Current status of deep learning applications in abdominal ultrasonography

Kyoung Doo Song

Department of Radiology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea

Deep learning is one of the most popular artificial intelligence techniques used in the medical field. Although it is at an early stage compared to deep learning analyses of computed tomography or magnetic resonance imaging, studies applying deep learning to ultrasound imaging have been actively conducted. This review analyzes recent studies that applied deep learning to ultrasound imaging of various abdominal organs and explains the challenges encountered in these applications.

Keywords: Abdominal; Deep learning; Ultrasound

Introduction

Artificial intelligence has been applied in many fields, including medicine. Deep learning has recently become one of the most popular artificial intelligence techniques in the medical imaging field, and it has been applied to various organs using different imaging modalities. In the abdomen, the main imaging modality is computed tomography (CT) for most organs [1]; however, deep learning research regarding abdominal ultrasonography (US) is ongoing. In this article, I review the current status of the application of deep learning to abdominal US and discuss the challenges involved.

Liver

US is one of the most commonly used imaging modalities for evaluating liver disease. In particular, it is used to screen for liver tumors, to evaluate liver status in patients with chronic liver disease, and to evaluate hepatic steatosis.

Diffuse Liver Disease

Ten studies applying deep learning to liver US imaging aimed to evaluate diffuse liver disease, especially hepatic fibrosis and steatosis [2–11]. These studies are summarized in Table 1.

In terms of the type of data used, B-mode image data is the most common. This is likely because B-mode images are the simplest, making it easier to acquire data. However, to evaluate liver fibrosis, some studies have used elastography. Wang et al. [6] used the full two-dimensional (2D) shear wave elastography (SWE) region of interest and included additional B-mode imaging areas from the surrounding area, and demonstrated that the deep learning method was more accurate than 2D-SWE measurements for assessing liver fibrosis. Xue et al. [2] reported that deep learning using

both B-mode and 2D-SWE images showed better performance than using only one of the two types of images. These results suggest that an analysis of the heterogeneity of intensity and texture of colored 2D-SWE and B-mode images can improve the accuracy of the assessment of liver fibrosis. The study of Han et al. [4] was the only one that used radiofrequency (RF) data (Fig. 1). RF signals are raw data obtained from US equipment that are used to generate B-mode images; however, some information is lost or altered during the conversion. Therefore, RF data contain more information than B-mode images and are less dependent on system settings and post-processing operations, such as the dynamic range setting or filtering operations. These characteristics of RF data may be advantageous

Table 1. Summary of studies applying deep learning to diffuse liver disease

Study	Task	Data used for deep learning	US system	Total no. of images (total no. of patients)	No. of validation set images (no. of patients)	Ground truth	Method
Xue et al. (2020) [2]	Fibrosis	B-mode+elastography	One	2,330 (466)	510 (102)	Pathology	CNN
Lee et al. (2020) [3]	Fibrosis	B-mode	≥4	14,583 (3,975)	300 (266) for internal validation 1,232 (572) for external validation	Pathology, elastography, clinical diagnosis	CNN
Wang et al. (2019) [6]	Fibrosis	Elastography	One	1,990 (398)	660 (132)	Pathology	CNN
Treacher et al. (2019) [7]	Fibrosis	B-mode elastography image	One	3,500 (326)	524 (N/A)	Shear wave velocity	CNN
Byra et al. (2018) [8]	Fibrosis	B-mode	One	550 (55)	Leave-one-out cross-validation	Pathology	CNN+SVM
Meng et al. (2017) [11]	Fibrosis	B-mode	N/A	279 (279)	77 (77)	Clinical diagnosis	CNN
Liu et al. (2017) [10]	Fibrosis	B-mode	One	91 (91)	3-fold cross-validation	Clinical diagnosis (Child-Pugh classification, CT, US)	CNN+SVM
Han et al. (2020) [4]	Steatosis	RF US data	One	2,560 RF signals per participant (204)	2,560 RF signals per participant (102)	MRI-derived proton density fraction	CNN
Cao et al. (2020) [5]	Steatosis	B-mode	One	1,092 (N/A)	240 (240)	US scoring system	CNN
Biswas et al. (2018) [9]	Steatosis	B-mode	One	63 (63)	10-fold cross-validation	Pathology	CNN

US, ultrasonography; CNN, convolutional neural network; N/A, not available; SVM, support vector machine; CT, computed tomography; RF, radiofrequency; MRI, magnetic resonance imaging.
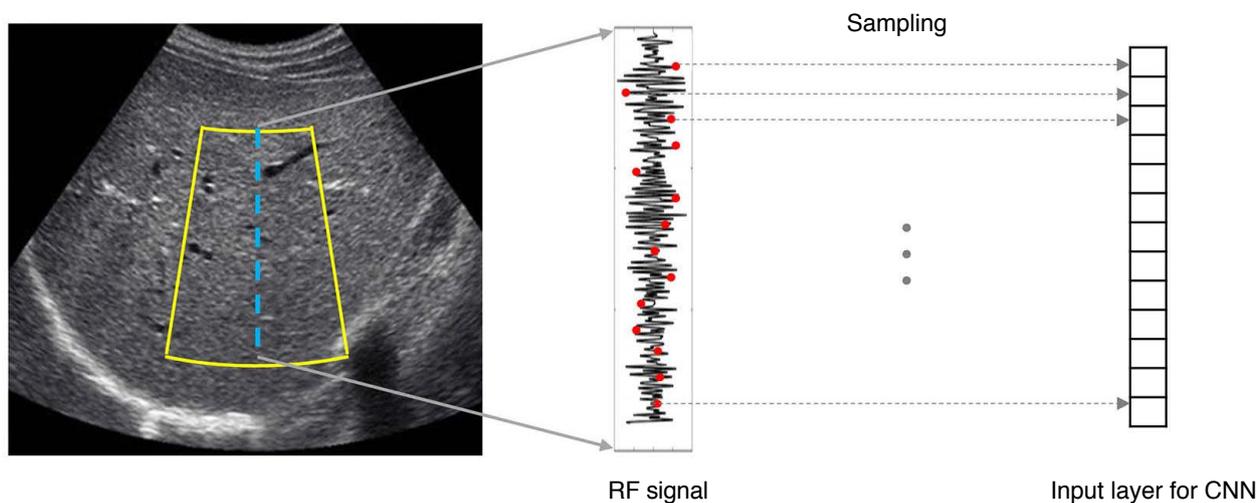

Fig. 1. **Deep learning using radiofrequency data.** The yellow outline indicates the region of interest for deep learning analysis. The radiofrequency signals corresponding to the blue line are downsampled. The downsampled signal values are used as input values to a convolutional neural network. RF, radiofrequency; CNN, convolutional neural network.

in terms of the generalizability of a deep learning model developed from RF data. However, further research is needed to determine whether using RF data will actually help.

Deep learning is basically a data-driven method. Deep learning can extract and learn nonlinear features from data; it does not extract features using domain expertise. In order to avoid over-fitting, developing a deep learning model requires a large amount of data. The amount of data used in the studies included in this review is relatively small compared to the large challenge databases of optical images such as the ImageNet Large Scale Visual Recognition Challenge. However, recently published studies have tended to use a larger amount of data. For example, in the study of Lee et al. [3], 14,583 total images were used to develop and validate a deep learning model for the evaluation of liver fibrosis.

Careful and meticulous confirmation of the clinical performance and utility of a developed model is required for it to be adopted in clinical practice. This involves more than just the completeness of a model or its performance evaluation during development. Robust clinical confirmation of a model's performance requires external validation. Furthermore, for the ultimate clinical verification of developed models, their effect on patient outcomes needs to be evaluated [12]. Of extant studies in this field, only Lee et al. [3] conducted external validation. In particular, it is desirable to perform external validation of a model's performance in a clinical cohort that represents the target population of the developed model using prospectively collected data [12]. However, Lee et al. [3] performed external validation using retrospectively collected data in a case-control group.

Focal Liver Disease

Several studies have applied deep learning to liver US imaging to detect or characterize focal liver lesions [13–17]. These studies are summarized in Table 2.

Compared to the number of studies applying deep learning to diffuse liver disease, the number of studies applying it to focal liver disease is small. The amount of available image data for focal liver disease is relatively small compared to that available for diffuse liver disease, and the US imaging findings of focal liver lesions often overlap. Additionally, in clinical practice, US imaging is usually used as a screening tool, not as a tool for disease confirmation. These factors likely explain why there is less activity in deep learning research targeting focal liver disease than there is in targeting diffuse liver disease.

In terms of the type of data used, contrast-enhanced US (CEUS) was used to develop deep learning models in three studies [13,15,17]. Among them, Liu et al. [17] and Pan et al. [15] used a 3D-convolutional neural network (CNN). CEUS imaging incorporates information regarding space as well as time. A 2D-CNN can only analyze the spatial features, such as texture and edge, from one frame of CEUS cine images, but a 3D-CNN can analyze temporal features as well (Fig. 2).

Schmauch et al. [16] used the dataset that was provided during a public challenge during the 2018 *Journées Francophones de Radiologie* in Paris, France. Although their model was tested on the dataset by the challenge organizers, no detailed information was provided as to how the dataset was collected or what lesions it contained. Except for this one study, external validation has not been performed in any other study that applied deep learning to focal liver disease.

Prostate

Most studies applying deep learning to prostate US imaging have focused on detecting and grading prostate cancer and the

Table 2. Summary of studies applying deep learning to focal liver disease

Study	Task	Data used for deep learning	US system	Total no. of images	No. of validation set images	Ground truth	Method
Liu et al. (2020) [17]	TACE response prediction	CEUS	Three	130 CEUS datasets	41 CEUS cine sets	mRECIST on CT/MRI	3D-CNN
Pan et al. (2019) [15]	Classification	CEUS	N/A	4,420 images from 242 tumors	10-Fold cross-validation	N/A	3D-CNN
Guo et al. (2018) [13]	Classification	CEUS	One	93 CEUS datasets	5-Fold cross-validation	Pathology, CT, MRI	DCCA-MKL
Schmauch et al. (2019) [16]	Detection and classification	B-mode	N/A	544 images	3-Fold cross-validation 177 Images for external validation	N/A	ResNet50
Hassan et al. (2017) [14]	Detection and classification	B-mode	N/A	110 images	10-Fold cross-validation	Unsupervised learning	SSAE+SVM

US, ultrasonography; TACE, transarterial chemoembolization; CEUS, contrast-enhanced ultrasonography; mRECIST, modified Response Evaluation Criteria in Solid Tumor; CT, computed tomography; MRI, magnetic resonance imaging; 3D, three-dimensional; CNN, convolutional neural network; N/A, not available; DCCA-MKL, deep canonical correlation analysis-multiple kernel learning; SSAE, stacked sparse auto-encoders; SVM, support vector machine.
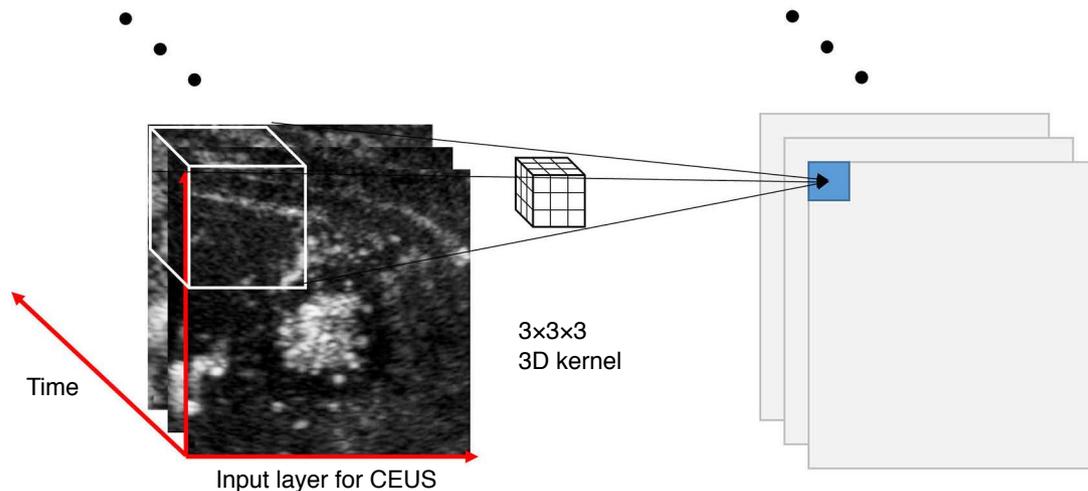
Fig. 2. A three-dimensional (3D) convolutional neural network (CNN) for contrast-enhanced ultrasound (CEUS). Using a 3D-CNN has the advantage of analyzing not only the spatial information of CEUS, but also the temporal information. In the 3D-CNN, multiple CEUS images arranged in temporal order form the input layer. In the 3D-CNN, a 3D kernel is applied. The kernel is not only applied to two-dimensional (2D) images as a 2D sliding convolution of a 2D-CNN, but also applied over consecutive images simultaneously in a 3D-CNN. As a result, the temporal correlation between each image could be captured by this 3D convolution.

segmentation of the prostate gland.

In the field of prostate cancer detection and its grading, a group of researchers has conducted several studies [18–24]. They used multi-parametric magnetic resonance imaging (MRI) data of the prostate gland, CEUS imaging data of suspicious lesions in multi-parametric MRI, and histopathologic results of MRI and transrectal US (TRUS)-fusion guided targeted biopsies. They applied deep learning to CEUS imaging data to classify prostate lesions or grade prostate cancers [18,19,21]. In one study, they integrated multi-parametric MRI and CEUS imaging data for the detection of prostate cancer [24].

TRUS is commonly used as a guiding imaging modality for prostate biopsies and for therapy of prostate cancer. An accurate delineation of the boundaries of the prostate gland on TRUS images is essential for the insertion of biopsy needles or cryoprobes, treatment planning, and brachytherapy. In addition, accurate prostate segmentation can assist in registration and image fusion of TRUS and MRI images. Manual segmentation of the prostate on TRUS imaging is time-consuming and often not reproducible. For these reasons, several studies have applied deep learning to automatically segment the prostate using TRUS imaging [25–31].

Kidney

Very few studies have applied deep learning to kidney US imaging. Zheng et al. [32] evaluated the diagnostic performance of deep learning to classify normal kidneys as well as congenital abnormalities of the kidney and urinary tract. Kuo et al. [33] and Yin et al. [34] used deep learning with kidney US imaging to predict kidney function and to segment the kidney, respectively.

Other Abdominal Organs

There are no reports of deep learning being applied to US imaging of other abdominal organs such as the pancreas or the spleen.

Challenges Applying Deep Learning to Abdominal US Imaging

Compared to CT or MRI, abdominal US imaging faces several challenges regarding the application of deep learning. First, US imaging is highly operator-dependent, both in terms of image acquisition and interpretation. In particular, obtaining images of abdominal US, in which the target organs are located deep inside the body, is more operator-dependent than US of other organs located superficially. Second, it is difficult to image organs that are found behind bone and air. Due to the rib cage and the air normally present in the bowels, abdominal US imaging windows are often limited, and it is therefore often difficult to obtain an image of an entire organ or even a good-quality image. Lastly, there is variability across US imaging systems from different manufacturers, and even those from the same company have version-specific variability. These various challenges make it difficult to standardize US images.

To overcome the challenges in applying deep learning to

abdominal US imaging, efforts to reduce operator and system differences and to improve imaging technology are needed. In this respect, certain studies are worth noting. As mentioned previously, Han et al. used US RF data, and using RF data is expected to reduce the variability among US systems. Camps et al. [35] applied deep learning to automatically assess the quality of transperineal US images in a way that may help to reduce the variability in image acquisition between operators. Finally, Khan et al. [36] proposed a deep learning-based beamformer to generate high-quality US images.

Summary

In this review, recent articles that applied deep learning to US imaging of various abdominal organs are analyzed. Many studies used databases of only a few hundred images or datasets; only a few studies surveyed used thousands of images. Most studies were case-control studies at the proof of concept level. Although several studies have conducted external validation, none have done external validation on prospective cohorts. Overall, the application of deep learning to abdominal US imaging is at an early stage. However, I expect that deep learning for US imaging will continue to progress, because it has many advantages compared to other imaging modalities and efforts are being made to overcome the existing challenges.

ORCID: Kyoung Doo Song: https://orcid.org/0000-0002-2767-3622

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

References

1. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. Med Image Anal 2017;42:60-88.
2. Xue LY, Jiang ZY, Fu TT, Wang QM, Zhu YL, Dai M, et al. Transfer learning radiomics based on multimodal ultrasound imaging for staging liver fibrosis. Eur Radiol 2020;30:2973-2983.
3. Lee JH, Joo I, Kang TW, Paik YH, Sinn DH, Ha SY, et al. Deep learning with ultrasonography: automated classification of liver fibrosis using a deep convolutional neural network. Eur Radiol 2020;30:1264-1273.
4. Han A, Byra M, Heba E, Andre MP, Erdman JW Jr, Loomba R, et al. Noninvasive diagnosis of nonalcoholic fatty liver disease and quantification of liver fat with radiofrequency ultrasound data using one-dimensional convolutional neural networks. Radiology 2020;295:342-350.
5. Cao W, An X, Cong L, Lyu C, Zhou Q, Guo R. Application of deep learning in quantitative analysis of 2-dimensional ultrasound imaging of nonalcoholic fatty liver disease. J Ultrasound Med 2020;39:51-59.
6. 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;68:729-741.
7. Treacher A, Beauchamp D, Quadri B, Fetzer D, Vij A, Yokoo T, et al. Deep learning convolutional neural networks for the estimation of liver fibrosis severity from ultrasound texture. Proc SPIE Int Soc Opt Eng 2019;10950:109503E.
8. Byra M, Styczynski G, Szmigielski C, Kalinowski P, Michalowski L, Paluszkiewicz R, et al. Transfer learning with deep convolutional neural network for liver steatosis assessment in ultrasound images. Int J Comput Assist Radiol Surg 2018;13:1895-1903.
9. 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;155:165-177.
10. Liu X, Song JL, Wang SH, Zhao JW, Chen YQ. Learning to diagnose cirrhosis with liver capsule guided ultrasound image classification. Sensors (Basel) 2017;17:149.
11. Meng D, Zhang L, Cao G, Cao W, Zhang G, Hu B. Liver fibrosis classification based on transfer learning and FCNet for ultrasound images. IEEE Access 2017;5:5804-5810.
12. Park SH, Han K. Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction. Radiology 2018;286:800-809.
13. 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;69:343-354.
14. Hassan TM, Elmogy M, Sallam ES. Diagnosis of focal liver diseases based on deep learning technique for ultrasound images. Arab J Sci Eng 2017;42:3127-3140.
15. Pan F, Huang Q, Li X. Classification of liver tumors with CEUS based on 3D-CNN. In: 2019 IEEE 4th International Conference on Advanced Robotics and Mechatronics (ICARM); 2019 Jul 3-5; Toyonaka, Japan. New York: Institute of Electrical and Electronics Engineers, 2019;845-849.
16. 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;100:227-233.
17. 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;30:2365-2376.

18. Azizi S, Imani F, Ghavidel S, Tahmasebi A, Kwak JT, Xu S, et al. Detection of prostate cancer using temporal sequences of ultrasound data: a large clinical feasibility study. Int J Comput Assist Radiol Surg 2016;11:947-956.

19. Azizi S, Bayat S, Yan P, Tahmasebi A, Nir G, Kwak JT, et al. Detection and grading of prostate cancer using temporal enhanced ultrasound: combining deep neural networks and tissue mimicking simulations. Int J Comput Assist Radiol Surg 2017;12:1293-1305.

20. Azizi S, Mousavi P, Yan P, Tahmasebi A, Kwak JT, Xu S, et al. Transfer learning from RF to B-mode temporal enhanced ultrasound features for prostate cancer detection. Int J Comput Assist Radiol Surg 2017;12:1111-1121.

21. 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;37:2695-2703.

22. Azizi S, Van Woudenberg N, Sojoudi S, Li M, Xu S, Abu Anas EM, et al. Toward a real-time system for temporal enhanced ultrasound-guided prostate biopsy. Int J Comput Assist Radiol Surg 2018;13:1201-1209.

23. Sedghi A, Pesteie M, Javadi G, Azizi S, Yan P, Kwak JT, et al. Deep neural maps for unsupervised visualization of high-grade cancer in prostate biopsies. Int J Comput Assist Radiol Surg 2019;14:1009-1016.

24. Sedghi A, Mehrtash A, Jamzad A, Amalou A, Wells WM 3rd, Kapur T, et al. Improving detection of prostate cancer foci via information fusion of MRI and temporal enhanced ultrasound. Int J Comput Assist Radiol Surg 2020;15:1215-1223.

25. Zeng Q, Samei G, Karimi D, Kesch C, Mahdavi SS, Abolmaesumi P, et al. Prostate segmentation in transrectal ultrasound using magnetic resonance imaging priors. Int J Comput Assist Radiol Surg 2018;13:749-757.

26. 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 (Bellingham) 2019;6:011003.

27. Karimi D, Zeng Q, Mathur P, Avinash A, Mahdavi S, Spadinger I, et al. Accurate and robust deep learning-based segmentation of the prostate clinical target volume in ultrasound images. Med Image Anal 2019;57:186-196.

28. Lei Y, Tian S, He X, Wang T, Wang B, Patel P, et al. Ultrasound prostate segmentation based on multidirectional deeply supervised V-Net. Med Phys 2019;46:3194-3206.

29. van Sloun RJ, Wildeboer RR, Mannaerts CK, Postema AW, Gayet M, Beerlage HP, et al. Deep learning for real-time, automatic, and scanner-adapted prostate (zone) segmentation of transrectal ultrasound, for example, magnetic resonance imaging-transrectal ultrasound fusion prostate biopsy. Eur Urol Focus 2021;7:78-85.

30. Wang Y, Dou H, Hu X, Zhu L, Yang X, Xu M, et al. Deep attentive features for prostate segmentation in 3D transrectal ultrasound. IEEE Trans Med Imaging 2019;38:2768-2778.

31. Orlando N, Gillies DJ, Gyacskov I, Romagnoli C, D'Souza D, Fenster A. Automatic prostate segmentation using deep learning on clinically diverse 3D transrectal ultrasound images. Med Phys 2020;47:2413-2426.

32. Zheng Q, Furth SL, Tasian GE, Fan Y. Computer-aided diagnosis of congenital abnormalities of the kidney and urinary tract in children based on ultrasound imaging data by integrating texture image features and deep transfer learning image features. J Pediatr Urol 2019;15:75.

33. Kuo CC, Chang CM, Liu KT, Lin WK, Chiang HY, Chung CW, et al. Automation of the kidney function prediction and classification through ultrasound-based kidney imaging using deep learning. NPJ Digit Med 2019;2:29.

34. Yin S, Zhang Z, Li H, Peng Q, You X, Furth SL, et al. Fully-automatic segmentation of kidneys in clinical ultrasound images using a boundary distance regression network. Proc IEEE Int Symp Biomed Imaging 2019;2019:1741-1744.

35. Camps SM, Houben T, Carneiro G, Edwards C, Antico M, Dunnhofer M, et al. Automatic quality sssessment of yransperineal ultrasound images of the male pelvic region, using deep learning. Ultrasound Med Biol 2020;46:445-454.

36. Khan S, Huh J, Ye JC. Deep learning-based universal beamformer for ultrasound imaging. In: Shen D, ed. Medical image computing and computer assisted intervention: MICCAI 2019. Lecture notes in computer science, Vol. 11768. Cham: Springer, 2019;619-627.