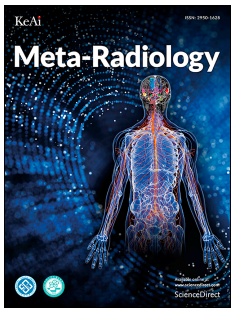
Journal Pre-proof



Application of deep learning on automated breast ultrasound: Current developments, challenges, and opportunities

Ruixin Wang, Zhiyuan Wang, Yuanming Xiao, Xiaohui Liu, Guoping Tan, Jun Liu

PII: S2950-1628(25)00006-2

DOI: https://doi.org/10.1016/j.metrad.2025.100138

Reference: METRAD 100138

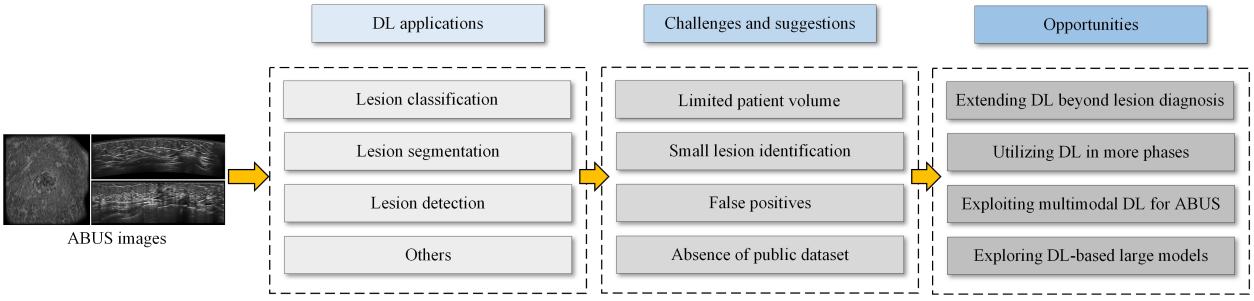
To appear in: *Meta-Radiology*

Received Date: 16 January 2025

Revised Date: 19 February 2025

Accepted Date: 13 March 2025

Please cite this article as: Wang R, Wang Z, Xiao Y, Liu X, Tan G, Liu J, Application of deep learning on automated breast ultrasound: Current developments, challenges, and opportunities, *Meta-Radiology*, https://doi.org/10.1016/j.metrad.2025.100138.



Application of deep learning on automated breast ultrasound: Current developments, challenges, and opportunities

Ruixin Wang [1], Zhiyuan Wang [2*], Yuanming Xiao [3], Xiaohui Liu [4], Guoping Tan [5], Jun Liu [167*]

[1]Department of Radiology, the Second Xiangya Hospital, Central South University, Changsha, China; rxwang1227@gmail.com (R.W.); junliu123@csu.edu.cn (J.L.)
[2]Department of Ultrasound, the Affiliated Cancer Hospital of Xiangya School of Medicine, Central South University, Changsha, China; wangzhiyuan@hnca.org.cn
[3]Health Management Medicine Center, the Third Xiangya Hospital, Central South University, Changsha, China; noriko_13@sina.com
[4]The First People's Hospital of Kunshan, Affiliated Kunshan Hospital of Jiangsu University, Suzhou, China; lxhmedicine123@163.com
[5]College of Computer and Software, Hohai University, Nanjing, China; gptan@hhu.edu.cn
[6]Clinical Research Center for Medical Imaging in Hunan Province, Changsha, China
[7]Department of Radiology Quality Control Center in Hunan Province, Changsha, China
*****Correspondence: wangzhiyuan@hnca.org.cn (Z.W.); junliu123@csu.edu.cn (J.L.)

Abstract:
Breast cancer is a major disease threatening the health of women worldwide. The advent of automated breast ultrasound (ABUS) has provided new possibilities for the early screening and diagnosis of breast cancer. Concurrently, artificial intelligence (Al)-based computer-aided diagnosis (CAD) systems, driven by deep learning (DL), have advanced significantly over the past decade. Unlike traditional handheld ultrasound (HHUS), ABUS enables the separation of scanning and diagnosis, increasing the demand for CAD systems that hold significant clinical value. In recent years, DL has become a dominant force in AI development, playing a crucial role in CAD for all kinds of medical imaging. However, despite its prominence in AI-driven medical image analysis, a comprehensive review of its applications in ABUS is still lacking. This paper provides a detailed analysis of the latest advancements, existing challenges, and future research opportunities in this rapidly evolving field.

Keywords:
Automated breast ultrasound; Breast cancer; Deep learning; Computer-aided diagnosis

1 Introduction

Breast cancer is the most common and deadliest cancer among women worldwide. As shown in Figure 1, according to the latest global cancer statistics by the International Agency for Research on Cancer (IARC)[1], in 2022, the incidence of breast cancer reached 2.29 million cases (23.8%), and the global breast cancer mortality reached 600,000 cases (15.4%). The incidence of breast cancer varies by continent, with Asia experiencing the highest burden, where both the incidence rate (42.9%) and mortality rate (47.3%) are significantly higher than those in other regions. Imaging techniques are central to the screening and diagnosis of breast cancer. Handheld ultrasound (HHUS),

mammography (X-ray imaging of the breast), and magnetic resonance imaging (MRI) are currently the most widely utilized imaging modalities for breast evaluation, each with distinct advantages and limitations. The advantages of HHUS include the absence of radiation, real-time imaging capabilities, and relatively high specificity. However, it also has notable drawbacks: it produces two-dimensional (2D) images, which can lead to missed diagnoses; the image acquisition process lacks standardization, resulting in poor reproducibility; and the diagnostic outcomes are heavily dependent on the skill and experience of ultrasound physicians.

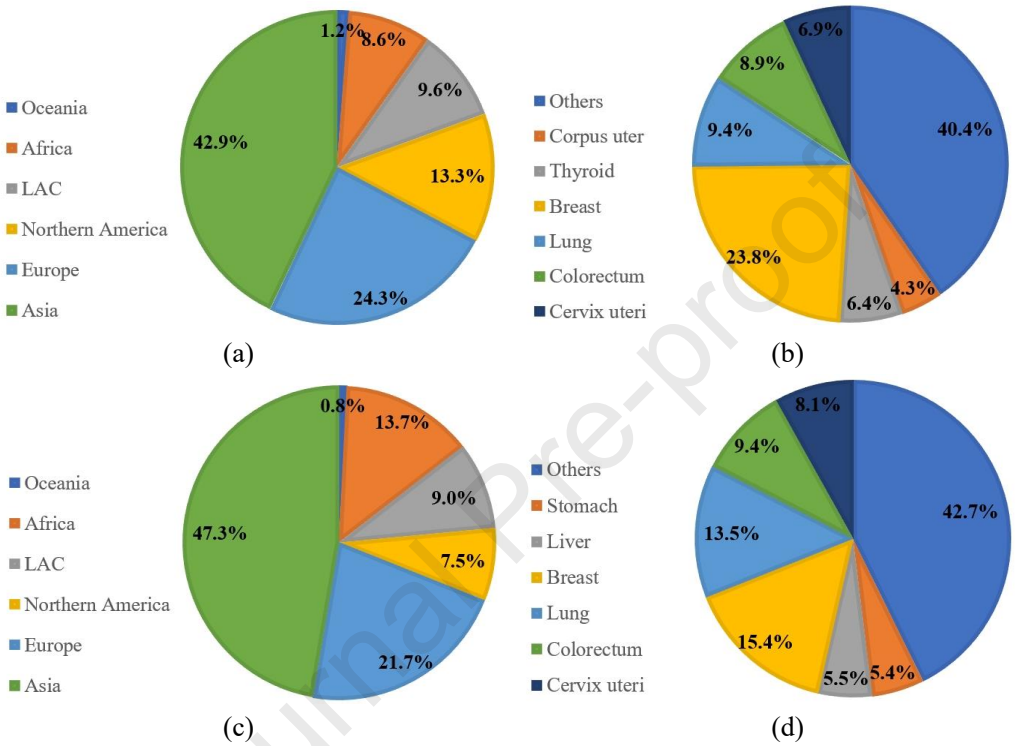


Figure 1. Global incidence and mortality of breast cancer in 2022. (a) Incidence by continent; (b) Incidence by cancer type; (c) Mortality by continent; (d) Mortality by cancer type.

Automated Breast Ultrasound (ABUS) is an innovative breast cancer imaging technology that generates three-dimensional (3D) ultrasound images of the whole breast. Compared to HHUS, ABUS provides automated, consistent imaging and 3D reconstruction, standardizing the ultrasound image acquisition process[2]. The 3D reconstructed images help reduce missed detections, while ABUS also alleviates operator fatigue and enhances examination efficiency. Furthermore, ABUS is particularly advantageous for high-risk or dense-breast patients, providing a more comprehensive and accurate evaluation compared to HHUS. Therefore, ABUS offers new opportunities for breast cancer screening and diagnosis. Specifically, it allows for the separation of scanning and diagnosis, making it particularly conducive to the integration of computer-aided diagnosis (CAD) technologies. CAD can enhance the efficiency of ABUS interpretation by physicians and improve diagnostic accuracy, which is essential for the broader adoption of ABUS. Over the past decade, deep learning (DL) technology, particularly DL-based computer vision techniques, has dominated the development of CAD[3]. A large volume of DL methods or models have been proposed to support the CAD across various imaging modalities, with progress also made in relation to ABUS. However, due to the limited adoption and the unique imaging characteristic of ABUS, targeted efforts are

necessary to develop specialized DL models for this modality. This article aims to provide a comprehensive analysis of the current state of DL for ABUS in the context of CAD, with a particular focus on DL methodologies. Additionally, the challenges encountered and the future research opportunities that warrant further exploration are also discussed.

2 Preliminaries about ABUS

As illustrated in Figure 2, the process of obtaining ABUS images typically involves the following steps: (1) Patient preparation: The patient is asked to remove their upper clothing and lie supine on the bed, with both arms either raised above the head or positioned to fully expose the breast area. (2) Initial settings: The technician adjusts the imaging depth according to the patient's breast size and selects the optimal scanning position. (3) Automated scanning: The technician activates the scanning switch, and the probe is automatically moved by a motor, performing continuous scans of multiple axial slices. (4) Data processing: Once the scanning is complete, the ABUS system reconstructs the acquired 2D axial images into two additional planes, i.e., the coronal and sagittal planes. The system also optimizes the image quality, including noise reduction and contrast enhancement. (5) 3D image display: The system displays the axial images, along with the reconstructed coronal and sagittal images, on the workstation for the physician's review.

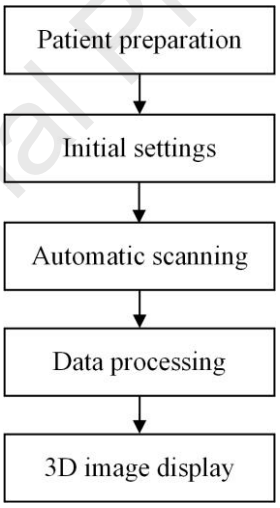


Figure 2. The image acquisition process of ABUS

As illustrated in Figure 3, the ABUS images consist of three orthogonal planes: axial, coronal, and sagittal planes. The axial plane refers to a transverse slice of the breast and is crucial for examining the tissue structure and evaluating the imaging characteristics of breast masses. The coronal plane is a slice perpendicular to the chest surface, offering a comprehensive view of the breast's structure. This plane is particularly useful for observing the anterior-posterior distribution of breast tissue and assessing the relationship between breast masses and surrounding glandular tissue, revealing key signs such as the "retraction phenomenon". The sagittal plane provides a longitudinal slice of the breast, and when combined with the axial and coronal planes, it facilitates multi-planar reconstruction and analysis, offering a more comprehensive understanding of the breast lesions.

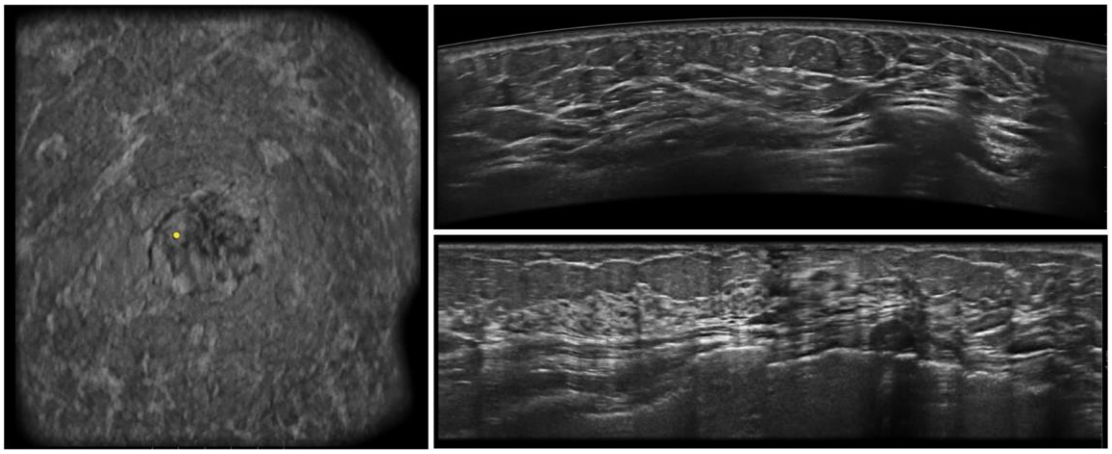


Figure 3. ABUS images. Left: coronal plane; upper right: axial plane; lower right: sagittal plane. The yellow point in the coronal plane: the nipple marked by the technician. The images were acquired with a GE Invenia ABUS 2.0.

The acquisition of ABUS images typically involves multiple standard scanning positions to ensure comprehensive coverage of the whole breast. As shown in Figure 4, the most common scanning positions are AP (Anterior-Posterior), LAT (Lateral), and MED (Medial). In the AP position, the probe scans from the front, primarily visualizing the anterior and upper areas of the breast. In the LAT position, the probe scans from the lateral (outer) side of the breast, allowing visualization of the outer breast tissue and the axillary region. In the MED position, the probe scans from the medial (inner) side of the breast, aiding in the visualization of the inner structures of the breast.

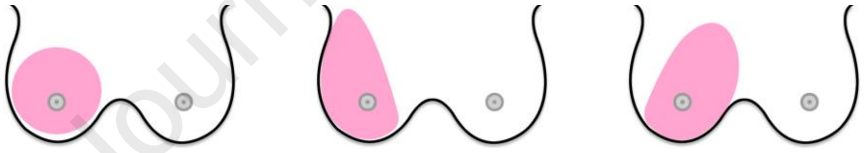


Figure 4. Scanning positions of ABUS. From left to right: AP, LAT, and MED

3 Methods

In this review, we conducted a comprehensive literature search across several databases, including PubMed, IEEE Xplore, Web of Science, and Google Scholar, to systematically identify relevant studies. The search utilized keywords such as “deep learning and automated breast ultrasound”, “automated breast ultrasound”, “deep learning and ABUS”, and “ABUS” to cover a broad range of research focused on the application of DL techniques in ABUS. We limited our search to papers published up to January 2025, to include the most current and relevant studies. Additionally, we reviewed the reference lists of selected papers to identify any other pertinent research. To maintain the quality of the review, we excluded the following types of literature: non-peer-reviewed articles, conference abstracts, editorials, commentaries, and letters to the editor; non-English language studies; and studies that focused solely on clinical research of DL models or systems, as this review specifically addresses DL methodologies.

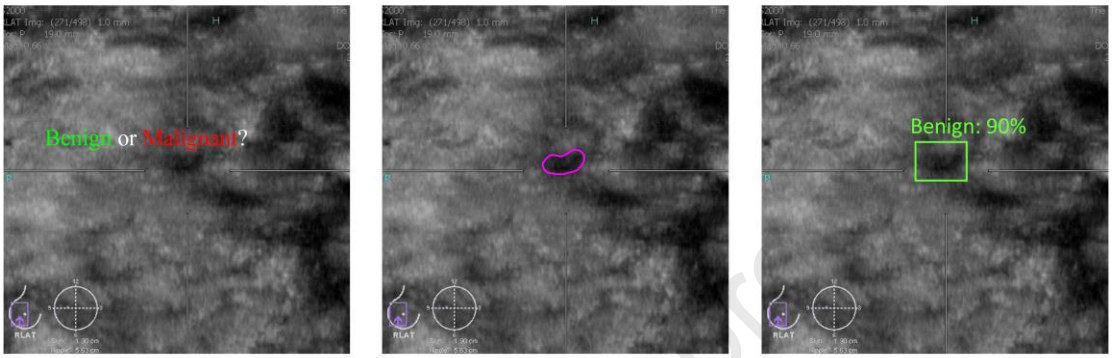
4 Results

Table I presents an overview of the reviewed literature, where breast lesion classification, segmentation, and detection are the most widely applied areas of DL in ABUS. As shown in Figure 5, breast lesion classification is to categorize lesions (such as lumps or nodules) based on their features, usually into benign or malignant categories. Classification typically relies on imaging features such as shape, boundary, and echogenicity to make the determination. Breast lesion segmentation refers to the automatic identification and extraction of the lesion region from the breast image. The result of segmentation is often a binary region map that outlines the lesion. Breast lesion detection automatically identifies whether there is a lesion region in the image and localizes it. This section will provide a detailed introduction to these pieces of literature.

Table I Literature overview

DL Application	Literature	Key DL technique	Data volume	Performance
Lesion classification	Wang et al.[4]	CNN, Transfer learning	263 patients (316 breast lesions)	Sensitivity: 88.6% Specificity: 87.6% AUC: 0.9468
	Hejduk et al.[5]	CNN	113 patients (55 high malignancy probability)	Accuracy: 90.9% AUC: 0.91
	Kim et al.[6]	Weak supervision, 3D CNN	363 patients (434 mass lesions)	Sensitivity: 87.75% Specificity: 93.75% AUC: 0.9491.
Lesion segmentation	Cao et al.[7]	U-Net, Dilated convolution, Focus loss	107 patients (170 volumes)	DSC: 0.6902 Jaccard index: 0.5661
	Pan et al.[8]	BLSTM, S-C attention, CNN	124 patients (170 volumes)	DSC: 0.8178 Recall: 80.67% Precision: 82.92%
	Lei et al.[9]	R-CNN	70 patients with breast cancer	DSC: 0.82±0.15
	Cheng et al.[10]	U-Net, Spatial and channel-wise self-attention	107 patients (170 volumes)	DSC: 0.8454 IoU: 0.7324
	Luo et al.[11]	3D U-Net, 3D CNN	397 patients (220 breast cancer)	Sensitivity: 92.7%(volume), 94.5%(lesion) and 96.5%(patient) FP: 2.4 per volume
Lesion detection	Li et al.[12]	YOLOv3, Multi-task learning	124 patients (181 tumor instances)	Sensitivity: 90% FP: 7.42 per volume
	Zhang et al.[13]	YOLOv4, Monte Carlo Dropout	124 patients (170 volumes)	Sensitivity: 88% FP: 0.19 per slice
	Zhang et al.[14]	YOLOv5, 3D ResNet, Transformer encoder	741 patients (3114 breast lesions)	Detection rate: 71.2%
	Oh et al.[15]	Faster R-CNN, U-Net	28 patients (168 scans)	Sensitivity: 93.65% FP: 8.6 per examination
	Li et al.[16]	YOLOX,	124 patients	Sensitivity: 90%

		Deep mutual learning	(238 volumes)	FP: 0.15 per slice
Breast anatomy segmentation	Lei et al.[17]	Deep supervision, Transfer learning	16 patients (16 volumes)	CD of lines (pixel): 3.9, 5.6 and 8.3 IoU of layers (%): 86.8, 72.2, 72.2 and 76.1
Biomarker estimation	Huang et al.[18]	3D ResNet, Multi-task learning	388 patients (388 volumes)	AUCs: 0.735 (ER), 0.767(PR), and 0.697(HER2)



Lesion classification (a) Lesion segmentation (b) Lesion detection (c)

Figure 5. Three most common DL applications on ABUS

4.1 Lesion classification

Wang et al.[4] developed a multi-view convolutional neural network (CNN) to address the multi-view nature of ABUS images and used transfer learning[19] for the classification of benign and malignant lesions in ABUS images. After training and evaluating the model on 316 breast lesions (135 malignant and 181 benign), the method achieved an area under the curve (AUC) value of 0.9468, with sensitivity and specificity reaching 88.6% and 87.6%, respectively. Furthermore, observer performance testing demonstrated that, by referencing the CNN's prediction results, the diagnostic performance of all human reviewers improved. Thus, this multi-view CNN, functioning as a secondary reviewer, showed significant potential in assisting breast cancer diagnosis.

Hejduk et al.[5] developed a post-processing technique based on CNN, which involves multi-step processing using several DL models to classify the malignancy of lesions, specifically distinguishing between BI-RADS 2/3 and BI-RADS 4/5 categories. For the region of interest (ROI) determined using a sliding window, Model 1 is first applied to classify background or breast tissue. Next, Model 2 categorizes the tissue as either normal or lesioned, and finally, Model 3 is employed for benign or malignant classification. The study utilized 645 ABUS scans from 113 patients for training and validation. The results showed that the accuracy for single image classification containing lesions was 79.7%, with an AUC of 0.91. On the whole ABUS dataset, the classification accuracy of the method improved to 90.9%, while the AUC remained consistent, demonstrating performance comparable to that of experienced radiologists.

Kim et al.[6] proposed a weakly supervised branch network, the Mask Branch Network, combined with a template mask to address the challenge of distinguishing between tumors and

background shadows in ABUS images. This method ingeniously utilizes the tumor feature information from radiology reports to generate a template mask, thereby enhancing the performance of the existing neural network classifier without requiring additional annotations. Experimental results demonstrated that the branch network with the template mask significantly improved classification accuracy. On a dataset comprising 363 patients (286 with 434 BI-RADS 2 or higher tumor lesions and 77 without tumor lesions), the network achieved a sensitivity of 87.75%, a specificity of 93.75%, and an AUC value of 0.9491.

4.2 Lesion segmentation

To address the challenges posed by significant variation in the shape and size of breast masses, as well as the relatively small training datasets in ABUS image segmentation, Cao et al.[7] designed a dilated densely connected U-Net[20] (D²U-Net) and an uncertainty-focused loss function. This method constructs a lightweight and efficient dense-connected segmentation network by integrating hybrid dilated convolutions[21] within the dense blocks of D²U-Net. This design allows the network to handle mass features at different scales, enhancing its ability to learn effectively from small medical image datasets. Additionally, the proposed uncertainty focus loss helps increase the model's attention on uncertain predictions, thereby improving the segmentation precision of blurry mass boundaries. The method was evaluated on a dataset of 170 volumes from 107 patients, with the experimental results showing a Dice Similarity Coefficient (DSC) of 0.6902 and a Jaccard index of 0.5661.

Pan et al.[8] proposed the SC-FCN-BLSTM model to address the issues of unclear tumor boundaries and significant shape variations in ABUS images. This method combines a bidirectional long short-term memory (BLSTM) network[22] with a spatial-channel attention (SC-attention) mechanism[23]. By incorporating finer-grained spatial information and rich semantic features, the method enhances the discriminative power of features. Additionally, the BLSTM module with SC-attention captures the inter-slice relationships, leveraging contextual information from the 3D data to assist segmentation and reduce false positives. The method was evaluated on a private ABUS dataset containing 124 patients and 170 volumes, achieving a DSC of 0.8178, a recall of 80.67%, and a precision of 82.92%.

To address the challenges faced in breast tumor segmentation from ABUS images, such as low image quality, speckle noise, shadowing effects, and low contrast, Lei et al.[9] proposed a Mask Scoring R-CNN approach based on Mask R-CNN[24]. This method enhances segmentation accuracy by incorporating a self-correlation strategy and establishing a direct link between the detection and segmentation subnetworks, ensuring the coherence of the final tumor segmentation contours. The method was evaluated on an ABUS dataset from 70 breast cancer patients diagnosed through biopsy, with 40 cases used for five-fold cross-validation and 30 cases for independent testing. The experimental results demonstrated that, compared to manually delineated contours, the proposed method achieved an average DSC of 0.85±0.10 in cross-validation, and 0.82±0.15 in an independent test set.

To address the challenges of lesion diversity, significant imaging artifacts, and blurred lesion boundaries, Cheng et al.[10] proposed a deepest semantically guided multi-scale feature fusion network (DSGMFFN). This method ensures the comprehensive utilization of deep semantic information during the decoding process. Additionally, a multi-scale feature fusion module (MFFM)

is introduced to capture information at various scales, while a novel mixed self-attention[25] mechanism (MSAM) is developed, combining both spatial and channel-wise self-attention to highlight the lesion areas. The method was evaluated on a dataset comprising 170 ABUS volumes with 3,742 slices. The experimental results demonstrated that DSGMFFN achieved a DSC of 0.8454 and an Intersection over Union (IoU) of 0.7324.

Luo et al.[11] addressed the challenge of balancing high sensitivity with low false positives in existing CAD systems by designing and optimizing a CNN based on a 3D U-Net architecture to improve lesion segmentation efficiency and accuracy in ABUS images. The network incorporated optimization strategies such as densely deep supervision (DDS) mechanism and threshold map (TM), and was further improved by utilizing a pre-trained C3D[26] model. The study employed a dataset comprising ABUS and HHUS data from 397 female subjects. Experimental results demonstrated that the proposed DL model achieved sensitivity rates of 93.8%, 97.2%, and 100% based on volume, lesion, and patient, respectively, in the training set, with an average of 1.9 false positives per volume. In the test set, sensitivity slightly decreased to 92.7%, 94.5%, and 96.5%, with a false positive rate of 2.4 per volume. For lesions larger than 1 cm^3, the sensitivity reached 98%, while for lesions less than or equal to 1 cm^3, the sensitivity was 87%. Additionally, the study revealed no significant impact of breast tissue composition, lesion morphology, or echogenic patterns on sensitivity and false positive rates.

4.3 Lesion detection

Compared to lesion segmentation, lesion detection in ABUS images focuses more on locating the lesion rather than delineating its exact boundaries. Li et al.[12] proposed a 3D tumor detection method based on CNN to address the challenges of high false-positive rates, difficulty in detecting small tumors, and insufficient utilization of spatial continuity between adjacent slices in ABUS images. To improve the detection accuracy, they enhanced the YOLOv3[27] detector by introducing a multi-task learning mechanism, enabling simultaneous tumor candidate region detection and tumor mask generation in 2D slices. This approach provided more accurate localization information. To tackle the issue of positional differences and scoring inconsistencies in 2D detection results, the authors proposed a rescoring algorithm and constructed a 3D volume-forming model. This model differentiates real tumor volumes from false-positive areas and further eliminates false positives based on slice length and average area. The method was tested on an ABUS dataset consisting of 124 patients with 181 tumor instances. Experimental results showed a sensitivity of 90%, 85%, 80%, 75%, and 70% under conditions of 7.42, 3.31, 1.62, 1.23, and 0.88 false positives per volume, respectively.

To improve tumor detection rates and reduce false positive regions in ABUS images, Zhang et al.[13] developed a tumor detection method based on a Bayesian YOLOv4[28]. To address challenges posed by speckle noise, shadowing, and significant variation in lesion echogenicity, the authors modified the YOLOv4 and incorporated Monte Carlo Dropout (MC-Drop)[29] to handle model prediction uncertainty. This approach enhanced the detection rate of challenging tumor regions while reducing false positives. The method was tested on a private dataset containing 170 ABUS tumor volumes, and the results showed a sensitivity of 88% and a false positive rate of 0.19 per slice, outperforming the original YOLOv4 and other mainstream object detectors.

Zhang et al.[14] developed an intelligent detection system called V-BUILDS (Volume-Breast

Ultrasound Intelligent Lesion Detection System) based on YOLOv5[30] to address the challenges in breast lesion detection, such as high operator dependence, poor diagnostic reproducibility, long examination times in ABUS images. The system utilizes an improved mosaic data augmentation technique for training and combines a 3D ResNet[31] to reduce false positives and a transformer encoder[25] module to enhance detection performance. Evaluated on a multicenter dataset with 741 cases and 2,538 volumes, the system achieved an overall detection rate of 71.2%. Specifically, it demonstrated high detection rates of 96.5% for BI-RADS 4/5 lesions and 95.8% for malignant tumors. However, the detection performance for small lesions (<10 mm) and benign lesions classified as BI-RADS 2 or 3 remains suboptimal.

Oh et al.[15] proposed a 3D breast nodule detection system to address the challenge of quickly interpreting the large volume of image data generated by ABUS. Their method combines the Faster R-CNN[32] detector with U-Net[20]. This approach uses 2D images from the axial, coronal, and sagittal planes to locate suspicious breast nodules, then converts the 2D detection results into 3D detection results. To reduce false positives, the method incorporates chest wall segmentation and aggregates the detection results. Experimental results on an internal dataset consisting of 168 scans from 28 female patients showed that the method achieved a sensitivity of 93.65% with 8.6 false positives per examination.

Li et al.[16] proposed the DML-YOLOX model, which is based on the YOLOX detector[33] and deep mutual learning (DML)[34], to address the issues of false positives and false negatives in tumor detection in ABUS images. To tackle the problem of overconfidence in single-model predictions, the method introduces exploration loss and consistency loss to enhance collaborative learning between two parallel networks. Additionally, to resolve the issue of overestimating bounding box sizes in traditional rotational data augmentation methods for ABUS images, the researchers designed a specialized rotational augmentation method based on the elliptical shape of tumor contours. They also incorporated rotational discriminative measurement to optimize regression loss and effectively mitigate the impact of label errors caused by large-angle rotations during training. The experimental results on a private ABUS dataset, which included 68 normal volumes and 68 tumor volumes, demonstrated that the DML-YOLOX model achieved a sensitivity of 90% and a false positive rate of 0.15 per slice.

4.4 Other Applications

Besides the above DL applications for lesion diagnosis, Lei et al.[17] investigated breast anatomical layer segmentation in ABUS images. To address issues such as inherent speckle noise, posterior acoustic shadows, and boundary-blurring caused by overlapping echogenicity spectra between different layers, they designed a boundary-regularized convolutional encoder-decoder network (ConvEDNet). The authors introduced deep boundary supervision (DBS), based on deep supervision[35] and adaptive domain transfer (ADT), based on transfer learning[19]. DBS enhances boundary recognition by providing geometric constraints with assistance from shallow networks, while ADT improves the network's adaptability to ultrasound image characteristics through a two-stage knowledge transfer approach. The method was cross-validated on an ABUS dataset consisting of 16 volumes and 3,134 2D images. Experimental results showed that ConvEDNet excelled in segmenting four anatomical structures: subcutaneous fat, breast parenchyma, muscle, and chest wall. The method effectively reduced false positives, thereby improving the efficiency of clinical image

interpretation.

Huang et al.[18] developed both a single-task model and a multi-task model based on 3D ResNet to noninvasively estimate three breast cancer biomarkers—estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2)—using ABUS images. The single-task model produces a prediction probability output, while in the multi-task model, an ABUS volume is input into the network to generate both a 3D segmentation probability map and biomarker classification scores. During training, the input volume is enhanced through an element-wise addition with the segmentation probability map from the previous iteration. In the study, which used a dataset of 388 ABUS volumes, with 23% (n=88) allocated for testing, the single-task and multi-task models achieved AUCs of 0.809 and 0.735 for ER, 0.688 and 0.767 for PR, and 0.626 and 0.697 for HER2, respectively. In the overall evaluation, the multi-task model outperformed the single-task model, achieving a higher macro AUC of 0.733 compared to 0.708 for the single-task model.

5 Discussions and conclusions

5.1 Challenges and suggestions

5.1.1 Limited patient volume

In existing research, the number of patients involved in training DL models for ABUS is limited. This may result from the massive amount of ABUS images, with each patient potentially generating thousands of images per breast, which imposes significant pressure on data processing, model development, and training. Moreover, the wide variety of breast diseases, including benign lesions, malignant tumors, cysts, and calcifications, leads to challenges arising from the limited data. Due to insufficient sample size, models may suffer from overfitting, resulting in limited generalization in clinical practice. Furthermore, the limited sample size can lead to data imbalance, where certain diseases (such as malignant tumors or small calcifications) have fewer samples, creating a large disparity in sample sizes across different types of lesions. This, in turn, affects the overall performance of the model and its ability to identify rare lesions.

To address this challenge, in addition to common techniques such as data augmentation and transfer learning, approaches like few-shot learning[36] and the use of generative models to create new data[37] can be applied. Few-shot learning aims to enable models to effectively learn from a limited number of samples, while generative models, such as Generative Adversarial Networks (GANs), can synthesize additional training data to alleviate the problem of insufficient data.

5.1.2 Small lesion identification

Existing studies[11][14] have indicated that the performance of DL models in segmenting or detecting small lesions in ABUS images requires further improvement. The characteristics of small lesions, including boundaries, morphology, and grayscale variations, are typically subtle and

difficult to differentiate from surrounding tissue, especially in high-noise, low-resolution images. ABUS images generally exhibit lower resolution compared to other imaging techniques, such as magnetic resonance imaging (MRI) or computed tomography (CT), which results in small lesions appearing less distinct, particularly in areas with dense breast tissue or thicker glands. This reduced resolution causes the lesion boundaries to become blurred, complicating the model's ability to accurately localize the lesion's position and morphology. In addition, some small lesions may exhibit morphological similarities to normal tissue or other benign lesions, which can cause them to be misidentified as benign structures, thereby increasing the risk of misdiagnosis. This challenge is particularly prominent in early-stage malignant tumors, microcalcifications, or small nodules in breast tissue, where lesions often lack clear boundaries or distinctive features, resulting in their potential oversight or misinterpretation as normal tissue in ABUS images.

To address the limitations in detecting small lesions in ABUS, techniques from the field of computer vision for small object recognition[38] can be leveraged. For instance, multi-scale learning[39] can enhance the model's ability to capture lesion features across different scales. By constructing neural networks with multi-scale feature extraction capabilities, the model can better accommodate variations in lesion size and improve its ability to detect small lesions. The attention mechanism[40] enables the model to focus on potential lesion areas by assigning greater weight to specific regions of interest and filtering out irrelevant background information, thus enhancing the sensitivity of lesion detection. Focal loss[41], an improved loss function that down-weights easily classified samples during training and directs the model's attention towards hard-to-classify small lesions, could be another remedy for reducing false negatives and false positives.

5.1.3 False positives

Due to the inherent limitations of ultrasound imaging, the resolution of breast tissue is generally lower, particularly in patients with dense glandular tissue, which may result in the misinterpretation of normal tissue as lesions. Additionally, breast lesions exhibit a broad range of types and morphological characteristics, including benign lumps, calcifications, and fibroadenomas, among others. Certain benign lesions, such as fibroadenomas or breast cysts, can exhibit shapes that are similar to those of malignant lesions, which may lead to misdiagnosis. The occurrence of false positives not only increases the consumption of medical resources, but also imposes a psychological burden on patients and contributes to rising healthcare costs.

To mitigate the issue of false positives, several technical improvements can be implemented, such as developing more advanced network architectures (e.g., Vision Mamba[42]) and more effective loss functions (e.g., exploration loss and consistency loss[16]). Additionally, incorporating clinical background information, such as age, family history, and medical history, can aid in distinguishing between benign and malignant lesions. For example, in patients with specific clinical profiles, lesions are more likely to be benign, allowing the model to adjust its prediction accordingly and reduce the rate of misdiagnosis. Lastly, multimodal detection, which integrates additional imaging data (e.g., X-ray, MRI), may also be considered.

5.1.4 Absence of public dataset

Public datasets serve as the foundation of the rapid development of DL, providing standardized and reproducible experimental data that are essential for model training and validation. However, the lack of publicly available ultrasound image datasets has long been a significant bottleneck in the development of DL applications of ultrasound imaging. Compared to other imaging modalities such as X-ray and CT, ultrasound image datasets are relatively limited, and there are currently no publicly available datasets for ABUS. This scarcity of public datasets complicates meaningful comparisons across studies and methods. When researchers train and test their models using proprietary datasets, they may encounter variations in data quality, annotation standards, and inherent biases, which hinder the ability to provide universally comparable performance metrics.

Addressing the lack of public datasets is crucial for advancing DL research and applications in ABUS. On one hand, hospitals, research institutions, and other organizations must enhance collaboration to establish and share high-quality datasets. On the other hand, the establishment of publicly available datasets would not only enhance the generalization and performance of DL models but also facilitate comparisons across different models, thereby promoting the standardization and advancement of technologies in this field.

5.2 Opportunities

5.2.1 Extending DL beyond lesion diagnosis

With the ongoing advancement of technology, the potential for integrating ultrasound and DL in breast cancer diagnosis is being explored across multiple domains. For instance, HHUS, in addition to being used for breast lesion diagnosis, has been applied in predicting lymph node metastasis[43][44], molecular subtype prediction[45], and upstaging of ductal carcinoma in situ (DCIS)[46]. While Huang et al.[18] have utilized DL to predict biomarkers (ER, PR, and HER2 expression) from ABUS images, the full potential of ABUS in these areas remains underexplored. Additionally, the use of DL and multimodal data including ABUS for predicting treatment efficacy[47] and prognostic evaluation in breast cancer warrants further investigation. These emerging research directions, which integrate imaging data with other clinical information such as genomic data and clinical indicators, present new opportunities for the development of personalized treatment strategies, enabling clinicians to make more precise decisions tailored to individual patient conditions.

5.2.2 Utilizing DL in more phases

DL can also be applied to other phases of ABUS-based breast examination. For example, in the quality control of ABUS images, DL models can assist in the automated identification of issues during the ABUS acquisition process, such as verifying proper nipple positioning, ensuring the regularity of breast shape, or assessing whether shadows behind the nipple are appropriately

controlled[48]. This can enhance image quality and improve diagnostic accuracy. In breast cancer screening, while clinicians commonly rely on BI-RADS classification, research on the automatic BI-RADS classification based on ABUS remains limited.

5.2.3 Exploiting multimodal DL for ABUS

Multiple studies[49][50] have demonstrated that combining ABUS with other imaging modalities can enhance diagnostic accuracy for breast cancer. Consequently, the development of multimodal AI systems represents a key direction for future research. In addition, the integration of ABUS with other ultrasound imaging techniques, such as Doppler and elastography[51] also warrants further exploration. Multimodal DL models can not only improve the detection performance for breast diseases but also expand the range of applications, as previously discussed.

5.2.4 Exploring DL-based large models

The advent of large model technologies has opened new possibilities for the application of DL in ABUS. For example, multimodal large models[52][53] which can simultaneously process diverse data types such as ultrasound images, clinical records, and genetic data, provide more comprehensive insights for integrated diagnosis. Additionally, universal segmentation models, such as SAM (Segment Anything Model)[54], exhibit robust foundational capabilities in image segmentation and can assist in accurately extracting lesion information from ABUS images, thereby improving both the accuracy and efficiency of disease diagnosis. With ongoing technological advancements, these emerging approaches are poised to significantly elevate the clinical utility of ABUS, improving the precision of breast cancer screening, diagnosis, and treatment.

Funding: This work was supported by National Natural Science Foundation of China (61971451, U22A20303), Innovative Province special construction foundation of Hunan Province (2019SK2131), the Science and Technology Innovation Program of Hunan Province (2021RC4016), Clinical Research Center for Medical Imaging in Hunan Province in China (2020SK4001), Hunan Cancer Hospital Climb Plan (ZX2021004) and Natural Science Foundation of Hunan Province (2023JJ60334)

Conflicts of Interest: The authors declare no conflicts of interest.

References

[1] International Agency for Research on Cancer (IARC). (2024). Cancer TODAY | IARC - Globocan 2022 (version 1.1). Retrieved from https://gco.iarc.who.int

[2] Vourtsis, A. (2019). Three-dimensional automated breast ultrasound: Technical aspects and first results. Diagnostic and Interventional Imaging, 100(10), 579–592.

[3] Sharma, B. P., & Purwar, R. K. (2024). Computer-aided detection and diagnosis of breast cancer: a review. ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal, 13, e31412-e31412.

[4] Wang, Y., Choi, E. J., Choi, Y., Zhang, H., Jin, G. Y., & Ko, S. B. (2020). Breast cancer classification in automated breast ultrasound using multiview convolutional neural network with transfer learning. Ultrasound in Medicine & Biology, 46(5), 1119-1132.

[5] Hejduk, P., Marcon, M., Unkelbach, J., Ciritsis, A., Rossi, C., Borkowski, K., & Boss, A. (2022). Fully automatic classification of automated breast ultrasound (ABUS) imaging according to BI-RADS using a deep convolutional neural network. European Radiology, 32(7), 4868–4878.

[6] Kim, D., Nam, C. M., Park, H., Jang, M., & Lee, K. J. (2022). Weakly supervised branch network with template mask for classifying masses in 3D automated breast ultrasound. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 3912-3919).

[7] Cao, X., Chen, H., Li, Y., Peng, Y., Wang, S., & Cheng, L. (2021). Dilated densely connected U-Net with uncertainty focus loss for 3D ABUS mass segmentation. Computer Methods and Programs in Biomedicine, 209, 106313.

[8] Pan, P., Chen, H., Li, Y., Cai, N., Cheng, L., & Wang, S. (2021). Tumor segmentation in automated whole breast ultrasound using bidirectional LSTM neural network and attention mechanism. Ultrasonics, 110, 106271.

[9] Lei, Y., He, X., Yao, J., et al. (2021). Breast tumor segmentation in 3D automatic breast ultrasound using mask scoring R-CNN. Medical Physics, 48(1), 204-214

[10] Cheng, Z., Li, Y., Chen, H., Zhang, Z., Pan, P., & Cheng, L. (2022). DSGMFFN: Deepest semantically guided multi-scale feature fusion network for automated lesion segmentation in ABUS images. Computer Methods and Programs in Biomedicine, 221, 106891.

[11] Luo, X., Xu, M., Tang, G., Wang, Y., Wang, N., Ni, D., Lin, X, Li, A. H. (2022). The lesion detection efficacy of deep learning on automatic breast ultrasound and factors affecting its efficacy: a pilot study. The British Journal of Radiology, 95(1130), 20210438.

[12] Li, Y., Wu, W., Chen, H., Cheng, L., & Wang, S. (2020). 3D tumor detection in automated breast ultrasound using deep convolutional neural network. Medical Physics, 47(11), 5669-5680.

[13] Zhang, Z., Li, Y., Wu, W., Chen, H., Cheng, L., & Wang, S. (2021). Tumor detection using deep learning method in automated breast ultrasound. Biomedical Signal Processing and Control, 68, 102677.

[14] Zhang, J., Tao, X., Jiang, Y., et al. (2022). Application of convolution neural network algorithm based on multicenter abus images in breast lesion det0065ction. Frontiers in Oncology, 12, 938413.

[15] Oh, K., Lee, S. E., & Kim, E. K. (2023). 3-D breast nodule detection on automated breast ultrasound using faster region-based convolutional neural networks and U-Net. Scientific Reports, 13(1), 22625.

[16] Li, Y., Zhang, Z., Sun, J., Chen, H., Chen, Z., & Wei, J. (2024). Tumor detection based on deep mutual learning in automated breast ultrasound. Multimedia Tools and Applications, 1-19.

[17] Lei, B., Huang, S., Li, R., Bian, C., Li, H., Chou, Y. H., & Cheng, J. Z. (2018). Segmentation of breast anatomy for automated whole breast ultrasound images with boundary regularized convolutional encoder–decoder network. Neurocomputing, 321, 178-186.

[18] Huang, Z., Zhang, X., Ju, Y., Zhang, G., Chang, W., Song, H., & Gao, Y. (2024). Explainable breast cancer molecular expression prediction using multi-task deep-learning based on 3D

whole breast ultrasound. Insights into Imaging, 15(1), 227.

[19] Torrey, L., & Shavlik, J. (2010). Transfer learning. In Handbook of research on machine learning applications and trends: algorithms, methods, and techniques (pp. 242-264). IGI global.

[20] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18 (pp. 234-241). Springer International Publishing.

[21] Yu, F. (2015). Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv:1511.07122.

[22] Zhang, S., Zheng, D., Hu, X., & Yang, M. (2015). Bidirectional long short-term memory networks for relation classification. In Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation (pp. 73-78).

[23] Oktay, O. (2018). Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999.

[24] He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision (pp. 2961-2969).

[25] Vaswani, A. (2017). Attention is all you need. Advances in Neural Information Processing Systems.

[26] Tran, D., Bourdev, L., Fergus, R., Torresani, L., & Paluri, M. (2015). Learning spatiotemporal features with 3d convolutional networks. In Proceedings of the IEEE International Conference on Computer Vision (pp. 4489-4497).

[27] Farhadi, A., & Redmon, J. (2018). Yolov3: An incremental improvement. In Computer Vision and Pattern Recognition (Vol. 1804, pp. 1-6). Berlin/Heidelberg, Germany: Springer.

[28] Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934.

[29] Lemay, A., Hoebel, K., Bridge, C. P., et al. (2022). Improving the repeatability of deep learning models with Monte Carlo dropout. npj Digital Medicine, 5(1), 174.

[30] Zhan, W., Sun, C., Wang, M., She, J., Zhang, Y., Zhang, Z., & Sun, Y. (2022). An improved Yolov5 real-time detection method for small objects captured by UAV. Soft Computing, 26, 361-373.

[31] Al-Khater, W., & Al-Madeed, S. (2024). Using 3D-VGG-16 and 3D-Resnet-18 deep learning models and FABEMD techniques in the detection of malware. Alexandria Engineering Journal, 89, 39-52.

[32] Ren, S., He, K., Girshick, R., & Sun, J. (2016). Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6), 1137-1149.

[33] Ge, Z. (2021). Yolox: Exceeding yolo series in 2021. arXiv preprint arXiv:2107.08430.

[34] Zhang, Y., Xiang, T., Hospedales, T. M., & Lu, H. (2018). Deep mutual learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 4320-4328).

[35] Lee, C. Y., Xie, S., Gallagher, P., Zhang, Z., & Tu, Z. (2015). Deeply-supervised nets. In Artificial Intelligence and Statistics (pp. 562-570). Pmlr.

[36] Parnami, A., & Lee, M. (2022). Learning from few examples: A summary of approaches to

few-shot learning. arxiv preprint arxiv:2203.04291.

[37] Celard, P., Iglesias, E. L., Sorribes-Fdez, J. M., Romero, R., Vieira, A. S., & Borrajo, L. (2023). A survey on deep learning applied to medical images: from simple artificial neural networks to generative models. Neural Computing and Applications, 35(3), 2291-2323.

[38] Liu, Y., Sun, P., Wergeles, N., & Shang, Y. (2021). A survey and performance evaluation of deep learning methods for small object detection. Expert Systems with Applications, 172, 114602.

[39] Jiao, L., Wang, M., Liu, X., et al. (2024). Multiscale Deep Learning for Detection and Recognition: A Comprehensive Survey. IEEE Transactions on Neural Networks and Learning Systems.

[40] Guo, M. H., Lu, C. Z., Liu, Z. N., Cheng, M. M., & Hu, S. M. (2023). Visual attention network. Computational Visual Media, 9(4), 733-752.

[41] Ross, T. Y., & Dollár, G. K. H. P. (2017, July). Focal loss for dense object detection. In proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2980-2988).

[42] Zhu, L., Liao, B., Zhang, Q., Wang, X., Liu, W., & Wang, X. (2024). Vision mamba: Efficient visual representation learning with bidirectional state space model. arXiv preprint arXiv:2401.09417.

[43] Zhou, L.-Q., Wu, X.-L., Huang, S.-Y., Wu, G.-G., Ye, H.-R., Wei, Q., Bao, L.-Y., Deng, Y.-B., Li, X.-R., Cui, X.-W., & Dietrich, C. F. (2020). Lymph Node Metastasis Prediction from Primary Breast Cancer US Images Using Deep Learning. Radiology, 294(1), 19–28.

[44] Gu, J., Tong, T., Xu, D., et al. (2023). Deep learning radiomics of ultrasonography for comprehensively predicting tumor and axillary lymph node status after neoadjuvant chemotherapy in breast cancer patients: A multicenter study. Cancer, 129(3), 356-366.

[45] Ma, M., Liu, R., Wen, C., et al. (2022). Predicting the molecular subtype of breast cancer and identifying interpretable imaging features using machine learning algorithms. European Radiology, 1-11.

[46] Qian, L., Lv, Z., Zhang, K., et al. (2021). Application of deep learning to predict underestimation in ductal carcinoma in situ of the breast with ultrasound. Annals of Translational Medicine, 9(4).

[47] Xie, Y., Chen, Y., Wang, Q., Li, B., Shang, H., & Jing, H. (2023). Early Prediction of Response to Neoadjuvant Chemotherapy Using Quantitative Parameters on Automated Breast Ultrasound Combined with Contrast-Enhanced Ultrasound in Breast Cancer. Ultrasound in Medicine & Biology, 49(7), 1638–1646.

[48] Chwaab, J., Diez, Y., Oliver, A., Martí, R., Zelst, J. V., Gubern-Mérida, A., Mourri, A. B., Gregori, J., & Günther, M. (2016). Automated quality assessment in three-dimensional breast ultrasound images. Journal of Medical Imaging, 3(2), 027002.

[49] Schäfgen, B., Juskic, M., Radicke, M., Hertel, M., Barr, R., Pfob, A., Togawa, R., Nees, J., Von Au, A., Fastner, S., Harcos, A., Gomez, C., Stieber, A., Riedel, F., Hennigs, A., Sohn, C., Heil, J., & Golatta, M. (2021). Evaluation of the FUSION-X-US-II prototype to combine automated breast ultrasound and tomosynthesis. European Radiology, 31(6), 3712–3720.

[50] Wilczek, B., Wilczek, H. E., Rasouliyan, L., & Leifland, K. (2016). Adding 3D automated breast ultrasound to mammography screening in women with heterogeneously and extremely dense breasts: Report from a hospital-based, high-volume, single-center breast cancer

screening program. European Journal of Radiology, 85(9), 1554–1563.

[51] Zhou, B. Y., Wang, L. F., Yin, H. H., et al. (2021). 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, 74.

[52] Tu, T., Azizi, S., Driess, D., et al. (2024). Towards generalist biomedical AI. NEJM AI, 1(3), AIoa2300138.

[53] Tanno, R., Barrett, D. G., Sellergren, A., et al. (2024). Collaboration between clinicians and vision–language models in radiology report generation. Nature Medicine, 1-10.

[54] Mazurowski, M. A., Dong, H., Gu, H., Yang, J., Konz, N., & Zhang, Y. (2023). Segment anything model for medical image analysis: an experimental study. Medical Image Analysis, 89, 102918.

Highlights

Application of deep learning on automated breast ultrasound: Current developments, challenges, and opportunities

Ruixin Wang, Zhiyuan Wang, Yuanming Xiao, Xiaohui Liu, Guoping Tan, Jun Liu

- Research highlights item 1：This article introduces the role of deep learning (DL) in the computer-aided diagnosis (CAD) for breast cancer based on Automated Breast Ultrasound (ABUS).
- Research highlights item 2：This article analyses the challenges faced with the DL-based CAD system for breast cancer when using ABUS.
- Research highlights item 3：The article explores future research opportunities in integrating DL with ABUS for breast cancer diagnosis.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The author *Click here to enter your name* is Choose an item for *Click here to enter the journal's name* and was not involved in the editorial review or the decision to publish this article.

☐The authors declare the following financial interests (e.g., any funding for the research project)/personal relationships (e.g., the author is an employee of a profitable company) which may be considered as potential competing interests:

Click here to enter your full declaration