore, it is of great importance to improve the accuracy of ultrasound-based ETE

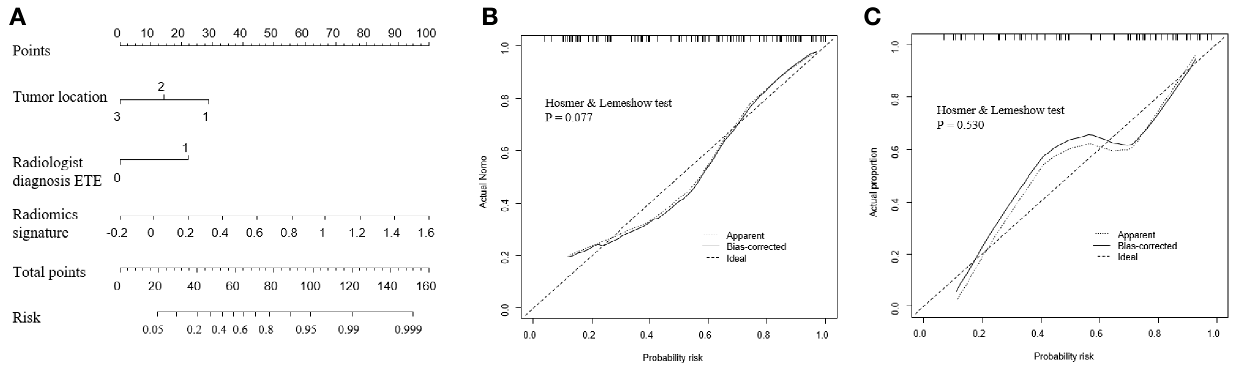


FIGURE 4 | Performance of the radiomic nomogram. **(A)** A nomogram based on the radiomic signature and clinical factors. **(B, C)** Calibration curve of the radiomic nomogram for the training cohort **(B)** and the validation cohort **(C)**. ETE, extrathyroidal extension.

TABLE 3 | Efficacies of the models predicting ETE in patients with PTC.

	Training cohort			Validation cohort		
	AUC (95% CI)	**Sensitivity (95% CI)**	**Specificity (95% CI)**	**AUC (95% CI)**	**Sensitivity (95% CI)**	**Specificity (95% CI)**
Radiomic signature	0.736 (0.633–0.838)	0.679 (0553–0.803)	0.732 (0.585–0.854)	0.760 (0.640–0.879)	0.703 (0.568–0.838)	0.741 (0.556–0.889)
Clinical model	0.768 (0.676–0.860)	0.625 (0.500–0.750)	0.829 (0.707–0.927)	0.741 (0.619–0.863)	0.514 (0.351–0.676)	0.889 (0.778–1.000)
Radiomic nomogram model	0.837 (0.756–0.919)	0.696 (0.571–0.821)	0.927 (0.853–1.000)	0.824 (0.723–0.925)	0.649 (0.486–0.784)	0.741 (0.593–0.889)

AUC, area under the curve; CI, confidence interval; ETE, extrathyroidal extension; PTC, papillary thyroid carcinoma.

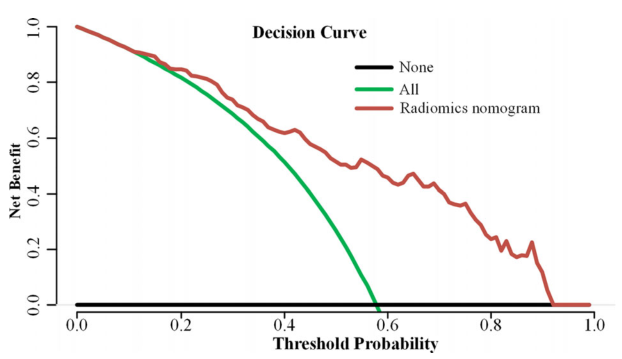


FIGURE 5 | Decision curve analysis of the radiomic nomogram model.

diagnosis in PTC. Some studies have suggested that radiomic is of great value in the diagnosis and prognosis assessment of many diseases (16, 19–23). In our study, we used the radiomic nomogram to identify ETE preoperatively. The constructed radiomic nomogram provides an easy-to-use diagnostic and predictive tool, which can prevent unnecessary surgery for patients without ETE. The higher AUC values of the radiomic nomogram indicate that the nomogram including the radiomic signature performs better than the clinical model in diagnosing ETE. The radiomic nomogram model was established using three variables, including the radiomic signature that incorporated six parameters extracted from a large number of image features by data dimension reduction. This model showed a high predictive value for the identification of ETE.

Ultrasound is the preferred imaging modality for the evaluation of PTC (24–27). It can show the degree of PTC contact with the adjacent thyroid capsule, but the diagnostic accuracy is low. Gweon et al. (7) showed that in 79 patients with PTC diagnosed by preoperative ultrasound, the accuracy rates were 60.8% and 66.2% for 2D and 3D ultrasound, respectively. Lee et al. (14) reported that if >50% of the PTC circumference was in contact with the adjacent thyroid capsule, ultrasound had a better AUC than CT (0.674 vs. 0.638, respectively) in the diagnosis of ETE, whereas combined ultrasound and CT had the highest accuracy (sensitivity 92.9%, specificity 70.4%, AUC 0.744). In the present study, the AUC of the radiomic nomogram model for the diagnosis of ETE was significantly higher than that reported by Lee et al. (0.824 vs. 0.744, respectively). This indicates that the newly developed radiomic nomogram model contained more information that is significantly related to ETE but not considered a traditional risk factor. For instance, PTC density and enhanced nonuniformity are characteristics that are difficult to quantify with the human eye. But these features are associated with tissue heterogeneity in PTC. The radiomic nomogram model takes the imaging biomarker PTC heterogeneity, a quantifiable feature related to the degree of malignancy in PTC, into account for the ETE evaluation. Thus, the quantitative nomogram model does not only overcome the subjectivity of the traditional ultrasound imaging diagnosis but also utilizes a lot of information that the naked eye cannot identify, thereby improving the accuracy of the ETE diagnosis.

Our study has several limitations. First, this is a retrospective study, causing a case selection bias that may have affected the study results. Second, our radiomic nomogram model for distinguishing ETE was established and validated in a single hospital. Third, in some PTC cases with unclear boundaries, the tumor was difficult to delineate. These cases were excluded from this study. Most of these cases were PTC with ETE, leading to a certain sample bias in this study. Fourth, the sample size included in this study is not large enough; further multicenter studies with larger sample sizes should be carried out. Forth, since the data in this study are from a single center and the same type of machine, the model will have some robust problems Finally, our radiomic nomogram model only used greyscale ultrasound images, and we will add radiomic characteristics of multimodal ultrasound to the nomogram in the future. In further studies, we will also use elastography and contrast-enhanced ultrasound images, which may contain more radiomic features than conventional 2D images.

In summary, a radiomic nomogram based on clinical risk factors and a radiomic signature was constructed for the prediction of ETE. This nomogram is expected to inform treatment strategies and assist clinical decision-making for a personalized ETE treatment of patients with PTC.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding authors.

ETHICS STATEMENT

Written informed consent was obtained from the individuals for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

XW, YR, and EA contributed equally to this study. XW, EA, HS, and GZ contributed to the conception and design of the study. JZ, FX, and XB organized the database. QZ, WL, and YR performed the statistical analysis. XW wrote the first draft of the manuscript. SH and XQ wrote sections of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was financially supported by National Natural Science Foundation of China (Project No.: 81771848,81971629) and Zhenjiang Commission of Science and Technology (Project No. SH2020046).

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin* (2019) 69(1):7–34. doi: 10.3322/caac.21551
2. Tran B, Roshan D, Abraham E, Wang L, Garibotto N, Wykes J, et al. An Analysis of The American Joint Committee on Cancer 8th Edition T Staging System for Papillary Thyroid Carcinoma. *J Clin Endocrinol Metab* (2018) 103 (6):2199–206. doi: 10.1210/jc.2017-02551
3. Hay ID, Johnson TR, Thompson GB, Sebo TJ, Reinalda MS. Minimal extrathyroid extension in papillary thyroid carcinoma does not result in increased rates of either cause-specific mortality or postoperative tumor recurrence. *Surgery* (2016) 159(1):11–21. doi: 10.1016/j.surg.2015.05.046
4. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, et al. Cancer statistics in China, 2015. *CA Cancer J Clin* (2016) 66(2):115–32. doi: 10.3322/caac.21338
5. Cooper DS, Doherty GM, Haugen BR, Kloos RT, Lee SL, Mandel SJ, et al. Revised American Thyroid Association Management Guidelines for Patients with Thyroid Nodules and Differentiated Thyroid Cancer. *Thyroid (New York NY)* (2009) 19(11):1167–214. doi: 10.1089/thy.2009.0110
6. Lee CY, Kim SJ, Ko KR, Chung K, Lee J. Predictive Factors for Extrathyroidal Extension of Papillary Thyroid Carcinoma Based on Preoperative Sonography. *J Ultrasound Med* (2014) 33(2):231–8. doi: 10.7863/ultra.33.2.231
7. Gweon HM, Son EJ, Youk JH, Kim J, Park CS. Preoperative Assessment of Extrathyroidal Extension of Papillary Thyroid Carcinoma. *J Ultrasound Med* (2014) 33(5):819–25. doi: 10.7863/ultra.33.5.819
8. Seo YL, Yoon DY, Lim KJ, Cha JH, Yun EJ, Choi CS, et al. Locally advanced thyroid cancer: can CT help in prediction of extrathyroidal invasion to adjacent structures? *Am J Roentgenol (1976)* (2010) 195(3):W240. doi: 10.2214/AJR.09.3965
9. Kim H, Kim J, Son EJ, Youk JH, Chung T, Park CS, et al. Preoperative prediction of the extrathyroidal extension of papillary thyroid carcinoma with ultrasonography versus MRI: A retrospective cohort study. *Int J Surg* (2014) 12(5):544–8. doi: 10.1016/j.ijsu.2014.03.003
10. Limkin EJ, Sun R, Dercle L, Zacharaki EI, Robert C, Reuzé S, et al. Promises and challenges for the implementation of computational medical imaging (radiomics) in oncology. *Ann Oncol* (2017) 28(6):1191–206. doi: 10.1093/annonc/mdx034
11. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* (2017) 14(12):749–62. doi: 10.1038/nrclinonc.2017.141
12. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RGPM, Granton P, et al. Radiomics: Extracting more information from medical images using advanced feature analysis. *Eur J Cancer* (2012) 48(4):441–6. doi: 10.1016/j.ejca.2011.11.036
13. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology* (2016) 278(2):563–77. doi: 10.1148/radiol.2015151169
14. Lee DY, Kwon T, Sung M, Kim KH, Hah JH. Prediction of Extrathyroidal Extension Using Ultrasonography and Computed Tomography. *Int J Endocrinol* (2014) 2014:1–6. doi: 10.1155/2014/351058
15. Choi JS, Kim J, Kwak JY, Kim MJ, Chang HS, Kim E. Preoperative staging of papillary thyroid carcinoma: comparison of ultrasound imaging and CT. *Am J Roentgenol (1976)* (2009) 193(3):871. doi: 10.2214/AJR.09.2386
16. Jiang M, Li C, Tang S, Lv W, Yi A, Wang B, et al. Nomogram Based on Shear-Wave Elastography Radiomics Can Improve Preoperative Cervical Lymph Node Staging for Papillary Thyroid Carcinoma. *Thyroid* (2020) 30(6):885–97. doi: 10.1089/thy.2019.0780
17. Tian X, Song Q, Xie F, Ren L, Zhang Y, Tang J, et al. Papillary thyroid carcinoma: an ultrasound-based nomogram improves the prediction of lymph node metastases in the central compartment. *Eur Radiol* (2020) 30(11):5881–93. doi: 10.1007/s00330-020-06906-6
18. Fitzgerald M, Saville BR, Lewis RJ. Decision curve analysis. *JAMA* (2015) 313:409–10. doi: 10.1001/jama.2015.37
19. Dong D, Tang L, Li ZY, Fang MJ, Gao JB, Shan XH, et al. Development and validation of an individualized nomogram to identify occult peritoneal metastasis in patients with advanced gastric cancer. *Ann Oncol* (2019) 30 (3):431–8. doi: 10.1093/annonc/mdz001
20. Wang W, Wu S, Zhang J, Xian M, Huang H, Li W, et al. Preoperative Pathological Grading of Hepatocellular Carcinoma Using Ultrasomics of Contrast-Enhanced Ultrasound. *Acad Radiol* (2020) S1076-332(20)30340–8. doi: 10.1016/j.acra.2020.05.033
21. Yu F, Wang J, Ye X, Deng J, Hang J, Yang B. Ultrasound-based radiomics nomogram: A potential biomarker to predict axillary lymph node metastasis in early-stage invasive breast cancer. *Eur J Radiol* (2019) 119:108658. doi: 10.1016/j.ejrad.2019.108658
22. Hu H, Wang Z, Huang X, Chen S, Zheng X, Ruan S, et al. Ultrasound-based radiomics score: a potential biomarker for the prediction of microvascular invasion in hepatocellular carcinoma. *Eur Radiol* (2019) 29(6):2890–901. doi: 10.1007/s00330-018-5797-0
23. Li W, Huang Y, Zhuang B, Liu G, Hu H, Li X, et al. Multiparametric ultrasomics of significant liver fibrosis: A machine learning-based analysis. *Eur Radiol* (2019) 29(3):1496–506. doi: 10.1007/s00330-018-5680-z
24. Domínguez JM, Nilo F, Martínez MT, Massardo JM, Muñoz S, Contreras T, et al. Papillary thyroid microcarcinoma: characteristics at presentation, and evaluation of clinical and histological features associated with a worse prognosis in a Latin American cohort. *Arch Endocrinol Metab* (2018) 62 (1):6–13. doi: 10.20945/2359-3997000000013
25. Woo CG, Sung CO, Choi YM, Kim WG, Kim TY, Shong YK, et al. Clinicopathological Significance of Minimal Extrathyroid Extension in Solitary Papillary Thyroid Carcinomas. *Ann Surg Oncol* (2015) 22(S3):728–33. doi: 10.1245/s10434-015-4659-0
26. Diker-Cohen T, Hirsch D, Shimon I, Bachar G, Akirov A, Duskin-Bitan H, et al. Impact of Minimal Extrathyroid Extension in Differentiated Thyroid Cancer: Systematic Review and Meta-Analysis. *J Clin Endocrinol Metab* (2018) 103(6):2100–6. doi: 10.1210/jc.2018-00081
27. Choi JS, Kim E, Moon HJ, Kwak JY. Higher body mass index may be a predictor of extrathyroidal extension in patients with papillary thyroid microcarcinoma. *Endocrine* (2015) 48(1):264–71. doi: 10.1007/s12020-014-0293-z

Conflict of Interest: WL was employed by Julei Technology Company.

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Oncology

ORIGINAL RESEARCH
published: 16 March 2021
doi: 10.3389/fonc.2021.614172

Using Deep Neural Network to Diagnose Thyroid Nodules on Ultrasound in Patients With Hashimoto's Thyroiditis

Yiqing Hou[1†], *Chao Chen*[2†], *Lu Zhang*[1], *Wei Zhou*[1], *Qinyang Lu*[2], *Xiaohong Jia*[1], *Jingwen Zhang*[1], *Cen Guo*[2], *Yuxiang Qin*[2], *Lifeng Zhu*[3], *Ming Zuo*[3], *Jing Xiao*[2], *Lingyun Huang*[2*] and *Weiwei Zhan*[1*]

[1] *Department of Ultrasound Diagnosis, Ruijin Hospital Affiliated to Shanghai Jiaotong University, Shanghai, China,* [2] *Ping An Technology (Shenzhen) Co., Ltd., Shenzhen, China,* [3] *Computer Centre, Ruijin Hospital Affiliated to Shanghai Jiaotong University, Shanghai, China*

Edited by:
Katsutoshi Sugimoto,
Tokyo Medical University, Japan

Reviewed by:
Hersh Sagreiya,
University of Pennsylvania,
United States
Yu Yoshimasu,
Tokyo Medical University, Japan

****Correspondence:***
Weiwei Zhan
shanghairuijin@126.com
Lingyun Huang
huanglingyun691@pingan.com.cn

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions,
a section of the journal
Frontiers in Oncology

Received: *05 October 2020*
Accepted: *28 January 2021*
Published: *16 March 2021*

Citation:
Hou Y, Chen C, Zhang L, Zhou W, Lu Q, Jia X, Zhang J, Guo C, Qin Y, Zhu L, Zuo M, Xiao J, Huang L and Zhan W (2021) Using Deep Neural Network to Diagnose Thyroid Nodules on Ultrasound in Patients With Hashimoto's Thyroiditis.
Front. Oncol. 11:614172.
doi: 10.3389/fonc.2021.614172

Objective: The aim of this study is to develop a model using Deep Neural Network (DNN) to diagnose thyroid nodules in patients with Hashimoto's Thyroiditis.

Methods: In this retrospective study, we included 2,932 patients with thyroid nodules who underwent thyroid ultrasonogram in our hospital from January 2017 to August 2019. 80% of them were included as training set and 20% as test set. Nodules suspected for malignancy underwent FNA or surgery for pathological results. Two DNN models were trained to diagnose thyroid nodules, and we chose the one with better performance. The features of nodules as well as parenchyma around nodules will be learned by the model to achieve better performance under diffused parenchyma. 10-fold cross-validation and an independent test set were used to evaluate the performance of the algorithm. The performance of the model was compared with that of the three groups of radiologists with clinical experience of <5 years, 5–10 years, >10 years respectively.

Results: In total, 9,127 images were collected from 2,932 patients with 7,301 images for the training set and 1,806 for the test set. 56% of the patients enrolled had Hashimoto's Thyroiditis. The model achieved an AUC of 0.924 for distinguishing malignant and benign nodules in the test set. It showed similar performance under diffused thyroid parenchyma and normal parenchyma with sensitivity of 0.881 *versus* 0.871 ($p = 0.938$) and specificity of 0.846 *versus* 0.822 ($p = 0.178$). In patients with HT, the model achieved an AUC of 0.924 to differentiate malignant and benign nodules which was significantly higher than that of the three groups of radiologists (AUC = 0.824, 0.857, 0.863 respectively, $p < 0.05$).

Conclusion: The model showed high performance in diagnosing thyroid nodules under both normal and diffused parenchyma. In patients with Hashimoto's Thyroiditis, the model showed a better performance compared to radiologists with various years of experience.

Keywords: thyroid nodule, ultrasound, deep learning, Hashimoto's thyroiditis, diagnosis

INTRODUCTION

Thyroid cancer has gained much attention because of its rapidly increasing incidence since the last decades though the increase in incidence is partially due to the improvements in diagnosis. It has become the 11th most common cancer in the world and the 5th most common cancer among female worldwide (1–3). Among all thyroid cancer, papillary thyroid cancer (PTC) is the most common histologic type, accounting for 80–90% of all thyroid cancer (4, 5). Hashimoto's thyroiditis (HT) is the most common auto-immune thyroiditis. The worldwide incidence was reported to range from eight to 46 cases per 1,000 each year depending on different inclusion criteria in various studies. It was at least eight times more prevalent in female, and its incidence is still increasing over time due to social and physical risk factors such as pressure, hormone disorder, and smoking (6, 7). It is considered a risk factor of PTC with an incidence of 0.5–30% in HT patients which is higher than the reported 14.2 per 100,000 person in the general population (4, 8).

Ultrasonography is the most common tool to diagnose thyroid disease, but the accuracy of the diagnosis usually depends on the experience of radiologists. Despite a higher incidence of PTC in patients with HT, it's more difficult to distinguish between benign and malignant nodules in these patients because they often present a coarse and heterogeneous thyroid parenchyma caused by the repetitive damage of chronic inflammation (9). It was reported that the underlying heterogeneous echogenicity can affect the ultrasound characteristics of thyroid nodule, especially the margin. Microlobulated or irregular margins were more frequently observed among benign nodules under heterogeneous thyroid parenchyma. Since these two features were considered as typical malignant features (10), benign nodules under heterogeneous parenchyma would more likely to be misdiagnosed as malignant nodules, thereby reducing the diagnostic performance of doctors, especially those with less experience. Park et al. (11) reported that in patients with heterogeneous thyroid parenchyma, the accuracy, specificity, and positive predictive rate for diagnosing malignancy were 77.6, 76.3, and 48.7% which were significantly lower than 84.4, 83.7, and 60.9% for patients with homogeneous parenchyma. That means more benign nodules will be misdiagnosed as thyroid cancer. Thus, overdiagnosis and overtreatment are more likely to occur in this part of the population because differential diagnosis between malignant and benign nodules is more challenging in patients with HT.

Computer aided diagnosis (CAD) system has made remarkable progress during these years. From the classic machine learning method (12) to the now prevailing deep learning model, the performance of the CAD system has greatly improved over time. In the traditional machine learning method (13), the explicit features such as size, shape, margin, echogenicity, microcalcification, and macrocalcification were extracted by algorithms or labeled by radiologists, and then sent into the classifiers for training. This kind of expert-knowledge-based system failed to meet the increasing demand for precision, generalization, and efficiency. Recently, deep neural network showed its competency in various tasks for medical image analysis, such as lesion detection and lesion pattern recognition (14, 15). DNN can extract more complex and implicit features and train classifiers synchronously in one unified framework. It can achieve better accuracy and ability of generalization not only because of its huge model capacity but also its deeper and more complex structure. In a recent study with a large training set containing 312,399 images (16), the DNN-based CAD system outperformed most of the radiologists. For these reasons, CAD was considered as a possible solution to reduce overdiagnosis of thyroid cancer. It can overcome the heterogeneity of human radiologists and has shown similar diagnostic performance to human radiologists in many studies (17).

However, no previous studies have been performed to develop a computer aided diagnosis (CAD) system in identifying PTC in HT patients which is believed to be a more challenging task. In this study, we aim to establish a CAD system using deep learning model and test its ability to differentiate malignant and benign thyroid nodules underlying diffused background of HT. Considering the complex heterogeneous echogenicity of thyroid parenchyma in HT patients, we trained and compared two DNN models, one focused only on the interior region of the nodule while another focused not only on the nodule area but also the parenchyma around the nodule. These two models are both pretrained with ImageNet Database.

MATERIALS AND METHODS

Study Design and Inclusion Criteria

This study was a retrospective study approved by the Institutional Review Board, with waiver of informed consent. We retrospectively included 2,932 patients who underwent thyroid ultrasonography from January 2017 to August 2019. 1,666 patients had HT and 1,266 patients had normal thyroid parenchyma. Among all patients, 80% were included as the training set and the rest 20% as the test set so that images in the training set do not appear in the test set.

All selected patients meet the following criteria for image quality control: (1) each nodule should have at least one image from at least two orthogonal planes, (2) the position and size match the ultrasound report and pathological report if pathological result is needed.

The requirement for pathological results depended on the grading of nodules. All nodules were graded using K-TIRADS in this study. Nodules with TIRADS 4A or above need to have definitive pathological results to be included in this study, while nodules graded TIRADS 2 or 3 were recognized as benign nodules and did not necessarily need pathological results.

The inclusion criteria for benign nodules are: (1) nodules graded TIRADS 2 or 3 with or without negative pathological results, (2) nodules graded TIRADS 4A or above with a negative cytological pathology result and Braf mutation verified by repeated FNA, (3) nodules graded TIRADS 4A or above with histological pathology proved to be benign. The inclusion criteria for malignant nodules are malignancy proved by cytological or histological pathology.

The inclusion criteria for HT were as follows: (1) thyroid parenchyma showed heterogeneous echogenicity under ultrasound; (2) serum TPOAb >5.61 IU/ml and/or TGAb >4.11 IU/ml; (3) TRAb within normal range(0–1.75 IU/L).

Image Acquisition and Evaluation

Ultrasound images were collected by radiologists with at least 3 years' clinical experience to ensure the quality of images. The ultrasound examinations were performed using MyLab 90, Esaote; iU22, Philips; Resona 7, Mindray; RS 80A Samsung; and Logic E9, GE Healthcare equipped with 7–12 MHz linear-array transducer. The original settings of thyroid mode were used to perform the examination. The region of interest (ROI) of the lesions was annotated using four crossed calipers.

All images included were graded according to K-TIRADS (18). Images in the test set were evaluated by three groups of doctors with clinical experience <5 years, 5–10 years and >10 years respectively. Each group consists of two doctors, and they were asked to give a consensus for whether a nodule was benign or malignant.

Development of Deep Learning Model

Our proposed model is illustrated in **Figure 1**. We chose DenseNet-161 pretrained with the ImageNet (19, 20) as our model backbone. DenseNet architecture explicitly differentiates between information that is added to the network and information that is preserved. Dense connections with feature maps being concatenated together are used, which are effective for feature exploration, thus DenseNets have made nearly the best performance on the general image classification tasks while substantially reducing the number of model parameters. We used one DenseNet structure with four dense blocks, which extracted features and gradually down-sampled the feature maps, and then input to the full connection layer. Finally, the model outputs the benign probability and the malignant probability of the input image. Then the pathology prediction result, benign or malignant, would be computed according to the probabilities and the threshold value,

We trained a baseline DNN model using only the region of nodule and a modified DNN model using features of both nodule and parenchyma. We expand the annotated nodule ROI according to the rules defined in **Table 1**, letting the model capture more features around the nodule edge and context information. Each ROI was padded with 0 if it reached the image boundary while expanding, and was rescaled without altering the original aspect ratio. To fit the input size of the pre-trained DenseNet-161 model, all training and testing images were resized to 224 × 224.

To avoid overfitting, data augmentation is also implemented. We adopted random horizontal flipping, random cropping and rotation within a small range for augmentation because excessive randomization cannot mimic the speckle noise in the ultrasound image. In specific, the range of random translation is not larger than 10% of the longer side length of ROI; the range of random rotation is not bigger than 12.5 angle degrees.

The diffused change information and pathology information on training data were given by the radiologists. Guided by the cross-entropy loss, we can learn the neural network end to end using deep learning framework.

In the training set, we used 10-fold cross validation for the identification of the optimal model, which was then used for the test set classification. We acquired the average classification performance for the test set, plotted in the receiver operating

TABLE 1 | Rules of expanding nodule ROI.

Longer side length of nodule ROI	Expanded square ROI size
$0 < len < 65$	$len + 256$
$65 < len < 150$	$len + 128$
$150 < len < 256$	$len + 64$
$len >= 256$	$len + 32$

Different ROI expansions adopted for nodule images of different sizes in pixels.

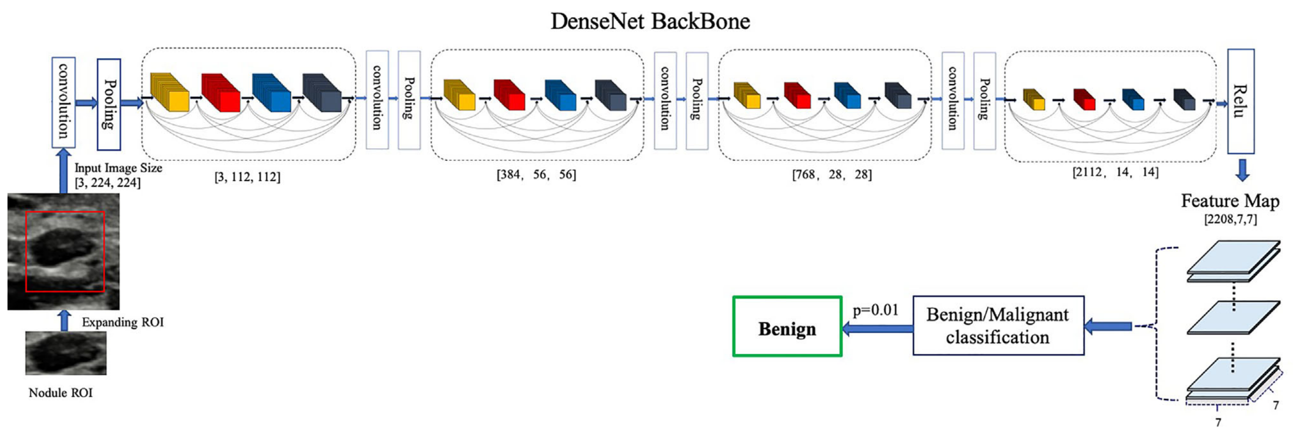


FIGURE 1 | Architecture of our proposed model. DenseNet-161(k = 48) used as the backbone. Different ROI expansions adopted for annotated nodule images of different sizes.

characteristic (ROC) curve. As usually recommended, the optimal threshold value was set at the highest Youden Index, or equivalently, the highest Sensitivity + Specificity (21). Accuracy, sensitivity, specificity, precision, and area under curve (AUC) of ROC curve were extracted from the 10 folds and presented as means ± SD.

Our proposed model was implemented using Python and DL toolkit Pytorch (22). We trained the network with stochastic gradient descent using Adam optimizer with a weight decay rate of 0•0005. All experiments were conducted on two workstations equipped with a 16-core 2.10 GHz Intel Core Processor (Skylake) and two NVIDIA Tesla V100 GPUs.

Statistical Analysis

General information such as the distribution of sex, age, and percentage of malignancy between training set and test set was calculated and compared between HT and normal groups. The group difference for age was calculated using t test. The group differences for qualitative data such as sex ratio and percentage of malignancy were analyzed using chi-square test.

Accuracy, sensitivity, specificity, precision, and AUC were exploited to evaluate the performance of our model *versus* radiologists. Statistical differences of AUCs between various diagnostic methods were compared using Delong test (23). Mann–Whitney U test was used for the comparison of the model's specificity, sensitivity, accuracy, and precision between HT subset and normal subset. Chi-square test was used for the comparison between model and radiologists in terms of specificity, sensitivity, accuracy, and precision.

Python was used to perform the Delong test and plot the ROC curve. The rest statistical analysis was performed by SPSS 24.0. p <0.05 was considered statistically significant.

RESULTS

Study Population

In total, 2,932 patients with 3,634 nodules and 9,106 images were included in this study. The images were split into training set and test set. All sets were partitioned strictly according to the criteria: images that belonged to the same patient were assigned to the same set. Test set contained 568 patients (710 nodules, 1,805 images) with 332 HT patients (58%). The training set had a total of 2,364 patients (2,924 nodules, 7,301 images), with 1,334 patients (56%) having HT. The baseline characteristics of the training set and test set were listed in **Table 2**.

Threshold Value and Comparison of Two Deep Neural Network Models

The ROC curve was plotted in **Figure 2** and the corresponding AUC demonstrated the diagnostic performance of our baseline and modified DNN model across all threshold values. The sensitivity, specificity, and Youden Index curve for modified DNN model were mapped in **Figure 3** to show the optimal threshold value. The maximum of Youden Index was 0.729, the corresponding threshold was 0.358. The performance metrics at

TABLE 2 | Baseline characteristics.

	Training Set	Test Set
Number of patients, n (%)	2,364	568
Patients with HT	1,334 (56.4%)	332 (58.5%)
Patients without HT	1,030 (43.6%)	236 (41.5%)
Number of images, n (%)	7,301	1,805
Images from patietns with HT	4,128 (56.5%)	1,086 (60.2%)
Images from patietns without HT	3,173 (43.5%)	722 (39.8%)
Number of nodules, n (%)	2,924	710
Benign nodules	1,920 (65.7%)	476 (67%)
malignant nodules	1,004 (34.3%)	234 (33%)
Nodule sizes (cm)		
Benign nodules	1.09 (0.86)	1.08 (0.89)
malignant nodules	1.08 (0.63)	1.06 (0.61)
Patient gender, n (%)		
Male	539 (22.8%)	136 (23.9%)
Female	1825 (77.2%)	432 (76.1%)
Mean age (years)	45.29 ± 12.45	45.09 ± 12.41

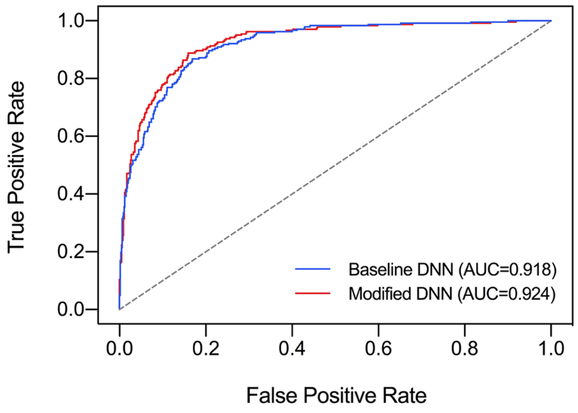


FIGURE 2 | Comparison of ROC curves and AUC of two DNN models. Baseline DNN model learned only the nodule area. Modified DNN model learned the nodule area as well as the surrounding parenchyma.

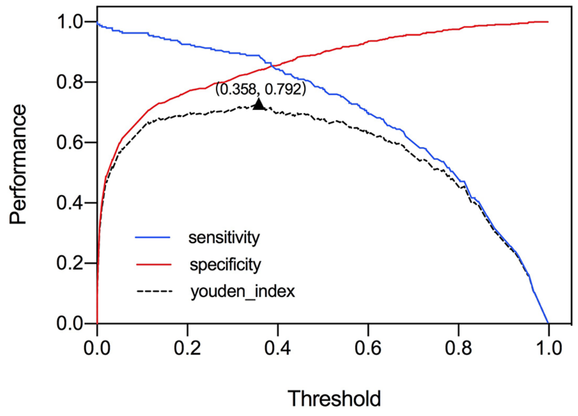


FIGURE 3 | Youden Index and threshold for modified DNN model.

the optimal threshold were compared between the two models. The AUC, sensitivity, and specificity for the baseline DNN model was 0.918, 0.874, 0.820 compared to 0.924, 0.881, 0.839 for the modified DNN model. The modified model showed a slightly better performance, and therefore we chose the modified DNN model as our CAD model in the following experiments.

Performance of Deep Neural Network Model on Test Sets

The performance metrics of our DNN model in distinguishing malignant and benign nodules on test set and the two subsets were listed in **Table 3**. It achieved similar AUC under the test set, HT subset and normal subset (AUC = 0.924, 0.924, 0.906 respectively).

When comparing the performance between HT subset and normal subset, the model showed similar accuracy, sensitivity, and specificity (p all >0.05). Only precision showed a significant difference (0.540 *vs* 0.784, p < 0.01). When stratified by nodule sizes, listed in **Table 4**, precision showed a notable decrease in the HT subset compared to normal subset among all nodule sizes, and it is more pronounced in nodules <5 mm.

The influence of nodule size on DNN model was demonstrated in **Table 4** and **Figure 4**. It was evaluated in normal subset and HT subset respectively. In both subsets, AUC values among nodules <5, 5–10, and 10–20 mm were similar while that for nodules >20 mm was slightly lower. However the ROC curves for nodules with different sizes were quite close as illustrated in **Figure 4**. For both HT subset and normal subset, the accuracy and specificity were similar among different nodule sizes while sensitivity for nodules >20 mm and precision for nodules <5 mm were greatly reduced. What's more, the precision for nodules >20 mm was also greatly reduced.

Performance of Deep Neural Network Model Compared to Radiologists Under Diffused Background

For HT subset, the DNN model achieved a higher AUC than that of the three groups of radiologists which showed significant difference as listed in **Table 5**. However, in the ROC curve (**Figure 5**), the operation points of the radiologists are close to the DNN model's ROC curve. For the other performance metrics, no significant difference was found for accuracy and precision between DNN model and radiologists. The model showed a higher sensitivity and a lower specificity but significant difference only exist between the DNN model and radiologists with <5 years of experience.

TABLE 4 | Performance metrics of DNN model in diagnosing malignant nodules of different sizes, evaluated on normal subset versus HT subset.

		HT Subset	Normal subset
Average size (SD)		0.975 (0.51)	1.25 (0.77)
<5 mm	AUC	0.915	0.895
	Accuracy	0.83	0.825
	Sensitivity	0.859	0.82
	Specificity	0.828	0.826
	Precision	0.327	0.651
5–10 mm	AUC	0.909	0.895
	Accuracy	0.82	0.846
	Sensitivity	0.902	0.868
	Specificity	0.794	0.822
	Precision	0.577	0.841
10–20 mm	AUC	0.883	0.907
	Accuracy	0.832	0.837
	Sensitivity	0.854	0.878
	Specificity	0.824	0.792
	Precision	0.652	0.827
>20 mm	AUC	0.871	0.845
	Accuracy	0.836	0.801
	Sensitivity	0.722	0.724
	Specificity	0.864	0.837
	Precision	0.594	0.688

AUC, Areas under the ROC curve. All metrics were the average of 10-folds.

Besides, in the test set and normal subset, the model also showed higher AUC but close operation point on the ROC curve compared to radiologists, as shown in **Table 5** and **Figure 5**. The difference of other metrics between model and radiologists was similar to that under HT subset.

DISCUSSION

Many studies have achieved satisfied results in nodule diagnosis by using DNN. Buda et al. reported the sensitivity and specificity of a DNN model trained under 1,230 cases (1377 nodules) to be 87 and 52% respectively, which were higher than those of the radiologists with experience ranging from 3 to 32 years and were similar to the consensus of three ACR experts (24). Li et al. included a training set with a total of 42,952 cases which contained the largest sample size so far. The trained DNN model was tested on one internal test set and two external test sets. The AUC of model under three test sets were 0.947, 0.912, and 0.908 respectively, which were significantly higher than those of the six experienced radiologists (16). On the contrary, Gao et al. found that the DNN model performed significantly lower than the radiologists (25). However, they chose AlexNet as their backbone which was different from ours. In our research,

TABLE 3 | Performance of model in diagnosing malignant nodules on test set and its subsets.

	AUC	Accuracy	Sensitivity	Specificity	Precision
Test set	0.924 (0.006)	0.851 (0.018)	0.881 (0.027)	0.839 (0.031)	0.673 (0.038)
HT subset	0.924 (0.010)	0.852 (0.026)	0.881 (0.035)	0.846 (0.036)	0.540 (0.053)
Normal subset	0.906 (0.010)	0.843 (0.011)	0.871 (0.033)	0.822 (0.029)	0.784 (0.024)
P-Value		0.587	0.938	0.178	<0.01

P-Value is that of diagnostic performance on HT subset versus normal subset; AUC, Areas under the ROC curve. All metrics were the average of 10-fold, presented as Mean (SD).

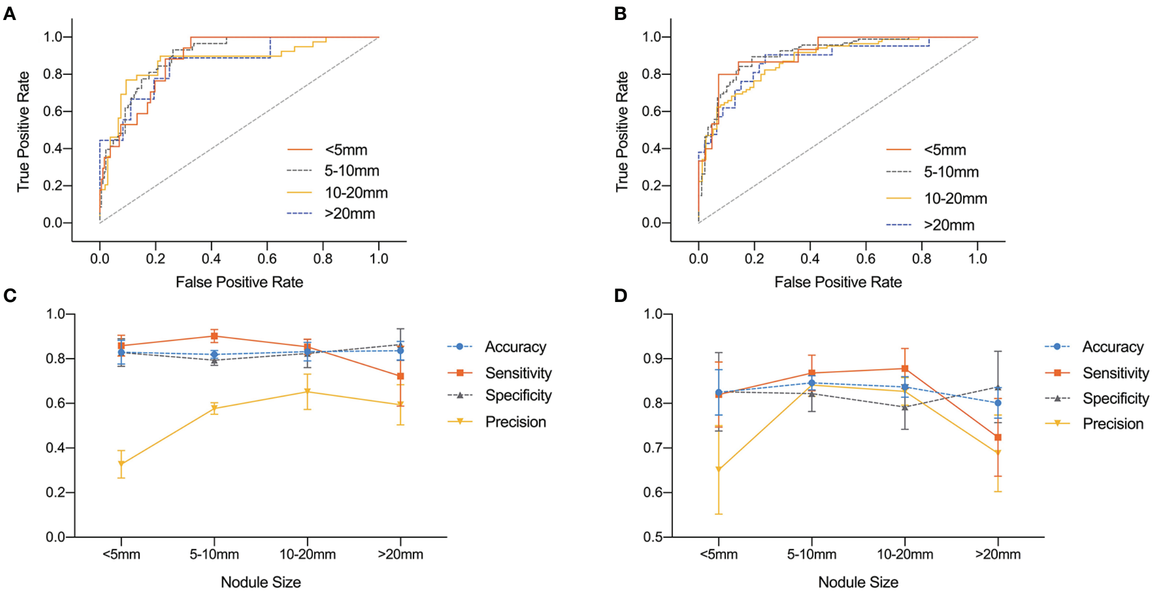


FIGURE 4 | Comparison of ROC curves and performance metrics of DNN model under different nodule sizes. **(A, C)** ROC curves and performance metrics for different nodule sizes under HT subset. **(B, D)** ROC curves and performance metrics for different nodule sizes under normal subset.

TABLE 5 | Performance of model *versus* radiologists of clinical experience <5 years, 5–10 years, and >10 years in diagnosing malignant nodules on the test set and its subsets.

	Diagnostic method	AUC	Accuracy	Sensitivity	Specificity	Precision
Test set	Model	0.924	0.851	0.881	0.839	0.673
	Radiologist <5 yr	0.818	0.868	0.707	0.928	0.784
	Radiologist 5–10 yr	0.843	0.864	0.798	0.888	0.726
	Radiologist >10 yr	0.848	0.858	0.826	0.87	0.701
	P-Value*	<0.01	0.781	<0.01	<0.01	0.016
	P-Value**	<0.01	1.000	0.001	0.04	0.346
	P-Value***	<0.01	0.733	0.777	0.3	0.752
HT subset	Model	0.924	0.852	0.881	0.846	0.540
	Radiologist <5 yr	0.824	0.897	0.723	0.924	0.588
	Radiologist 5–10 yr	0.857	0.875	0.831	0.882	0.514
	Radiologist >10 yr	0.863	0.863	0.862	0.863	0.487
	P-Value*	<0.01	0.401	0.001	0.003	0.226
	P-Value**	<0.01	0.928	0.060	0.312	0.811
	P-Value***	<0.01	0.787	0.486	1.000	0.874
Normal subset	Model	0.906	0.843	0.871	0.822	0.784
	Radiologist <5 yr	0.825	0.842	0.712	0.938	0.893
	Radiologist 5–10 yr	0.846	0.853	0.797	0.894	0.847
	Radiologist >10 yr	0.844	0.85	0.804	0.885	0.837
	P-Value*	<0.01	0.603	0	0	0.017
	P-Value**	<0.01	0.916	0.01	0.035	0.179
	P-Value***	<0.01	0.833	0.015	0.072	0.272

P-Value is that of model versus radiologist with <5 years' clinical experience; P-Value** is that of model versus radiologist with 5–10 years' clinical experience; P-Value*** is that of model versus radiologist with >10 years' clinical experience; AUC, Areas under the ROC curve.*

All metrics were the average of 10-folds.

DenseNet was chosen as the backbone for its higher performance on the general image classification tasks while substantially reducing the number of model parameters (19, 20).

It is worth mentioning that there was no research revealing the performance of DNN model under diffused thyroid background, and our research filled in this gap. We designed a modified DNN model for diffused background, which learned nodule features as well as background features using border extension. We compared the baseline model which analyzed only the nodule area with our modified DNN model to see how learning thyroid parenchyma helped nodule diagnosis. It turned out that the baseline DNN model showed a slightly lower AUC

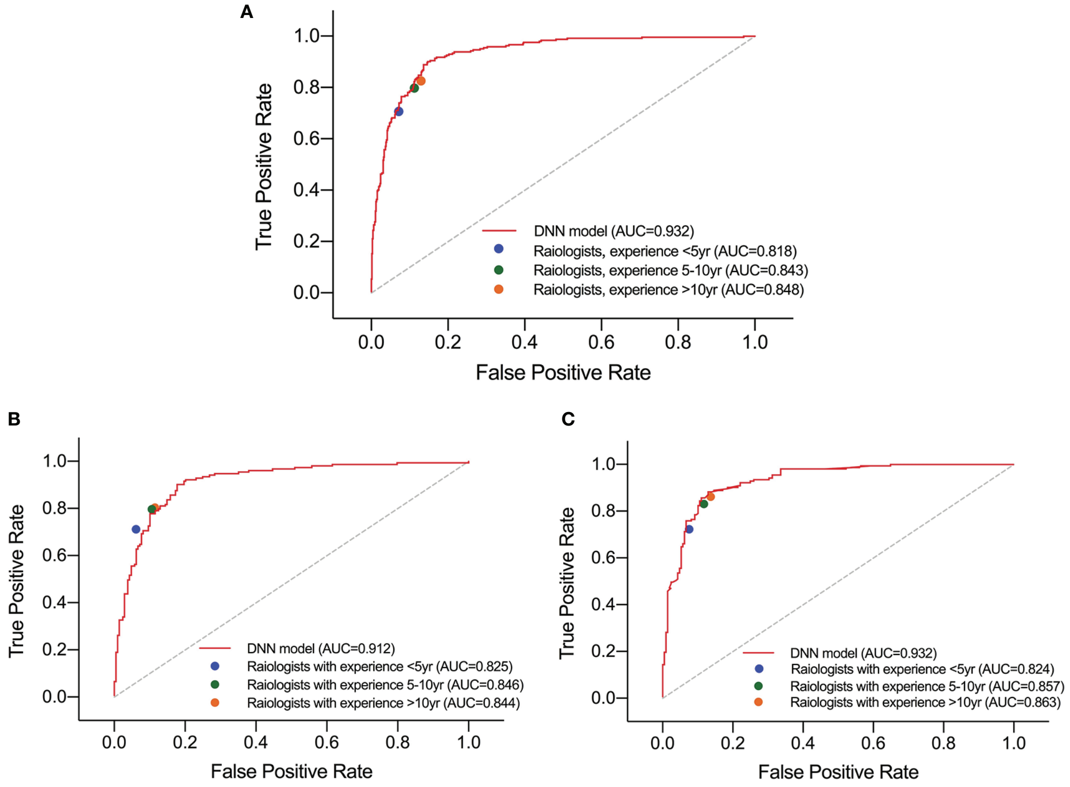


FIGURE 5 | Performance of DNN model and three groups of radiologists in diagnosing malignant nodules under test set **(A)**, normal subset **(B)**, and HT subset **(C)**.

compared to the modified DNN model. In **Figure 2**, there was an obvious separation between the two ROC curves in the upper left area of the figure which means the modified DNN model had higher sensitivity as well as specificity. This result further supported our hypothesis that learning thyroid parenchyma can help improve the diagnostic accuracy of CNN under heterogeneous background.

The border extension design of the modified DNN model was enlightened by our clinical experience that heterogeneous thyroid parenchyma may affect nodules' sonographic features. This idea was supported by a series of literatures. Park et al. found that benign nodules in this background are more likely to show vague boundaries (11) which contribute to decrease of accuracy and specificity in differentiating malignant and benign nodules in HT patients. Malignant nodules could also have a more obscure boundary and irregular margin under diffused thyroid parenchyma (26). It could be concluded from the literature that diffused parenchyma affects the nodule's feature mainly by its border. Therefore, it is reasonable to include parenchyma features around the border using boundary extension so that the influence of parenchyma on nodule border can be considered when diagnosing nodules' malignancy under diffused background. The rule of ROI expansion we proposed in **Table 1** was based on the fact that a small nodule usually contains less features inside the nodule due to a limited nodule area. Therefore, more border and background information should be taken into consideration during the diagnosis process. For large nodules, there were sufficient features within the nodule area so border information can be less emphasized. What's more, all images would undergo size normalization process after border expansion before given to the model. For a large nodule, whose image size was already larger than the required input size, the details inside the nodule area would be compressed as image being zoomed out during size normalization. To keep the original sonographic features as unchanged as possible, we should not expand ROI of large nodules too much for it would aggravate the loss of detail.

During the data collection process, our research included HT patients as well as non-HT patients. There are two reasons for this design. First, containing non-HT cases can make our model more generalized. Second, patients with normal parenchyma can serve as a comparison to the HT patients in the test set. Furthermore, the images used in this study were scanned by ultrasound machines from six different companies, which further increased the diversity of data and were closer to the clinical reality.

In patients with HT, the trained DNN model showed a significantly higher AUC value than human but as shown in the ROC curve, dots of radiologists are not too far from the DNN model's ROC curve. Also, the model showed a higher sensitivity but a lower specificity compared to human radiologists. This indicates that the DNN model outperformed humans in distinguishing malignant and benign nodules mainly due to a higher sensitivity. However in a real-world setting, the overall performance should be considered according to different clinical tasks. Therefore, the model would be more suitable than human radiologists for screening malignancy in a large population especially in HT patients. But in other clinical scenario it may not have that much remarkable advantage compared to human radiologists. Another advantage of DNN model is its diagnostic homogeneity. Ultrasound diagnosis is subjective, and it greatly depends on clinical experience. In patients with HT, the heterogeneous background could affect the margin of nodule and thus further decrease inter-observer and intra-observer agreement, especially between less experienced radiologists (11). On the contrary, DNN model can extract image features quantitatively and output a consistent conclusion through standardized processing methods. Therefore, DNN has a higher reproducibility compared to human radiologists.

However, the precision of our model in the HT subset is lower than in the normal subset which means the trained DNN model is less confident about predicting malignancy within HT subsets. One possible explanation is that the sonograms of benign nodules under HT parenchyma were more suspicious, and thus the model had a higher chance to misdiagnose benign nodules as malignant ones. This hypothesis was supported by a clinical research by Park M. et al. (11) who discovered that benign nodules under HT parenchyma showed more malignant features resulting in a lower positive predictive rate in diagnosing malignancy. Another possible reason is that the nodules in the HT subset are smaller than those in the normal subset. We analyzed the average diameter of nodules under each subset, and we found that the average size of nodules in the HT subset was smaller than that in the normal subset, although not significant. Smaller nodules tend to have less features than big nodules which can cause the model to be less confident in the HT subset. There was also literature supporting the negative effect of nodule size on the performance of model (27) which further supports this hypothesis. It was also notable that the difference of precision between two subsets is more obvious for smaller nodules. This could also be explained by the influence of nodule sizes. Since larger nodules had sufficient features for the model to make reliable predictions, they would be less affected by the heterogeneous parenchyma than the smaller nodules.

The parenchyma had little influence on the performance for our modified DNN model, while the size of the nodules had certain impact on its diagnostic ability. The precision of nodules <5 mm was significantly reduced, while the diagnostic sensitivity of nodules >20 mm was also significantly reduced. As previously reported, Wang et al. also discovered a similar trend (27). One possible reason was that the ROI of small nodules contained less features than big nodules. Therefore, the model was not as confident in the diagnosis of small nodules as in the big ones. The decrease in sensitivity for large nodules might be due to the fact that follicular carcinoma accounts for a greater proportion in nodules >20 mm than in the other three groups. The ultrasound features of follicular carcinoma were similar to benign nodules. However, follicular lesions account for a very low proportion in our training and test sets, and therefore the models didn't get enough training on identifying this kind of nodule.

Studies have shown that diagnostic accuracy and specificity of doctors in diffused background were reduced (11). In our study, the performance of radiologists didn't seem to decrease in the HT subset which is contrary to what was reported before. We speculated that one possible reason for this paradoxical situation is that the HT subset may contain slightly more TIRADS 2 and TIRADS 5 nodules due to selection bias when collecting images, which unfortunately made the HT subset relatively easier to diagnose. This was a limitation of our study and could be avoided by stratified sampling according to TIRADS grades in the future study. Another possible reason is that coexisting HT may increase the false negative rate of FNA for subcentimeter thyroid nodules (26). In our study, to avoid false negative cases as much as possible, nodules graded TIRADS 4A or above with a negative cytological results without repeated FNA were eliminated. As a result, there would be a higher chance that subcentimeter nodules graded TIRADS 4A or above in HT patients were excluded. However those nodules are rather difficult to distinguish between malignant and benign. This could also explain why radiologists did better in the HT subset.

This study had several limitations. First, the training set and test set of this study were from the same hospital, lacking external test set. The performance of our model needed to be verified further more by external trials. Second, the data set contained slightly more benign nodules. However, due to the large amount of data used in this study, it should not be considered as a significant deviation. Third, PTC is the main pathological type for malignancy in this study. Only a small portion was follicular carcinoma. The model couldn't get enough training samples on identifying follicular lesions. Therefore, the model cannot accurately distinguish follicular carcinoma from benign nodules. Another limitation is that the nodules graded as TIRADS 2 and 3 do not necessarily have pathological results. There may be inter-observer variation in nodules with lower TIRADS grading, so it is possible to include very few malignant nodules as benign nodules.

CONCLUSION

In conclusion, our modified DNN model performed slightly better than the radiologists with different years of experience in diagnosing thyroid nodules underlying Hashimoto Thyroiditis. It showed higher sensitivity compared to the radiologists. It was also capable of diagnosing malignant nodules in normal patients. Thus, the DNN model might be a possible solution for screening malignant thyroid nodules in the large population.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding authors.

ETHICS STATEMENT

This study is a retrospective study and was approved by the Institutional Review Board of Shanghai Ruijin Hospital with waiver of informed consent.

AUTHOR CONTRIBUTIONS

YQH, CC, and WZ were major contributors in writing the manuscript, conducting and designing the study. YQH, LZ, WZ, XHJ, JWZ, and WWZ participated in the image annotation, evaluation and study design. QYL, YXQ, CC, LYH, JX, and CG provided technical support. LFZ and MZ provided clinical information on all cases. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by Smart Health Fund (No. 2018ZHYL0106) by the Shanghai Municipal Health Commission and National Natural Science Foundation of China (Nos. 81671688 and 82071923).

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* (2018) 68 (6):394–424. doi: 10.3322/caac.21492
2. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, et al. Cancer statistics in China, 2015. *CA Cancer J Clin* (2016) 66(2):115–32. doi: 10.3322/caac.21338
3. Morris LG, Tuttle RM, Davies L. Changing Trends in the Incidence of Thyroid Cancer in the United States. *JAMA Otolaryngol Head Neck Surg* (2016) 142(7):709–11. doi: 10.1001/jamaoto.2016.0230
4. Lim H, Devesa SS, Sosa JA, Check D, Kitahara CM. Trends in Thyroid Cancer Incidence and Mortality in the United States, 1974-2013. *JAMA* (2017) 317 (13):1338–48. doi: 10.1001/jama.2017.2719
5. Kitahara CM, Sosa JA. The changing incidence of thyroid cancer. *Nat Rev Endocrinol* (2016) 12(11):646–53. doi: 10.1038/nrendo.2016.110
6. Caturegli P, De Remigis A, Rose NR. Hashimoto thyroiditis: clinical and diagnostic criteria. *Autoimmun Rev* (2014) 13(4-5):391–7. doi: 10.1016/j.autrev.2014.01.007
7. Noureldine SI, Tufano RP. Association of Hashimoto's thyroiditis and thyroid cancer. *Curr Opin Oncol* (2015) 27(1):21–5. doi: 10.1097/CCO.0000000000000150
8. Konturek A, Barczyński M, Wierzchowski W, Stopa M, Nowak W. Coexistence of papillary thyroid cancer with Hashimoto thyroiditis. *Langenbecks Arch Surg* (2013) 398(3):389–94. doi: 10.1007/s00423-012-1021-x
9. Takashima S, Matsuzuka F, Nagareda T, Tomiyama N, Kozuka T. Thyroid nodules associated with Hashimoto's thyroiditis: assessment with US. *Radiology* (1992) 185:125–30. doi: 10.1148/radiology.185.1.1523294
10. Tessler FN, Middleton WD, Grant EG, Hoang JK, Berland LL, Teefey SA, et al. ACR Thyroid Imaging, Reporting and Data System (TI-RADS): White Paper of the ACR TI-RADS Committee. *J Am Coll Radiol* (2017) 14(5):587–95. doi: 10.1016/j.jacr.2017.01.046
11. Park M, Park SH, Kim EK, Yoon JH, Moon HJ, Lee HS, et al. Heterogeneous echogenicity of the underlying thyroid parenchyma: how does this affect the analysis of a thyroid nodule? *BMC Cancer* (2013) 13:550. doi: 10.1186/1471-2407-13-550
12. Choi SH, Kim EK, Kwak JY, Kim MJ, Son EJ. Interobserver and intraobserver variations in ultrasound assessment of thyroid nodules. *Thyroid* (2010) 20 (2):167–72. doi: 10.1089/thy.2008.0354
13. Lim KJ, Choi CS, Yoon DY, Chang SK, Kim KK, Han H, et al. Computer-aided diagnosis for the differentiation of malignant from benign thyroid nodules on ultrasonography. *Acad Radiol* (2008) 15(7):853–8. doi: 10.1016/j.acra.2007.12.022
14. Huynh BQ, Li H, Giger ML. Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. *J Med Imaging* (2016) 3:034501. doi: 10.1117/1.JMI.3.3.034501
15. Christodoulidis S, Anthimopoulos M, Ebner L, Christe A, Mougiakakou S. Multisource transfer learning with convolutional neural networks for lung pattern. *Anal IEEE J Biomed Health* (2017) 21:76–84. doi: 10.1109/JBHI.2016.2636929
16. Li X, Zhang S, Zhang Q, Wei X, Pan Y, Zhao J, et al. Diagnosis of thyroid cancer using deep convolutional neural network models applied to sonographic images: a retrospective, multicohort, diagnostic study. *Lancet Oncol* (2019) 20(2):193–201. doi: 10.1016/S1470-2045(18)30762-9
17. Xu L, Gao J, Wang Q, Yin J, Yu P, Bai B, et al. Computer-Aided Diagnosis Systems in Diagnosing Malignant Thyroid Nodules on Ultrasonography: A Systematic Review and Meta-Analysis. *Eur Thyroid J* (2020) 9(4):186–93. doi: 10.1159/000504390
18. Shin JH, Baek JH, Chung J, Ha EJ, Kim JH, Lee, et al. Ultrasonography Diagnosis and Imaging-Based Management of Thyroid Nodules: Revised Korean Society of Thyroid Radiology Consensus Statement and Recommendations. *Korean J Radiol* (2016) 17(3):370–95. doi: 10.3348/kjr.2016.17.3.370
19. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, HI, USA (2017). pp. 2261–9. doi: 10.1109/CVPR.2017.243
20. He K, Girshick R, Dollár P. Rethinking imagenet pre-training. In: *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. Seoul, Korea (South) (2019). pp. 4917–26. doi: 10.1109/ICCV.2019.005023
21. Krzanowski WJ, Hand DJ. *ROC Curves for Continuous Data*. Boca Raton: Chapman & Hall/CRC (2009).
22. Paszke A, Gross S, Chintala S, Chanan G, Yang E, Devito Z, et al. *Automatic differentiation in pytorch*. NIPS Autodiff Workshop (2017).
23. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* (1988) 44(3):837–45. doi: 10.2307/2531595
24. Buda M, Wildman-Tobriner B, Hoang JK, Thayer D, Tessler FN, Middleton WD, et al. Management of Thyroid Nodules Seen on US Images: Deep Learning May Match Performance of Radiologists. *Radiology* (2019) 292 (3):695–701. doi: 10.1148/radiol.2019181343
25. Gao L, Liu R, Jiang Y, Song W, Wang Y, Liu J, et al. Computer-aided system for diagnosing thyroid nodules on ultrasound: A comparison with radiologist-based clinical assessments. *Head Neck* (2018) 40(4):778–83. doi: 10.1002/hed.25049
26. Gao L, Ma B, Zhou L, Wang Y, Yang S, Qu N, et al. The impact of presence of Hashimoto's thyroiditis on diagnostic accuracy of ultrasound-guided fine-needle aspiration biopsy in subcentimeter thyroid nodules: A retrospective study from FUSCC. *Cancer Med* (2017) 6(5):1014–22. doi: 10.1002/cam4.997
27. Wang L, Yang S, Yang S, Zhao C, Tian G, Gao Y, et al. Automatic thyroid nodule recognition and diagnosis in ultrasound imaging with the YOLOv2 neural network. *World J Surg Oncol* (2019) 17(1):12. doi: 10.1186/s12957-019-1558-z

Conflict of Interest: Authors CC, QYL, CG, YXQ, JX, and LYH were employed by the company Ping An Technology Company of China, Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Oncology

ORIGINAL RESEARCH
published: 26 March 2021
doi: 10.3389/fonc.2021.544979

Machine Learning-Based Ultrasomics Improves the Diagnostic Performance in Differentiating Focal Nodular Hyperplasia and Atypical Hepatocellular Carcinoma

Wei Li[1†], *Xiao-Zhou Lv*[2†], *Xin Zheng*[1], *Si-Min Ruan*[1], *Hang-Tong Hu*[1], *Li-Da Chen*[1], *Yang Huang*[1], *Xin Li*[3], *Chu-Qing Zhang*[4], *Xiao-Yan Xie*[1], *Ming Kuang*[1,5], *Ming-De Lu*[1,5], *Bo-Wen Zhuang*[1*] and *Wei Wang*[1*]

[1] Department of Medical Ultrasonics, Institute of Diagnostic and Interventional Ultrasound, Ultrasomics Artificial Intelligence X-Lab, The First Affiliated Hospital of Sun Yat-Sen University, Guangzhou, China, [2] Department of Traditional Chinese Medicine, The First Affiliated Hospital of Sun Yat-sen University, Guangzhou, China, [3] Research Center, GE Healthcare, Shanghai, China, [4] Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou, China, [5] Department of Hepatobiliary Surgery, The First Affiliated Hospital of Sun Yat-Sen University, Guangzhou, China

Edited by:
Youyong Kong,
Southeast University, China

Reviewed by:
Yuming Jiang,
Stanford University, United States
Jinghao Duan,
Shandong University, China
Yan Wu,
Second Affiliated Hospital of Harbin Medical University, China

***Correspondence:**
Wei Wang
wangw73@mail.sysu.edu.cn
Bo-Wen Zhuang
zhuangbw3@mail.sysu.edu.cn

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *23 March 2020*
Accepted: *03 March 2021*
Published: *26 March 2021*

Citation:
Li W, Lv X-Z, Zheng X, Ruan S-M, Hu H-T, Chen L-D, Huang Y, Li X, Zhang C-Q, Xie X-Y, Kuang M, Lu M-D, Zhuang B-W and Wang W (2021) Machine Learning-Based Ultrasomics Improves the Diagnostic Performance in Differentiating Focal Nodular Hyperplasia and Atypical Hepatocellular Carcinoma. Front. Oncol. 11:544979. doi: 10.3389/fonc.2021.544979

Background: The typical enhancement patterns of hepatocellular carcinoma (HCC) on contrast-enhanced ultrasound (CEUS) are hyper-enhanced in the arterial phase and washed out during the portal venous and late phases. However, atypical variations make a differential diagnosis both challenging and crucial. We aimed to investigate whether machine learning-based ultrasonic signatures derived from CEUS images could improve the diagnostic performance in differentiating focal nodular hyperplasia (FNH) and atypical hepatocellular carcinoma (aHCC).

Patients and Methods: A total of 226 focal liver lesions, including 107 aHCC and 119 FNH lesions, examined by CEUS were reviewed retrospectively. For machine learning-based ultrasomics, 3,132 features were extracted from the images of the baseline, arterial, and portal phases. An ultrasomics signature was generated by a machine learning model. The predictive model was constructed using the support vector machine method trained with the following groups: ultrasomics features, radiologist's score, and combination of ultrasomics features and radiologist's score. The diagnostic performance was explored using the area under the receiver operating characteristic curve (AUC).

Results: A total of 14 ultrasomics features were chosen to build an ultrasomics model, and they presented good performance in differentiating FNH and aHCC with an AUC of 0.86 (95% confidence interval [CI]: 0.80, 0.89), a sensitivity of 76.6% (95% CI: 67.5%, 84.3%), and a specificity of 80.5% (95% CI: 70.6%, 85.9%). The model trained with a combination of ultrasomics features and the radiologist's score achieved a significantly higher AUC (0.93, 95% CI: 0.89, 0.96) than that trained with the radiologist's score (AUC: 0.84, 95% CI: 0.79, 0.89, $P < 0.001$). For the sub-group of HCC with normal AFP value, the model trained with a combination of ultrasomics features, and the radiologist's score remain achieved the highest AUC of 0.92 (95% CI: 0.87, 0.96) compared to that with the ultrasomics features (AUC: 0.86, 95% CI: 0.74, 0.89, $P < 0.001$) and radiologist's score (AUC: 0.86, 95% CI: 0.79, 0.91, $P < 0.001$).

Conclusions: Machine learning-based ultrasomics performs as well as the staff radiologist in predicting the differential diagnosis of FNH and aHCC. Incorporating an ultrasomics signature into the radiologist's score improves the diagnostic performance in differentiating FNH and aHCC.

Keywords: ultrasonography, machine learning, focal nodular hyperplasia, hepatocellular carcinoma, ultrasomics

INTRODUCTION

The typical enhancement pattern of hepatocellular carcinoma (HCC) on contrast-enhanced ultrasound (CEUS) is characterized by hyper-enhancement in the arterial phase and wash out during the portal venous and late phases (1). However, atypical variations occur, especially in some well-differentiated tumors, accounting for 5–41% of HCC cases; such lesions may show sustained hyper-/iso-enhancement in the portal venous and late phases and are defined as atypical HCC (aHCC) (2–4). Meanwhile, most benign focal liver lesions show complete hyper- or iso-enhancement in the portal venous and late phases, making differential diagnosis both crucial and challenging (5, 6). This diagnostic difficulty could be resolved using CEUS techniques, such as micro-flow imaging to further characterize the enhancement features in the arterial phase, e.g., a spoke-wheel artery for focal nodular hyperplasia (FNH) and chaotic vessel for HCC (7–10). However, the interpretation of features involves the experience of radiologists, making inter-reader variability inevitable.

In contrast to the traditional practice of treating medical images as pictures intended solely for visual interpretation, radiomics features could reflect not only the macroscopic manifestation but also the cellular and molecular nature of tissues (11–13). Radiomics offers a vast scale of imaging biomarkers that could potentially assist in detecting and diagnosing, evaluating the prognosis and predicting the therapeutic response, and monitoring the disease status of cancer (11, 12, 14–16). Machine learning-based ultrasomics approaches, derived from radiomics, involve the analysis and transformation of ultrasound images into large sets of quantitative data and have been identified as potential alternatives to detect and classify lesions (17, 18).

Recently, few applications of machine learning in HCC diagnosis have been reported (19, 20). Most machine learning systems have demonstrated excellent diagnostic performance, with the area under the receiver operating characteristic curve (AUC) of 0.89-0.97 for HCC characterization (19, 20). Gatos et al. applied radiomics to segment and classify focal liver lesions on non-enhanced T2-weighted images, providing a non-invasive method for assessing liver lesions (21). Some studies have shown that multi-modal ultrasound images also perform well for the detection and classification of focal liver lesions (19, 22, 23). However, most studies have only compared the diagnostic performance between machine learning systems and radiologists. The influence of the performance of these systems on radiologists when used in clinical practice has not been evaluated. Thus, the added clinical value of machine learning systems to observers is necessary to determine and validate.

The purpose of our research was to develop a machine learning-based ultrasomics approach to assess ultrasomics features for improving the diagnostic performance in differentiating FNH and aHCC.

PATIENTS AND METHODS

Patients

This retrospective analysis obtained ethical approval and waived the informed consent requirement. From December 2013 to January 2018, 119 patients with FNH and 107 patients with aHCC lesions were included in the study based on the inclusion and exclusion criteria. The inclusion criteria were as follows: (a) CEUS was performed; (b) lesions were visually hyper-enhanced during the arterial phase and sustained hyper- or iso-enhanced during the portal venous and late phases; (c) HCC was diagnosed by pathological examinations and FNH was confirmed by pathological examinations or supported by CT or MRI findings with a minimum 1 year follow-up; and (d) no treatment was conducted before CEUS. Patients were excluded if they had multiple tumors. Baseline clinical trial data, including age, gender, and some blood test, such as hepatitis background and alpha-fetoprotein (AFP), were performed no more than 7 days before or after the CEUS examination.

Image Acquisition

US examinations were performed using an Aplio 500 scanner (Canon Medical Systems, Tokyo, Japan), equipped with a 375BT convex transducer (frequency, 3.5 MHz) and an Aixplorer scanner (Supersonic, Paris, France) with an SC6-1 curvilinear transducer (frequency, 1–6 MHz). Contrast harmonic imaging (CHI) and contrast pulse sequencing (CPS) were used with a mechanical index of 0.06–0.10. Baseline ultrasonography was performed to scan the liver thoroughly before CEUS. Additionally, the target lesions were identified and observed carefully during the baseline observation in B mode. The imaging settings, such as the gain, depth, and focus, were optimized for each examination. After the CHI or CPS mode was activated, a bolus intravenous injection of 2.4 mL of SonoVue (Bracco, Milan, Italy) was administered, followed by flushing with 5 mL of saline. The targeted lesion was observed continuously for 5 minutes. The arterial, portal venous, and late phases were defined as 10–30 seconds, 31–120 seconds, and 121–300 seconds after injection, respectively. CEUS examinations were performed by one of two radiologists (WW and X-YX) with at least 10 years' experience of performing CEUS. Three images from the same section, which showed the maximum observation of the target lesion, were taken from each patient, (a) a baseline ultrasound; (b) an arterial phase image at the enhancement time of 25–30 seconds; and (c) a portal phase image at the enhancement time of 60–70 seconds.

Radiologist's Scoring

Two staff radiologists (B-WZ and L-DC) reviewed ultrasound images and videos retrospectively, and they had more than 5 years of experience in assessing liver CEUS data. The radiologists were not involved in the feature extraction process below. All patient identification information from the images was removed, and the researchers were unaware of all the clinicopathological information. The diagnostic criteria for HCC and FNH were based on the 2012 guidelines issued by the European Federation of Societies for Ultrasound in Medicine and Biology (EFSUMB) (2). The diagnostic criteria for HCC were the manifestation of basket pattern and/or chaotic vessels (**Figure 1A**) and non-enhanced areas (**Figure 1B**). The diagnostic criteria of CEUS feature for FNH were centrifugal enhancement (Video 1), spoke-wheel artery (**Figure 1C**), unenhanced central scar (**Figure 1D**), and feeding artery.

For each group, the diagnostic confidence was scored using a subjective three-point scale (grade 1, definitely or most likely FNH; grade 2, indeterminate; and grade 3, most likely or definitely HCC). If there was inconsistency, we performed a consensus reading, and the consensus data were used for subsequent analysis.

Ultrasomics Feature Extraction

Digital imaging and communications in medicine (DICOM) images were used to extract ultrasomics features using the in-

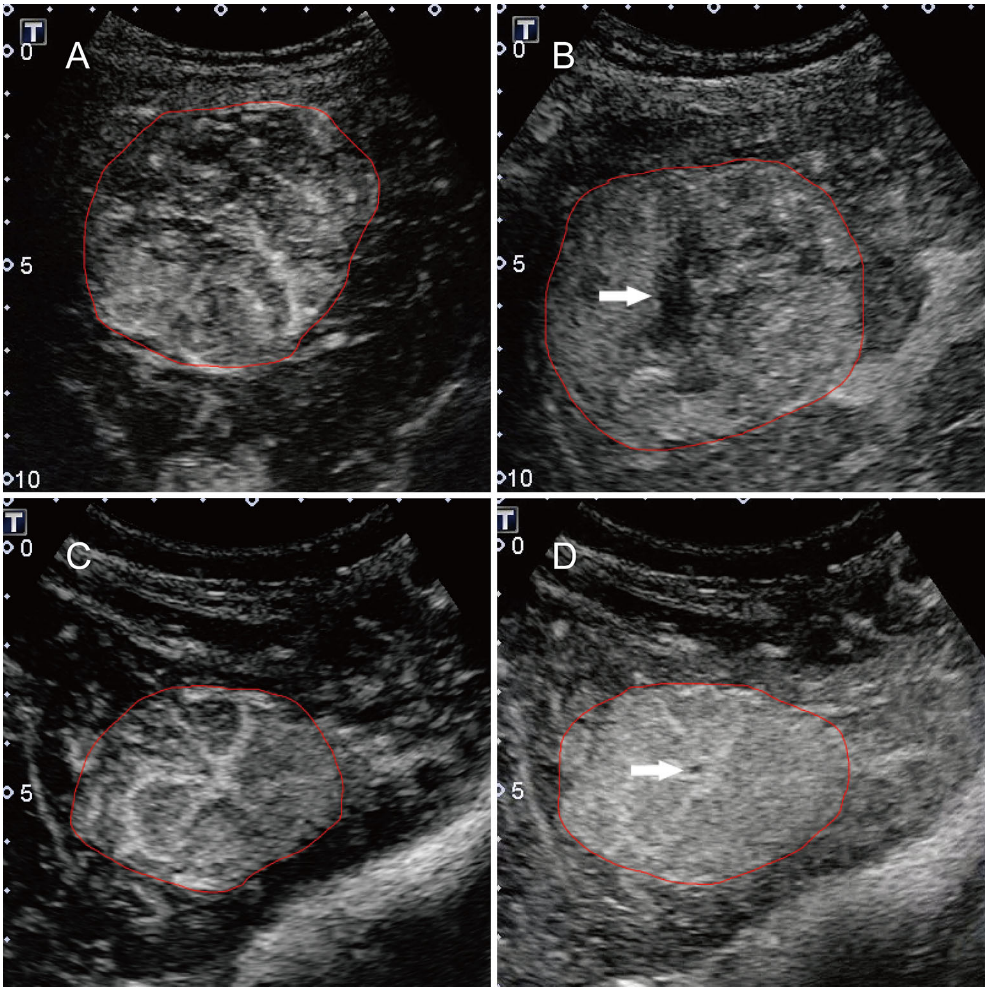


FIGURE 1 | Typical features for HCC and FNH lesions. **(A)** the basket pattern and/or chaotic vessels; **(B)** non-enhancing areas (arrow); **(C)** spoke-wheel arteries; and **(D)** unenhanced central scar (arrow). Annotations of the ROI generated by the radiologists around the tumor outline are delineated in red.

house designed Ultrasomics-Platform software (Version 1.0; Ultrasomics Artificial Intelligence X-lab, Guangzhou, China). After an image was imported, the radiologist drew a region of interest (ROI) on the largest cross-section along the tumor contour. Next, the software automatically extracted the features from the ROIs. In total, 1,044 features could be extracted from a single image. These 1,044 features extracted from a single image consisted of five categories of features: histogram parameters, textural parameters, form factor parameters, grey-level co-occurrence matrix (GLCM) parameters, and run length matrix (RLM) parameters. Detailed information on the features is provided in **Supplementary Material S1**. Finally, 3,132 features were extracted from the baseline US, arterial phase and portal phase of CEUS images of each patient. Initially, two radiologists (WL and YH, with at least 5 years of experience in performing US examinations) were required to trace out ROIs on the selected images. The inter-observer and intra-observer reproducibility in feature extraction were assessed and are described in **Supplementary Material S2**. The remaining images were delineated by the first radiologist.

Feature Selection and Model Development for Prediction

Of the 3,132 features from each patient, many were highly redundant, which could degrade the classification. We eliminated redundant features by using a two-step feature selection method. First, if two features were highly-correlated with a correlation coefficient higher than 0.95, one of the features was removed. Second, we eliminated features with an AUC less than 0.6. According to the Harrell guidelines for multivariate analysis, the number of events should be at least 10 times greater than that of the included covariates (24). The least absolute shrinkage and selection operator (LASSO) regression was used to perform the ultrasomics features selection in the training dataset. All ultrasomics feature values were normalized by using the mean and variance of the feature values to be within similar dynamic ranges.

A support vector machine (SVM) based on the radial basis function (RBF) kernel was trained from the selected feature subset produced by the preceding steps. The entire data set was randomly divided into a training dataset (comprising 80% of subjects) and a validation dataset (comprising the remaining 20% of subjects). The training dataset was used to construct a model, which was then evaluated using the validation dataset. A 10-fold cross-validation method was adopted to ensure the robustness of the classifiers to training and testing data. All processes were repeated 10 times with random seeds, generating 10 different training and validation datasets. We built the model using the training dataset and then evaluated it using the validation dataset repeatedly. Subsequently, the model with the best classification performance was selected as the best model.

Statistical Analysis

Descriptive statistics are summarized as the mean ± standard deviation (SD) or median and interquartile range. Comparisons between groups were tested using Student's t test or the Mann-Whitney test for quantitative variables and the chi-squared test or Fisher's test for qualitative variables.

A weighted kappa statistics test was used to assess the two radiologists' scores. We evaluated the reproducibility of the ultrasomics feature extraction using the "irr" package in R. The LASSO regression was performed using the "glmnet" package.

All observations of patients with known outcomes were classified into three datasets (1): radiologist's score, (2) ultrasomics features, and (3) a combination of ultrasomics features and radiologist's score. The diagnostic performance of the radiologist's score was evaluated by plotting receiver operating characteristic (ROC) curves. The diagnostic performance in discriminating between FNH and aHCC is expressed as the AUC. The ultrasomics features and the combination of the ultrasomics features and radiologist's score were further compared through an SVM classifier using the "rattle" package in R. The performance of the SVM model was tested using the AUC. Paired comparisons of AUC values were performed by a two-sided Wilcoxon signed-rank test at a significance level of 5%. The predictive sensitivity (SEN), specificity (SPE), positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (+LR), and negative likelihood ratio (-LR) were calculated at a cut-off point that maximized the value of the Youden index. Comparisons among the three datasets were performed using the Delong test. Decision curve analysis (DCA) was performed with the "dca.R" function. All statistical tests were two-sided tests, and $P< 0.05$ indicated statistical significance. All Statistical analyses were performed using R version 3.3.3 (http://www.r-project.org/).

RESULTS

Clinical Characteristics

The clinical characteristics are listed in **Table 1**. The study included 226 patients; 107 (47.3%) patients (mean age, 54.0 ± 11.9 years old) had a final diagnosis of HCC; and the remaining 119 (52.7%) patients (mean age, 34.5 ± 11.7 years old) had a final diagnosis of FNH. 20 FNH lesion were confirmed by pathological examinations (11 by biopsy, 9 by surgery), while 99 cases were supported by CT or MRI findings with a minimum one-year follow-up. No significant difference was found in the tumor number between the two groups (P=0.118). The average lesion size of FNH and HCC was 3.3 ± 1.8 cm (range: 0.8-10.2 cm) and 4.8 ± 3.4 cm (range: 0.8-18.6 cm), respectively.

Ultrasomics Signature Construction and Validation

After the feature selection and dimensional reduction process, 14 selected features were taken as the input of the SVM to train a prediction model, including 6 features derived from baseline US images, 3 from arterial phase images, and 4 from portal phase images (**Figure 2**, **Supplementary Material S3**). All feature values were normalized to achieve similar dynamic ranges. The parameter C which is used to control the error-margin trade-off was set at 1, and the kernel width sigma was 0.012. Next, the training and validation procedures for tumor classification were employed with 10-fold cross-validation.

TABLE 1 | Clinical Characteristics and Laboratory Information of the Patients.

Patients	FNH(N=119)	aHCC(N =107)	P value
Gender (male/female)	61/58	92/15	<0.001
Age (years)	34.5 ± 11.7	54.0 ± 11.9	<0.001
HBsAg (IU/ml)			<0.001
≤0.05	116 (97.5)	15 (14.0)	
0.05–250	1 (0.8)	34 (31.8)	
>250	2 (1.7)	58 (54.2)	
HBV-DNA (IU/mL)			<0.001
<100	118 (99.2)	46 (43.0)	
100–10^5	0	44 (41.1)	
>10^5	1 (0.8)	17 (15.9)	
HCV-Ab (S/CO)			0.212
<1.0	119 (100)	105 (98.1)	
≥1.0	0	2 (1.9)	
AFP (μg/L)			<0.001
<20	117 (98.3)	38 (35.5)	
20–400	2 (1.7)	40 (37.4)	
>400	0	29 (27.1)	
Tumor number			0.118
1	112 (94.1)	92 (86.0)	
2	4 (3.4)	8 (7.5)	
≥3	3 (2.5)	7 (6.5)	
Tumor size (cm)	3.34 ± 1.80	4.76 ± 3.35	<0.01
<3	64 (53.8)	37 (34.6)	
3-5	35 (29.4)	35 (32.7)	
>5	20 (16.8)	35 (32.7)	

Data are the number of patients, with the percentage in parentheses unless indicated. aHCC, atypical hepatocellular carcinoma.

Diagnostic Performance of Ultrasomics Features and Radiologist's Score Models

A total of 83 FNHs were correctly identified while 36 lesions were incorrectly identified as HCC, leading to a specificity of 69.8% by the radiologists. For combined model, 100 FNHs were correctly classified whereas 19 lesions were incorrectly assigned to HCC, resulting in a specificity of 84.0%. Comparing the performance of the radiologists' score and the combined model, twenty-four cases have a different result; consequently, the combined model leads to an additional 15 FNHs and 5 HCCs being correctly classified.

The model trained with the combination of the ultrasomics features and radiologist's score performed significantly better (AUC: 0.93, 95% CI: 0.89, 0.96) than that trained with the ultrasomics features (AUC: 0.86, 95% CI: 0.80, 0.89, $P < 0.001$) and radiologist's score (AUC: 0.84, 95% CI: 0.79, 0.89, $P < 0.001$). Overall, the model based on the radiologist's score had the highest diagnostic SEN of 94.4% (95% CI: 88.2%, 97.9%) but the lowest SPE of 69.8% (95% CI: 60.7%, 77.8%) compared with the combined model (SEN: 93.5% [95% CI: 87.0%, 97.3%], SPE: 84.9% [95% CI: 77.1%, 90.8%]), and the ultrasomics features model (SEN: 76.6% [95% CI: 67.5%, 84.3%], SPE: 80.5% [95% CI: 70.6%, 85.9%]). Furthermore, when the ultrasomics features were combined with the radiologist's score, the diagnostic performance was significantly improved in terms of the AUC, SPE, and PPV and + LR (AUC: 0.93, SPE: 84.9%, and PPV: 84.7%, +LR 6.2) compared with the performance of the other two models. The performance measurements of each dataset are reported in **Table 2** based on each ROC curves to distinguish between FNH and aHCC (**Figure 3**).

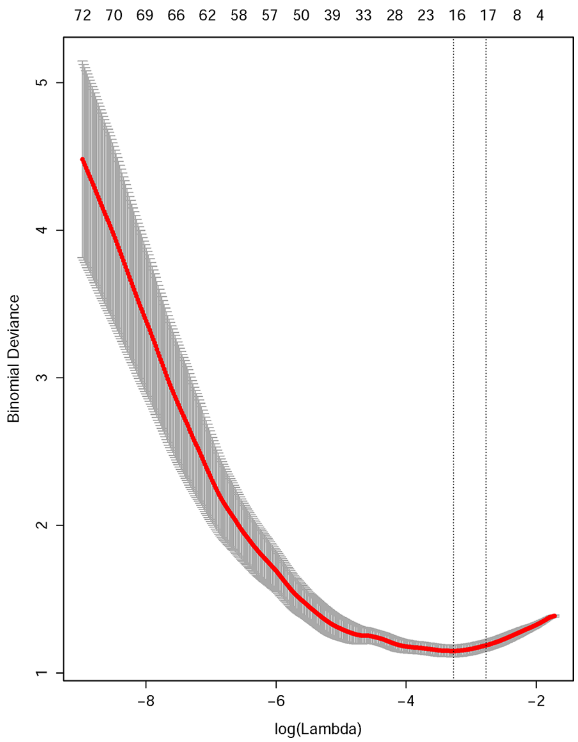


FIGURE 2 | Radiomics feature selection using the least absolute shrinkage and selection operator (LASSO) regression model. The 10-fold cross-validation (CV) process was repeated 50 times to generate the optimal penalization coefficient lambda (λ) in the LASSO model. The value of λ that produced the minimum average binomial deviance was used to select features. Dotted vertical lines were drawn at the optimal values using the minimum criteria and the 1 standard error of the minimum criteria (the 1-SE criteria). A λ value of 0.043 was chosen (the 1-SE criteria) according to 10-fold CV, where optimal λ resulted in 14 nonzero coefficients.

TABLE 2 | Diagnostic Performance of the Three Models in Differentiating Focal Nodular Hyperplasia and Atypical Hepatocellular Carcinoma.

	Ultrasomics score	Radiologist's score	Combined
Sensitivity (%)	76.6 (67.5-84.3)	94.4 (88.2-97.9)	93.5 (87.0-97.3)
Specificity (%)	80.5 (70.6-85.9)	69.8 (60.7-77.8)	84.9 (77.1-90.8)
PPV (%)	76.6 (67.5-84.3)	73.7 (65.5-80.9)	84.7 (77.0-90.7)
NPV (%)	79.0 (70.6-85.9)	93.3 (85.9-97.5)	93.5 (87.1-97.3)
+LR	3.7 (3.2-4.2)	3.1 (2.7-3.5)	6.2 (5.6-6.8)
-LR	0.3 (0.2-0.5)	0.1 (0.04-0.2)	0.1 (0.03-0.2)
AUC of training set	0.94 (0.89-0.99)	0.93 (0.85-0.98)	0.99 (0.94-1.00)
AUC of validation set	0.86 (0.80-0.89)	0.84 (0.79-0.89)	0.93 (0.89-0.96)

Data in parentheses are 95% confidence interval. PPV, positive predictive value; NPV, negative predictive value; +LR, positive likelihood ratio; -LR, negative likelihood ratio; AUC, area under the curve.

The DCA shows that within most reasonable threshold probability ranges, the combined model showed the highest overall net benefit than the radiologist's score or ultrasound feature model. The DCA results for the three models are presented in **Figure 4**.

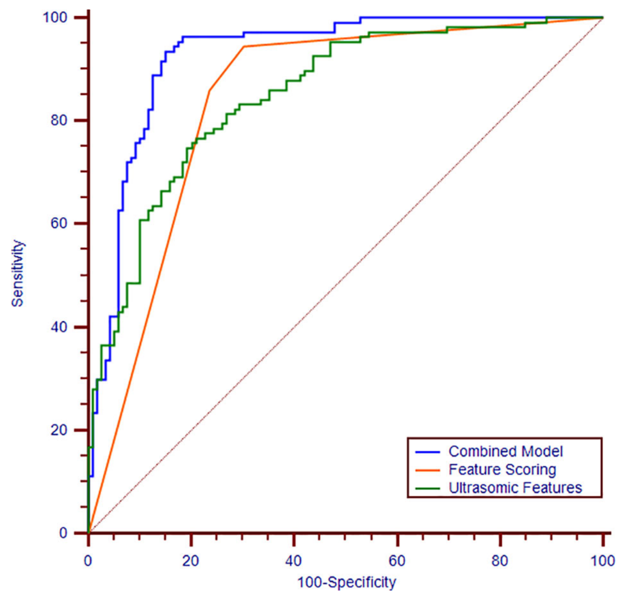


FIGURE 3 | Receiver operating characteristic curves of the combination of ultrasomics features and radiologist's score (blue curve), ultrasomics features (green curve), and radiologist's score (orange curve). The areas under the curves are 0.93, 0.86, 0.84, respectively.

Validation in the Sub-group of HCC With Normal AFP Level

Forty patients were confirmed to develop HCC with normal AFP level. The model trained above performed significantly better (AUC: 0.92, 95% CI: 0.87, 0.96) with the combination of the ultrasomics features and radiologist's score than that with the ultrasomics features (AUC: 0.86, 95% CI: 0.74, 0.89, $P < 0.001$) and radiologist's score (AUC: 0.86, 95% CI: 0.79, 0.91, $P < 0.001$) (**Table 3**).

DISCUSSION

In this study, we derived and validated an ultrasomics-based machine learning approach to analyze ultrasound images for the preoperative individualized diagnosis of FNH and aHCC. Our analysis reveals that the diagnostic performance of ultrasomics is comparable to that of a staff radiologist in differentiating between FNH and aHCC. Furthermore, when adding ultrasomics to the radiologist's classification, the diagnostic performance was improved significantly with an AUC ranging from 0.84 to 0.93 ($P < 0.001$). Our study shows that ultrasomics may increase the diagnostic confidence of radiologists in CEUS examinations and potentially improve their accuracy when facing atypical features. Thus, clinicians would benefit from this decision-making process in the diagnosis of HCC.

In clinical practice, when radiologists face a lesion that shows hyper-enhancement in the arterial phase and sustained enhancement in the portal vein and late phases, it is difficult to diagnose HCC. However, the high sensitivity and negative predictive values of the radiologist would be useful in clinical practice for excluding disease; thus, HCC would be excluded by the radiologist if the result was considered to be FNH by ultrasomics. Therefore, this system would help to reduce unnecessary biopsies or active clinical treatment requested by experienced radiologists.

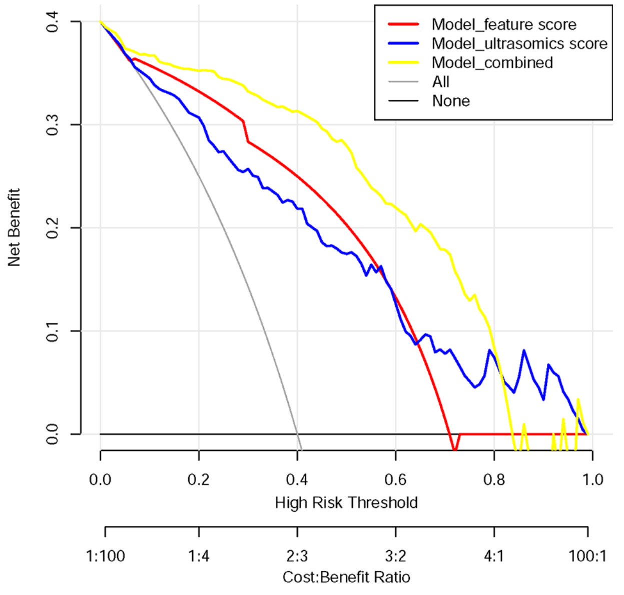


FIGURE 4 | Decision curve analysis for each model. The y-axis measures the net benefit. The net benefit was calculated by summing the benefits (true positive results) and subtracting the harms (false-positive results), weighting the latter by a factor related to the relative harm of undetected cancer compared with the harm of unnecessary treatment. The combined model (yellow line) had the highest net benefit compared with the other two models (blue line and red line) and simple strategies, such as the follow-up of all patients (grey line) or no patients (horizontal black line), across the full range of threshold probabilities at which a patient would choose to undergo a follow-up imaging examination.

In contrast, ultrasomics, referred to as high-throughput computing, extracts innumerable quantitative features from US images (18). By transforming digital medical images into mineable high-dimensional data, ultrasomics yields features, such as textural features, that could objectively reflect the homogeneity or heterogeneity of an image. These patterns could represent enhancement features just as heterogeneity might represent chaotic vessels and necrosis. Focal liver lesions can be featured by typical features in the arterial phase and wash-out during the portal and late phases. In this study, the features displayed in the arterial phase could provide a major benefit for the diagnosis of liver tumors. In previous studies, we utilized a maximum intensity projection technique of micro-flow imaging and achieved higher spatial resolution and higher temporal resolution when detecting vessel contours. Compared to conventional CEUS features (AUC: 0.84), micro-flow imaging technology provided significant improvements over the detection rates achieved for the staff radiologists (AUC: 0.89) (10). In this study, ultrasomics features alone can achieve a similar diagnostic performance (AUC: 0.86) as

TABLE 3 | Validation in the sub-group of HCC with normal AFP value.

	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	+LR	-LR	AUC
Ultrasomics score	77.5 (61.5-89.1)	80.5 (72.2-87.2)	60.0 (43.1-75.3)	89.8 (81.5-95.2)	4.4 (3.4-5.8)	0.3 (0.2-0.6)	0.86 (0.74- 0.89)
Radiologist's score	92.5 (79.6-98.3)	77.1 (68.5-84.3)	57.8 (44.7-70.2)	96.8 (90.9-99.3)	4.0 (3.5-4.6)	0.1 (0.03-0.3)	0.86 (0.79-0.91)
Combined	95.0 (83.0-99.2)	82.2 (74.1-88.6)	64.4 (50.9 - 76.4)	98.0 (92.9-99.7)	5.3 (4.8-6.0)	0.06 (0.01-0.2)	0.92 (0.87- 0.96)

Data in parentheses are 95% confidence intervals. PPV, positive predictive value; NPV, negative predictive value; +LR, positive likelihood ratio; -LR, negative likelihood ratio; AUC, area under the curve.

micro-flow imaging (AUC: 0.868–0.873). Ultrasomics could reach such achievement because it analyzes textural features objectively and quantitatively to describe the intrinsic characteristics of tumors, in particular heterogeneous tumors. Ultrasomics analysis has already been applied to various types of disease, such as HCC, liver fibrosis, and breast cancer (18, 25–27). The potential of ultrasomics has already been demonstrated for liver imaging in some studies (17, 18).

In this study, we additionally evaluated the benefit of ultrasomics in assisting doctors with the interpretation of medical images. As an interdisciplinary technology, it combines elements of imaging generation, digital image processing, statistical imaging, and knowledge engineering to manage the volume of information related to the diagnostic process and outcome prediction (17, 18, 25). In our study, the additional information in the combined model led to improved diagnostic performance (AUC: 0.93) and higher specificity of 84.9% compared with the ultrasomics (AUC: 0.86, specificity: 80.5%) and radiologist's score (AUC: 0.84, specificity: 69.8%) models. The combined model was also comparable to the ML model based on multi-modal ultrasound images (AUC: 0.94, sensitivity: 91.0%, specificity: 86.0%) (19). The results are also comparable to and even better than those of MRI (AUC: 0.89, sensitivity: 82.2%, specificity: 71.4%), as previously reported (20). However, the use of artificial intelligence is not intended to replace expert diagnosticians because no solution is guaranteed and knowledge-based maintenance is required. Artificial intelligence is also affected by several elements, such as the source of images and the cognition of disease. Presently, most domains of large data have not tapped the full potential of artificial intelligence technology. However, rapid developments in the area will add more potential to the advantages. Therefore, the most important role of artificial intelligence is to help improve diagnostic accuracy and assist rather than replace clinicians in making treatment decisions. It is worth noting that the combined model greatly improved the diagnostic ability of radiologists. A similar conclusion was obtained in another study (28).

Our research has some limitations. First, this study was retrospective and conducted in one center. This may cause potential variations and selection bias in the patient population and imaging methods, which is difficult to generalize the outcomes to other agencies. Second, due to the relative rarity of aHCC, the sample size is relatively small, which may cause over-fitting to this particular population. Hence, large-scale multicenter studies are necessary for the future to validate the results. Third, only two radiologists were involved in the assessment of the basic imaging features and feature extraction. All outcomes were based on the features extracted by one radiologist, which may not be generalizable to all radiologists. Fourth, the machine and imaging settings in this study were inconsistent, which may affect the ultrasomics features (29, 30).

In conclusion, an ultrasomics approach was developed to investigate the association between the quantitative ultrasound features and pathological characteristics of tumors effectively and objectively. We evaluated the added value of ultrasomics to the radiologist, and this approach improved the performance of CEUS by providing quantitative and standardized criteria to radiologists, thereby enabling the more confident application of CEUS in detecting HCC to achieve better treatment planning. Our findings can assist clinicians in the differential diagnosis between FNH and aHCC accurately using CEUS images, and this allows for early and precise medical management and treatment.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Institutional Review Board of the First Affiliated Hospital of Sun Yat-Sen University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

WW, M-DL, MK, and X-YX contributed conception and design of the study. XZ collected data from clinical trials. WL and B-WZ drafted and revised the manuscript. X-ZL, WW, L-DC, and S-MR revised the manuscript for important intellectual content. X-ZL, XL, H-TH, YH, and C-QZ gave technical support. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by grants from the National Natural Science Foundation of China (No. 81601500 and 81701701).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.544979/full#supplementary-material

REFERENCES

1. Martie A, Sporea I, Popescu A, Sirli R, Danila M, Serban C, et al. Contrast enhanced ultrasound for the characterization of hepatocellular carcinoma. *Med Ultrason* (2011) 13(2):108–13.
2. Claudon M, Dietrich CF, Choi BI, Cosgrove DO, Kudo M, Nolsoe CP, et al. Guidelines and good clinical practice recommendations for contrast enhanced ultrasound (CEUS) in the liver–update 2012: a WFUMB-EFSUMB initiative in cooperation with representatives of AFSUMB, AIUM, ASUM, FLAUS and ICUS. *Ultraschall Med* (2013) 34(1):11–29. doi: 10.1055/s-0032-1325499
3. Nicolau C, Catala V, Vilana R, Gilabert R, Bianchi L, Sole M, et al. Evaluation of hepatocellular carcinoma using SonoVue, a second generation ultrasound contrast agent: correlation with cellular differentiation. *Eur Radiol* (2004) 14 (6):1092–9. doi: 10.1007/s00330-004-2298-0
4. Jang HJ, Kim TK, Burns PN, Wilson SR. Enhancement patterns of hepatocellular carcinoma at contrast-enhanced US: comparison with histologic differentiation. *Radiology* (2007) 244(3):898–906. doi: 10.1148/radiol.2443061520
5. Dietrich CF. Contrast-Enhanced Ultrasound of Benign Focal Liver Lesions. *Ultraschall der Med* (2018) 40(1):12–29. doi: 10.1055/a-0668-5746
6. Friedrich-Rust M, Klopffleisch T, Nierhoff J, Herrmann E, Vermehren J, Schneider MD, et al. Contrast-Enhanced Ultrasound for the differentiation of benign and malignant focal liver lesions: a meta-analysis. *Liver Int* (2013) 33 (5):739–55. doi: 10.1111/liv.12115
7. Wang W, Chen LD, Lu MD, Liu GJ, Shen SL, Xu ZF, et al. Contrast-enhanced ultrasound features of histologically proven focal nodular hyperplasia: diagnostic performance compared with contrast-enhanced CT. *Eur Radiol* (2013) 23(9):2546–54. doi: 10.1007/s00330-013-2849-3
8. Yen YH, Wang JH, Lu SN, Chen TY, Changchien CS, Chen CH, et al. Contrast-enhanced ultrasonographic spoke-wheel sign in hepatic focal nodular hyperplasia. *Eur J Radiol* (2006) 60(3):439–44. doi: 10.1016/j.ejrad.2006.06.007
9. Xu HX, Liu GJ, Lu MD, Xie XY, Xu ZF, Zheng YL, et al. Characterization of focal liver lesions using contrast-enhanced sonography with a low mechanical index mode and a sulfur hexafluoride-filled microbubble contrast agent. *J Clin Ultrasound* (2006) 34(6):261–72. doi: 10.1002/jcu.20234
10. Li W, Wang W, Liu GJ, Chen LD, Wang Z, Huang Y, et al. Differentiation of Atypical Hepatocellular Carcinoma from Focal Nodular Hyperplasia: Diagnostic Performance of Contrast-enhanced US and Microflow Imaging. *Radiology* (2015) 275(3):870–9. doi: 10.1148/radiol.14140911
11. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology* (2016) 278(2):563–77. doi: 10.1148/radiol.2015151169
12. Sun R, Limkin EJ, Vakalopoulou M, Dercle L, Champiat S, Han SR, et al. A radiomics approach to assess tumor-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: an imaging biomarker, retrospective multicohort study. *Lancet Oncol* (2018) 19(9):1180–91. doi: 10.1016/S1470-2045(18)30413-3
13. Banerjee S, Wang DS, Kim HJ, Sirlin CB, Chan MG, Korn RL, et al. A computed tomography radiogenomic biomarker predicts microvascular invasion and clinical outcomes in hepatocellular carcinoma. *Hepatology (Baltimore Md)* (2015) 62(3):792–800. doi: 10.1002/hep.27877
14. Limkin EJ, Sun R, Dercle L, Zacharaki EI, Robert C, Reuze S, et al. Promises and challenges for the implementation of computational medical imaging (radiomics) in oncology. *Ann Oncol* (2017) 28(6):1191–206. doi: 10.1093/annonc/mdx034
15. Huang Y, Liu Z, He L, Chen X, Pan D, Ma Z, et al. Radiomics Signature: A Potential Biomarker for the Prediction of Disease-Free Survival in Early-Stage (I or II) Non-Small Cell Lung Cancer. *Radiology* (2016) 281(3):947–57. doi: 10.1148/radiol.2016152234
16. Prasanna P, Patel J, Partovi S, Madabhushi A, Tiwari P. Radiomic features from the peritumoral brain parenchyma on treatment-naive multi-parametric MR imaging predict long versus short-term survival in glioblastoma multiforme: Preliminary findings. *Eur Radiol* (2017) 27(10):4188–97. doi: 10.1007/s00330-016-4637-3
17. Wang K, Lu X, Zhou H, Gao Y, Zheng J, Tong M, et al. Deep learning Radiomics of shear wave elastography significantly improved diagnostic performance for assessing liver fibrosis in chronic hepatitis B: a prospective multicentre study. *Gut* (2018) 68(4):729–41. doi: 10.1136/gutjnl-2018-316204
18. Li W, Huang Y, Zhuang BW, Liu GJ, Hu HT, Li X, et al. Multiparametric ultrasomics of significant liver fibrosis: A machine learning-based analysis. *Eur Radiol* (2018) 29(3):1496–506. doi: 10.1007/s00330-018-5680-z
19. Yao Z, Dong Y, Wu G, Zhang Q, Yang D, Yu JH, et al. Preoperative diagnosis and prediction of hepatocellular carcinoma: Radiomics analysis based on multi-modal ultrasound images. *BMC Cancer* (2018) 18(1):1089. doi: 10.1186/s12885-018-5003-4
20. Wu J, Liu A, Cui J, Chen A, Song Q, Xie L. Radiomics-based classification of hepatocellular carcinoma and hepatic haemangioma on precontrast magnetic resonance images. *BMC Med Imaging* (2019) 19(1):23. doi: 10.1186/s12880-019-0321-9
21. Gatos I, Tsantis S, Karamesini M, Spiliopoulos S, Karnabatidis D, Hazle JD, et al. Focal liver lesions segmentation and classification in nonenhanced T2-weighted MRI. *Med Phys* (2017) 44(7):3695–705. doi: 10.1002/mp.12291
22. Gatos I, Tsantis S, Spiliopoulos S, Skouroliakou A, Theotokas I, Zoumpoulis P, et al. A new automated quantification algorithm for the detection and evaluation of focal liver lesions with contrast-enhanced ultrasound. *Med Phys* (2015) 42(7):3948–59. doi: 10.1118/1.4921753
23. Acharya UR, Koh JEW, Hagiwara Y, Tan JH, Gertych A, Vijayananthan A, et al. Automated diagnosis of focal liver lesions using bidirectional empirical mode decomposition features. *Comput Biol Med* (2018) 94:11–8. doi: 10.1016/j.compbiomed.2017.12.024
24. Helmreich JE. *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis. 2nd edn*. Switzerland: Springer International Publishing (2015). doi: 10.1007/978-3-319-19425-7
25. Hu HT, Wang Z, Huang XW, Chen SL, Zheng X, Ruan SM, et al. Ultrasound-based radiomics score: a potential biomarker for the prediction of microvascular invasion in hepatocellular carcinoma. *Eur Radiol* (2019) 29 (6):2890–901. doi: 10.1007/s00330-018-5797-0
26. Chen LD, Ruan SM, Liang JY, Yang Z, Shen SL, Huang Y, et al. Differentiation of intrahepatic cholangiocarcinoma from hepatocellular carcinoma in high-risk patients: A predictive model using contrast-enhanced ultrasound. *World J Gastroenterol* (2018) 24(33):3786–98. doi: 10.3748/wjg.v24.i33.3786
27. Guo B, Ouyang F, Ouyang L, Liu Z, Lin S, Meng W, et al. Development and validation of an ultrasound-based nomogram to improve the diagnostic accuracy for malignant thyroid nodules. *Eur Radiol* (2018) 29(3):1518–26. doi: 10.1007/s00330-018-5715-5
28. Wang J, Wu CJ, Bao ML, Zhang J, Wang XN, Zhang YD. Machine learning-based analysis of MR radiomics can help to improve the diagnostic performance of PI-RADS v2 in clinically relevant prostate cancer. *Eur Radiol* (2017) 27(10):4082–90. doi: 10.1007/s00330-017-4800-5
29. Yu X, Guo Y, Huang SM, Li ML, Lee WN. Beamforming effects on generalized Nakagami imaging. *Phys Med Biol* (2015) 60(19):7513–31. doi: 10.1088/0031-9155/60/19/7513
30. Byra M, Wan L, Wong JH, Du J, Shah SB, Andre MP, et al. Quantitative Ultrasound and B-Mode Image Texture Features Correlate with Collagen and Myelin Content in Human Ulnar Nerve Fascicles. *Ultrasound Med Biol* (2019) 45(7):1830–40. doi: 10.1016/j.ultrasmedbio.2019.02.019

Conflict of Interest: XL was employed by GE Healthcare.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Oncology

METHODS
published: 29 March 2021
doi: 10.3389/fonc.2021.627556

Automatic Detection of Gastric Wall Structure Based on Oral Contrast-Enhanced Ultrasound and Its Application on Tumor Screening

An Sui[1†], *Zhaoyu Hu*[1†], *Xuan Xie*[1], *Yinhui Deng*[1], *Yuanyuan Wang*[1], *Jinhua Yu*[1*] *and Li Shen*[2*]

[1] Electronic Engineering Department, Fudan University, Shanghai, China, [2] Department of Ultrasound, Chongming Branch, Xinhua Hospital Affiliated to Shanghai Jiaotong University School of Medicine, Shanghai, China

Edited by:
Katsutoshi Sugimoto,
Tokyo Medical University, Japan

Reviewed by:
Yohei Koyama,
Tokyo Medical University, Japan
Naoki Matsumoto,
Nihon University School of Medicine,
Japan

***Correspondence:**
Jinhua Yu
jhyu@fudan.edu.cn
Li Shen
shenlicm@126.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: 09 November 2020
Accepted: 15 March 2021
Published: 29 March 2021

Citation:
Sui A, Hu Z, Xie X, Deng Y, Wang Y, Yu J and Shen L (2021) Automatic Detection of Gastric Wall Structure Based on Oral Contrast-Enhanced Ultrasound and Its Application on Tumor Screening. Front. Oncol. 11:627556. doi: 10.3389/fonc.2021.627556

Gastric cancer is the second most lethal type of malignant tumor in the world. Early diagnosis of gastric cancer can reduce the transformation to advanced cancer and improve the early treatment rate. As a cheap, real-time, non-invasive examination method, oral contrast-enhanced ultrasonography (OCUS) is a more acceptable way to diagnose gastric cancer than interventional diagnostic methods such as gastroscopy. In this paper, we proposed a new method for the diagnosis of gastric diseases by automatically analyzing the hierarchical structure of gastric wall in gastric ultrasound images, which is helpful to quantify the diagnosis information of gastric diseases and is a useful attempt for early screening of gastric cancer. We designed a gastric wall detection network based on U-net. On this basis, anisotropic diffusion technology was used to extract the layered structure of the gastric wall. A simple and useful gastric cancer screening model was obtained by calculating and counting the thickness of the five-layer structure of the gastric wall. The experimental results showed that our model can accurately identify the gastric wall, and it was found that the layered parameters of abnormal gastric wall is significantly different from that of normal gastric wall. For the screening of gastric disease, a statistical model based on gastric wall stratification can give a screening accuracy of 95% with AUC of 0.92.

Keywords: gastric cancer, ultrasound, U-net, anisotropic diffusion, edge detection

INTRODUCTION

Gastric cancer is one of the common malignant tumors. The incidence and mortality of gastric cancer in China account for almost half of the world's annual rate (1). The prognosis of gastric cancer is closely related to the timing of diagnosis and treatment. The 5-year survival rate of patients with advanced gastric cancer is still less than 30% even if they receive comprehensive treatment mainly by surgery (2–5). Early diagnosis of gastric cancer can make the clinical stage of the tumor move forward, reduce the transformation to advanced cancer, improve the early treatment rate and the overall cure rate of gastric cancer, which can not only save but also improve the consumption quality of medical resources (6). Therefore, early diagnosis and treatment of gastric cancer has great clinical value.

At present, the medical imaging methods used in the diagnosis of gastric cancer mainly include gastroscopy, CT, MRI and gastric ultrasound. Histopathological diagnosis of gastric mucosa biopsy under gastroscope is the gold standard for the diagnosis of gastric cancer. Gastroscopy and biopsy of gastric mucosa are highly valued and recommended all over the world (7). However, the early diagnostic rate of gastric cancer in developing countries is still unsatisfactory. In China, the early diagnosis and treatment rate of gastric cancer is only about 10% (8). Moreover, as an invasive examination method, gastroscopy has poor acceptability in the population and is difficult to be popularized as a screening method for gastric cancer. CT has high spatial resolution and clear anatomical structure, which is an important examination method for gastric diseases. But the ionizing radiation of CT is harmful to human body. In addition to the long scanning time and expensive price, MRI is also easy to be affected by the difference of pre-scanning disposition and type, field strength, sequence and parameters, which leads to the unsatisfactory imaging stability of MRI in gastric cancer and cannot be widely used in clinical screening.

In recent years, the application of gastric ultrasound is becoming more and more popular, which has unique advantages. Ultrasound imaging is non-invasive, painless, cheap, convenient and real-time. Because the ultrasound beam can penetrate the gastric wall and display the various levels of gastric wall structure, gastric ultrasound has great application value in the diagnosis of gastric diseases, and has a higher detection rate of gastric wall thickening lesions. As a non-invasive and efficient diagnostic method, gastric ultrasound can provide clinicians with a lot of valuable information, timely detect the changes of gastric wall in terms of morphology and thickness, help to estimate the extent of invasion of gastric wall and understand the metastasis and diffusion of various organs around the stomach (9). And it has been preliminarily proved that trans-abdominal ultrasonography can detect gastric cancer early from histopathology and ultrasound physical characteristics (10–12).

In this paper, we proposed a new method based on U-net to automatically identify the gastric wall area in the gastric ultrasound image. The anisotropic diffusion filter and edge detection method are used to stratify the gastric wall structure and calculate the ratio of each layer, which can be used as a reference to diagnose the disease. It's a preliminary report on diagnosis of gastric diseases by using the ratio of thickness of each layer of gastric wall. It is helpful to quantify the diagnostic information of gastric diseases, improve the accuracy of ultrasound diagnosis of gastric cancer, and is expected to improve the screening efficiency of gastric cancer.

MATERIALS AND METHODS

Materials

In this study, we collected 251 gastric ultrasound images from 106 patients, including 47 male patients and 59 female patients. 32 cases were diagnosed with gastric disease, 11 cases were diagnosed with gastric cancer, and the rest of cases are normal. Gastric diseases included 10 cases of gastric ulcer, 12 cases of chronic gastritis, 8 cases of acute gastritis, 1 case of gastric stromal tumor and 1 case of gastric polyps. And the sites of diseases included gastric body, gastric antrum and gastric horn. In 11 cases of gastric cancer, there were 9 cases of early gastric cancer and 2 cases of advanced gastric cancer. For the cases of early gastric cancer, the sites of canceration included gastric body and gastric antrum. One case of gastric body cancer and one case of gastric antrum cancer were pathologically diagnosed as intramucosal cancer, and the other 7 cases were adenocarcinoma. The other two cases of advanced gastric cancer were gastric cardia cancer and gastric body cancer, and both were pathologically diagnosed as adenocarcinoma. The patient characteristics of three cohorts are summarized in **Table 1**. It is worth pointing out that these 11 cases of gastric cancer were found to be abnormal by ultrasound examination for the first time, and they were finally confirmed to be gastric cancer through surgery and pathology.

Ultrasound Scanning Method

The scanning equipment we used is WISONIC Clover 60 portable color Doppler ultrasound diagnostic instrument (Huasheng Medical Technology Co., Ltd, Shenzhen, China) and Ge LOGIQ E9 (GE company, America). Convex array probes are routinely used with a frequency of 5.0MHz. The center frequency will be appropriately adjusted according to the weight of the patient. The patient obeyed the requirement and fasted for 8 hours before the examination to ensure the gastric cavity was empty. During the examination, the patient drinks 500-700 mL warm water with 48g of ultrasonic contrast agent (Xin Zhang®, Huqingyutang Pharmaceutical Company, Hangzhou, China). The stomach body, gastric angle, gastric antrum, pylorus and duodenal bulb were routinely screened when patients are in the standing position, and the cardia and gastric fundus were checked in the supine position. If necessary, take the left, right and semi-recumbent positions as a supplementary examination position. The examiner observes the cardia, the fundus of the stomach, the body of the stomach, the corner of the stomach, the antrum, and the pylorus in turn. If gastric lesions are found, perform local image magnification or use high-frequency probes to observe the hierarchical structure of the stomach wall, the shape of the lesion, size range and its relationship with neighboring organs, etc.

TABLE 1 | Patient characteristics of three cohorts.

Characteristics	Normal Cohort	Benign Lesions Cohort	Gastric Cancer Cohort
Age (Mean ± SD)	53.04 ± 14.96	58.04 ± 18.06	71.43 ± 9.33
Sex			
Male	7	8	5
Female	16	16	2
Total	23	24	7

To be specific, scanning can be divided into five steps:

1. Scan the cardia, ask the patient to lie on his back and move the probe from xiphoid process to the left costal arch.
2. The gastric fundus is scanned by placing the probe in the 10th intercostal space.
3. Scan the cross section of gastric fundus, body and antrum, ask the patient to lie on the right side, and move the probe from the left costal arch along the contour of the stomach.
4. Scan the coronal plane of gastric fundus, body and antrum, ask the patient to lie on the right side, take the probe tail as the fulcrum, rotate the probe along the left rib arch, and tilt the probe 45° at the same time.
5. Scan the gastric antrum and pylorus, ask the patient to lie on his back, and place the probe at the right vertical costal arch.

Methods

The flow chart of the method is shown in **Figure 1**. First, a U-net-based gastric wall detection network is used to detect the region-of-interest (ROI) area of the gastric wall. Then, in the detected ROI area, anisotropic diffusion technology is used to extract the layered structure of the gastric wall. By calculating and counting the thickness of the five-layer structure of the gastric wall, a simple and useful gastric cancer screening model was obtained.

We labeled the regions of gastric wall in 251 gastric ultrasound images, and generated a mask corresponding to every single data as the input of U-net. After training, we will get a model which can automatically identify the ROI of gastric wall.

The gastric wall has five layers, which are mucosa, muscularis mucosa, submucosa, muscularis propria and serosa. The change of thickness ratio of each layer can be used as the basis for diagnosing gastric diseases (**Figure 2**).

The image output from the model is filtered by speckle reduced anisotropic diffusion (SRAD) to make the hierarchical structure of gastric wall more obvious. The edge detection algorithm is used to find the four boundaries among the five levels, and then the five-layer structure is obtained. The ratio of thickness between layers will be calculated, and we can distinguish them by comparing the result of normal gastric wall with that of abnormal gastric wall.

Gastric Wall Detection Based on U-Net

Because the structure of the stomach is fixed and the semantic information is not rich, it is basically the stomach cavity and the stomach wall, so it is necessary to refer to the high-level semantic information and the low-level semantic information in the work of automatic identification of gastric wall. In addition, the data acquisition of medical images is much more difficult than other images, so the model we designed should not be too large because of the small amount of data. Otherwise, too many parameters will easily lead to over fitting and poor prediction effect. Based on the above reasons, we chose U-net structure to establish the model of automatic identification of gastric wall (**Figure 3**). It can meet the needs of referencing low-level and high-level semantic information at the same time, and solve the problem that the amount of data is small and it is difficult to build an accurate model.

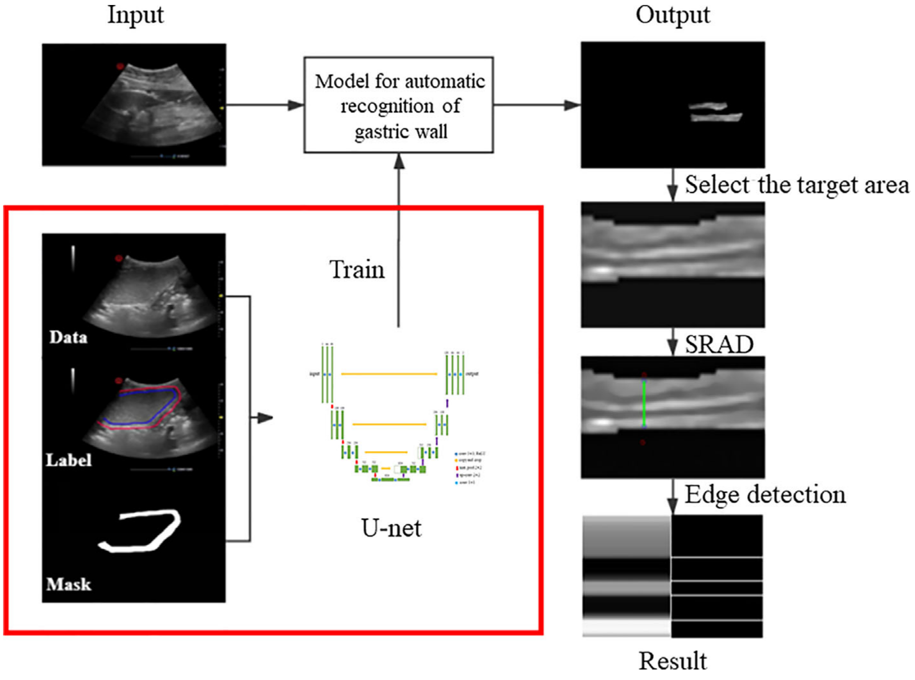


FIGURE 1 | Flow chart of the method.

U-net consists of two paths, the contraction path on the left and the expansion path on the right. In the contraction path, there are 10 times of 3 × 3 convolutions (the size of convolution kernel is 3 × 3), the relu activation layer, and four times of 2 × 2 max pooling. In each down sampling process, the size of the image is reduced, the resolution is reduced, and the number of characteristic channels is doubled. Correspondingly, each step of deconvolution in the expansion path will reduce the number of channels by half, and copy and cross with the previously saved low-level feature map of the same scale. Then, the obtained results are sampled again, and the process is repeated until the image is restored to the original scale. This structure is also called encoder decoder structure. The contraction path corresponds to the encoder, and the expansion path corresponds to the decoder. The encoder part of U-net downsamples 4 times, and the decoder part of u-net upsamples 4 times. The feature image obtained from the down sampling of left encoder is restored to the resolution of the original image. Finally, the final output segmentation image is obtained by softmax (13).

U-net adopts splicing fusion mode, which is completely different from other common segmentation networks such as full convolution network. It stitches the features together in the dimension of channels, which is equivalent to doubling the number of channels to form thicker features. In the case of full convolution network fusion, the corresponding points are added together, and the dimensions will not change, and no thicker features will be formed.

Compared with FCN and deeplab, U-net performs four upsampling, and uses the method which connects the low-level feature map to the high-level feature map. U-net does not

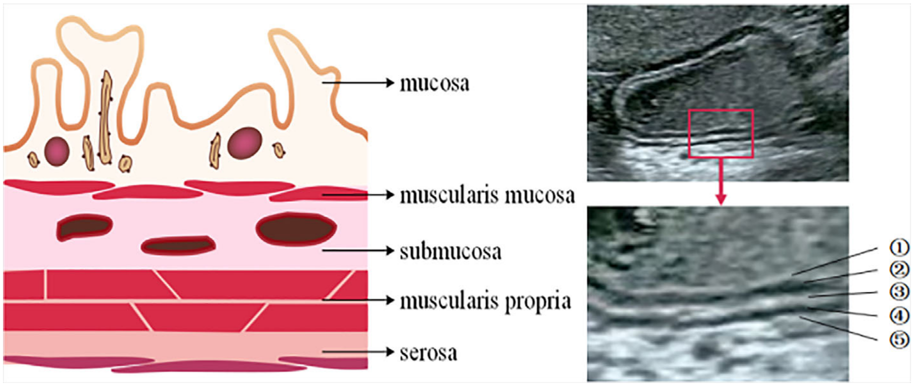


FIGURE 2 | The structure of gastric wall. The gastric wall has five layers.

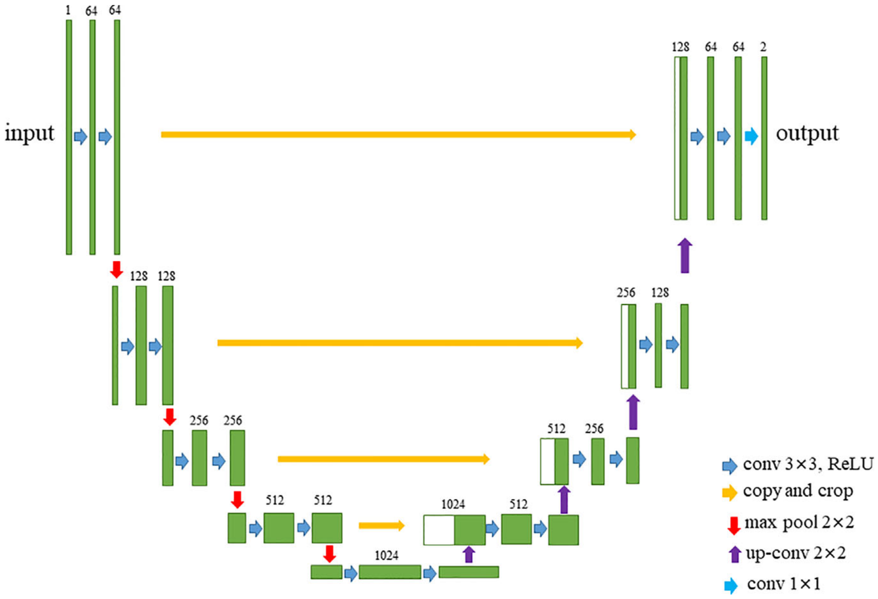


FIGURE 3 | The structure of U-Net.

directly carry out the back propagation of supervision and loss function on the high-level semantic feature map, which not only ensures that the recovered feature map integrates more low-level features, but also makes the features of different scales get fusion and reference, so as to make better prediction and more fine edge information.

Gastric Wall Stratification Based on SRAD (Speckle Reducing Anisotropic Diffusion)

Given an image $I_0(x,y)$ with finite energy and no zero intensity value, the output image $I(x,y;t)$ is obtained by the following partial differential equation:

$$\begin{cases} \frac{\partial I(x,y;t)}{\partial t} = div[c(q) \nabla I(x,y;t)] \\ I(x,y;0) = I_0(x,y), (\frac{\partial I(x,y;t)}{\partial \overrightarrow{\Omega}}) \mid \partial \Omega = 0 \end{cases} \tag{1}$$

It is called the *SRAD PDE*, in the same form as anisotropic diffusion. The diffusion coefficient is defined as:

$$c(q) = \frac{1}{1 + \frac{q^2(x,y;t) - q_0^2(t)}{q_0^2(t)[1+q_0^2(t)]}} \tag{2}$$

or

$$c(q) = \exp[-\frac{q^2(x,y;t) - q_0^2(t)}{q_0^2(t)[1+q_0^2(t)]}] \tag{3}$$

The instantaneous coefficient of variation is defined as:

$$q(x,y;t) = \sqrt{\frac{\frac{1}{2}(\frac{|\nabla I|}{I})^2 - \frac{1}{16}(\nabla^2 I/I)^2}{(1 + \frac{1}{4}\frac{\nabla^2 I}{I})^2}} \tag{4}$$

$q(x,y;t)$ represents the degree of dispersion between pixels, which is large at the edge and small in the homogeneous region. $q_0(t)$ is a speckle scale function.

SRAD encourages isotropic diffusion in homogeneous regions, where q fluctuates and $C(q)$ is about 1. In addition, it is necessary to manually select a homogeneous region to determine the value (14). After SRAD, the edge structure of gastric wall of the image is more clear.

Quantitative Measurement of Gastric Wall

After we get the image of gastric wall, because the gastric wall has five layers of staggered structure, after the SRAD anisotropic diffusion filtering processing, there will be a more obvious difference between the light and the dark, that is, the edge is strengthened by SRAD and become more obvious. In this case, the edge detection algorithm can be used to find out the junction of each layer, and the thickness of each layer is the difference of the positions of each layer. In this paper, Sobel operator is used for edge detection (15).

The following method is proposed to determine whether the patient with gastric ultrasound image has gastric disease:

When the gastric wall of one patient can be divided into five layers, the proportions for the five layers in the entire gastric wall are calculated and combined to be recorded as x which is a vector with 5 dimensions. Then the standard value recorded as s can be obtained by averaging the x for all the patients and we can have $s = (0.278, 0.133, 0.154, 0.154, 0.280)$ in our study. The distance d between s and the x for one patient is defined as following

$$d = \sqrt{(x_1 - s_1)^2 + (x_2 - s_2)^2 + (x_3 - s_3)^2 + (x_4 - s_4)^2} \tag{5}$$

d is the value that we use to determine the situation of gastric wall of one patient. If the value of d is large, it might be regarded as abnormal gastric wall.

RESULTS

Detection Results

In our study, there are three methods of labeling applied for obtaining the ground truth of gastric wall which are as following.

1. Apply one rectangle to label the part of the gastric wall area shown as **Figure 4A**
2. Apply two rectangles trying to cover more parts of the gastric wall area compared with the method with one rectangle, which is shown in **Figure 4B**
3. Label the entire area of gastric wall shown as **Figure 4C**.

The principle of the labeling is to cover the gastric wall as much as possible. However due to the complexity of the ultrasound image representation of the gastric wall in clinical practice, it is not always possible to accurately select the entire region of gastric wall. This is the reason that we apply the above three methods for the labeling in our study. It should be denoted that the labeling results based on all of them are considered as the ground truth of gastric wall for the task of gastric wall detection.

Intersection over Union (IoU) is applied here as the metric to quantitatively evaluate the performance of the proposed model for the gastric detection, which is defined as:

$$IoU = \frac{A \cap B}{A \cup B} \tag{6}$$

where A denotes the detection result and B is the ground truth of gastric wall. The symbol of ∩ denotes the intersection of the two regions and the symbol of ∪ denotes the union of the two regions. The high value of the IoU denotes the good performance of model detection. The detailed results are shown in **Table 2**.

It can be seen that the proposed deep model for the detection of gastric wall demonstrates its relatively effectiveness to some extent. The fact that the largest IoU value comes from the method of labeling the entire area of gastric wall denotes that the proposed model effectively learns to represent the gastric wall. Therefore, the detection result may achieve better IoU when the ground truth is the true region of the gastric wall. It should be also noted that the performance of the detection part should be ultimately evaluated only by the results of the classification to quantitatively determine the situation of the patient as the following section since it is the only purpose of our study. **Figure 5** gives two examples of gastric wall detection based on U-net.

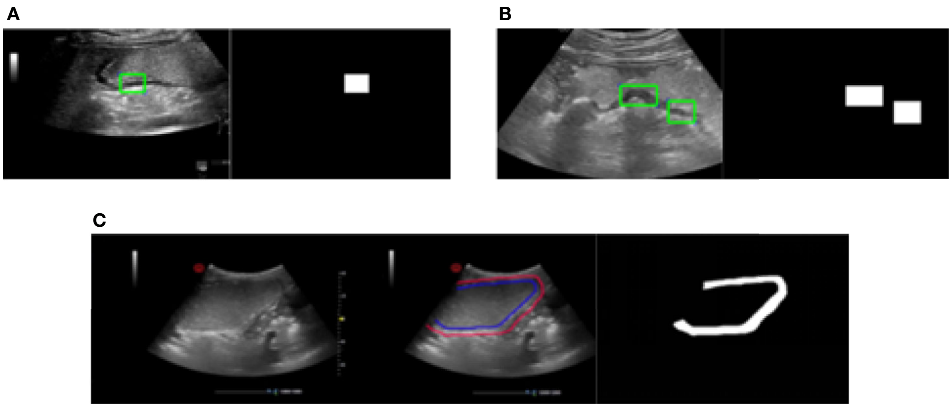


FIGURE 4 | Three methods of labeling. **(A)** Apply one rectangle to label the part of the gastric wall area. **(B)** Apply two rectangles trying to cover more parts of the gastric wall area compared with the method with one rectangle. **(C)** Label the entire area of gastric wall.

TABLE 2 | Gastric wall detection results.

Labeling method	IoU
One rectangle	0.36
Two rectangles	0.32
Label the entire area	0.43

Quantitative Stratification Results

We select a column in the image, corresponding to the green line in the image (**Figure 6A**). The ratio is obtained by edge detection of pixels in the column. Each line corresponds to a column of pixel values (**Figure 6B**). All the white lines in the image represent the position of the edge detected by Sobel edge detection (**Figure 6C**). Four edges of five-layer structure are detected, which is consistent with our expectation. It is easy to get the vertical position of each white line. We can calculate the relative thickness of each layer of structure and the ratio between five layers of structure by making a simple subtraction.

We select a number of such columns and average the results as the final ratio result. In this example, the proportional relationship between the five structures is 0.358: 0.189: 0.116: 0.2: 0.137.

For the three cohorts in our study, we calculated the *d* value defined as Eq. 5 for all the patients. **Figure 7** demonstrates the distributions of *d* values and it can be seen that the difference among normal, benign lesions and gastric cancer is quite obvious.

For the classification of gastric walls among normal, benign lesions and gastric cancer, the corresponding experiments were performed and the results are demonstrated in **Table 3**. We randomly select 70% of the data in the dataset as the training set and 30% as the test set. In the training set, we find the threshold which can best distinguish normal and abnormal cases, and apply this threshold to the test set. Results were validated by quantitative indexes including Accuracy (ACC), Sensitivity (SENS), Specificity (SPEC), Positive Predictive Value (PPV), Negative Predictive Value (NPV), Matthew's Correlation Coefficient (MCC), F1 score and P-value between two target groups. Supposing TP, TN, FP, FN represent true positive, true negative, false positive and false negative, then accuracy (ACC), sensitivity (SENS), specificity (SPEC), positive predictive value (PPV), negative predictive value (NPV), Matthew's correlation coefficient (MCC) and F1 score can be defined and calculated as:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$

$$SENS = \frac{TP}{TP + FN}$$

$$SPEC = \frac{TN}{TN + FP}$$

$$PPV = \frac{TP}{TP + FP}$$

$$NPV = \frac{TN}{TN + FN}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

It can be seen that the proposed quantitative method achieves excellent results for the classification of gastric walls among normal, benign lesions and gastric cancer.

CONCLUSIONS AND DISCUSSION

The main method for gastric cancer diagnosis and screening in developed countries is gastroscopy, which has significantly increased the early detection rate of gastric cancer and

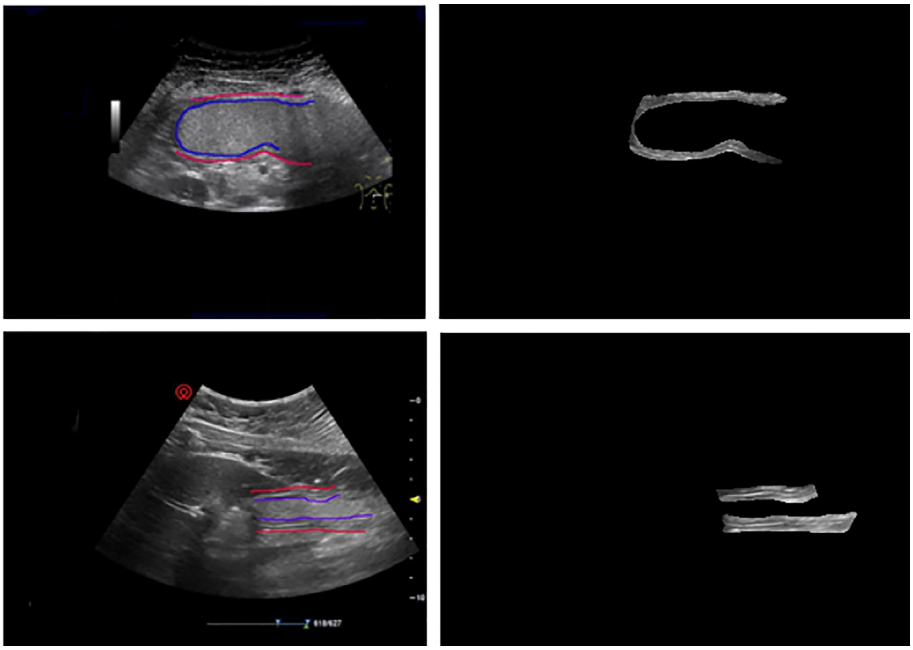


FIGURE 5 | The prediction results of gold standard and our model. After training, the model can accurately predict the gastric wall area in gastric ultrasound images.

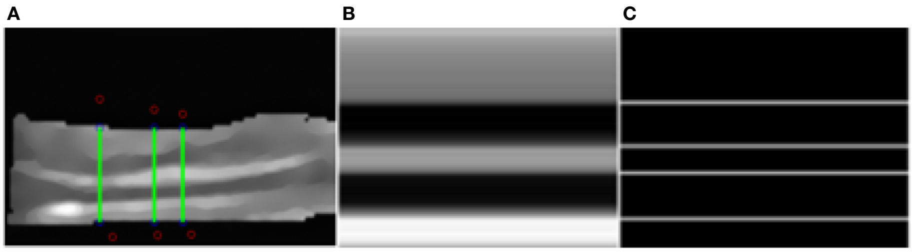


FIGURE 6 | Segmentation results. **(A)** The green lines. **(B)** The pixel value of one line. **(C)** Detected edges.

improved the survival rate of patients. Gastroscopy and gastric mucosal biopsy techniques are also highly recommended in China. However, the early diagnosis rate of gastric cancer is still very poor, with the early diagnosis and treatment rate of gastric cancer is only about 10%. China is a developing country with more than half of the rural population. Not only is there an inherent causal relationship between certain traditional eating habits of rural residents and the occurrence of gastric cancer, but the low self-care awareness of rural population is a relatively high-risk factor for gastric cancer. What is even more unfavorable is that the number of physicians who can perform gastroscopy in rural areas is severely insufficient, and rural residents have low compliance with gastroscopy, resulting in serious practical difficulties in gastroscopy screening for gastric cancer (16). Therefore, it is urgent to explore and establish a gastric cancer screening and diagnosis strategy with Chinese characteristics.

Among many clinical imaging techniques, ultrasound imaging has the advantages of high soft tissue resolution, easy operation, safety and painlessness. In China, the penetration rate of ultrasound equipment is extremely high, and ultrasound equipment at all types of medical institutions must be equipped. For gastric cancer, ultrasound can show the location, size, level of invasion of the stomach wall, and whether there are swollen lymph nodes in the stomach.

The stomach is a hollow organ when without filling, so it is indeed hard to obtain accurate measurement of gastric wall structure by ultrasound. Therefore, we used oral contrast trans-abdominal ultrasonography, which has obvious advantages. The contrast agent fills the stomach cavity to form

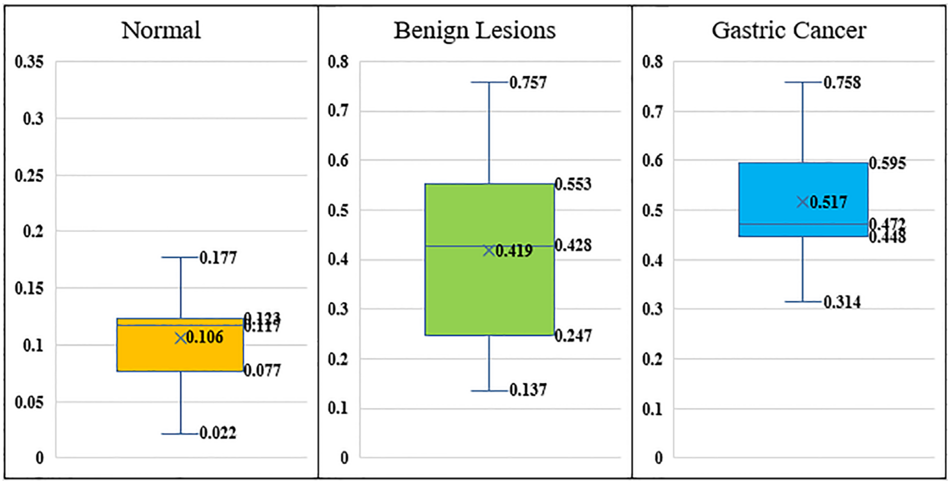


FIGURE 7 | The distributions of *d* values.

TABLE 3 | Classification results based on diagnostic model.

Experiment	AUC	ACC	SENS	SPEC	PPV	NPV	MCC	F1score	P-value
Normal vs Benign Lesions	0.90	0.94	1.00	0.80	0.93	1.00	0.86	0.96	<0.0001
Normal vs Gastric Cancer	0.93	0.92	1.00	0.86	0.83	1.00	0.85	0.91	<0.0001
Normal vs Benign Lesions &Gastric Cancer	0.92	0.95	1.00	0.83	0.93	1.00	0.88	0.97	<0.0001

a high-quality "acoustic window", which is effective in improving the ultrasound imaging ability of the stomach wall structure. For most of the conventional ultrasound diagnostic apparatuses, the imaging quality can improve the hierarchical structure and continuity of the stomach wall, and the recognition ability can meet the detection of most gastric cancer.

We have proposed a U-net based model for automatic recognition of gastric wall region from gastric ultrasound images. We use speckle reduced anisotropic diffusion to make the hierarchical structure of gastric wall more obvious. By dividing the five layers of gastric wall and calculating the ratio of each layer, the normal gastric wall and abnormal gastric wall can be accurately distinguished. This method is also a useful attempt for early screening of gastric cancer. In the existing cases, the detection accuracy of abnormal gastric wall is 95%.

As far as we know, there is no quantitative analysis of the thickness of five layers of gastric wall before. This article is a preliminary report on this aspect. In normal and abnormal cases, the ratio of gastric wall thickness is different. The change of gastric wall thickness often means the occurrence of gastric diseases, and our experimental results have confirmed it. The experimental results have shown that the proposed detection and calculation method of gastric wall is helpful to quantify the diagnosis information of gastric diseases, and is expected to improve the efficiency of ultrasound screening for gastric cancer.

In the future, with the advantages of noninvasive safety, cost-effectiveness, high equipment penetration rate and inspection compliance, gastric ultrasound will play more unique roles and advantages in the diagnosis and screening of gastric cancer.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethic Committee of Xinhua Hospital Affiliated to Shanghai Jiaotong University. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

AS and ZH wrote the article. YD provided the technical guidance. AS and XX made the Figure. YW has made substantial contribution to the work. JY and LS conceived and designed the idea. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by Shanghai science and technology action innovation plan (19441903100).

REFERENCES

1. Kim GH, Liang PS, Bang SJ, Hwang JH. Screening and surveillance for gastric cancer in the United States: Is it needed? *Gastrointest Endosc* (2016) 84(1):18–28. doi: 10.1016/j.gie.2016.02.028
2. Veisani Y, Delpisheh AJ. Survival rate of gastric cancer in Iran; a systematic review and meta-analysis. *Gastroenterol Hepatol Bed Bench* (2016) 9(2):78–86. doi: 10.31557/APJCP.2019.20.1.13
3. Akhondi-Meybodi M, Ghane M, Akhondi-Meybodi S, Dashti G. Five-year Survival Rate for Gastric Cancer in Yazd Province, Central Iran, from 2001 to 2008. *Middle East J Dig Dis* (2017) 9(1):39–48. doi: 10.15171/mejdd.2016.50
4. Modarres S-R, Gholizadeh-Pasha A, Khatibi S, Siadati S. Estimating postoperative survival rate of gastric cancer patients and its effective factors in Babol, northern Iran: 2006-2011. *Caspian J Intern Med* (2017) 8(3):190–5. doi: 10.22088/cjim.8.3.190
5. Hitoshi K, Takashi I, Kohei A, Yoh I, Isao M, Ichiro O, et al. Five-year survival analysis of surgically resected gastric cancer cases in Japan: a retrospective analysis of more than 100,000 patients from the nationwide registry of the Japanese Gastric Cancer Association, (2001-2007). *Gastric Cancer* (2018) 21(1):144–54. doi: 10.1007/s10120-017-0716-7
6. Sugano KJ. Screening of gastric cancer in Asia. *Best Pract Res Clin Gastroenterol* (2015) 29(6):895–908. doi: 10.1016/j.bpg.2015.09.013
7. Ping G, Yanjun G, Lin L, Xiaoxia B, Chunhua Z, Shumin H, et al. Clinical study of serum ELISA combined with gastroscopy in screening early gastric cancer. *Appl mod Drugs China.* (2017) 11(2):19–21. doi: 10.14164/j.cnki.cn11-5581/r.2017.02.008
8. Changfeng M, Yuntao M, Wang X, Hu M, Xiaojun Yang J. Construction of a scoring model for high risk factors of gastric cancer and its value in selection of screening time and scheme. *Chin J Cancer* (2019) 29(7):501–7. doi: 10.19401/j.cnki.1007-3639.2019.07.004
9. Xin J, Jie W. The value of gastric wall thickness in gastric cancer screening Electronic. *J Clin Med Lit* (2015) 2(013):2596–7. doi: 10.16281/j.cnki.jocml.2015.13.103
10. Zhang L, Shen L, Lu D, Chen YJ. Preliminary report on health economic evaluation of oral contrast-enhanced ultrasound combined with gastroscopy in screening gastric cancer. *Surg Theory Pract* (2014) 19(5):464–6. doi: 10.3969/j.issn.1007-9610.2014.05.020
11. Shen L, Zhou C, Liu L, Zhang L, Lu D, Cai J, et al. Application of oral contrast trans-abdominal ultrasonography for initial screening of gastric cancer in rural areas of China. *Dig Liver Dis* (2017) 49(8):918–23. doi: 10.1016/j.dld.2017.04.008
12. Liu Z, Guo J, Wang S, Zhao Y, Li J, Ren W, et al. Evaluation of transabdominal ultrasound after oral administration of an echoic cellulose-based gastric ultrasound contrast agent for gastric cancer. *BMC Cancer* (2015) 15(1):932. doi: 10.1186/s12885-015-1943-0
13. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Med Image Comput Comput Assisted Intervention Munich* (2015) 234–41. doi: 10.1007/978-3-319-24574-4_28
14. Yu Y, Acton ST. Speckle reducing anisotropic diffusion. *IEEE Trans Image Process* (2002) 11(11):1260–70. doi: 10.1109/TIP.2002.804276
15. Zhang Y, Han X, Zhang H, Zhao L. Edge detection algorithm of image fusion based on improved Sobel operator. *IEEE 3rd Information Technology and Mechatronics Engineering Conference (ITOEC).* Chongqing. (2017) 457–61. doi: 10.1109/ITOEC.2017.8122336
16. Wang FH, Shen L, Li J, Zhou ZW, Liang H, Zhang XT, et al. The Chinese Society of Clinical Oncology (CSCO): clinical guidelines for the diagnosis and treatment of gastric cancer. *Cancer Commun (Lond)* (2019) 39(1):10. doi: 10.1186/s40880-019-0349-9

ACKNOWLEDGMENTS

We would like to thank all of the patients for their participation in this study.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

MINI REVIEW
published: 10 June 2021
doi: 10.3389/fonc.2021.631813

Articles That Use Artificial Intelligence for Ultrasound: A Reader's Guide

Ming Kuang[1,2*†], Hang-Tong Hu[1†], Wei Li[1], Shu-Ling Chen[1] and Xiao-Zhou Lu[3]

[1] Department of Medical Ultrasonics, Ultrasomics Artificial Intelligence X-Lab, Institute of Diagnostic and Interventional Ultrasound, The First Affiliated Hospital of Sun Yat-Sen University, Guangzhou, China, [2] Department of Hepatobiliary Surgery, The First Affiliated Hospital of Sun Yat-sen University, Guangzhou, China, [3] Department of Traditional Chinese Medicine, The First Affiliated Hospital of Sun Yat-sen University, Guangzhou, China

Artificial intelligence (AI) transforms medical images into high-throughput mineable data. Machine learning algorithms, which can be designed for modeling for lesion detection, target segmentation, disease diagnosis, and prognosis prediction, have markedly promoted precision medicine for clinical decision support. There has been a dramatic increase in the number of articles, including articles on ultrasound with AI, published in only a few years. Given the unique properties of ultrasound that differentiate it from other imaging modalities, including real-time scanning, operator-dependence, and multi-modality, readers should pay additional attention to assessing studies that rely on ultrasound AI. This review offers the readers a targeted guide covering critical points that can be used to identify strong and underpowered ultrasound AI studies.

Keywords: ultrasound, artificial intelligence, machine learning, deep learning, radiomics

OPEN ACCESS

Edited by:
Hui-Xiong Xu,
Tongji University, China

Reviewed by:
Jun Shi,
Shanghai University, China
Xin-Wu Cui,
Huazhong University of Science and Technology, China

***Correspondence:**
Ming Kuang
kuangm@mail.sysu.edu.cn

†These authors share first authorship

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: 21 November 2020
Accepted: 12 May 2021
Published: 10 June 2021

Citation:
Kuang M, Hu H-T, Li W, Chen S-L and Lu X-Z (2021) Articles That Use Artificial Intelligence for Ultrasound: A Reader's Guide. Front. Oncol. 11:631813. doi: 10.3389/fonc.2021.631813

INTRODUCTION

By looking into pixels not readily visible to the human naked eyes, artificial intelligence (AI) has led medical imaging into the era of big data (1). Articles using conventional machine learning (ML) algorithms and deep learning, especially convolutional neural networks (CNN), have also become more numerous over the past several years. Studies have reported the use of AI in X-rays, computerized tomography (CT), magnetic resonance imaging (MRI), ultrasound, and other types of scans, and they have reported superior performance of AI to that of conventional methods in disease detection, characterization, and patient prognosis prediction (2–4).

Working groups of the Consolidated Standards of Reporting Trials-Artificial Intelligence (CONSORT-AI) and the Standard Protocol Items: Recommendations for Interventional Trials-Artificial Intelligence (SPIRIT-AI) have developed an extension to the core CONSORT 2010 items and 2013 SPIRIT statement that serves as a guidance for medical AI studies (5, 6). Given the rapid expansion of the literature published, *JAMA* has provided a reader's guide to assessing clinical AI articles (7), which reviewed the basics of machine learning and aspects of the clinical implementation of AI. The editorial board of *Radiology* also highlighted several crucial considerations meant to formalize AI methodology in medical imaging studies (8). However, when AI is used with ultrasound, issues become complicated for the current existing guides.

Ultrasound uses the reflection of the ultrasonic beam to reveal tissue structure. It is one of the most widely used methods of imaging in clinical practice. It serves as a mainstay in obstetricians, cardiology, interventional therapy guidance and post-treatment surveillance (9). Ultrasound-based radiomics studies, called ultrasomics (10), follow the standard three-step AI process for medical imaging: data preparation, model development and testing, and evaluation of clinical effectiveness (11). However, given ultrasound's unique properties of real-time scanning, operator-dependence, and multi-modality, some specific issues may influence the performance of AI models and the generalizability of a study's results. For example, operator dependence may influence the use of expert-dataset-based model training to the resident-dataset-based model testing and use in primary hospitals. In this minireview, we aim to provide the readers with an overview of how to assess medical imaging AI articles, including some specific points regarding ultrasound AI studies.

OBJECTIVE: IS THE CLINICAL SCENARIO CLEARLY DEFINED?

The objective of a medical imaging AI study should comply with two principles: first, it must be derived from clinical practical needs, and second, it must be applicable to AI technique. For example, un-enhanced ultrasound is recommended for monitoring populations at high risk of liver cancer (12), so it would be a risk stratification tool. An unenhanced ultrasound AI tool would ideally increase the detection rate of liver lesions and assist in risk assessment. When transformed into AI tasks, target recognition and classification are both technically feasible.

MATERIALS AND METHODS: IS THERE AN INDEPENDENT TESTING DATASET BESIDES THE TRAINING AND VALIDATION SETS?

AI models are prone to overfitting. Both conventional ML and CNN algorithms can vary greatly in performance across different data sources (13). After a model is trained using the training set, its hyperparameters must be tuned in the validation set (also called the tuning set) for better generalizability. If multiple models had been trained, the validation set could also be used to select models. Once a model is finalized, its performance must be evaluated in a testing set, which has no overlap with the training or validation sets. Ideally, the testing set comes from other centers, which involves data from different ultrasound devices and vendors, and patients with different demographic characteristics. A study that reports generalizable results in an independent testing dataset would be much more valuable than a study that relies on internal validation or single-dataset-based cross-validation.

MATERIALS AND METHODS: IS THE IMAGE PROCESSING PROCEDURE CLEARLY DESCRIBED?

A clear description of the image processing procedure is vital for the assessment of study repeatability and reproducibility. Readers should pay attention to the ultrasound data acquisition process and the validity of the data range. Questions below should be raised when acquiring such information. Is the data collected retrospectively or prospectively? Which modality does the study apply? Is it radio frequency signal, grayscale, elastography, doppler imaging, contrast-enhanced ultrasound (CEUS), or transferring between modalities (14, 15)? Also, the number of pictures per patient enrolled for the training or testing and whether the patients' clinical data are involved in the AI development should be inspected.

In terms of ultrasound data preprocessing, each step should be presented clearly. Ultrasound images are derived from various devices produced by different radiologists. Ultrasound is highly operator-dependent (16, 17), which causes variations in image quality, target lesion identification, and selection of representative sections. Cropping is widely adopted in image processing in medical AI studies, and it filters out most irrelevant, non-lesion information, and for the ultrasound, reduces image heterogeneity by adjusting size and depth. Augmentation can enrich data diversity, and it can simulate the common causes of image heterogeneity as observed under real-world conditions in ultrasound examinations (18, 19). For example, resizing reduces resolution variation of different devices, rotation simulates scanning from different angles and sections, and contrast adjustment simulates variation in gain and dynamic range.

MATERIALS AND METHODS: IS THE ALGORITHM FOR MODELING SUITABLE?

Conventional ML algorithms such as logistic regression, support vector machine (SVM), random forest, and Naïve Bayes have much fewer parameters than deep learning algorithms. For example, SVM has only 13 parameters to be adjusted, while the ResNet-50 has an amount of 2.3×10^7 parameters. Thus, conventional ML algorithms require far less training than deep learning algorithms do (20). With a limited sample size, such as a set of only hundreds of images (not videos), conventional ML algorithms are preferred (21). However, with thousands or millions of images, deep learning algorithms, principally CNN in imaging analysis, are recommended. The minimum number of training images needed varies across different tasks and algorithms and may only be determined by evaluating the relationship between its increase and changes in model performance.

Algorithms' clinical intelligibility, which means the level of understandability of an algorithm in a clinical way, should also be considered. There has not been any ultrasound-specific imaging analysis algorithms reported. Instead, model algorithm selection is primarily based on the type of task. Ultrasound has multiple modalities. CEUS videos record a lesion's hemodynamic information revealed by the dynamic perfusion of microbubble contrast agents. Multi-phase image features can be extracted by simply analyzing frames from each phase but the time sequencing features were missing. Recurrent neural networks (RNNs) such as long short-term memory (LSTM) or gated recurrent units can be incorporated to these time-dimension-related tasks (18). Previous studies using LSTM in CEUS reported excellent performance (22, 23). The application of clinically explicable AI algorithms to modeling renders the study findings more clinically acceptable.

MATERIALS AND METHODS: IS THE AI ALGORITHM PUBLICLY AVAILABLE?

Even being generalizable among different datasets in a given study, especially for studies carried out in a single center, AI performance still needs a broad verification. The existing public medical imaging data sets are minimal (24), and no public ultrasound dataset exists. Authors are encouraged to make their AI models publicly available *via* such websites as GitHub (https://github.com/) to allow independent validation, fine-tuning, and updating. A study reporting publicly available AI algorithms may improve its results' reliability in this way.

RESULTS: HOW DO THE RESULTS PRODUCED BY THE AI MODEL COMPARE TO THOSE PRODUCED BY EXPERT RADIOLOGISTS?

Medical AI must be evaluated against the performance of radiology experts (8). The value of a prospectively designed AI performance testing procedure can be determined by comparing its performance to that of human experts under real-world conditions. In retrospectively designed studies, missing data, and data mismatch regarding the target lesion are unavoidable in datasets collected from clinical practice, considering which is beyond AI's ability (25). Radiologists make ultrasound diagnosis in real time during face-to-face examinations, where they receive far more information than retrospective image review does. The common study design usually underestimates radiologists' performance and renders meaningful evaluation of medical AI difficult.

Combing clinician experience and AI's advantages can render imaging more efficient and accurate than either alone (26). Because ultrasound offers diagnosis in real time and is heavily dependent on the operator, ultrasound AI's performance should be compared to that of radiologists with varied experiences to develop a viable human-AI interaction strategy (27). Ideally, this strategy would involve dynamic assessment during an ultrasound examination. A specific application scenario based AI developing and testing study would have considerable practical value.

RESULTS: ARE THE EVALUATION INDEXES SUITABLE?

For detection and classification purposes, an AI model is first evaluated by the receiver operating characteristic curve (ROC) or precision-recall curve (PRC), and further by its accuracy, error rate or F1 value. However, in medical imaging analysis programs, performance is assessed based on indicators of clinical significance, such as sensitivity and specificity for diagnosis and prediction programs (28, 29), detection rate for disease screening and lesion detection (30, 31), κ and dice coefficient for inter-annotator agreement and overlapping in radiotherapy planning (32, 33). For example, for a screening task model, detection rate and sensitivity would be the primary indexes for model evaluation, while for diagnostic tasks, high specificity or positive predictive value would be the top priority. A specifically preferred high evaluation index can be achieved using an appropriate cutoff value for AI outputs but not necessarily by the default of 0.5 or the Youden index.

DISCUSSION: ARE THE RESULTS COMPARED TO STATE-OF-ART REPORTS?

AI results should be compared to state-of-art reports, both the previous studies of the same design and these using other imaging modalities, traditional methods, or guideline recommendations. Readers should keep in mind that results without independent tests or internally validated results are not comparable to studies reporting independently tested results, no matter how good the statistics are relative to state-of-art results. A well-designed study with practical results is much more valuable than studies with flawed design but with good statistical results.

DISCUSSION: WHAT IS THE UNSOLVED PROBLEM OF THE PRESENT WORK?

Limitations of medical AI studies are often the challenge of future work. For example, what situation wouldn't the AI system be implemented when considering that AI performance errors and failure cases could influence clinical practice decision-making? What are the latent factors keeping AI systems from generalizing to other centers and populations, given the hardware requirements, algorithm versions, data quality, and processing procedures? How can these be solved in further study? Is the sample size large enough to build a robust model? The relationship between the training dataset size and model performance should be evaluated, as Dunnmon et al. (34) in the research reporting that the AI performance benefited little after a certain number of images were used for training.

CONCLUSION

Given ultrasound's unique properties, readers should pay additional attention when assessing an AI study that relies on ultrasound than those that rely on other imaging modalities. Here, we list several crucial points to help readers distinguish strong ultrasound AI articles from underpowered articles. With more formalized standards for medical AI studies published in the future, ultrasound AI studies may better benefit the clinical practice.

AUTHOR CONTRIBUTIONS

Conception, design, and final approval of the manuscript, MK. Preparing the main manuscript, H-TH. Editing and review of the manuscript, WL, S-LC. Revision of the manuscript, X-ZL. All authors contributed to the article and approved the submitted version.

FUNDING

This work is supported by the National Natural Science Foundation of China (NO. 81971630).

ACKNOWLEDGMENTS

We thank LetPub (www.letpub.com) for its linguistic assistance during the preparation of this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found onlineat: https://www.frontiersin.org/articles/10.3389/fonc.2021.631813/full#supplementary-material

REFERENCES

1. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial Intelligence in Radiology. *Nat Rev Cancer* (2018) 18(8):500–10. doi: 10.1038/s41568-018-0016-5
2. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: Extracting More Information From Medical Images Using Advanced Feature Analysis. *Eur J Cancer (Oxford Engl 1990)* (2012) 48(4):441–6. doi: 10.1016/j.ejca.2011.11.036
3. Carin L, Pencina MJ. On Deep Learning for Medical Image Analysis. *Jama* (2018) 320(11):1192–3. doi: 10.1001/jama.2018.13316
4. Dong Y, Wang QM, Li Q, Li LY, Zhang Q, Yao Z, et al. Preoperative Prediction of Microvascular Invasion of Hepatocellular Carcinoma: Radiomics Algorithm Based on Ultrasound Original Radio Frequency Signals. *Front Oncol* (2019) 9:1203. doi: 10.3389/fonc.2019.01203
5. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK. Reporting Guidelines for Clinical Trial Reports for Interventions Involving Artificial Intelligence: The CONSORT-AI Extension. *Nat Med* (2020) 26(9):1364–74. doi: 10.1038/s41591-020-1034-x
6. Cruz Rivera S, Liu X, Chan AW, Denniston AK, Calvert MJ. Guidelines for Clinical Trial Protocols for Interventions Involving Artificial Intelligence: The SPIRIT-AI Extension. *Nat Med* (2020) 26(9):1351–63. doi: 10.1038/s41591-020-1037-7
7. Liu Y, Chen PC, Krause J, Peng L. How to Read Articles That Use Machine Learning: Users' Guides to the Medical Literature. *Jama* (2019) 322(18):1806–16. doi: 10.1001/jama.2019.16489
8. Bluemke DA, Moy L, Bredella MA, Ertl-Wagner BB, Fowler KJ, Goh VJ, et al. Assessing Radiology Research on Artificial Intelligence: A Brief Guide for Authors, Reviewers, and Readers-From the Radiology Editorial Board. *Radiology* (2019) 294(3):487–9. doi: 10.1148/radiol.2019192515
9. Muse ED, Topol EJ. Guiding Ultrasound Image Capture With Artificial Intelligence. *Lancet (London England)* (2020) 396(10253):749. doi: 10.1016/S0140-6736(20)31875-4
10. Li W, Huang Y, Zhuang BW, Liu GJ, Hu HT, Li X, et al. Multiparametric Ultrasomics of Significant Liver Fibrosis: A Machine Learning-Based Analysis. *Eur Radiol* (2019) 29(3):1496–506. doi: 10.1007/s00330-018-5680-z
11. Yin R, Jiang M, Lv WZ, Jiang F, Li J, Hu B, et al. Study Processes and Applications of Ultrasomics in Precision Medicine. *Front Oncol* (2020) 10:1736. doi: 10.3389/fonc.2020.01736
12. Morgan TA, Maturen KE, Dahiya N, Sun MRM, Kamaya AAmerican College of Radiology Ultrasound Liver I, et al. Us LI-RADS: Ultrasound Liver Imaging Reporting and Data System for Screening and Surveillance of Hepatocellular Carcinoma. *Abdominal Radiol* (2018) 43(1):41–55. doi: 10.1007/s00261-017-1317-y
13. De Fauw J, Ledsam JR, Romera-Paredes B, Nikolov S, Tomasev N, Blackwell S, et al. Clinically Applicable Deep Learning for Diagnosis and Referral in Retinal Disease. *Nat Med* (2018) 24(9):1342–50. doi: 10.1038/s41591-018-0107-6
14. Han X, Wang J, Zhou W, Chang C, Ying S, Shi J. Deep Doubly Supervised Transfer Network for Diagnosis of Breast Cancer with Imbalanced Ultrasound Imaging Modalities. In: A.L Martel, et al. (eds) *Medical Image Computing and Computer Assisted Intervention - MICCAI 2020. MICCAI 2020. Lecture Notes in Computer Science*. 12266 doi: 10.1007/978-3-030-59725-2_14
15. Fei X, Shen L, Ying S, Cai Y, Zhang Q, Kong W, et al. Parameter Transfer Deep Neural Network for Single-Modal B-Mode Ultrasound-Based Computer-Aided Diagnosis. *Cogn Comput* (2020) 12(6):1252–64. doi: 10.1007/s12559-020-09761-1
16. Todsen T, Tolsgaard MG, Olsen BH, Henriksen BM, Hillingsø JG, Konge L, et al. Reliable and Valid Assessment of Point-of-Care Ultrasonography. *Ann Surg* (2015) 261(2):309–15. doi: 10.1097/SLA.0000000000000552
17. Chou R, Cuevas C, Fu R, Devine B, Wasson N, Ginsburg A, et al. Imaging Techniques for the Diagnosis of Hepatocellular Carcinoma: A Systematic Review and Meta-Analysis. *Ann Internal Med* (2015) 162(10):697–711. doi: 10.7326/M14-2509
18. Ramachandram D, Taylor GW. Deep Multimodal Learning: A Survey on Recent Advances and Trends. *IEEE Signal Process Magazine* (2017) 34(6):96–108. doi: 10.1109/MSP.2017.2738401
19. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A Guide to Deep Learning in Healthcare. *Nat Med* (2019) 25(1):24–9. doi: 10.1038/s41591-018-0316-z
20. Bi WL, Hosny A, Schabath MB, Giger ML, Birkbak NJ, Mehrtash A, et al. Artificial Intelligence in Cancer Imaging: Clinical Challenges and Applications. *CA: Cancer J Clin* (2019) 69(2):127–57. doi: 10.3322/caac.21552
21. Zhang H, Guo L, Wang D, Wang J, Bao L, Ying S, et al. Multi-Source Transfer Learning Via Multi-Kernel Support Vector Machine Plus for B-Mode Ultrasound-Based Computer-Aided Diagnosis of Liver Cancers. *IEEE J Biomed Health Inf* (2021). PP(99):1–1. doi: 10.1109/JBHI.2021.3073812
22. Azizi S, Bayat S, Yan P, Tahmasebi A, Kwak JT, Xu S, et al. Deep Recurrent Neural Networks for Prostate Cancer Detection: Analysis of Temporal Enhanced Ultrasound. *IEEE Trans Med Imaging* (2018) 37(12):2695–703. doi: 10.1109/TMI.2018.2849959
23. Sharma H, Droste R, Chatelain P, Drukker L, Papageorghiou AT, Noble JA. Spatio-Temporal Partitioning and Description of Full-Length Routine Fetal Anomaly Ultrasound Scans. *Proc IEEE Int Symposium Biomed Imaging* (2019) 16:987–90. doi: 10.1109/ISBI.2019.8759149
24. Soffer S, Ben-Cohen A, Shimon O, Amitai MM, Greenspan H, Klang E. Convolutional Neural Networks for Radiologic Images: A Radiologist's Guide. *Radiology* (2019) 290(3):590–606. doi: 10.1148/radiol.2018180547

25. Doshi-Velez F, Perlis RH. Evaluating Machine Learning Articles. *Jama* (2019) 322(18):1777–9. doi: 10.1001/jama.2019.17304
26. Steiner DF, MacDonald R, Liu Y, Truszkowski P, Hipp JD, Gammage C, et al. Impact of Deep Learning Assistance on the Histopathologic Review of Lymph Nodes for Metastatic Breast Cancer. *Am J Surg Pathol* (2018) 42(12):1636–46. doi: 10.1097/PAS.0000000000001151
27. Moga TV, Popescu A, Sporea I, Danila M, David C, Gui V, et al. Is Contrast Enhanced Ultrasonography a Useful Tool in a Beginner's Hand? How Much Can a Computer Assisted Diagnosis Prototype Help in Characterizing the Malignancy of Focal Liver Lesions? *Med Ultrasonography* (2017) 19(3):252–8. doi: 10.11152/mu-936
28. Wang K, Lu X, Zhou H, Gao Y, Zheng J, Tong M, et al. Deep Learning Radiomics of Shear Wave Elastography Significantly Improved Diagnostic Performance for Assessing Liver Fibrosis in Chronic Hepatitis B: A Prospective Multicentre Study. *Gut* (2019) 68(4):729–41. doi: 10.1136/gutjnl-2018-316204
29. Eun NL, Kang D, Son EJ, Park JS, Youk JH, Kim JA, et al. Texture Analysis With 3.0-T MRI for Association of Response to Neoadjuvant Chemotherapy in Breast Cancer. *Radiology* (2020) 294(1):31–41. doi: 10.1148/radiol.2019182718
30. Wang P, Berzin TM, Glissen Brown JR, Bharadwaj S, Becq A, Xiao X, et al. Real-Time Automatic Detection System Increases Colonoscopic Polyp and Adenoma Detection Rates: A Prospective Randomised Controlled Study. *Gut* (2019) 68(10):1813–9. doi: 10.1136/gutjnl-2018-317500
31. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, van Ginneken B, Karssemeijer N, Litjens G, et al. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. *Jama* (2017) 318(22):2199–210. doi: 10.1001/jama.2017.14580
32. Oktay O, Nanavati J, Schwaighofer A, Carter D, Bristow M, Tanno R, et al. Evaluation of Deep Learning to Augment Image-Guided Radiotherapy for Head and Neck and Prostate Cancers. *JAMA Network Open* (2020) 3(11): e2027426. doi: 10.1001/jamanetworkopen.2020.27426
33. Liang S, Tang F, Huang X, Yang K, Zhong T, Hu R, et al. Deep-Learning-Based Detection and Segmentation of Organs at Risk in Nasopharyngeal Carcinoma Computed Tomographic Images for Radiotherapy Planning. *Eur Radiol* (2019) 29(4):1961–7. doi: 10.1007/s00330-018-5748-9
34. Dunnmon JA, Yi D, Langlotz CP, Ré C, Rubin DL, Lungren MP. Assessment of Convolutional Neural Networks for Automated Classification of Chest Radiographs. *Radiology* (2019) 290(2):537–44. doi: 10.1148/radiol.2018181422

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 07 October 2021
doi: 10.3389/fonc.2021.611436

A Computer-Aided Diagnosis System and Thyroid Imaging Reporting and Data System for Dual Validation of Ultrasound-Guided Fine-Needle Aspiration of Indeterminate Thyroid Nodules

Xiaowen Liang[1†], *Yingmin Huang*[1†], *Yongyi Cai*[2], *Jianyi Liao*[1] and *Zhiyi Chen*[1,3*]

[1] Department of Ultrasound Medicine, The Third Affiliated Hospital of Guangzhou Medical University, Guangzhou, China, [2] Department of Ultrasound, Liwan Center Hospital of Guangzhou, Guangzhou, China, [3] The First Affiliated Hospital, Medical Imaging Centre, Hengyang Medical School, University of South China, Hengyang, China

Edited by:
Hui-Xiong Xu,
Tongji University, China

Reviewed by:
Romany Mansour,
The New Valley University, Egypt
Xiao-Wan Bo,
Tongji University, China

***Correspondence:**
Zhiyi Chen
zhiyi_chen@gzhmu.edu.cn

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *29 September 2020*
Accepted: *16 September 2021*
Published: *07 October 2021*

Citation:
Liang X, Huang Y, Cai Y, Liao J and Chen Z (2021) A Computer-Aided Diagnosis System and Thyroid Imaging Reporting and Data System for Dual Validation of Ultrasound-Guided Fine-Needle Aspiration of Indeterminate Thyroid Nodules. Front. Oncol. 11:611436. doi: 10.3389/fonc.2021.611436

Purpose: The fully automatic AI-Sonic computer-aided design (CAD) system was employed for the detection and diagnosis of benign and malignant thyroid nodules. The aim of this study was to investigate the efficiency of the AI-Sonic CAD system with the use of a deep learning algorithm to improve the diagnostic accuracy of ultrasound-guided fine-needle aspiration (FNA).

Methods: A total of 138 thyroid nodules were collected from 124 patients and diagnosed by an expert, a novice, and the Thyroid Imaging Reporting and Data System (TI-RADS). Diagnostic efficiency and feasibility were compared among the expert, novice, and CAD system. The application of the CAD system to enhance the diagnostic efficiency of novices was assessed. Moreover, with the experience of the expert as the gold standard, the values of features detected by the CAD system were also analyzed. The efficiency of FNA was compared among the expert, novice, and CAD system to determine whether the CAD system is helpful for the management of FNA.

Result: In total, 56 malignant and 82 benign thyroid nodules were collected from the 124 patients (mean age, 46.4 ± 12.1 years; range, 12–70 years). The diagnostic area under the curve of the CAD system, expert, and novice were 0.919, 0.891, and 0.877, respectively ($p < 0.05$). In regard to feature detection, there was no significant differences in the margin and composition between the benign and malignant nodules ($p > 0.05$), while echogenicity and the existence of echogenic foci were of great significance ($p < 0.05$). For the recommendation of FNA, the results showed that the CAD system had better performance than the expert and novice ($p < 0.05$).

Conclusions: Precise diagnosis and recommendation of FNA are continuing hot topics for thyroid nodules. The CAD system based on deep learning had better accuracy and feasibility for the diagnosis of thyroid nodules, and was useful to avoid unnecessary FNA.

The CAD system is potentially an effective auxiliary approach for diagnosis and asymptomatic screening, especially in developing areas.

Keywords: thyroid nodule, ultrasound, computer-aided diagnosis, TI-RADS, fine-needle aspiration

INTRODUCTION

Approximately 95% of endocrine cancers involve the thyroid, which contributes to the continually increasing incidence of thyroid cancer (1, 2). Ultrasound (US) is widely used as a non-invasive and effective screening modality for the detection of thyroid nodules. However, although various diagnostic standards and guidelines are available, the diagnostic specificity remains unsatisfactory (3). Since first proposed in 2009, the Thyroid Imaging Reporting and Data System (TI-RADS) has become among the most widely applied approaches for US-based diagnosis of thyroid diseases, as this system allows the physician to analyze the composition, echogenicity, shape, orientation, margin, calcification, presence of a halo, and type of vascularization for more accurate diagnosis and treatment (4). In 2017, the American College of Radiology launched the final version of the TI-RADS, which uses a scoring method to optimize and standardize US-guided fine-needle aspiration (FNA) (5). However, this modality is still limited by subjectivity and inconsistencies when applied clinically. The specificity of FNA for the detection of thyroid nodules is reportedly only 60–70%, suggesting a high occurrence of non-diagnostic results (6, 7). Moreover, since there is currently no consensus on the standardization of US-guided FNA (8), it is difficult to unify all physicians to follow the same diagnostic criteria, even within the same department or hospital. The use of TI-RADS has changed the focus of diagnostic subjectivity to the detection of features, which can decrease but not completely avoid intra-observer variability (9). Therefore, a consistently effective and accurate method for the diagnosis of thyroid nodules with good repeatability is urgently needed.

Over the past decade, artificial intelligence (AI)-aided US techniques, which integrate US and computer science, have become increasingly employed for the detection and diagnosis of thyroid diseases (10, 11). From traditional machine learning to deep learning methods, many algorithms, such as the Support Vector Machine (12), GoogleNet (13), and a convolutional neural network (14), have been shown to be effective for US-based diagnosis of thyroid nodules. These advanced methods have become increasingly used in recent years due to advances in commercial software applications, such as AmCad-UT (AmCad BioMed Corporation, Taipei City, Taiwan) (15), S-Detect (Samsung Medison Co., Ltd., Seoul, Korea) (3), and AI-SONIC (Demetics Medical Technology, Zhejiang, China). AI-SONIC is a fully automatic diagnosis system based on deep learning, which includes a training set consisting of more than 60,000 US images of the thyroid. A cascade convolutional neural network (CNN) is a type of hybrid deep learning model based on a special splitting method and two different CNN architectures (one with 15 convolutional layers and two pooling layers for segmentation, and another with four convolutional layers and four pooling layers for detection). In this study, 10-fold cross-validation was performed with a training set, validation set, and testing set ratio of 8:1:1. The detection and diagnostic efficiency of AI-SONIC are reportedly very good, as evidenced by a Dice score of 0.9224 with a diagnostic area under the curve (AUC) of 0.98 (16, 17). Besides the excellent diagnostic accuracy, other advantages of this software include an automatic recommendation of a follow-up plan, including FNA, and the provision of detailed information about the features of the thyroid detected by the software itself to render the recommendations more understandable and acceptable. However, as far as we know, there is no report on the interpretability of CAD in thyroid ultrasound diagnosis, and the ability of CAD in improving novices' diagnostic efficiency based on AI-SONIC needs to be proved.

In our center, a dual-verification process is used to improve the reliability of the results. In order to validate the use of this approach in clinical practice in areas lacking medical resources, the diagnostic efficiency was compared among an expert, a novice, and a computer-aided diagnosis (CAD) system. Furthermore, the relationship between the pathology results and US features detected by the CAD system were analyzed to determine whether the CAD system can improve the efficiency of US-guided FNA.

MATERIALS AND METHODS

Patients

A total of 138 thyroid nodules were collected from 124 patients who received treatment at the Third Affiliated Hospital of Guangzhou Medical University and Liwan Center Hospital of Guangzhou from January 2016 to May 2019. The final diagnosis of the thyroid nodules was confirmed with the use of specimens collected during surgery or FNA. Detailed patient data was retrieved from the Hospital Information Manage System. The inclusion criteria were (1) complete and clear US images before surgery or FNA and (2) the inclusion of a single nodule in a two-dimensional US image, while the exclusion criteria were (1) a lack of pathology results, (2) a history of partial thyroidectomy, and (3) any complication of a diffuse thyroid disease. The study protocol was approved by the local ethics committee, and written informed consent was obtained from all patients prior to study inclusion.

Equipment and Data Analysis

The US examinations were performed with the use of different brands of equipment, which included the Phillips IU22/IE33/CX50 systems (Philips Healthcare, Eindhoven, Netherlands), Hitachi Hi Vision Preirus/Ascendus systems (Hitachi Ltd., Tokyo, Japan), GE Logiq E9/S6/S8/E6/E8 systems (GE Healthcare, Milwaukee, WI, USA), Siemens S1000/S2000 systems (Siemens Healthineers, Munich, German), Resona 7 (Mindary, Shenzhen, China), and Aplio 300/500 systems (Toshiba Corporation, Tokyo, Japan). An expert with 8 years of experience and a novice with 1 year of experience in US examinations of the thyroid reviewed the nodules and diagnosis recommended by the TI-RADS, respectively, and recorded the final score of the nodules. To ensure the validity of the assessments, the score of the TI-RADS as judged by the expert was confirmed by another expert with 15 years of experience.

Detection and Diagnosis With the CAD System

The CAD software was loaded into a computer, and the US images were uploaded into the software. After confirmation that there was only one nodule on the center of the screen, the software could detect the nodule automatically and quantitatively analyzed the margin, composition, echogenicity, and the existence of echogenic foci (**Figure 1**). This process only took 1–2 s. If the CAD system failed to identify a nodule, the physician defined the region of interest manually and the analysis was repeated. The probability of a malignant *versus* benign nodule and a recommendation of a follow-up were provided. Furthermore, the software would output a six-in-one interface, which included information of the shape, margin, composition, echogenicity, and the existence and degree of echogenic foci, which were used to apply the results systematically.

Recommendation of FNA

The recommendation of FNA by the TI-RADS was based on the following classification criteria of thyroid nodules: 0 points, TR1; 2 points, TR2; 3 points, TR3; 4–6 points, TR4; and ≥7 points, TR5. A score of >4 points indicated a 5–20% chance of malignancy. TR1–2 indicated no need for FNA. If the largest diameter of a TR3 nodule was ≥1.5 cm, TR4 nodule ≥1.0 cm, or TR5 nodule ≥0.5 cm, the patient should be followed up. If the largest diameter of a TR3 nodule was ≥2.5 cm, TR4 nodule ≥1.5 cm, or TR5 nodule ≥1.0 cm, FNA was needed. The CAD system provided advice and scoring according to the probability value (**Figure 2**). The cytological results of FNA and histological results of surgery were considered as reference standards.

Statistical Analysis

All statistical analyses were performed using IBM SPSS Statistics for Windows, version 25.0. (IBM Corporation, Armonk, NY, USA). Quantitative data are expressed as the mean ± standard deviation. The pathology results were regarded as the gold standard. Receiver operator characteristic curve analysis was performed to evaluate the diagnostic performance and compared in terms of sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and accuracy. The AUCs were computed to assess the diagnostic performances of the CAD system, while performance of the TI-RADS was assessed by the expert and novice. Paired comparisons of sensitivity and specificity were evaluated using the chi-squared test and the McNemar test. A two-sided probability (*p*) value of <0.05 was considered statistically significant.

RESULTS

Population Characteristics

Of the 138 thyroid nodules collected from 126 patients (including 27 males and 99 females), 56 were malignant (54 papillary thyroid carcinomas, one follicular thyroid carcinoma, and one indeterminate malignant tumor) and 82 were benign

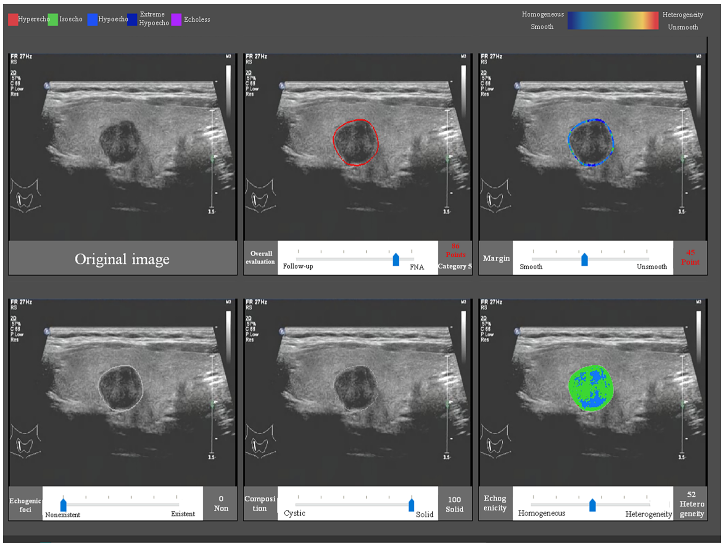


FIGURE 1 | The CAD interface for detection and diagnosis of thyroid nodules. Upper row (from left to right): original image, overall evaluation (score and category), and margin (score and category). Bottom row (from left to right): echogenic foci (percentage and classification), composition (percentage and classification), and echogenicity (percentage and classification).

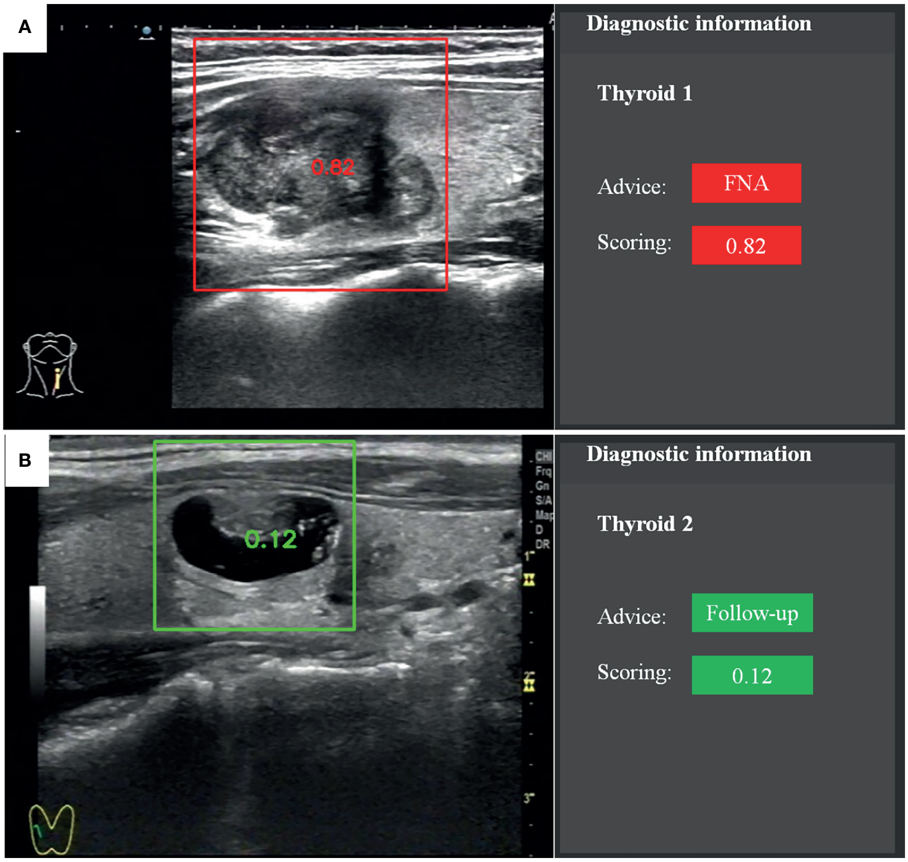


FIGURE 2 | Recommendation and scoring provided by the fully automatic CAD system. **(A)** Papillary thyroid carcinoma. **(B)** Nodular goiter.

(79 nodular goiters and three adenomas). The mean age of the patients was 46.4 ± 12.1 (range, 12–70) years.

Analysis of Diagnostic Accuracy

The diagnostic accuracy of the CAD system, expert, and novice is shown in **Table 1**. The results show that the CAD system had better diagnostic sensitivity, specificity, PPV, NPV, and accuracy than the expert and novice ($p < 0.05$). The AUCs of the CAD system, expert, and novice were 0.919, 0.891, and 0.877, respectively (**Figure 3**).

Correlations Among Features Detected by the CAD System and the Pathology Results

To validate the diagnostic efficiency of features detected by the CAD system, differences in the features of echogenic foci, margin, composition, and echogenicity between benign and malignant nodules were analyzed (**Table 2**). The results revealed significant differences in the existence of echogenic foci and the echogenicity of the nodules between groups ($p < 0.05$).

TABLE 1 | Diagnosis accuracy of CAD system, expert and novice in the study.

	Sensitivity	Specificity	PPV	NPV	Accuracy
CAD system	0.642	0.930	0.929	0.676	0.761
Expert	0.605	0.923	0.929	0.585	0.725
Novice	0.565	0.913	0.929	0.512	0.681

PPV, Positive Predictive Value; NPV, Negative Predictive Value.

Efficiency of FNA Recommendation

The accuracy of a recommendation of FNA was compared between the expert (using TI-RADS) and the CAD system (**Table 3**). According to the diagnoses of the expert, there were five TR1-2 nodules, which did not require follow-up or FNA. Corresponding to the recommendations of the TI-RADS (i.e., TR3 nodule ≥2.5 cm, TR4 nodule ≥1.5 cm, and TR5 nodule ≥1.0 cm), of a total of 60 thyroid nodules, 31 were benign and 29 were malignant. According to the results obtained with the CAD system, FNA was recommended for a total of 58 thyroid nodules (10 benign and 48 malignant). Hence, there were significant differences between the expert and CAD system ($p < 0.001$).

DISCUSSION

Since most thyroid nodules have indolent behavior with positive prognoses, patients are potentially receiving excessive care.

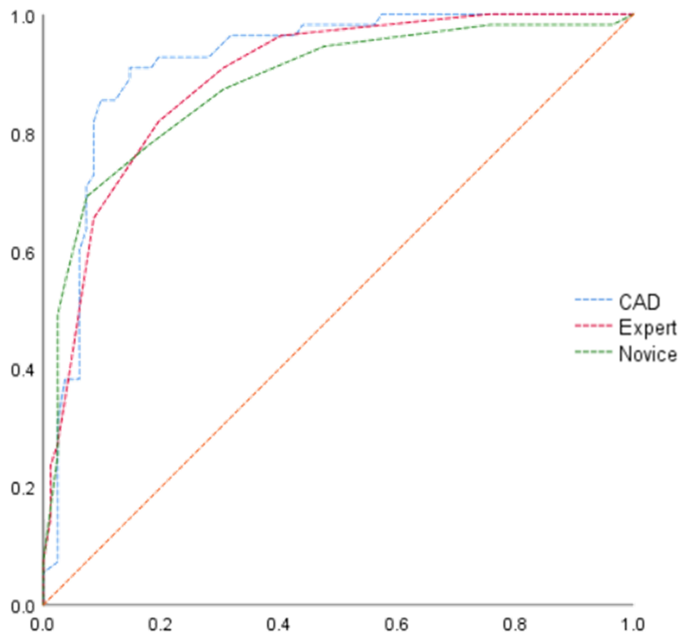


FIGURE 3 | ROC curves of the CAD system, expert, and novice for US diagnosis of thyroid nodules.

TABLE 2 | Correlation among diagnostic features detected by CAD and pathology result.

Diagnostic features		Benign	Malignant	*P* value
Echogenic foci	Existent	40	17	0.031*
	Nonexistent	42	39	
Margin	Smooth	26	18	0.957
	Unsmooth	56	38	
Composition	Cystic (or almost)	0	3	0.065
	Solid (or almost)	82	53	
Echogenicity	Homogeneous	10	17	0.008*
	Heterogeneity	72	39	

**P < 0.05.*

TABLE 3 | FNA recommendation by expert (using TI-RADS) and CAD system.

Recommendation	Groups	Total	Benign	Malignant
FNA	ACR TI-RADS	60	31	29
	CAD system	58	10	48
Follow-up	ACR TI-RADS	78	51	27
	CAD system	80	72	8

FNA, Final-needle aspiration.

The key point to an accurate diagnosis is to improve efficiency and avoid unnecessary FNA as much as possible (18, 19). Although AI techniques continue to rapidly develop, further investigations of efficiency and interpretability are needed before applications for the diagnosis of thyroid nodules. In this study, the efficiency of a CAD system (i.e., AI-SONIC) was validated for the diagnosis of thyroid nodules by comparisons of an expert and novice using the TI-RADS. Also, the features detected by the CAD system, which can lead to explicable results, were analyzed. Furthermore, the efficiency of the recommendation of FNA by the CAD system was investigated. It seems that the CAD system, based on deep learning, had better performance and thus is suitable as an aid for the diagnosis and screening asymptomatic thyroid nodules, especially in areas lacking medical resources.

Performance of the CAD System for US Diagnosis of Thyroid Nodules

Regardless of the high incidence of thyroid nodules, there is a relatively low risk of malignancy. Accurate diagnosis is not only helpful for a diagnosis, but a proper recommendation of FNA can also be obtained. The first time AI was used for US-based diagnosis of thyroid nodules dates back to 1988 (20). AI techniques for US-based diagnosis of thyroid nodules can be divided into two parts: recognition (or segmentation) and classification. In recent years, various advanced algorithms have been proposed to improve the diagnostic accuracy, including support vector machines (21), random forest learning algorithm (22), and artificial neural network algorithm (23), among others. The development of deep learning is another great advancement, which has led to faster, more accurate, and fully automatic diagnoses. However, clinical application is far different from laboratory investigations, thus an objective assessment is necessary. Therefore, since there are differences in data, diagnostic criteria, and equipment, it is impractical to directly compare the diagnostic efficiency of a particular system among different studies. For example, Szczepanek-Parulska et al. compared the performance of the S-Detect System and the

European Thyroid Imaging Reporting and Data System for the diagnosis of 133 cases, and found that the CAD system had better performance (3). However, in another study using the same commercial software, the results showed that the CAD system had lower specificity than an experienced radiologist (41.2 *vs* 83.5%, respectively) (24). In the present study, the sensitivity, specificity, PPV, NPV, and accuracy of the CAD system (AUC = 0.919) were higher than the expert (AUC = 0.891) and novice (AUC = 0.877), indicating that the CAD system based on deep learning could improve the diagnostic efficiency of a novice, similar to the findings of a previous study (25). However, the CAD system is more likely to be an auxiliary approach for diagnosis and asymptomatic screening, rather than a replacement of experienced experts. On the other hand, low specificity remains problematic for US-based diagnoses with a CAD system (26). Notably, in the present study, the highest specificity was only 93.0%. There are several possible explanations for this finding. First, the classification of thyroid nodules with the TI-RADS by the expert and novice was based on scoring criteria, while a diagnosis with the CAD system was closely related to the quality of the image data. Second, since this was a retrospective study, it was inevitable to avoid selecting more positive cases with high-quality images. Third, all patients in this study had undergone surgery or FNA, which may have led to selection bias. The required time for diagnosis among the groups was not analyzed, since it was obvious that the CAD system can output the results and conclusions much faster.

Interpretability of the CAD System

With the continued applications in the medical field, the interpretability of AI is gradually playing an important role in the acceptance by clinicians (27), especially for software based on deep learning. AI-SONIC software provides detailed feature information detected automatically about the existence of echogenic foci, margin, composition, and echogenicity of the thyroid nodules. Here, analyses of the differences in features and pathology revealed that only the existence of echogenic foci and echogenicity of the thyroid nodules were relevant. Theoretically speaking, the existence of echogenic, especially punctate, foci was highly related to papillary thyroid carcinoma (28). However, the results showed that echogenic foci detected by the CAD system were more common with benign, rather than malignant, nodules, likely because spongiform-type nodules associated with goiter were not excluded. In addition, the CAD system was unable to accurately differentiate macrocalcifications from microcalcifications. Besides the failure to exclude lesions associated with nodular goiter, the insignificance of composition features might also relate to liquefactive necrosis in rapidly growing malignant tumors. Furthermore, to better validate the robustness of the CAD system, the validation set had not unified the standard plane of thyroid nodules; thus, some of the obtained images were cross-sections, while others were longitudinal sections, which could have led to uncertainties of boundaries of benign nodules. This was potentially one of the reasons why there were no significant differences in margin features between groups.

Application of CAD in FNA Recommendation

Optimally, the diagnosis and treatment of thyroid nodules should avoid unnecessary FNA (29). Since a CAD system can reduce intra- and inter-observer variability, some studies have validated the value of a CAD system to improve the efficiency of a recommendation of FNA (30–32). In the present study, the performance of the AI-SONIC based on deep learning was much better for a recommendation of FNA than the expert using the TI-RADS ($p < 0.001$). As compared with the TI-RADS evaluated by expert, the CAD system recommended FNA for more malignant thyroid nodules and follow-up for more benign nodules. Because most of the decisions for FNA are made by experts, the efficiency of a recommendation of FNA between the CAD system and novice was not compared.

Limitations

There were several limitations to this study that should be addressed. First, this was a retrospective study; thus, there was selection bias and a lack of standardization. Although this study was conducted in two centers, the sample size was relatively small. In addition, the features automatically detected by the software were not completely identical to those of the TI-RADS, especially the shapes of the nodules. More specific situations should be considered to further improve the CAD system, especially accurate identification of echogenic foci of malignant nodules and spongiform-type nodules associated with goiter. Lastly, intra-observer variability could not be ruled out. Hence, future studies to address these issues are warranted.

CONCLUSIONS

The application of a CAD system is now changing the approaches for the diagnosis and treatment of thyroid nodules. In the present study, the diagnosis on thyroid nodules was improved with the CAD system, although many problems remain that must be addressed. We believe that with the continued development of technology and medical science, the CAD system based on deep leaning and larger datasets will become a suitable, rapid, and high-quality approach for US-based diagnosis and screening of thyroid nodules, especially in areas lacking medical resources.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

XL and YH are responsible for the substantial contributions to conception and design, as well as drafting the article. XL, YC, and JL are responsible for the acquisition of data and the analysis of data. ZC is responsible for the final approval of the version to be published. All authors contributed to the article and approved the submitted version.

FUNDING

This work is supported by the Research Projects of the National Natural Science Foundation of China (No. 82102054), Major Research Projects of Universities in Guangdong Province (No. 2019KZDZX1032), Youth Foundation of Scientific Research of Third Affiliated Hospital of Guangzhou Medical University (No. 2018Q18).

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer Statistics. *CA Cancer J Clin* (2020) 70:7–30. doi: 10.3322/caac.21590
2. Siu JM, McCarty JC, Gadkaree S, Caterson E, Randolph G, Witterick I, et al. Association of Vessel-Sealant Devices vs Conventional Hemostasis With Postoperative Neck Hematoma After Thyroid Operations. *JAMA Surg* (2019) 154:e193146. doi: 10.1001/jamasurg.2019.3146
3. Szczepanek-Parulska E, Wolinski K, Dobruch-Sobczak K, Antosik A, Ostalowska A, Krauze A, et al. S-Detect Software vs. EU-TIRADS Classification: A Dual-Center Validation of Diagnostic Performance in Differentiation of Thyroid Nodules. *J Clin Med* (2020) 9(8):2495. doi: 10.3390/jcm9082495
4. Grant EG, Tessler FN, Hoang JK, Langer J, Beland M, Berland L, et al. Thyroid Ultrasound Reporting Lexicon: White Paper of the ACR Thyroid Imaging, Reporting and Data System (TIRADS) Committee. *J Am Coll Radiol* (2015) 12:1272–9. doi: 10.1016/j.jacr.2015.07.011
5. Tessler FN, Middleton WD, Grant EG, Hoang J, Berland L, Teefey S, et al. ACR Thyroid Imaging, Reporting and Data System (TI-RADS): White Paper of the ACR TI-RADS Committee. *J Am Coll Radiol* (2017) 14:587–95. doi: 10.1016/j.jacr.2017.01.046
6. Peng JY, Pan FS, Wang W, Wang Z, Shan Q, Lin J, et al. Malignancy Risk Stratification and FNA Recommendations for Thyroid Nodules: A Comparison of ACR TI-RADS, AACE/ACE/AME and ATA Guidelines. *Am J Otolaryngol* (2020) 41:102625. doi: 10.1016/j.amjoto.2020.102625
7. Wolinski K, Stangierski A, Ruchala M. Comparison of Diagnostic Yield of Core-Needle and Fine-Needle Aspiration Biopsies of Thyroid Lesions: Systematic Review and Meta-Analysis. *Eur Radiol* (2017) 27:431–6. doi: 10.1007/s00330-016-4356-9
8. Lauria Pantano A, Maddaloni E, Briganti SI, Beretta Anguissola G, Perrella E, Taffon C, et al. AACE/ACE/AME and ACR TI-RADS Ultrasound Classifications Performance in Identifying Cytological High-Risk Thyroid Nodules. *Eur J Endocrinol* (2018) 178:595–603. doi: 10.1530/EJE-18-0083
9. Tan L, Tan YS, Tan S. Diagnostic Accuracy and Ability to Reduce Unnecessary FNAC: A Comparison Between Four Thyroid Imaging Reporting Data System (TI-RADS) Versions. *Clin Imaging* (2020) 65:133–7. doi: 10.1016/j.clinimag.2020.04.029
10. Saw PE, Jiang SP. The Significance of Interdisciplinary Integration in Academic Research and Application. *Bio Integration* (2020) 1:2–5. doi: 10.15212/bioi-2020-0005
11. Liang XW, Cai YY, Yu JS, Liao JY, Chen ZY. Update on Thyroid Ultrasound: A Narrative Review From Diagnostic Criteria to Artificial Intelligence Techniques. *Chin Med J (Engl)* (2019) 132:1974–82. doi: 10.1097/CM9.0000000000000346
12. Chen SJ, Chang CY, Chang KY, Tzeng JE, Chen YT, Lin CW, et al. Classification of the Thyroid Nodules Based on Characteristic Sonographic Textural Feature and Correlated Histopathology Using Hierarchical Support Vector Machines. *Ultrasound Med Biol* (2010) 36:2018–26. doi: 10.1016/j.ultrasmedbio.2010.08.019
13. Chi J, Walia E, Babyn P, Wang J, Groot G, Eramian M. Thyroid Nodule Classification in Ultrasound Images by Fine-Tuning Deep Convolutional Neural Network. *J Digit Imaging* (2017) 30:477–86. doi: 10.1007/s10278-017-9997-y
14. Sun C, Zhang Y, Chang Q, Liu T, Zhang S, Wang X, et al. Evaluation of a Deep Learning-Based Computer-Aided Diagnosis System for Distinguishing Benign From Malignant Thyroid Nodules in Ultrasound Images. *Med Phys* (2020) 47(9):3952–60. doi: 10.1002/mp.14301
15. Chen KY, Chen CN, Wu MH, Ho MC, Tai HC, Kuo WH, et al. Computerized Quantification of Ultrasonic Heterogeneity in Thyroid Nodules. *Ultrasound Med Biol* (2014) 40:2581–9. doi: 10.1016/j.ultrasmedbio.2014.06.009
16. Ma J, Wu F, Jiang T, Zhu J, Kong D. Cascade Convolutional Neural Networks for Automatic Detection of Thyroid Nodules in Ultrasound Images. *Med Phys* (2017) 44(5):1678–91. doi: 10.1002/mp.12134
17. Ma J, Wu F, Jiang T, Zhao Q, Kong D. Ultrasound Image-Based Thyroid Nodule Automatic Segmentation Using Convolutional Neural Networks. *Int J Comput Assist Radiol Surg* (2017) 12:1895–910. doi: 10.1007/s11548-017-1649-7
18. McIver B, Hay ID, Giuffrida DF, Dvorak CE, Grant CS, Thompson GB, et al. Anaplastic Thyroid Carcinoma: A 50-Year Experience at a Single Institution. *Surgery* (2001) 130:1028–34. doi: 10.1067/msy.2001.118266
19. Castellana M, Virili C, Paone G, Scappaticcio L, Piccardo A, Giovanella L, et al. Ultrasound Systems for Risk Stratification of Thyroid Nodules Prompt Inappropriate Biopsy in Autonomously Functioning Thyroid Nodules. *Clin Endocrinol (Oxf)* (2020) 93:67–75. doi: 10.1111/cen.14204
20. Hannequin P, Liehn JC, Maes B, Delisle MJ. Multivariate Analysis in Solitary Cold Thyroid Nodules for the Diagnosis of Malignancy. *Eur J Cancer Clin Oncol* (1988) 24:881–8. doi: 10.1016/0277-5379(88)90197-6
21. Kwon MR, Shin JH, Park H, Cho H, Kim E, Hahn SY. Radiomics Based on Thyroid Ultrasound Can Predict Distant Metastasis of Follicular Thyroid Carcinoma. *J Clin Med* (2020) 9:2156. doi: 10.3390/jcm9072156
22. Zhang B, Tian J, Pei S, Chen Y, He X, Dong Y, et al. Machine Learning-Assisted System for Thyroid Nodule Diagnosis. *Thyroid* (2019) 29:858–67. doi: 10.1089/thy.2018.0380
23. Liu C, Xie L, Kong W, Lu X, Zhang D, Wu M, et al. Prediction of Suspicious Thyroid Nodule Using Artificial Neural Network Based on Radiofrequency Ultrasound and Conventional Ultrasound: A Preliminary Study. *Ultrasonics* (2019) 99:105951. doi: 10.1016/j.ultras.2019.105951
24. Xia S, Yao J, Zhou W, Dong Y, Xu S, Zhou J, et al. A Computer-Aided Diagnosing System in the Evaluation of Thyroid Nodules-Experience in a Specialized Thyroid Center. *World J Surg Oncol* (2019) 17:210. doi: 10.1186/s12957-019-1752-z
25. 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) 19:665–72. doi: 10.3348/kjr.2018.19.4.665
26. Gao L, Liu R, Jiang Y, Song W, Wang Y, Liu J, et al. Computer-Aided System for Diagnosing Thyroid Nodules on Ultrasound: A Comparison With Radiologist-Based Clinical Assessments. *Head Neck* (2018) 40:778–83. doi: 10.1002/hed.25049
27. Tosun AB, Pullara F, Becich MJ, Taylor DL, Fine JL, Chennubhotla SC. Explainable AI (xAI) for Anatomic Pathology. *Adv Anat Pathol* (2020) 27:241–50. doi: 10.1097/PAP.0000000000000264
28. Tahvildari AM, Pan L, Kong CS, Desser T. Sonographic-Pathologic Correlation for Punctate Echogenic Reflectors in Papillary Thyroid Carcinoma: What Are They? *J Ultrasound Med* (2016) 35:1645–52. doi: 10.7863/ultra.15.09048
29. Zhao CK, Chen SG, Alizad A, He YP, Wang Q, Wang D, et al. Three-Dimensional Shear Wave Elastography for Differentiating Benign From Malignant Thyroid Nodules. *J Ultrasound Med* (2018) 37:1777–88. doi: 10.1002/jum.14531

30. Ha EJ, Baek JH. Application of Machine Learning and Deep Learning to Thyroid Imaging: Where do We Stand? *Ultrasonography* (2020) 40(1):23–9. doi: 10.14366/usg.20068
31. Delibasis KK, Asvestas PA, Matsopoulos GK, Zoulias E, Tseleni-Balafouta S. Computer-Aided Diagnosis of Thyroid Malignancy Using an Artificial Immune System Classification Algorithm. *IEEE Trans Inf Technol BioMed* (2009) 13:680–6. doi: 10.1109/TITB.2008.926990
32. Thomas J, Haertling T. AIBx, Artificial Intelligence Model to Risk Stratify Thyroid Nodules. *Thyroid* (2020) 30:878–84. doi: 10.1089/thy.2019.0752

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read
for greatest visibility
and readership

FAST PUBLICATION
Around 90 days
from submission
to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative,
and constructive
peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers
acknowledged by name
on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data
and methods to enhance
research reproducibility

DIGITAL PUBLISHING
Articles designed
for optimal readership
across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics
track visibility across
digital media

EXTENSIVE PROMOTION
Marketing
and promotion
of impactful research

LOOP RESEARCH NETWORK
Our network
increases your
article's readership