Artificial intelligence based on automated breast volume scanner (ABVS) in breast nodules: A deep learning model for diagnosis

Wei Jiang (jiangwei1966nian@aliyun.com)
Huazhong University of Science and Technology Union Shenzhen Hospital

Yi Yang
Huazhong University of Science and Technology Union Shenzhen Hospital

Ting Zhu
Huazhong University of Science and Technology Union Shenzhen Hospital

Xiaoji Deng
Huazhong University of Science and Technology Union Shenzhen Hospital

Jing Fang
Huazhong University of Science and Technology Union Shenzhen Hospital

Research article

Keywords: artificial intelligence, breast, automated breast volume scanner, deep learning, ultrasound

Posted Date: February 18th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1085352/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
**Abstract**

**Objectives** We aimed to explore the use of artificial intelligence (AI) in differential diagnosis of benign and malignant breast nodules with automated breast volume scanner (ABVS).

**Methods** The AI model used in this study is based on the VNet, trained with a collection of breast nodules. Another collection of breast nodules were used to test the diagnostic accuracy among AI model, radiologists with different experiences, and AI-assisted junior radiologist. Sensitivity, specificity, accuracy, positive predictive value, negative predictive value and area under the curve (AUC) were used as indices to evaluate the efficiency on distinguishing between benign and malignant breast nodules of four different diagnostic groups.

**Results** The training cohort includes 100 breast nodules, the validation cohort includes 47 breast nodules. Among these 147 cases, 92 cases were benign (62.6%), 55 cases were malignant (37.4%). The AUC values of AI diagnosis, independent diagnosis by senior radiologist, independent diagnosis by junior radiologist, and AI-assisted diagnosis by junior radiologist were 0.839(95%CI, 0.703-0.930), 0.956(95%CI, 0.853-0.994), 0.814(95%CI, 0.674-0.912), and 0.963(95%CI, 0.863-0.996), respectively. The diagnostic efficiency of junior radiologist improved with the assistance of AI (AUC = 0.814 vs 0.963, Z=2.772, p <0.01). No significant difference in diagnostic ability was found between the AI-assisted junior radiologist and the senior radiologists (AUC = 0.963 vs 0.956, Z=0.158, p > 0.05).

**Conclusions** The AI model could distinguish between benign and malignant breast nodules with ABVS. AI model is the fastest diagnosis method with equivalent diagnosis efficiency with experienced radiologists, and could improve the diagnostic ability of junior radiologist.

**Background**

Breast cancer has been the most common malignant tumor, as well as the second leading cause of cancer death in women in the world[1]. Early detection, diagnosis and treatment are the most important management for breast cancer. Ultrasound has been used as an effective method for breast cancer screening and diagnosis, but ultrasound would be influenced by operator’s experience [2]. Automatic breast volume scanner (ABVS) is a fully automatic three-dimensional breast volume imaging. The streamlined operation method and standardized scanning mode are independent, which allows repeat review. Studies have shown that the consistency of ABVS diagnosis of breast mass is higher than ultrasound and the diagnostic performance of the ABVS is similar with that of ultrasound [3, 4]. ABVS could obtain arbitrary planar images through three-dimensional reconstruction, which helps to comprehensively and clearly display the morphology, margins, microcalcifications and ductal dilatation of breast lesions.

In recent years, artificial intelligence has gradually been applied to the ultrasound-assisted diagnosis of breast lesions with its advantages of high efficiency, objectivity, and strong reproducibility. It has been gradually applied in breast cancer screening and classification of breast nodules [5]. Early computer-
Aided diagnosis (CAD) is mainly considered to be traditional machine learning (ML), which requires humans to actively extract and encode image features. Deep learning (DL) avoids the dependence of machine learning algorithms on manual definition and extraction of features, and could autonomously extract massive, abstract and define image features to achieve end-to-end learning[6–8].

As we all know, ultrasound diagnosis has strong human dependence and weak radiologist consistency in diagnosis. In order to overcome the above shortcomings, this study puts forward the concept of combining ABVS technology with AI-assisted diagnosis. The goal is to build a stable AI model for breast nodule diagnosis.

**Methods**

**Participants**

The data included in our retrospective study are all from university-affiliated hospital. Patients in our hospital from August 2019 to February 2021 were enrolled. The research project was reviewed and approved by our hospital's ethics review board (No.LW-2021-006) and follows the Standards for Reporting Diagnostic Accuracy Studies [9]. With acknowledge, all participants had signed the informed consent for the specific purpose of data analysis in this study.

The inclusion criteria were (a) nodules diagnosed as BI-RADS 3 or above by conventional ultrasound (b) breast suitable for ABVS, no prosthesis, ulceration or infection, could tolerate pressure. (c) pathological results were confirmed.

The exclusion criteria were (a) nodule protruding from the surface of the skin (b) no final diagnosis

**ABVS protocol**

The research processes for the developing cohort and validation cohort were identical. Two radiologists (with 3 and 20 years of experience, respectively) practice the ABVS system and diagnose according to the images.

The ABVS equipment was Acuson Oxana 2 ABVS US system (Siemens Medical Solutions, Inc., Pleasonton, CA, USA), mounting a 5-14 MHz linear transducer (Siemens 14L5BV) on a flexible mechanical arm. Frequency was set according to breast cup size, cup A : 11 MHz, cup B : 10 MHz, cup C : 9 MHz, cup D and above: 9 MHz. The optimal preset condition is adjusted according to the patient's breast (A-D cup).

During the examination, the patient laid in the supine position with the side under examination placed overhead. With patient's hands lift upwards, the transducer was covered with a replaceable membrane to achieve adequate contact with the skin, and then applied to each breast with a gentle compression. Each breast was scanned with lateral, medial and anterior-posterior views. Additional views were acquired if better anatomical coverage of larger breasts was needed. For the purpose of correct orientation and image reformation, the nipple was used as a reference point in each scan view. Scan parameters for each
view were as follows, field of view 15.4 x 16.8 cm, depth 6 cm, image thickness 0.5 mm, transducer speed 16.8 cm/min, reconstructed slice thickness 0.5–1 mm, acquisition time 65 s. The total time to finish a three-view bilateral breast scan was 6 min. On average, the time for bilateral examination was 15 min (3-views bilateral scan time plus patient preparation time and probe positioning time). ABVS image database were saved in DICOM format.

Deep Learning Architecture

(1) ABVS Labelling

Import ABVS DICOM data into the itk-snap software, six-dimensional volumetric images of the patient’s bilateral breasts were used for labelling as follows: (1) amplify the suspicious breast nodules and recognize ‘the nodule’s margin, segment the nodule in layers of 0.5mm’s thickness on the direction from skin to muscle. (2) use the itk-snap software to automatically reconstruct the three-dimensional structure of the nodule (3) put the labelling data together with the original data into one file. The labelling results are shown in Fig. 1. The labelling is practiced by two radiologists, with 3 and 20 years of experience, respectively. If the reviewer is confused about the labelling, another radiologist with 20 years of experience would help to make the decision.

(2) Segmentation and Classification:

The segmentation and classification model used in this study is based on the VNet[10]. The network structure is as shown in Fig. 2:

This model is a multi-task deep learning model. The tasks of both segmentation and classification are carried out within the same network, and share the weight of this network. During the process of training, the tasks of segmentation and classification could generate some disturbances with each other, effectively limiting the over-fitting and improving the generalization ability of the model. The classification task of this model is to extract the features of the results of the four-layer encoder and decoder, and then perform the fully connected operation to map it into a probability vector to obtain the final classification result.

\[
Loss = \lambda_1 Loss_{cls} + (1 - \lambda_1) Loss_{seg}
\]

Among the parameters, Loss represents final loss, \(Loss_{cls}\) represents loss in classification task, \(Loss_{seg}\) represents loss in segmentation task, \(\lambda\) is a constant.
\[ \text{Loss}_{\text{cls}} = -\frac{1}{n} \sum_{i=1}^{n} (\alpha y_i (1 - p_i)^r \log(p) + (1 - \alpha)(1 - y_i)p_i^r \log(1 - p_i)) \]

\( \text{Loss}_{\text{cls}} \) adopts a focal loss, which is a dynamic scaling cross-entropy loss function. As the confidence of correct classification increases, the scaling factor continues to decrease. The value of \( \lambda \) is set as 2 to achieve a satisfied result. As the number of malignant and benign samples is not balanced, the ratio of benign to malignant is about 3.0, \( \alpha \) could correct the imbalance of the sample by amplifying the weight of the small sample loss. \( n \) is the total number of samples and \( y_i \) is the predicted category. This study is a two-category hypothesis, so \( y_i \) is 0 or 1, \( p_i \) is the probability value of the prediction as a positive sample.

\[ \text{Loss}_{\text{seg}} = \lambda_2 \sum_{i=1}^{n} \left(1 - \frac{2y_i \hat{y}_i + \varepsilon}{y_i + \hat{y}_i + \varepsilon}\right) + (1 - \lambda_2) \sum_{i=1}^{n} (-y_i \log y_i - (1 - y_i) \log(1 - y_i)) \]

The segmentation \( \text{Loss}_{\text{seg}} \) combines Dice loss and BCE loss, \( \lambda_2 \) is a constant, \( n \) is the total number of samples, \( y_i \) represents the predicted pixel category, \( \hat{y}_i \) represents the real pixel category, and \( \varepsilon \) is a very small perturbation coefficient to prevent the divisor from being 0.

This model sets up an attention module on the basis of VNet. Through the previous result of training cohort, the attention module could normalized and sum with the original data, the difference between the nodule area and the non-nodule area is significantly increased, to help the model focus on the nodule area.

**Image Analysis**

During the course of the study, the independent diagnosis results of the senior and junior radiologists on ABVS and diagnose time were recorded. The time that AI and radiologists spend, as well as the diagnosis results were recorded. The sensitivity, specificity, positive predictive value, negative predictive value, accuracy, and area under the ROC curve (AUC) of various diagnostic groups are calculated.

**Reference standard**

All breast nodules were sampled for biopsy using a cutting needle or surgery. The samples were fixed in 10% formalin, stained with hematoxylin-eosin and analyzed with immunohistochemistry to get the
pathological result. The radiologists, pathologists, and laboratory technicians were blinded to the results of images.

Statistical Analysis

Statistical analysis was performed with SPSS, version 23.0. The data were expressed in x±s. The sensitivity, specificity, accuracy and other parameters were compared using the X² test, and constructing ROC curve. The AUCs between different evaluators were compared by using Z test. P< 0.05 indicated a significant difference.

Results

Participant Characteristics

A total of 146 patients (mean age, 42.7±11.9, ranging from 18 to 72) with 147 nodules were eligible for the study; Among these 147 cases, 92 cases were benign (62.6%), the majority was fibroadenoma, 55 cases were malignant (37.4%), and the majority was invasive ductal carcinoma(Table 1, Fig. 4, Fig. 5); 147 original ABVS images of breast nodules were acquired.

<table>
<thead>
<tr>
<th>Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall Group(n=147)</td>
</tr>
<tr>
<td>TC(n=100)</td>
</tr>
<tr>
<td>Lesion size(cm)</td>
</tr>
<tr>
<td>Parallel diameter</td>
</tr>
<tr>
<td>Vertical diameter</td>
</tr>
<tr>
<td>Final diagnosis</td>
</tr>
<tr>
<td>Malignant</td>
</tr>
<tr>
<td>Benign</td>
</tr>
<tr>
<td>TC: Training cohort, VC: Validation cohort</td>
</tr>
</tbody>
</table>

Diagnostic Performance

The AI takes 15.6 minutes in diagnosis. The senior and junior radiologists takes 32.8 minutes and 60 minutes in diagnosis, respectively. The results of AI diagnosis, independent diagnosis by senior radiologist, independent diagnosis by junior radiologist, and AI-assisted diagnosis by junior radiologists are summarized, as shown in Table 2.
Table 2
Comparison of diagnosis results in each group

<table>
<thead>
<tr>
<th></th>
<th>Junior radiologist</th>
<th>Senior radiologist</th>
<th>AI</th>
<th>AI-assisted junior radiologist</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>85.00%(17/20)</td>
<td>95.00%(19/20)</td>
<td>90.00% (18/20)</td>
<td>100.00%(20/20)</td>
</tr>
<tr>
<td>Specificity</td>
<td>77.78%(21/27)</td>
<td>96.30%(26/27)</td>
<td>77.78% (21/27)</td>
<td>92.59%(25/27)</td>
</tr>
<tr>
<td>PPV</td>
<td>73.91%(17/23)</td>
<td>95.00%(19/20)</td>
<td>75.00% (18/24)</td>
<td>90.91%(20/22)</td>
</tr>
<tr>
<td>NPV</td>
<td>87.50%(21/24)</td>
<td>96.30%(26/27)</td>
<td>91.30% (21/23)</td>
<td>100.00%(25/25)</td>
</tr>
<tr>
<td>Accuracy</td>
<td>80.85%(38/47)</td>
<td>95.74%(45/47)</td>
<td>85.11% (41/47)</td>
<td>95.74%(45/47)</td>
</tr>
</tbody>
</table>

In this study, The ROC curves of four diagnostic groups are shown in Fig. 5, and the area under the diagnostic curve (AUC) of each group is also obtained. The AUC values of AI diagnosis, independent diagnosis by senior radiologist, independent diagnosis by junior radiologist, and AI-assisted diagnosis by junior radiologist were 0.839(95%CI, 0.703-0.930), 0.956(95%CI, 0.853-0.994), 0.814(95%CI, 0.674-0.912), and 0.963(95%CI, 0.863-0.996), respectively.

The AUC (0.956) diagnosed by senior radiologist was greater than the AUC(0.814) with junior radiologist(Z=2.778, p=0.0055) and the AUC(0.839) with AI diagnosis(Z=2.019, p =0.0435), but was similar with the AUC(0.963) diagnosed by AI-assisted junior radiologist(Z=0.158, p =0.8744). The AUC (0.963) diagnosed by AI-assisted junior radiologist was greater than the AUC (0.814) with the independent diagnosis by junior radiologist(Z=2.772, p =0.0056). The AUC (0.963) diagnosed by AI-assisted junior radiologist was greater than the AUC (0.814) with AI diagnosis(Z=2.534, p =0.0113). The AUC (0.839) diagnosed by AI was similar with the AUC (0.814) with the independent diagnosis by junior radiologist(Z=1.000, p =0.3173).

**Discussion**

In 2020, there were 279,100 new cases of breast cancer worldwide, replacing lung cancer, becoming the world’s largest cancer burden [1]. Ultrasound is an important breast imaging modality for diagnostic work-up to distinguish solid masses from cysts, and for screening in dense breasts. ABVS image data can also be stored in a standardized way, and the workstation system can be accessed at any time, which can be used for remote consultation and is also conducive to docking with AI technology.

AI technology has been applied to breast ultrasound in various applications.[11, 12]. It has been a valuable tool in breast medicine, it is mostly used in two-dimensional breast ultrasound image recognition. FU JoKa[13] use the automatic two-dimensional ultrasound imaging techniques based on AI...
Deep learning (DL) is the core technology of AI and its mainstream framework utilize an important algorithm of early machine learning, the artificial neural networks. Its principle is to simulate the multi-layer neural network structure of the human brain and make judgments. DL is roughly divided into two categories: (1) supervised learning (2) unsupervised learning. The difference between the two lies in whether the training data has been labelled when the relevant feature is extracted. If the data has labels, it is called supervised learning, and unlabeled is unsupervised learning. The convolutional neural network (CNN) in supervised learning is a research hotspot. This is a feed-forward artificial neural network that includes an input layer, a convolutional layer, an activation function, and a pooling layer and the fully connected layer, has the powerful function of extracting a series of abstract image features, and most of the cutting-edge research methods in ultrasound medicine also take this as the core.

CNN has great potential in the classification, detection, and segmentation of medical images. The CNN proposed by Krizhevsky et al [17] has attracted much attention in the ImageNet large-scale competition because its image recognition performance is better than traditional methods. Jiao et al. [18] developed a CNN-based CAD system in 2016 to classify breast cancer lumps. Moon et al. [19] constructed a 3D CNN network, and used a sliding window to extract VOI, and used focus loss and integrated learning methods to reduce data imbalance and other problems, reducing the false positives and false negatives of CAD target detection.

The model used in this study is also based on the 3D CNN network, which is improved under the framework of VNet. It is a multi-task model that has the functions of segmentation and classification at the same time, so that the two can promote each other to achieve the best classification effect. Another innovation of this research is to encode an operation interface. The radiologist only needs to input ABVS pictures, and according to the operation process, the malignant possible value of the mass can be obtained within a few seconds.

From the results of this study, in terms of independent diagnosis, the AI diagnostic statistical indicators were not statistically different from those of junior radiologist, proving that AI diagnosis efficiency is equivalent to that of junior radiologist. Under the same diagnostic efficiency, AI diagnosis takes only 15.6 minutes, and junior diagnosis takes 60 minutes, which proves that AI diagnosis is faster and more efficient, and it is worthy of promotion in heavy clinical work. The diagnostic specificity (96.3%) of senior radiologist is higher than AI diagnosis (77.78%) and junior radiologist diagnosis (77.78%). This may be related to senior radiologists being more cautious in the diagnosis of malignant lesions, but it is similar
with that of AI-assisted junior radiologist. In terms of diagnostic efficiency, the diagnostic AUC value (0.956) of senior radiologist is higher than the AUC value of AI (0.839). This is consistent with the actual clinical work situation. Senior radiologists have rich clinical experience, so the diagnosis is more accurate and the diagnostic efficiency is higher, and result also verified that AI could compensate for the deficiency.

In terms of AI-assisted diagnosis, since the diagnostic efficiency of senior radiologists is higher than those of AI and junior radiologist's diagnosis, the diagnosis results of senior radiologists assisted by AI have not been analyzed. The AUC value diagnosed by AI-assisted junior radiologist was significantly greater than that of independent diagnosis by AI or junior radiologist, proving the AI diagnosis result could provide certain reference for junior radiologists. Relevant literature has similar results with our study. Xu et al[20] found that the CAD system can effectively improve the accuracy of breast cancer diagnosis and reduce the reading time. CAD is more beneficial to radiologists who have no experience in ABVS diagnosis. Zelst et al[21] compared the diagnosis time of doctors with and without CAD assistance, and the specificity and positive predictive value were improved. Fujioka et al[22] used the GoogLeNet-based CNN model to identify benign and malignant breast masses in ultrasound examinations, and explored its diagnostic efficacy with three radiologists (with 4 years, 8 years, and 20 years of work experience, respectively). The results showed that compared with radiologists The CNN model has the same or better diagnostic performance. CAD is more valuable for assisted diagnosis by radiologists with less working years. In summary, AI has a certain clinical diagnostic value, and is more valuable for assisted diagnosis by younger radiologists.

In short, this study uses retrospective data and the existing AI model to conduct beneficial clinical trials on the artificial and intelligent assisted diagnosis of breast masses, and obtain certain valuable results, but out study also have some limitations. First, the caseload of breast nodules in our study was relatively small and larger population is needed for further studies, especially the training cohort. Second, the AI model used in this study was selected by AI researchers based on experience, and it would be better to apply multiple models to these cases for careful comparison and analysis. Finally, the effect of nodule size to the results could be further explored.

**Conclusions**

The AI model has ideal performance for the diagnosis of benign and malignant breasts in ABVS images, and can provide a certain reference value for clinical ultrasound diagnosis of breast nodules.

**Abbreviations**


<table>
<thead>
<tr>
<th>ABVS</th>
<th>automated breast volume scanner</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI</td>
<td>artificial intelligence</td>
</tr>
<tr>
<td>AUC</td>
<td>area under the curve</td>
</tr>
<tr>
<td>CAD</td>
<td>computer-aided diagnosis</td>
</tr>
<tr>
<td>ML</td>
<td>machine learning</td>
</tr>
<tr>
<td>DL</td>
<td>Deep learning</td>
</tr>
<tr>
<td>CNN</td>
<td>convolutional neural network</td>
</tr>
</tbody>
</table>

**Declarations**

**Ethics approval and consent to participate**

The research project was reviewed and approved by ethics review board (4403055625484) of Shenzhen Hospital, Huazhong University of Science and Technology Union (No.LW-2021-006). With acknowledge, all participants had signed the informed consent for the specific purpose of data analysis in this study.

**Consent for publication**

Not applicable

**Availability of data and materials**

The datasets of the current study are available from the corresponding author on reasonable request pending the approval of the institution(s) and trial/study investigators who contributed to the dataset.

**Competing interests**

Not applicable

**Funding**

Not applicable

**Authors’ contributions**

Wei Jiang, Yi Yang contributed to the design of the concept. Ting Zhu, Xiaoji Deng, Jing Fang contributed in collecting and reviewing the data. All authors contributed to the writing and reviewing of the paper. All authors read and approved the final manuscript. Wei Jiang is the corresponding author for this manuscript.

**Acknowledgments**
We sincerely appreciate all the patients who participated in this study.

References

219.


Figures

Figure 1
Labelling data of breast nodules

Figure 2
Structure of the multi-task deep learning network

Its basic structure consists of an encoder and a decoder: the left encoder is convolutional and downsampling extracts the input 3D image features, the decoder on the right restores the image resolution through upsampling and convolution, and at the same time realizes the residual connection with the encoder output of the same layer by skip connection.

Figure 3
(A) ABVS breast imaging (B) AI model diagnosis, malignant probable value is 0.72 (C) immunohistochemistry reveals Invasive ductal carcinoma.

Figure 4

(A) ABVS breast imaging (B) AI model diagnosis, malignant probable value is 0.37 (C) immunohistochemistry reveals fibroadenoma.

Figure 5
ROC of Al diagnosis, senior radiologist diagnosis, junior radiologist diagnosis, junior radiologist diagnosis assisted with Al.