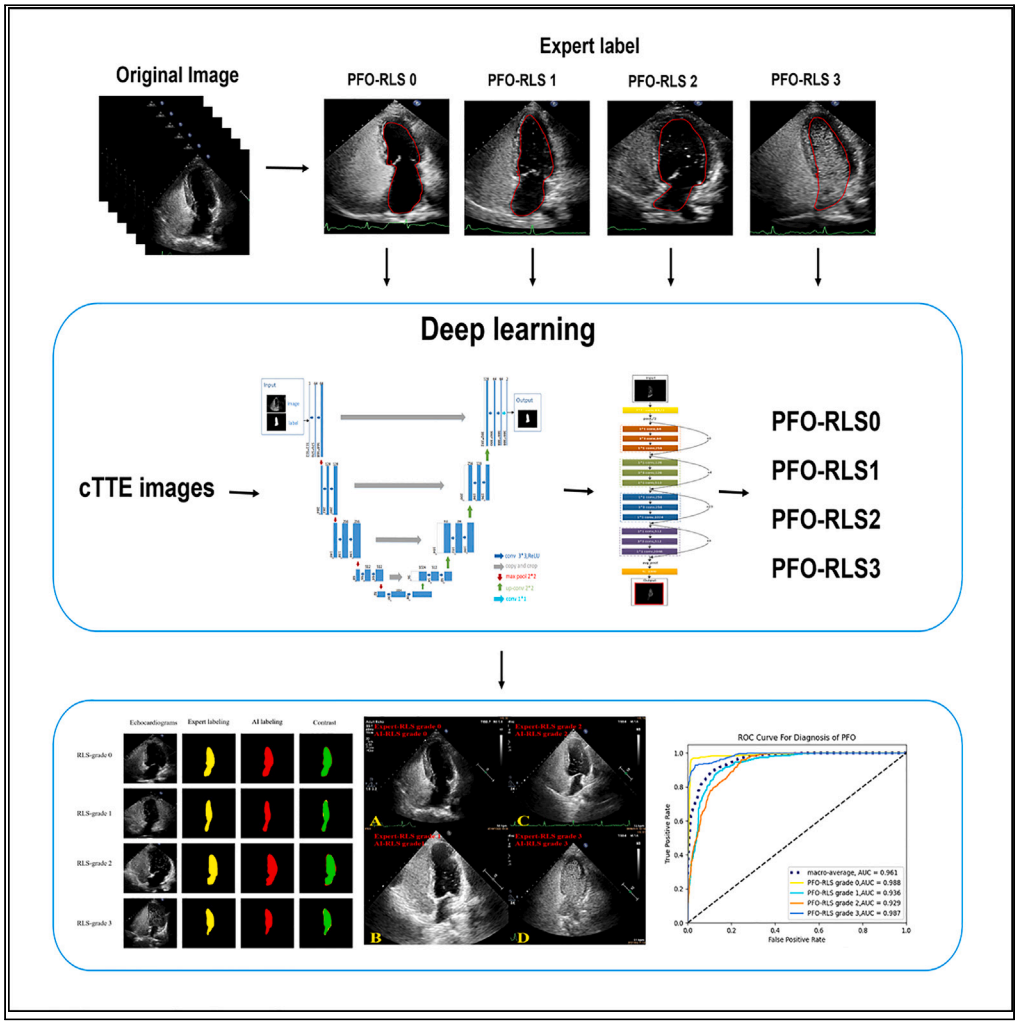
iScience

CellPress
OPEN ACCESS

Article

Artificial intelligence diagnosis of patent foramen ovale in contrast transthoracic echocardiography



Yuanyuan Sheng, Lixin Chen, Mengjie Gu, ..., Jinfeng Xu, Zhaohui Ning, Yingying Liu

jinfengxu@ext.jnu.edu.cn (J.X.)
ningzhaohui@haust.edu.cn (Z.N.)
yingyingliu@ext.jnu.edu.cn (Y.L.)

Highlights

AI model developed to detect the presence of PFO automatically in both cTTE images and videos

AI-based PFO-RLS classification model on par with intermediate-level physicians assessment

Sheng et al., iScience 27, 111012
November 15, 2024
https://doi.org/10.1016/j.isci.2024.111012

iScience

CellPress
OPEN ACCESS

Article

Artificial intelligence diagnosis of patent foramen ovale in contrast transthoracic echocardiography

Yuanyuan Sheng,[1,4] Lixin Chen,[1,4] Mengjie Gu,[2] Shuyu Luo,[1] Yuxiang Huang,[1] Xiaoxuan Lin,[1] Xiaohua Liu,[1] Qian Liu,[1] Xiaofang Zhong,[1] Guijuan Peng,[1] Jian Li,[1] Bobo Shi,[1] Lin Wang,[2] Jinfeng Xu,[1,*] Zhaohui Ning,[3,*] and Yingying Liu[1,5,*]

SUMMARY

Artificial intelligence (AI) is rarely directly used in patent foramen ovale (PFO) diagnosis. In this study, an AI model was developed to detect the presence of PFO automatically in both contrast transthoracic echocardiography (cTTE) images and videos. The whole intelligent diagnosis neural network framework consists of two functional modules of image segmentation (Unet, *n* = 1866) and image classification (ResNet 101, *n* = 9152). Finally, another test databases, including 20 cTTE videos (4609 cTTE images), was used to compare the RLS classification model accuracy between AI model and different levels of physicians. The Dice similarity coefficient of left chamber segmentation model of cTTE images was 91.41%, the accuracy of PFO-RLS classification model of cTTE images was 83.55%, the accuracy of PFO-RLS classification model of cTTE videos was 90%. Besides, the AI diagnosis time was significantly shorter than doctors (at only 1.3 s).

INTRODUCTION

The foramen ovale is an indispensable component of the fetal circulation. If it has failed closure after birth, the remaining fissure-like channel is called the patent foramen ovale (PFO). PFO has been reported to be closely associated with a variety of diseases, including migraine,[1] transient ischemic attack (TIA),[2] cryptogenic stroke,[3] peripheral arteriovenous embolism, and decompression sickness.[4] The incidence of PFO in young people with cryptogenic stroke is up to 46%[5] compared with 20–25% in the general population.[6] Recently, four large randomized controlled trials (RCTs)[3,7–9] have shown that transcatheter PFO closure is superior to medical therapy alone in reducing the risk of stroke recurrence. Therefore, it is essential to diagnose PFO quickly and accurately.

PFO is mainly examined and diagnosed by echocardiography combined with contrast of right heart, including contrast transthoracic echocardiography (cTTE), contrast transoesophageal echocardiography (cTEE), contrast-enhanced transcranial Doppler (cTCD).[6] Although cTEE examination can clearly show the structure (of atrial septum around the fossa ovalis), size and type of the PFO, and is considered the "gold standard" for the diagnosis of PFO.[6] But cTEE is a semitraumatic and painful examination requiring oropharyngeal intubation. Besides, during the examination, it is difficult for patients to cooperate when carrying out the Valsalva maneuver (VM). The cTEE RLS detection rate and shunt are lower than cTTE.[10,11] The sensitivity of cTCD is 68%–100% in the diagnosis of RLS. The disadvantage of cTCD is difficult to distinguish the source of RLS, which is often used to rule out a diagnosis of negative PFO in clinical practice. cTTE can display the cardiac structure and achieve high diagnostic sensitivity in PFO detection with sufficiently successful VM. Moreover, cTTE can distinguish the source of RLS according to the time of microbubble appearance in the left chambers, which has been widely used in clinical practice.[6]

However, the accuracy of these examination results depends too much on the diagnostic experience of echocardiographers, and there are differences in the diagnosis accuracy among different level of physicians. It is well known that it has a shortage of physicians in China. In 2020, there were only 2.41 licensed physicians per thousand residents,[12] while the international average was 3.6 per thousand residents.[13] By the end of 2020, only 5.7% of practicing physicians had senior professional titles in China, and 15.6% of practicing physicians had associate professional titles (data from the 2021 China Health Statistical Yearbook[12]). These findings established that there is an even greater deficit of experienced senior physicians in China. Moreover, performing simultaneous hierarchical diagnosis of PFO-RLS is time-consuming, requiring physicians to playback videos frame by frame, which greatly increases the time of diagnosis. Fortunately , with the rapid development of

[1]Shenzhen Medical Ultrasound Engineering Center, Department of Ultrasound, Shenzhen People's Hospital (The Second Clinical Medical College, Jinan University, The First Affiliated Hospital, Southern University of Science and Technology), Shenzhen 518020, China
[2]School of Information Engineering, Henan University of Science and Technology, Luoyang 471023, China
[3]The First Afffiliated Hospital, Henan University of Science and Technology, Luoyang 471003, China
[4]These authors contributed equally
[5]Lead contact
*Correspondence: jinfengxu@ext.jnu.edu.cn (J.X.), ningzhaohui@haust.edu.cn (Z.N.), yingyingliu@ext.jnu.edu.cn (Y.L.)
https://doi.org/10.1016/j.isci.2024.111012

Table 1. Basic information of cTTE images and videos

Variable	cTTE images	cTTE videos
Disease [PFO (%)/Normal (%)]	1389(74.4%)/477(25.6%)	15(75%)/5(25%)
Sex [Male (%)/Female (%)]	272(39.2%)/422(60.8%)	4(20%)/16(80%)
Age (Year)	31.0 ± 8.76	42.1 ± 9.58
Manufacturer	Philips EPIQ 7C (Probe S5-1,1-5MHz)	Philips EPIQ 7C (Probe S5-1,1-5MHz)
Heart beat rate	79.34 ± 9.46	78.85 ± 8.51
Systolic/Diastolic blood pressure (mmHg)	128.43/80.23	124.10/81.95
Patient's height(cm)	167.35 ± 6.47	163.25 ± 5.67
Patient's weight (Kg)	56.25 ± 7.56	59.75 ± 8.01
Spatial Size of 2D images(pixels)	800 × 600	800 × 600
Video length(frames)	/	230 ± 120

Data are represented as mean ± SD.
Unless otherwise specified, data in parentheses are percentages.

artificial intelligence, AI-assisted diagnosis has also been integrated into ultrasonic diagnosis field, which can replicate the diagnosis and treatment experience of senior physicians, thus helping physicians achieve higher diagnostic efficiency, save time and optimize the whole diagnosis process.

Now, AI-assisted ultrasonic diagnosis with the help of deep learning and computer vision has become a hot research topic. For example, ref. [14] proposed an ultrasound localization microscopy (ULM) method based on sub-pixel CNN to promote the performance of localizing microbubbles and achieve good results. Pu[15] proposed an automatic fetal ultrasound standard plane recognition model based the fusion of CNN and RNN in the IIoT environment and realized precise localization and tracking of fetal organs across frames. Ning[16] proposed a saliency-guided morphology-aware Unet model for lesion segmentation in breast ultrasound images and extensive experiments on five datasets demonstrate high performance and robustness. Hassan[17] proposed an automated prostate cancer classification algorithm based on CNN, which generates a maximum accuracy of 97% on ultrasound images and 80% on MRI images. Rhyou[18] proposed an end-to-end multi-step network called CirrhosisNet that includes two transfer-learned convolutional neural networks for semantic segmentation and classification tasks, which achieved an accuracy of 99.95%, a sensitivity of 100%, and a specificity of 99.9% on the Samsung Medical Center dataset using 8 × 8 pixels-sized μ-patches.

The previous related work has provided useful research ideas. To achieve higher accuracy and stability of PFO diagnosis, we established a team of echocardiographers to provide example labeling and final validation of the AI training. We propose an AI model based on Unet and Resnet101 network models for diagnosing PFO based on cTTE images, which provides a tool for assisting diagnosis of PFO in the clinical.

RESULTS

The PFO machine learning database included 1866 cTTE images and 20 cTTE videos. The basic information is shown in Table 1.

Target image determination (cTTE image left chamber segmentation)

The average Dice similarity coefficient of the evaluation index for left chamber segmentation model in the cTTE images was 91.41%, the average accuracy was 99.09%, the average recall was 89.55%, the average positive predictive value (PPV) was 93.97%, and the average input/output unit (IOU) percentage was 84.45%. See Figure 1 for the schematic diagram comparing left chamber expert annotation and AI segmentation. See the Videos S1, S2, S3, S4, S5, and S6 for the demonstration of the left chamber segmentation model effect of video in the test set.

Classification model of disease degree (cTTE image RLS grade)

The AI rate of correct PFO-RLS classification model of cTTE images was 83.55%. The classification model evaluation indexes of different PFO-RLS categories are shown in Table S1. The classification model accuracy for PFO-RLS grades 0 and 3 was higher, at 96.1% and 93.4%, respectively, while the classification model accuracy for PFO-RLS grades 1 and 2 was relatively low, at 73.9% and 70.8%, respectively. The confusion matrix of the cTTE image test group is shown in Table 2, and the ROC for the diagnosis of PFO is shown in Figure 2. Figure 3 shows the comparison between experts and AI in terms of the original cTTE images accurately identified by RLS classification model diagnosis and those that were incorrectly diagnosed.

Classification model of disease degree (cTTE video RLS grade)

Based on the PFO-RLS grade of the cTTE video judged by experts as the gold standard, the accuracy rate of the AI for cTTE video classification model was 90%, and the time required for diagnosis was significantly shorter in comparison to the time required by experts

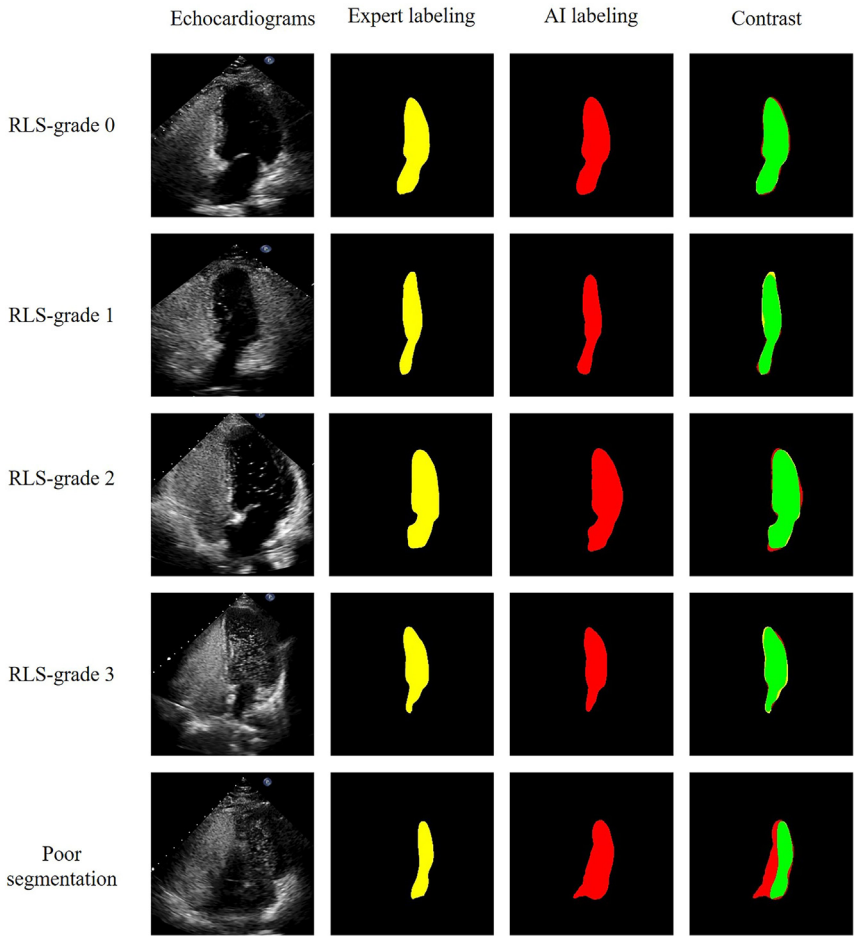


Figure 1. Schematic diagram of left chambers segmentation model of different grade of PFO-RLS cTTE images by AI

Yellow represents the region of the left chambers by the expert, red represents the region of the left chambers by the neural network, and green represents the right region of the left chambers predicted by the neural network.

(1.214 ± 0.648s vs. 37.43 ± 29.52s). The diagnosis accuracy rate of intermediate-level physicians was 90%, with a diagnosis time of 32.44 ± 25.60s, while the accuracy rate of junior physicians was 80%, with a diagnosis time of 56.49 ± 40.29s (as shown in Table 3). The confusion matrix of the cTTE video test group is shown in Table S2.

The physicians needed different amounts of time to diagnose different grades of PFO-RLS. It took them less time to determine PFO-RLS grades 0 and 3 and more time to determine PFO-RLS grades 1 and 2, up to 186.80 s. The same trend was observed in different groups of physicians with different levels of experience (as shown in Figure 4). The AI was consistent with intermediate-level physicians in terms of diagnostic accuracy, but the time required for the AI to make a diagnosis was significantly shorter: the diagnosis could be completed within 1.3 s, the repetition performance was good, and the performance did not differ by the PFO-RLS grade.

Interobserver and intraobserver consistency

Both the experts and the AI had a high degree of consistency in judging the PFO-RLS grade, with a kappa value of 0.800. The kappa value for the experts and the intermediate-level physicians was 0.867. The kappa value for the expert and junior physicians was 0.733. The expert intraobserver kappa value was 0.867, and the results of the two AI grades were identical, with a kappa value of 1.

DISCUSSION

In this study, a deep learning model based on cTTE images for segmentation of the left chambers and PFO-RLS classification was established. The segmentation model accuracy of the target left chambers was 91.41%, the RLS classification model accuracy of cTTE images was 83.55%, and the RLS classification model accuracy of cTTE videos was 90%. The AI was able to make the diagnosis of the PFO-RLS grade within 1.3 s, which was highly consistent with the expert diagnosis results. Therefore, AI has the potential not only to greatly reduce the diagnosis time but

Table 2. PFO-RLS classification model results using a deep learning network

		AI Prediction			
–	–	RLS-grade 0	RLS-grade 1	RLS-grade 2	RLS- grade 3
–	RLS-grade 0	438	18	0	0
PFO	RLS-grade 1	24	337	86	9
–	RLS-grade 2	1	73	323	59
–	RLS-grade 3	0	0	30	426
–	–	–	–	–	–

Confusion matrix presenting PFO-RLS classification model results of the 1824 echocardiograms included in the study. Every row of the matrix corresponds with the actual view presented to the deep learning algorithm and each row sums up to a total of 456 subjects. The values in each column correspond to the view prediction output by the deep learning network.

also to improve the diagnosis accuracy in clinical practice, which provides a more efficient and accurate method for use by clinicians (especially inexperienced junior physicians) when diagnosing PFO.

AI has been widely used in various parts of the medical field, and some preliminary explorations have been made in the specialized field of echocardiography. James et al. and other experts from several institutions have successfully constructed an AI model to automate the measurement of end-diastolic left ventricular anterior and posterior diameter (LVIDd), and the measurement error is close to the level of expert measurement.[19] Ivar [20] succeeded in automatically identifying three standard apical views, tracking the myocardium, performing motion estimation, and measuring GLS without any operator input and with fully automated measurements based on AI. AI was also used to segment the mitral annulus in 3D transoesophageal echocardiography.[21] In terms of disease diagnosis, a recent large-scale multicentre study used two-dimensional and color Doppler video learning to assist in the diagnosis of heart valve stenosis and insufficiency,[22] and this system achieved good recognition and diagnostic effects. At present, AI in the field of echocardiography mainly focuses on the identification and measurement of cardiac chamber diameter and function indexes, while there are few studies on AI used directly for disease diagnosis. In this study, DL model based on cTTE images for PFO diagnosis was proposed. The accuracy of our DL model for the segmentation of left chambers was 91.41%, and the accuracy of disease classification was 90%. There was a previous study on the diagnosis of PFO based on cTTE videos.[23] An EchoCP dataset was established in that study, and part of images in the cTTE video was captured to segment the four chambers. The UNet neural network was used for training, and its segmentation model accuracy was 89% and RLS classification accuracy was 67%. Compared with the EchoCP dataset, our study has higher image segmentation and classification accuracy. The analytical reasons may include the following points: (1) The EchoCP dataset adopts gray values for classification model, while in our study, the original image is directly selected to reduce image loss, and the AI model can obtain more original information from the image. (2) In our model, a series of complete and continuous static frame images in cTTE video is selected to retain more original information. (3) Our model more closely mimics the way clinicians think about disease diagnosis. Therefore, AI model obtain deeper image information and avoid network degradation phenomena by introducing channel attention and spatial attention mechanisms. Our AI model, which is based on classic ResNet-101 as a network, considering both the importance of different channels of pixels and the importance of different locations on the same channel pixels to achieve the purpose of characteristic elaboration. Thus, relatively high classification model accuracy can be obtained.

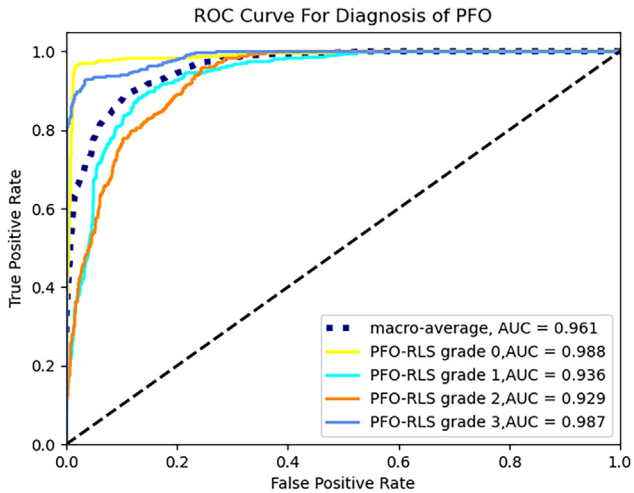


Figure 2. Receiver operating characteristic curves for the DL model's predictions of negative and positive diagnosis of PFO, and PFO-RLS class

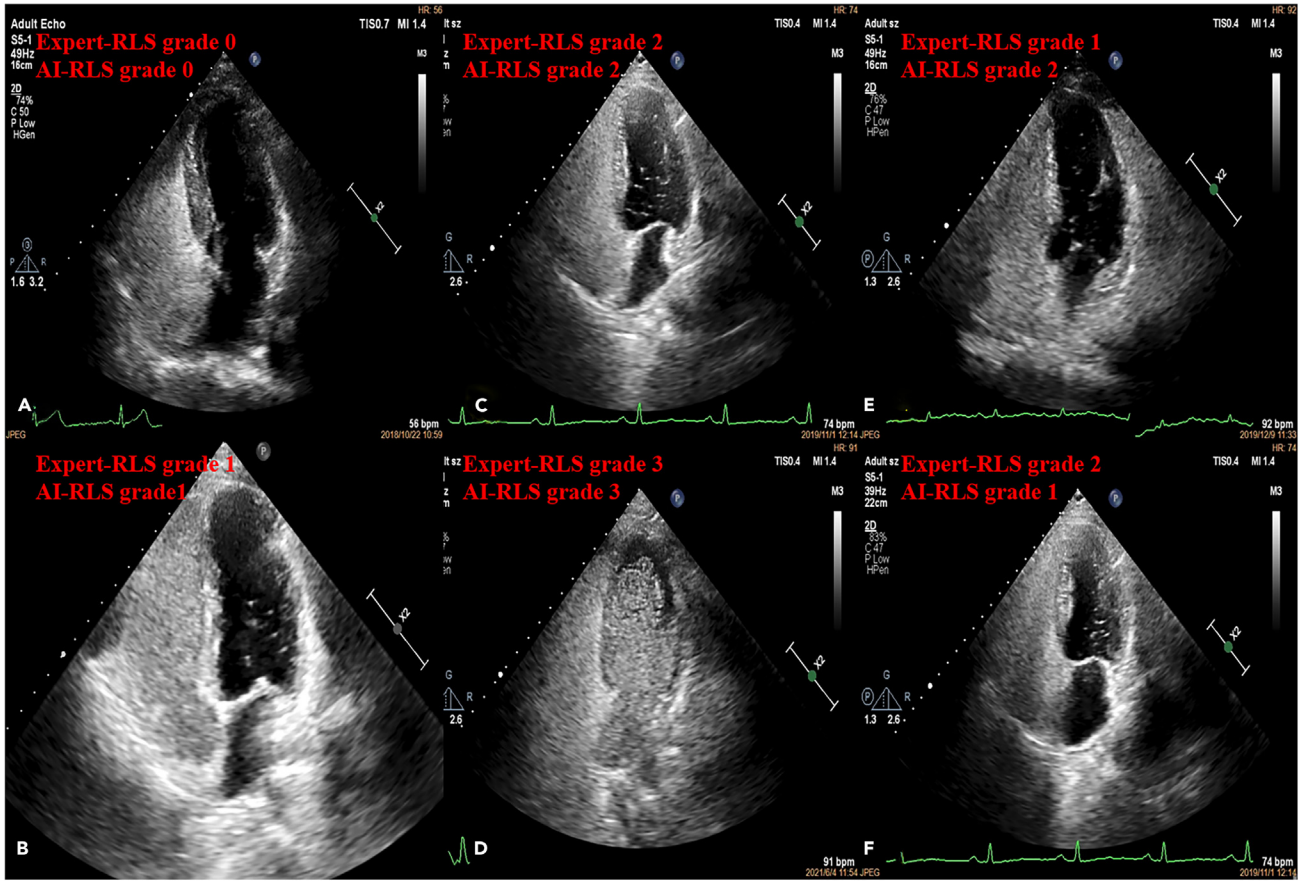


Figure 3. The original image comparison diagram of RLS classification model of cTTE static images by experts and AI

(A–D) indicates that AI is consistent with RLS grade results judged by experts; E and F indicates that the AI is inconsistent with the RLS grade results judged by experts.

We used the Unet to segment cTTE images. This choice is based on the excellent design of Unet in the field of biomedical image pixel-level segmentation, especially adept at dealing with image challenges such as low contrast, high noise, and complex structures, which are precisely reflected in ultrasound images. The skip connection mechanism of Unet skillfully transfers the features of the encoder directly to the corresponding layers of the decoder, ensuring the complete preservation of key edge and texture information during image reconstruction. This is particularly crucial for accurately identifying the heart boundaries in ultrasound images. In addition, compared to other complex

Table 3. Diagnostic accuracy and diagnostic time of PFO-RLS classification model by different levels of physicians

Group	Accuracy	Average Diagnosis time (s)	PFO grade 0 Diagnosis time (s)	PFO grade 1 Diagnosis time (s)	PFO grade 2 Diagnosis time (s)	PFO grade 3 Diagnosis time (s)
AI	17/20(85%)	1.214 ± 0.648[c,b,a]	1.241 ± 0.840[c,b,a]	1.205 ± 0.631[c,b,a]	1.211 ± 0.687[c,b,a]	1.198 ± 0.494[c,b]
Experts	20/20(100%)	37.43 ± 29.52[c]	17.37 ± 12.94[c]	73.21 ± 19.36	51.79 ± 9.67[c]	7.17 ± 4.26[c]
intermediate-Level physicians	18/20(90%)	32.44 ± 25.60[c]	14.94 ± 8.79[c]	41.37 ± 13.43[c]	52.28 ± 30.64[c]	21.18 ± 25.25[c]
junior physicians	16/20(80%)	56.49 ± 40.29	28.91 ± 15.06	70.10 ± 25.79	84.59 ± 51.94	42.35 ± 35.98

Experts:>10 years of experience, intermediate-level physicians: 5–10 years of experience, and junior physicians:<5 years of experience.
Diagnosis time data are represented as mean ± SD.
Unless otherwise specified, data in parentheses are percentages.
[a]means that it takes less time than expert, [a]$p < 0.05$.
[b]means that it takes less time than intermediate-level physicians, [b]$p < 0.05$.
[c]means that it takes less time than junior physicians,[c]$p < 0.05$.

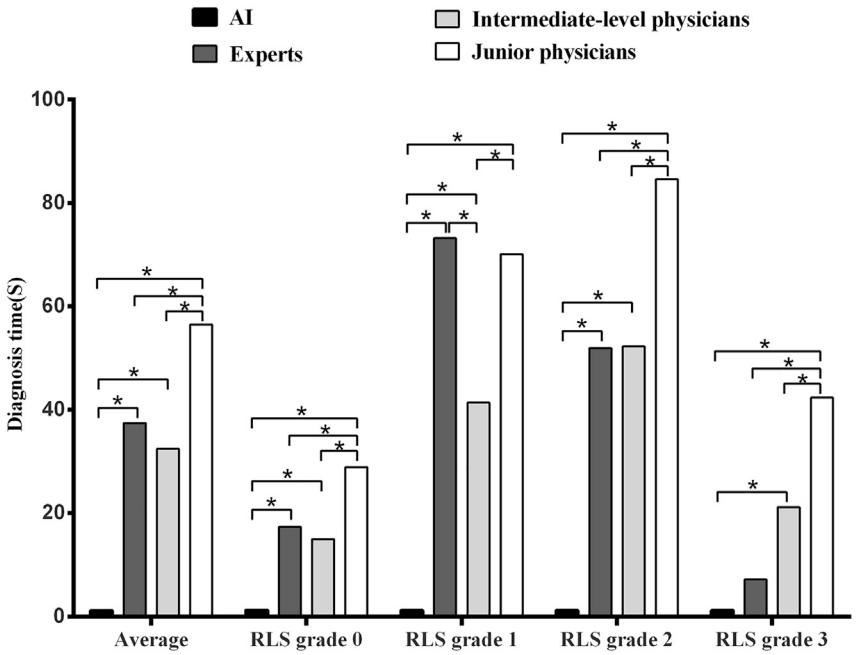


Figure 4. Comparison of the time required for diagnosis of PFO-RLS classification model between physicians with different seniority and AI

* indicates a statistically significant difference between the two groups.

network structures, Unet is simpler and more direct, effectively reducing the training time and computational resource consumption. In the PFO classification task, we adopted the ResNet-101. It successfully solved the problem of vanishing gradients by introducing residual blocks, making it possible to train deep neural networks with hundreds of layers. The design of residual blocks allows the input to be directly passed to the output through skip connections, effectively preventing performance degradation in deep networks and significantly improving training efficiency and model performance. Especially in image classification tasks, ResNet has shown excellent performance. Using the pre-trained ResNet-101 model, we achieved fast and high-precision image classification with transfer learning technology, which greatly saved training time and improved classification accuracy. We also explored various other architectures during our research, including Attention UNet, VGG, EfficientNet, EfficientNetV2, ResNet-34, ResNet-50, etc. After a comprehensive comparison of model performance, including accuracy, training time, and resource consumption, we ultimately selected Unet and ResNet101 as our preferred models.

In this study, the accuracy rate of the AI model for judging PFO-RLS classification in cTEE videos reached 90%, which was better than junior physician group and equal to intermediate-level physician group. Although there is no great advantage in the accuracy of PFO diagnosis, the AI model does greatly shorten the required diagnosis time. When physicians evaluate the PFO-RLS grade, it takes a short time for them to judge the RLS grades 0 and 3 because when there are no microbubbles or when there are a large number of microbubbles in the left chambers, the image characteristics are obvious, and easy to identify, and it is unnecessary to spend time counting the microbubbles in the left chambers. We confirmed this in our study. The classification model accuracy of our AI model in judging PFO-RLS grades 0 and 3 is extremely high and is equal to expert. However, for RLS grades 1 and 2, it takes a long time for physicians to make a judgment, up to 3 min, because when the number of microbubbles is intermediate, clinicians need to make repeated comparisons when judging RLS grade, and must select the image with the largest number of microbubbles in the left chambers in the video and make a final judgment by counting the microbubbles. The classification model accuracy in judging PFO-RLS grades 1 and 2 was 73.9% and 70.8%, respectively, which were significantly reduced. Reviewing the incorrectly classified images or the corresponding videos, we find that it is most difficult to determine the classification when there are approximately 10 left chambers microbubbles and because the images lose the correlation between the frames before and after the video. It is difficult to identify microbubbles when they are clustered or close to the left ventricular endocardial surface, and the small diameter weak echo of microbubbles also increases the difficulty of identification. Therefore, the average image classification accuracy in diagnosing PFO-RLS at all grades was not significantly better than middle and senior-level physicians. Influenced by many factors, the consistency of the two diagnoses made by the same physician will also decrease. It is mentioned earlier in this article that the intraobserver consistency kappa value was 0.867, and that the AI model was basically able to reach 100%. This highlights one of the great strengths of AI, which is that the time it takes to make a diagnosis is independent of these extrinsic factors and is reproducible. The average diagnosis time was approximately 1.3 s, and the repeatability test results produced basically the same result. We found 2 cases of incorrect diagnosis made by the AI model, which may be due to left atrial ultrasound artifacts or to the interference of large chordae tendineae and papillary muscle echo in the left ventricular endocardial muscle trabeculae and left cardiac cavity, leading to the incorrect identification of microbubbles. At the same time, we also found that the AI correctly classified it in one of the cTTE videos after PFO closure,

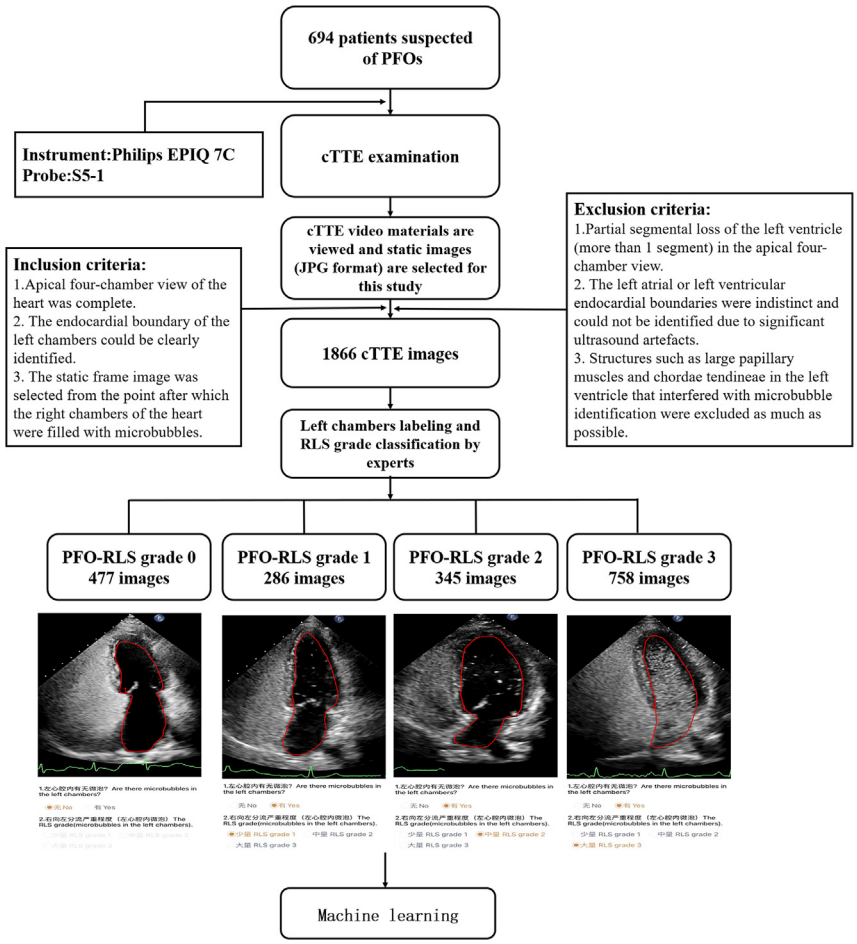


Figure 5. Flow chart of database establishment and image annotation

which indicates good prospects for the DL model to also be applicable to patients after PFO closure in the future, which needs to be further explored in large-scale studies.

Limitations of the study

There are some limitations in this study. First, the establishment of the PFO database from a single research center, but multicentre large sample data would be more helpful to improve the accuracy of the machine learning. Second, the segmentation model and RLS classification model of cTTE images of the left chambers were based on the results of expert labeling and diagnosis as the gold standard. In the cTTE video test set, physicians were able to observe the preceding and subsequent frames to better identify microbubbles, while the AI judgment results were only based on static images without a continuous relationship between images. Third, this study only used ultrasound instruments from a single manufacturer to collect, screen, and establish a cTTE images database, which may mean the findings are not applicable to ultrasound instruments made by different manufacturers. Finally, the AI model was only used to perform simple RLS classification model for PFO, but it could not accurately show the number of microbubbles on the image for accurate quantitative analysis, which is an aspect we intend to address with the direction of our future research.

Conclusions

In this study, an AI model was proposed for PFO diagnosis based on cTTE images. The accuracy rate of left chamber segmentation model was 91.41%, the accuracy rate of RLS classification model of cTTE images was 83.55%, the accuracy rate of RLS classification model of cTTE videos was 90%, and the AI diagnosis time was significantly shorter than doctors (at only 1.3 s).

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Yingying Liu (yingyingliu@ext.jnu.edu.cn).

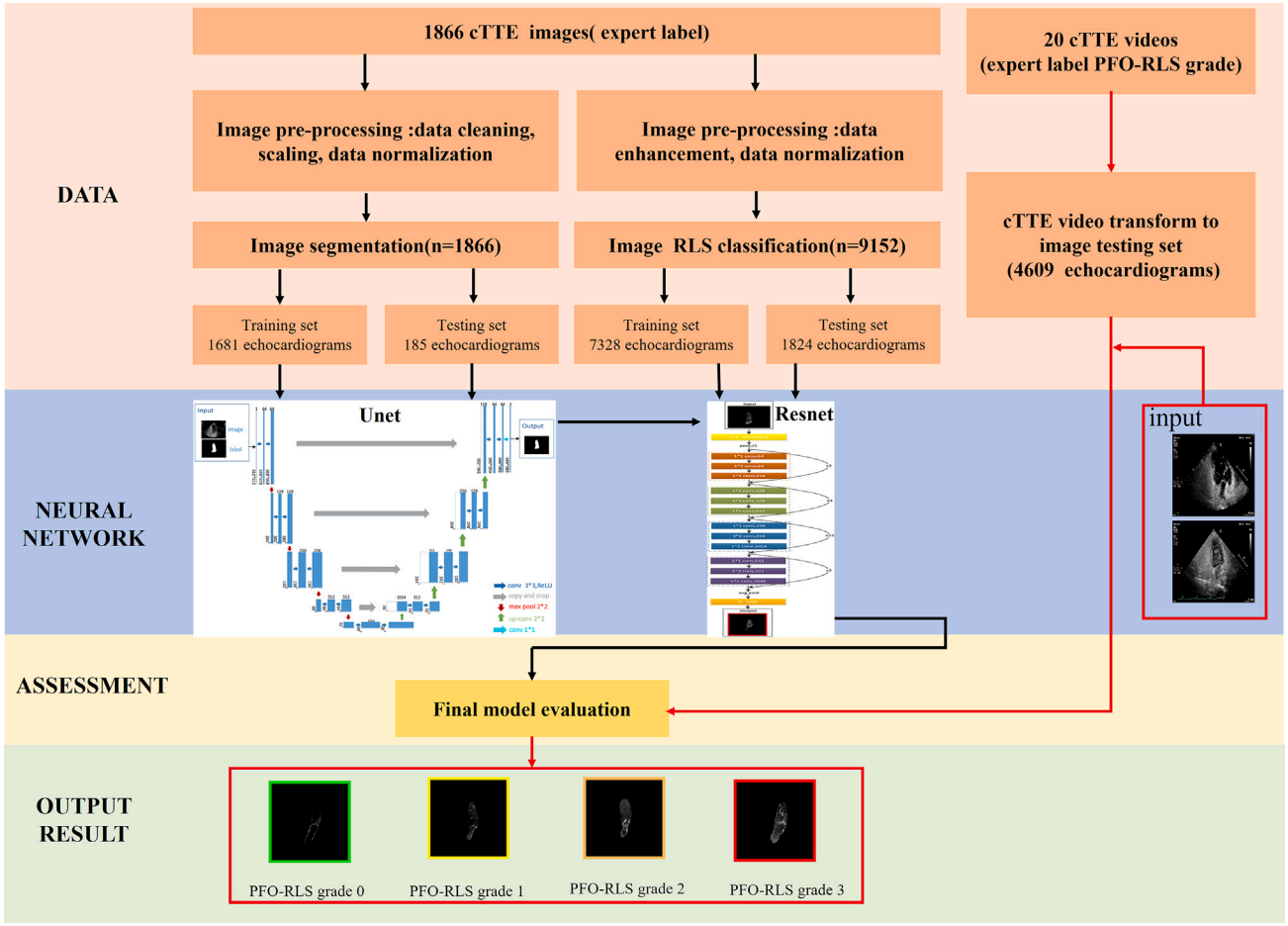


Figure 6. System pipeline

The neural network contains 1866 cTTE images. In the image segmentation model part, 1681 images were used for training the network and 185 images were used for testing the network. For the image classification model part, 7328 images were used for training the network and 1824 images were used for testing the network. Finally, we evaluated the performance of the network on a dataset of 20 cTTE videos marked by 2 experts.

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Data are not publicly shared but is available upon reasonable request from the lead contact.
- All original code has been deposited at Mendeley Data and is publicly available as of the date of publication. DOIs are listed in the key resources table (https://github.com/Mjiegu/PFO_AIdis).
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

This research was funded by the National Natural Science Foundation of China (Grant No. 82102041); and a Project of Innovation of the Science and Technology Commission of Shenzhen City (JCYJ20210324113804013); a Project of International cooperative research of the Science and Technology Commission of Shenzhen City(GJHZ20210705142205017, GJHZ20210705142206019). Shenzhen People's Hospital Clinnician Scientist Training Plan (SYWGSCGZH202302). Natural Science Foundation of Guangdong Province (2024A1515030143).

AUTHOR CONTRIBUTIONS

J.X.: conceptualization, investigation, supervision, funding acquisition. Z.N.: conceptualization, methodology, investigation, supervision. Y.L.: conceptualization, methodology, analyzed the data, writing—review and editing, supervision, funding acquisition. Y.S.: resources, methodology, analyzed the data, writing—original draft. M.G. and L.C.: resources, analyzed the data, methodology. S.L.: analyzed the data, methodology. Y.H. and X.L.: analyzed the data, resources. X.L., Q.L., and X.Z.: software, analyzed the data. M.G. and G.P.: software, analyzed the data. J.L., B.S., and L.W.: analyzed the data, resources, methodology.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Establishment of the cTTE images database
  - Labeling of images
  - Dataset allocation
  - Deep learning model
  - Image segmentation-Unet neural network model
  - Image classification–ResNet 101 neural network model
  - Model testing
  - left chamber segmentation and PFO-RLS classification
  - Evaluations of experts and physicians
- QUANTIFICATION AND STATISTICAL ANALYSIS
- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.111012.

Received: April 14, 2024
Revised: September 5, 2024
Accepted: September 18, 2024
Published: September 21, 2024

REFERENCES

1. Liu, K., Wang, B.Z., Hao, Y., Song, S., and Pan, M. (2020). The Correlation Between Migraine and Patent Foramen Ovale. Front. Neurol. *11*, 543485. https://doi.org/10.3389/fneur.2020.543485.
2. Mazzucco, S., Li, L., Binney, L., and Rothwell, P.M.; Oxford Vascular Study Phenotyped Cohort (2018). Prevalence of patent foramen ovale in cryptogenic transient ischaemic attack and non-disabling stroke at older ages: a population-based study, systematic review, and meta-analysis. Lancet Neurol. *17*, 609–617. https://doi.org/10.1016/s1474-4422(18)30167-4.
3. Lee, P.H., Song, J.K., Kim, J.S., Heo, R., Lee, S., Kim, D.H., Song, J.M., Kang, D.H., Kwon, S.U., Kang, D.W., et al. (2018). Cryptogenic Stroke and High-Risk Patent Foramen Ovale: The DEFENSE-PFO Trial. J. Am. Coll. Cardiol. *71*, 2335–2342. https://doi.org/10.1016/j.jacc.2018.02.046.
4. Honek, J., Sramek, M., Honek, T., Tomek, A., Sefc, L., Januska, J., Fiedler, J., Horvath, M., Novotny, S., Brabec, M., et al. (2022). Screening and Risk Stratification Strategy Reduced Decompression Sickness Occurrence in Divers With Patent Foramen Ovale. JACC Cardiovasc Imaging *15*, 181–189. https://doi.org/10.1016/j.jcmg.2021.06.019.
5. Lamy, C., Giannesini, C., Zuber, M., Arquizan, C., Meder, J.F., Trystram, D., Coste, J., and Mas, J.L. (2002). Clinical and imaging findings in cryptogenic stroke patients with and without patent foramen ovale: the PFO-ASA Study. Atrial Septal Aneurysm. Stroke *33*, 706–711. https://doi.org/10.1161/hs0302.104543.
6. Silvestry, F.E., Cohen, M.S., Armsby, L.B., Burkule, N.J., Fleishman, C.E., Hijazi, Z.M., Lang, R.M., Rome, J.J., and Wang, Y.; American Society of Echocardiography; Society for Cardiac Angiography and Interventions (2015). Guidelines for the Echocardiographic Assessment of Atrial Septal Defect and Patent Foramen Ovale: From the American Society of Echocardiography and Society for Cardiac Angiography and Interventions. J. Am. Soc. Echocardiogr. *28*, 910–958. https://doi.org/10.1016/j.echo.2015.05.015.
7. Mas, J.L., Derumeaux, G., Guillon, B., Massardier, E., Hosseini, H., Mechtouff, L., Arquizan, C., Béjot, Y., Vuillier, F., Detante, O., et al. (2017). Patent Foramen Ovale Closure or Anticoagulation vs. Antiplatelets after Stroke. N. Engl. J. Med. *377*, 1011–1021. https://doi.org/10.1056/NEJMoa1705915.
8. Sondergaard, L., Kasner, S.E., Rhodes, J.F., Andersen, G., Iversen, H.K., Nielsen-Kudsk, J.E., Settergren, M., Sjostrand, C., Roine, R.O., Hildick-Smith, D., et al. (2017). Patent Foramen Ovale Closure or Antiplatelet Therapy for Cryptogenic Stroke. N. Engl. J. Med. *377*, 1033–1042. https://doi.org/10.1056/NEJMoa1707404.
9. Saver, J.L., Carroll, J.D., Thaler, D.E., Smalling, R.W., MacDonald, L.A., Marks, D.S., and Tirschwell, D.L.; RESPECT Investigators (2017). Long-Term Outcomes of Patent Foramen Ovale Closure or Medical Therapy after Stroke. N. Engl. J. Med. *377*, 1022–1032. https://doi.org/10.1056/NEJMoa1610057.
10. Yang, X., Wang, H., Wei, Y., Zhai, N., Liu, B., and Li, X. (2020). Diagnosis of Patent Foramen Ovale: The Combination of Contrast Transcranial Doppler, Contrast Transthoracic Echocardiography, and Contrast Transesophageal Echocardiography. BioMed Res. Int. *2020*, 8701759. https://doi.org/10.1155/2020/8701759.
11. Zuber, M., Cuculi, F., Oechslin, E., Erne, P., and Jenni, R. (2008). Is transesophageal echocardiography still necessary to exclude patent foramen ovale? Scand. Cardiovasc. J. *42*, 222–225. https://doi.org/10.1080/14017430801932832.
12. National Health Commission of the People's Republic of China (2021). China Health Statistics Yearbook (Chinese Peking Union Medical College Press).
13. Health at a Glance (2021). OECD Indicators OECD (Publishing Paris). https://www.oecd-ilibrary.org/social-issues-migration-health/health-at-a-glance-2021_ae3016b9-en.
14. Liu, X., Zhou, T., Lu, M., Yang, Y., He, Q., and Luo, J. (2020). Deep learning for ultrasound localization microscopy. IEEE Trans. Med. Imag. *39*, 3064–3078. https://doi.org/10.21227/jdgd-0379.
15. Pu, B., Li, K., Li, S., and Zhu, N. (2021). Automatic Fetal Ultrasound Standard Plane Recognition Based on Deep Learning and IIoT. IEEE Trans. Industr. Inform. *17*, 7771–7780. https://doi.org/10.1109/tii.2021.3069470.
16. Ning, Z., Zhong, S., Feng, Q., Chen, W., and Zhang, Y. (2022). SMU-Net: Saliency-Guided Morphology-Aware U-Net for Breast Lesion Segmentation in Ultrasound Image. IEEE Trans. Med. Imag. *41*, 476–490. https://doi.org/10.1109/tmi.2021.3116087.
17. Hassan, M.R., Islam, M.F., Uddin, M.Z., Ghoshal, G., Hassan, M.M., Huda, S., and Fortino, G. (2022). Prostate cancer classification from ultrasound and MRI images using deep learning based Explainable Artificial Intelligence. Future Generat. Comput. Syst. *127*, 462–472. https://doi.org/10.1016/j.future.2021.09.030.
18. Rhyou, S.Y., and Yoo, J.C. (2023). Aggregated micropatch-based deep learning neural

network for ultrasonic diagnosis of cirrhosis. Artif. Intell. Med. *139*, 102541. https://doi.org/10.1016/j.artmed.2023.102541.

19. Howard, J.P., Stowell, C.C., Cole, G.D., Ananthan, K., Demetrescu, C.D., Pearce, K., Rajani, R., Sehmi, J., Vimalesvaran, K., Kanaganayagam, G.S., et al. (2021). Automated Left Ventricular Dimension Assessment Using Artificial Intelligence Developed and Validated by a UK-Wide Collaborative. Circ. Cardiovasc. Imaging *14*, e011951. https://doi.org/10.1161/circimaging.120.011951.
20. Salte, I.M., Østvik, A., Smistad, E., Melichova, D., Nguyen, T.M., Karlsen, S., Brunvand, H., Haugaa, K.H., Edvardsen, T., Lovstakken, L., and Grenne, B. (2021). Artificial Intelligence for Automatic Measurement of Left Ventricular Strain in Echocardiography. JACC. Cardiovasc. Imaging *14*, 1918–1928. https://doi.org/10.1016/j.jcmg.2021.04.018.
21. Andreassen, B.S., Veronesi, F., Gerard, O., Solberg, A.H.S., and Samset, E. (2020). Mitral Annulus Segmentation Using Deep Learning in 3-D Transesophageal Echocardiography. IEEE J. Biomed. Health Inform. *24*, 994–1003. https://doi.org/10.1109/JBHI.2019.2959430.
22. Yang, F., Chen, X., Lin, X., Chen, X., Wang, W., Liu, B., Li, Y., Pu, H., Zhang, L., Huang, D., et al. (2022). Automated Analysis of Doppler Echocardiographic Videos as a Screening Tool for Valvular Heart Diseases. JACC. Cardiovasc. Imaging *15*, 551–563. https://doi.org/10.1016/j.jcmg.2021.08.015.
23. Wang, T.C., Li, Z.H., Huang, M.P., Zhuang, J., Bi, S.S., Zhang, J.W., Shi, Y.Y., Fei, H.W., and Xu, X.Y. (2021). EchoCP: An Echocardiography Dataset in Contrast Transthoracic Echocardiography for Patent Foramen Ovale Diagnosis. In MICCAI (Springer), pp. 506–515.
24. van Rossum, G. (1995). Python Reference Manual (Department of Computer Science).
25. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al. (2019). Pytorch: An imperative style, high-performance deep learning library. Adv. Neural Inf. Process. Syst. *32*.
26. Bradski, G. (2000). The opencv library. Dr. Dobb's J. Softw. Tools Prof. Program. *25*, 120–123.
27. Harris, C.R., Millman, K.J., Van Der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., et al. (2020). Array programming with NumPy. Nature *585*, 357–362.
28. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. *12*, 2825–2830.
29. Gu, M.J. (2024). Code for PFO. https://github.com/Mjiegu/PFO_AIdis.
30. Ben-Assa, E., Rengifo-Moreno, P., Al-Bawardy, R., Kolte, D., Cigarroa, R., Cruz-Gonzalez, I., Sakhuja, R., Elmariah, S., Pomerantsev, E., Vaina, L.M., et al. (2020). Effect of Residual Interatrial Shunt on Migraine Burden After Transcatheter Closure of Patent Foramen Ovale. JACC Cardiovasc. Interv. *13*, 293–302. https://doi.org/10.1016/j.jcin.2019.09.042.
31. Martin, F., Sanchez, P.L., Doherty, E., Colon-Hernandez, P.J., Delgado, G., Inglessis, I., Scott, N., Hung, J., King, M.E., Buonanno, F., et al. (2002). Percutaneous transcatheter closure of patent foramen ovale in patients with paradoxical embolism. Circulation *106*, 1121–1126. https://doi.org/10.1161/01.cir.0000027819.19722.ee.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
Python	van Rossum,[24]	https://www.python.org/
Pytorch	Paszke et al.[25]	https://pytorch.org/
opencv	Bradski et al.[26]	https://opencv.org/
numpy	Harris et al.[27]	https://numpy.org/
Scikit-learn	Pedregosa et al.[28]	https://scikit-learn.org/stable/index.html
Code for this paper	Gu.[29]	https://github.com/Mjiegu/PFO_Aldis

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

The cTTE images and videos used in this study originated from patients clinically suspected of PFO with underwent cTTE and cTEE examination in the outpatient or inpatient department of Shenzhen People's Hospital between 2018 and 2021. This study was approved by the ethics committee of Shenzhen People's Hospital, and participants provided written informed consent.

METHOD DETAILS

Establishment of the cTTE images database

Our study is of the retrospective nature. The measurement instrument was a Phillip EPIQ7C system (Philips Ultrasound, Bothell, WA, USA) equipped with an S5-1 probe, frequency 1–5 MHz, and an X7-2T probe, frequency 2–7 MHz. A total of 694 persons with cTTE videos were selected, including 386 PFO-RLS-positive and 308 PFO-RLS-negative patients. Two experts analyzed 694 cTTE videos frame by frame and randomly selected images with adequate quality to establish the cTTE images database for training and testing the DL networks. Additionally, according to the PFO-RLS grading results from 694 cTTE videos, two experts randomly selected 5 cTTE videos from each grade (20 videos in total) as another test database. Data from the cTTE video database were not included in the establishment of the cTTE image database.

Inclusion criteria

1. Apical four-chamber view of the heart was complete. 2. The endocardial boundary of the left chambers could be identified. 3. The image was selected from the point after which the right chambers were filled with microbubbles.

Exclusion criteria

1. Partial segmental loss of the left ventricle (more than 1 segment) in the apical four-chamber view. 2. The left atrial or left ventricular endocardial boundaries were indistinct and could not be identified due to significant ultrasound artifacts. 3. Structures such as large papillary muscles and chordae tendineae in the left ventricle that interfered with microbubble identification were excluded as much as possible.

For PFO-RLS-negative patients, 1–2 images were randomly captured in each video, while for PFO-RLS-positive patients, 3–4 images showing more microbubbles in the left chambers were randomly captured in each video. Finally, a total of 1866 cTTE images were included in the database (see Figure 5 for details). A cTEE examination was selected as the gold standard for the diagnosis of PFO.

Labeling of images

A multi-person online collaborative labling system based on IOS system has been developed to support the special requirements of the ultrasonic diagnostic images. The original images were uploaded to the system in JPG format, and 2 experts annotated segmentation of the left chambers and classified the left chamber microbubbles in the 1866 images. The left chambers were labeled to divide the heart cavity region of the left atrium and left ventricle along the endocardial surface. The chordae tendineae, papillary muscles, and muscle trabeculae that may occur in the mitral valve and left ventricle were all divided into the left chambers. Left ventricular microbubbles were graded by the number of microbubbles in the left chambers on a single frame image at rest.[30,31] The grade was as follows: PFO-RLS grade 0 indicated no microbubbles in the left chambers; PFO-RLS grade 1 indicated <10 microbubbles/frame in the left chambers; PFO-RLS grade 2 indicated 10–30 microbubbles/frame in the left chambers; and PFO-RLS grade 3 indicated >30 microbubbles/frame in the left chambers, or that the left chambers were almost full of microbubbles and the heart chamber appeared cloudy. In this study, the grading assigned by the two experts resulted in the following quantities of images: PFO-RLS grade 0, 477 images; PFO-RLS grade 1, 286 images; PFO-RLS grade 2,345 images; and PFO-RLS grade 3, 758 images (see Figure 5 for details).

The 20 cTTE videos were divided into RLS categories by two experts. The experts reviewed the videos frame by frame in QVue software (Philips Ultrasound, Bothell, WA, USA), and identified the iframe with the largest number of microbubbles, recorded the number of microbubbles observed in the left heart cavity of the frame to perform RLS classification model, and recorded the time required for diagnosis. Where there were doubts about the classification, the decision was discussed by the two experts and made in consensus.

Dataset allocation

cTTE images target left chamber segmentation

From the 1866 images, 185 images were randomly selected according to the different RLS grades as the test set to test the performance of the segmentation model, and the rest were used as the training set to develop the AI model.

cTTE images PFO-RLS classification

The 1866 segmented images augmentation through translation, rotation, and other operations to obtain 2288 images of RLS for each grade, and the final 9152 images. From the 9152 images, 456 images were randomly selected according to the different RLS grades (1824 images in total) as the test set to test the performance of the classification model, and the rest were used as the training set to develop the AI model.

Deep learning model

The whole intelligent diagnosis neural network framework consists of two functional modules of image segmentation and image classification.

In the image segmentation stage, we first cleaned the obtained data to eliminate duplicate data and obtained the final 1866 cTTE images. Then, to improve the convergence speed of the model, we normalized and scaled the cleaned data. As the ResNet-101 network we used was a deep network with complex parameters, more data are needed for image classification model to avoid overfitting. Therefore, the segmented images augmentation through translation, rotation, and other operations to obtain the final 9152 images. Similarly, we normalized and scaled the data to improve the convergence speed of the model.

Image segmentation-Unet neural network model

First, the cTTE images were input to the network coding layer, and the features of the left chambers in the image were extracted by the convolution pooling operation to generate a series of feature maps. Then, these feature maps were decoded. During this process, to improve the accuracy of feature extraction, Unet splices the feature maps of the coding layer and the decoding layer by jumping connections so that it can merge the feature maps from different stages. The final output of the segmented left chamber region binary map is shown in the Figure S1.

Image classification–ResNet 101 neural network model

First, the segmented left chamber image was input into ResNet-101. The image was convolved through the convolutional layer to extract the regional features of the left chambers. Second, multiple deep convolutional blocks were used to learn the acquired features. At the same time, the short connections in the network were able to span several layers, and the input features could be directly mapped to the output so that the network was not degraded during the training process. Then, in the last layer of the convolution block, the weight of the acquired feature information was reallocated. Finally, the extracted features were classified by average pooling (see Figure S2).

Figure 6 shows the whole train and test process of image segmentation model and image classification model. DL was performed with the Python 3.7 programming language.

Model testing

Image segmentation

The 1866 images were randomly divided into 10 groups, and each group contained 4 levels of different RLS grades. The 10-fold cross-validation method was used, among which 9 groups were used as the training set for developing the model, and the others were used as the test set for testing the performance of the model. This process was iterated, and the 10 performance evaluation metrics obtained are averaged.

left chamber segmentation and PFO-RLS classification

A test dataset is a set of images that the AI has not encountered during training. For this study, the test images were obtained by static frame conversion from 20 cTTE videos from the echocardiography room of our hospital. The test images were selected independently of the training set and were graded by the two experts according to the PFO-RLS grading categories. Each cTTE video was converted into static images by using QVue software. Twenty cTTE videos were converted into 4609 images, and then the converted images of each cTTE video were entered into the final neural network model for left chamber segmentation and PFO-RLS classification. The final result was the classification number for at least 3 consecutive frames and the identification of the category with the highest classification number. The final output was the image with graded label color added, and the diagnosis time required for processing the cTTE video was recorded.

Evaluations of experts and physicians

The evaluations of the experts, the intermediate-level physicians, and the junior physicians recruited for this study were tested against the evaluations of the AI model. The 6 ultrasound physicians who were recruited for this task were divided into different levels according to their experience, including (from the most to least experienced) expert physicians (>10 years of experience), intermediate-level physicians (5–10 years of experience), and junior physicians (<5 years of experience). We randomly selected 20 cTTE videos from the ultrasound medical center of Shenzhen People's Hospital to compare and evaluate differences in diagnostic performance between the AI and different levels of physicians.

QUANTIFICATION AND STATISTICAL ANALYSIS

Continuous variables are expressed as the mean ± SD. Categorical variables are presented as a number and percentages (%). One-way analysis of variance (ANOVA) was used to compare the differences in PFO diagnosis time between different the different physician expertise groups and the AI. Interobserver and intraobserver consistency was evaluated by using the kappa test. $p < 0.05$ was considered to indicate statistical significance. Statistical analyses were performed using IBM SPSS V.20.0 software. The evaluation index of the DL model and receiver operating characteristic (ROC) curve production were performed by using Python 3.7 software.

ADDITIONAL RESOURCES

This study did not use any type of experimental models. The study was conducted with the approval of the Medical Ethics Committee of Shenzhen People's Hospital (reference number: LL-KY-202182). During the retrospective collection process, personal information that could identify individuals was eliminated. The informed consent was signed by all patients.