Technology and Health Care 32 (2024) S457–S464
DOI 10.3233/THC-248040
IOS Press

Deep learning-based differentiation of ventricular septal defect from tetralogy of Fallot in fetal echocardiography images

Xia Yu[a,b], Liyong Ma[b,c,*], Hongjie Wang[a,b], Yong Zhang[d], Hai Du[c], Kaiyuan Xu[c] and Lianfang Wang[c]

[a] *Weihai Maternal and Children Health Hospital, Weihai, Shandong, China*

[b] *Weihai Key Laboratory of Precision Medical Technology, Weihai, Shandong, China*

[c] *School of Information Science and Engineering, Harbin Institute of Technology, Weihai, Shandong, China*

[d] *School of Ocean Engineering, Harbin Institute of Technology, Weihai, Shandong, China*

Abstract.

BACKGROUND: Congenital heart disease (CHD) seriously affects children's health and quality of life, and early detection of CHD can reduce its impact on children's health. Tetralogy of Fallot (TOF) and ventricular septal defect (VSD) are two types of CHD that have similarities in echocardiography. However, TOF has worse diagnosis and higher morality than VSD. Accurate differentiation between VSD and TOF is highly important for administrative property treatment and improving affected factors' diagnoses.

OBJECTIVE: TOF and VSD were differentiated using convolutional neural network (CNN) models that classified fetal echocardiography images.

METHODS: We collected 105 fetal echocardiography images of TOF and 96 images of VSD. Four CNN models, namely, VGG19, ResNet50, NTS-Net, and the weakly supervised data augmentation network (WSDAN), were used to differentiate the two congenital heart diseases. The performance of these four models was compared based on sensitivity, accuracy, specificity, and AUC.

RESULTS: VGG19 and ResNet50 performed similarly, with AUCs of 0.799 and 0.802, respectively. A superior performance was observed with NTS-Net and WSDAN specific for fine-grained image categorization tasks, with AUCs of 0.823 and 0.873, respectively. WSDAN had the best performance among all models tested.

CONCLUSIONS: WSDAN exhibited the best performance in differentiating between TOF and VSD and is worthy of further clinical popularization.

Keywords: Congenital heart disease, fetal echocardiography images, tetralogy of Fallot, ventricular septal defect, deep learning

1. Introduction

Congenital heart disease (CHD) is an umbrella term that covers all heart defects present at birth due to cardiovascular malformation in the fetal period. It is also one of the leading causes of death in newborns

[*] Corresponding author: Liyong Ma, School of Information Science and Engineering, Harbin Institute of Technology, Weihai, Shandong, China. E-mail: maly@hitwh.edu.cn.

and infants. CHD annually affects 1.35 million infants worldwide. One-third of them are critically ill, with approximately 4.5% dying in the womb and 21% dying soon after birth [1]. CHD is hazardous or even lethal, and misdiagnosis or missed diagnosis may damage infant growth and development [2].

Fetal echocardiography is the preferred method for diagnosing CHD, as it does not expose people to ionizing radiation, has a low cost, and allows for real-time observation of the fetal heart. Fetal echocardiography was used to observe all nine fetal cardiac views. Due to these advantages, fetal echocardiography has become the most important screening and diagnostic tool for CHD [3]. Deep learning-based artificial intelligence approaches have been applied to diagnose CHD in recent years. Zhu et al. designed a convolutional neural network (CNN) that incorporated generative adversarial networks (GANs) for recognizing fetal cardiac abnormalities in the four-chamber view on fetal echocardiography images at end-systole. GAN's image generation and data augmentation functions were combined with transfer learning to increase the detection accuracy rate. The final accuracy rate reached 85.0%, compared to an average of 81.0% among physicians [4]. Komatsu et al. trained the model using only cardiac data from normal fetuses and applied the model to recognize cardiac structural abnormalities in the four-chamber view plus three-vessel and trachea view on fetal echocardiograms. The AUCs for the model were 0.787 and 0.891, respectively, and the detection probability of the model-aided approach to visualization was also reported [5]. Based on echocardiographic view recognition, Arnaout et al. used the rule-based classifier to differentiate between normal heart and CHD and reported an AUC of 0.99 and a specificity of 0.96 [6].

The application of deep learning as a visual assistant for diagnosing CHD is still in its infancy, and few studies have been conducted thus far in this field. A few studies have described the binary classification of normal and abnormal heart structures based on fetal echocardiography. However, there is a lack of studies differentiating between easily confusing CHDs.

Ventricular septal defects (VSDs) are among the most common CHDs [7,8]. Presenting as a hole in the wall that separates the right and left ventricles of the heart, VSD can cause a shunt of blood between the ventricles. If left untreated, VSD may lead to fetal malnourishment after birth and impair infant growth and development. VSD may close independently during infancy or early childhood if the defect is small. Larger ones can be repaired by surgeries or internal medicine intervention. The prognosis of VSD is usually good after proper treatment.

Tetralogy of Fallot (TOF) is a common cyanotic CHD, and if left untreated, most infants with TOF will die in infancy [9,10]. On fetal echocardiography, TOF usually presents with VSD, overriding aorta, and pulmonary artery stenosis. A normal aorta begins at the left ventricle and connects with the left ventricular outflow tract. However, in the case of an overriding aorta, the aorta is displaced above the VSD and obtains blood from both the left and right ventricles of the heart. Pulmonary artery stenosis is a heart defect that causes the narrowing of the pulmonary artery, making the pulmonary artery even thinner than the aorta.

TOF and VSD share some similarities. A structural defect can be observed on the ventricular septal plane in the left ventricular outflow tract view on a fetal echocardiogram. VSD can show interruption of echo or continuity of ventricular septum in echocardiography, and the diameter of interruption is different. Because of long-term left to right shunt, it can also show right ventricle enlargement. TOF is a complex congenital heart defect, which mainly includes four pathological changes: VSD, pulmonary artery stenosis, right ventricle hypertrophy and right ventricle outflow funnel muscular cone. In addition to the interruption or continuity of ventricular septum echo similar to VSD, TOF in echocardiography can also show the stenosis of pulmonary valve ring, including the stenosis of pulmonary valve and pulmonary artery trunk, the thickening of right ventricle wall, and the thickening of funnel muscle and cone structure

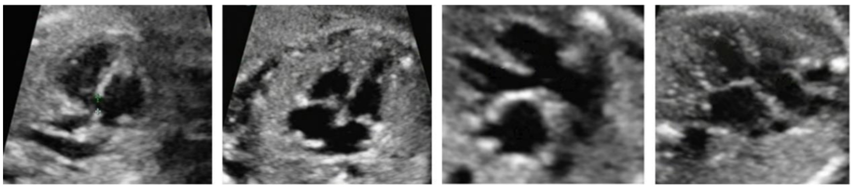

Fig. 1. Fetal VSD and TOF images.

abnormality of right ventricle outflow tract. Therefore, TOF is a much more severe condition with a worse prognosis and higher mortality than simple VSD. Accurate differentiation between VSD and TOF is highly important for administering proper treatment and improving affected infants' prognoses. To the best of our knowledge, no reports have been published on using deep learning-based approaches for differentiating between VSD and TOF, and the present study was intended to fill this gap.

2. Method

2.1. Fetal echocardiography images

The fetal echocardiography images used in this study came from Weihai Maternal and Child Health Care Hospital. The present study was approved by the ethics committee of Weihai Maternal and Child Health Care Hospital (WFEY-QR-CR-857). The images showed the left ventricular outflow tract view on fetal echocardiography of fetuses affected by VSD or TOF at 22 to 24 weeks of gestation. We collected 96 fetal echocardiography images of VSD and 106 images of TOF. These images were preprocessed by performing region of interest (ROI) delineation. Figure 1 shows the results of ROI delineation for VSD and TOF.

2.2. Deep learning models

Four deep learning models were used to differentiate between VSD and TOF, namely, VGG19, RestNet50, NTS-Net, and WSDAN.

VGG19 and RestNet50 are two classic image classification models that are also widely used in the classification of medical images. VGG model is constructed in an alternating manner of convolutional layers and fully connected layers. VGG architecture became a standard architecture when it was proposed, and promoted the development of deeper convolutional neural network. ResNet solves the vanishing gradient problem in deep neural networks. Its core is the residual block. A residual learning unit is inserted after the traditional convolution layer. ResNet50 is a variant of ResNet, with a deeper depth than VGG and a 50 layer structure. NTS-Net and WSDAN are two fine-grained classification models. The fine-grained classification model is suitable for situations where subtle differences in images need to be captured, while also possessing the ability to handle different scales. Due to the need to distinguish subtle differences between the two types of congenital heart disease, these two fine-grained classification models were adopted.

The VGG19 model uses several 3x3 convolutional kernels. With multiple nonlinear layers, VGG19 can learn more complex patterns [11,12]. In addition, VGG19 requires fewer parameters than the deep learning models that use larger convolutional kernels, engendering less computational cost.

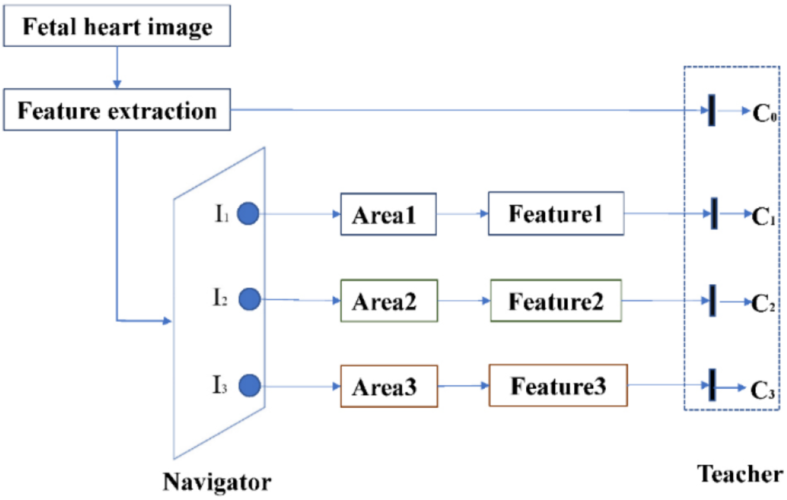

Fig. 2. Architecture of NTS-Net (Navigator-Teacher-Scrutinizer Network).

The RestNet50 model has been introduced to solve the problems associated with increasing model depth. Deep learning models are believed to have stronger extraction and classification abilities with more layers. However, the model's classification performance will no longer be improved beyond a certain depth. Rather, the model tends to converge slowly and exhibits poor performance. The RestNet50 model is known for its shortcut connection that "skips over" some layers, thus converting a regular network to a residual network. In this way, the mode's classification performance can be further improved as the number of layers increases [13,14,15].

The NTS-Net model, proposed by Yang et al., consists of Navigator, Teacher, and Scrutinizer modules [16]. This model can be viewed as multi-agent cooperation: The Navigator detects the most informative regions under guidance from the Teacher. After that, Scrutinizer scrutinizes the proposed regions from Navigator and makes predictions. The NTS-Net model extracts the most informative K regions from the original images, combined with the full image as input to train the Scrutinizer network. In other words, these K regions are used to facilitate classification. Figure 2 displays the navigator network and the teacher network with $K = 3$. Convolutional layers compute the feature hierarchy layer by layer to obtain a series of feature maps of different spatial resolutions. The informativeness of regions among different scales and ratios is generated using multiscale feature maps from different layers. Each region has a score denoting the informativeness of the region, and different regions are sorted based on this score. Nonmaximum suppression is applied to the regions based on their informativeness to reduce region redundancy. The top informative K regions are taken and resized to the predefined size. Then, they are fed into the teacher network to obtain their confidence through mapping. The teacher network outputs confidence as teaching signals to help the navigator network learn. The NTS-Net model can improve classification accuracy by extracting the most informative K regions, containing richer information with higher probabilities. In addition, a novel loss function is designed so that the Teacher is enabled to guide Navigator to localize the most informative regions in an image.

WSDAN (weakly supervised data augmentation network), proposed by Hu et al. [17], achieves data augmentation by generating attention maps. As shown in Fig. 3, WSDAN consists of two stages: weakly supervised attention learning and attention-guided data augmentation.

In the first stage, attention maps are first generated by feature extraction and convolutional operations. Reinforcement learning is implemented on local important features, followed by pooling and splicing

Table 1
Comparison of the classification performance of different models

Model	ResNet50	VGG19	NTS-Net	WSDAN
Sensitivity, %	78.95	84.21	80.00	85.00
Accuracy, %	80.00	80.49	82.93	87.80
Specificity, %	80.95	77.27	85.71	90.48

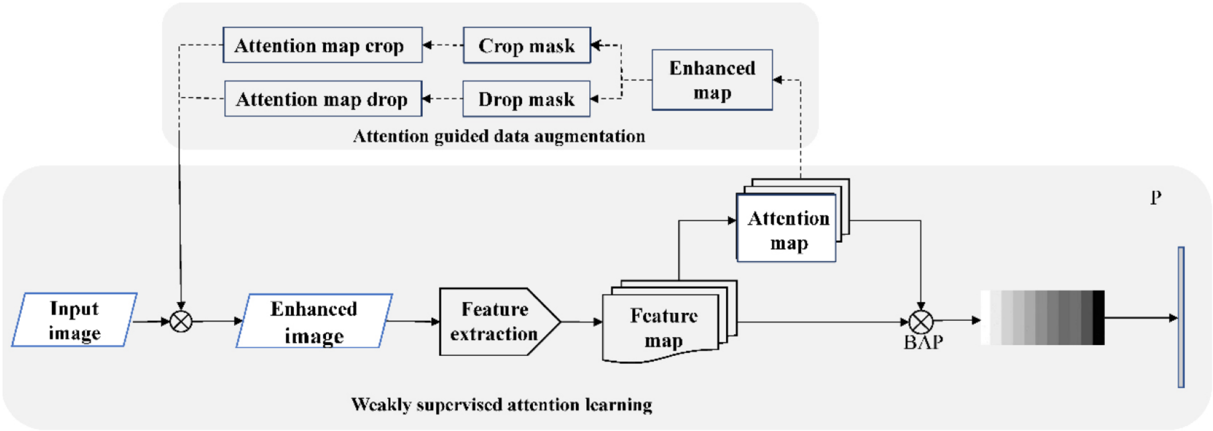

Fig. 3. Architecture of WSDAN.

operations before the features are fed to the classification layer. In the second stage, one attention map is randomly selected to augment this image, including attention cropping and attention dropping. Finally, the raw and augmented data are trained as input data.

2.3. Method

The dataset was divided into a training set and a test set at an 8:2 ratio. The training set consisted of 77 VSD images and 84 TOF images; the test set consisted of 19 VSD images and 21 TOF images. The four models were first trained on the training set and then tested on the test set. The classification results were evaluated, and the performance indicators were calculated. All four models were realized using Python 3.7 and PyTorch 1.3.1. Receiver operating characteristic (ROC) curves were drawn for the four models, and the area under the curve (AUC) was calculated. The sensitivity, specificity, and accuracy of the models were calculated. Python 3.7 software and the statistical toolbox Scipy 1.3.1 were used. χ^2 test was performed to evaluate the classification performance of the models on the test set. A p-value below 0.05 was statistically significant.

3. Results

Table 1 compares different models' sensitivity, accuracy, specificity, and AUC. ResNet50 and VGG19 had similar performances. The sensitivity of VGG19 was higher than that of ResNet50, but the specificity of the former was lower than that of the latter. NTS-Net outperformed both ResNet50 and VGG19 in accuracy and specificity, but its sensitivity came closer to that of ResNet50 but was lower than VGG19. Generally, NTS-Net was superior to ResNet50 and VGG19. WSDAN had the best performance in each indicator among all models tested, which were statistically significant ($P < 0.01$).

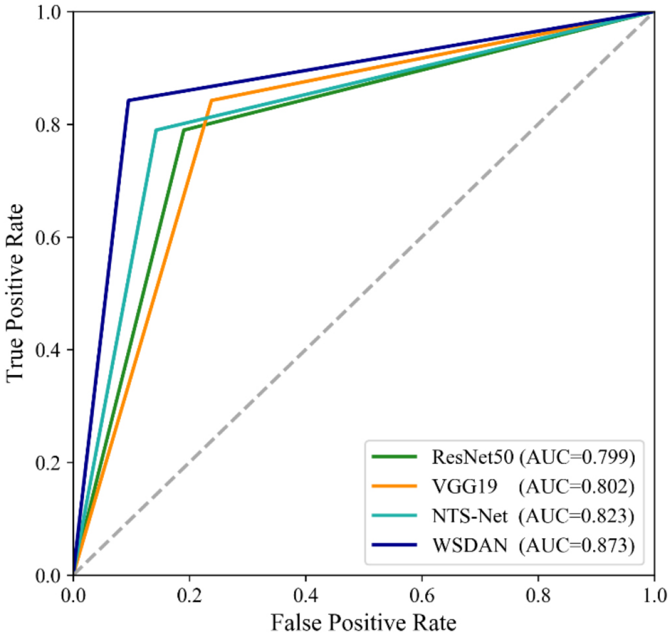

Fig. 4. Comparison of ROC for different models.

Figure 4 shows a comparison of the ROCs for different models. Similar conclusions can be drawn from Table 1 and Fig. 4. The AUC of ResNet50 was very similar to that of VGG19. NTS-Net had a higher AUC than ResNet50 and VGG19. WSDAN had the highest AUC among all models, indicating the highest ability to differentiate between two CHDs.

4. Discussion

Among all diagnostic models for CHD used in this study, WSDAN performed similarly to that reported in the literature [4,5]. However, the performance of ResNet50 and VGG19 in this study was lower than in previous reports. The above existing studies generally involved discrimination between abnormal and normal heart structures. Our work here was to discriminate between two CHDs that are easily confused with each other, which was a more difficult task.

The two types of congenital heart disease, TOF and VSD, have very similar images in fetal echocardiography, with only differences in connection details. One major differentiating feature between TOF and VSD is the overriding aorta. However, an overriding aorta is not usually a prominent feature in fetal echocardiography, which adds to the difficulty of neural network-based differentiation between the two conditions. The commonly used convolutional neural network model cannot effectively distinguish such details. In order to distinguish these subtle differences, fine-grained classification model is employed. Given the minor differences between the two CHDs in cardiac structure, we need a high-efficiency fine-grained categorization algorithm. The class differences are usually large enough in a classification task, while those between the subclasses are very small. Fine-grained image categorization can differentiate between subclasses barely differentiable from each other. The fine-grained classification model uses convolutional neural network to automatically extract features, and complex network model design and careful loss function design to capture the details, shape, texture and other local features in the image.

These features are the key to distinguishing between TOF and VSD. The experimental results also indicate that the fine-grained classification model is more effective in distinguishing TOF and VSD. NTS-Net and WSDAN are techniques specific for fine-grained image categorization, and both achieved better performance in differentiating between VSD and TOF.

WSDAN model is a weak supervised learning model, which can use a small amount of labeled data for training, but does not need a large number of labeled data. This has lower costs and higher efficiency compared to other models that require a large amount of annotated data. WSDAN model enhances the diversity of training data with data augmentation, thereby improving the model's generalization ability. In fine-grained image categorization, discriminative regions for differentiating between subclasses only cover a small area on the image, and the main task of the network model is to localize such discriminative regions. In terms of network architecture, WSDAN model adopts bilinear attention pooling, which can effectively capture key information regions in the image, suppress irrelevant regions, and improve the model's classification ability. WSDAN model also uses a center loss like attention mechanism to help the model better learn the internal structure of categories. This attention mechanism can make the model pay more attention to the differences within categories, thereby improving the classification accuracy of the model. To sum up, WSDAN model has unique characteristics and advantages in weak supervised learning, data enhancement, bilinear attention pooling, and center loss like attention mechanism, which makes the model of WSDAN perform well in fine-grained classification tasks compared with NTS-Net.

The present study had some intrinsic limitations, such as its single-center design and limited sample size. The models and the methodology need further validation.

5. Conclusion

Four conventional deep-learning models were used to analyze fetal echocardiography images to differentiate between two congenital heart diseases, tetralogy of Fallot and ventricular septal defect. Of these models, ResNet50 and VGG19 had similar performance, but their diagnostic performance was the worst among the four models. NTS-Net and WSDAN, specifically designed for fine-grained image categorization, had better performance, and the differentiating ability of WSDAN was the highest. This study proved the potential clinical value of artificial intelligence approaches for differentiating between congenital heart diseases

Acknowledgments

This work was supported by the Shandong Province Medical and Health Science and Technology Development Plan Project (2019WS221) and the Shandong Province Natural Science Foundation (ZR2022MH047).

Conflict of interest

None to report.

References

[1] Brankovic J, Boschetto C, Masini A, et al. Changed outcomes of fetuses with congenital heart disease. Journal of Cardiovascular Medicine. 2015; 16(8): 568-575.

[2] Donofrio M, Moon-Grady A, Hornberger L, Copel J, Sklansky M, Abuhamad A, et al. Diagnosis and treatment of fetal

cardiac disease a scientific statement from the American heart association. Circulation. 2014; 129(21): 2183-2242.

[3] Verdurmen K, Eijsvoogel N, Lempersz C, Vullings R, Schroer C, van Laar J, Oei S. A systematic review of prenatal screening for congenital heart disease by fetal electrocardiography. International Journal of Gynecology & Obstetrics. 2016; 135(2): 129-134.

[4] Zhou X, Zhang Y, Zhang Y, et al. Application of artificial intelligence in screening the four-chamber view of fetal echocardiography. Chinese Journal of Ultrasonography. 2020; 29(8): 668-672.

[5] Komatsu M, Sakai A, Komatsu R, et al. Detection of cardiac structural abnormalities in fetal ultrasound videos using deep learning. Applied Sciences. 2021; 11(1): 371-374.

[6] Arnaout R, Curran L, Zhao Y, et al. An ensemble of neural networks provides expert-level prenatal detection of complex congenital heart disease. Nature Medicine. 2021; 27(5): 882-891.

[7] Adan A, Eleyan L, Zaidi M, Ashry A, Dhannapuneni R, Harky A. Ventricular septal defect: diagnosis and treatments in the neonates: a systematic review. Cardiology in the Young. 2021; 31(5): 756-761.

[8] Miyake T. A review of isolated muscular ventricular septal defect. World Journal of Pediatrics. 2020; 16(2): 120-128.

[9] Forman J, Beech R, Slugantz L, Donnellan A. A review of Tetralogy of Fallot and Postoperative Management. 2019; 31(3): 315.

[10] Krieger E, Valente A. Tetralogy of Fallot. Cardiology Clinics. 2020; 38(3): 365.

[11] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. 3rd International Conference on Learning Representations (ICLR) 2015. May 7-9 2015, San Diego, CA, United States.

[12] Han X, Liang G. Echocardiographic features of patients with coronary heart disease and angina pectoris under deep learning algorithms. Scientific Programming, 2021; 8336959.

[13] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), December 9, 2016, Las Vegas, NV, United States, pp. 770-778.

[14] Ma L, Ma C, Liu Y, Wang X. Thyroid diagnosis from SPECT images using convolutional neural network with optimization. Computational Intelligence and Neuroscience. 2019; 6212759.

[15] Sun M, Ma L, Su X, Gao X, Liu Z, Ma L. Channel separation-based network for the automatic anatomical site recognition using endoscopic images. Biomedical Signal Processing and Control. 2022; 71: 103167.

[16] Yang Z, Luo T, Dong W, et al. Learning to navigate for fine-grained classification. European Conference on Computer Vision, Munich, Germany. 2018; 438-454.

[17] Hu T, Qi H, Huang Q, et al. See better before looking closer: Weakly supervised data augmentation network for fine-grained visual classification. IEEE Transactions on Pattern Analysis & Machine Intelligence. 2019; 18(1): 247-263.