Decision on submission to Informatics in Medicine Unlocked

From: Informatics in Medicine Unlocked (em@editorialmanager.com)

To: rialuthfan@yahoo.com

Date: Monday, February 22, 2021 at 05:55 AM GMT+7

Manuscript Number: IMU-D-21-00049

Automated Image Segmentation for Cardiac Septal Defect based on Contour Region with Convolutional Neural Networks

Dear Dr Nova,

Thank you for submitting your manuscript to Informatics in Medicine Unlocked.

I have completed my evaluation of your manuscript. The reviewers recommend reconsideration of your manuscript following major revision. I invite you to resubmit your manuscript after addressing the comments below. Please resubmit your revised manuscript by May 07, 2021.

When revising your manuscript, please consider all issues mentioned in the reviewers' comments carefully: please outline every change made in response to their comments and provide suitable rebuttals for any comments not addressed. Please note that your revised submission may need to be re-reviewed.

To submit your revised manuscript, please log in as an author at https://www.editorialmanager.com/imu/, and navigate to the "Submissions Needing Revision" folder.

Informatics in Medicine Unlocked values your contribution and I look forward to receiving your revised manuscript.

pleasantly,
Ed

Edward J Ciaccio
Editor-in-Chief
Informatics in Medicine Unlocked

Editor and Reviewer comments:

Reviewer #1: Authors developed U-Net type CNN to make segmentation of 2D echocardiogram image from four patient groups (ASD,VSD,AVSD,normal). Authors should modify the following points.

(1)The number of patients is small in this study and authors did not apply discriminant analysis to discriminate among patient groups (ASD,VSD,AVSD,normal) due to the small number. From these reasons, the title of this paper should include the term "A preliminary study".
(2) Authors should add possible explanation for the reason why U-Net architecture was superior to V-Net in Discussion section.
(3) Why authors focused on the apical A4C and A5C view and did not use subcostal and suprasternal view?
(4) Explain Fig. 3 and Fig.4 more in detail.
(5) Indicate white holes in Fig.7.
(6) Function f must be missing in Eq.(1).
(7) 1/nci must be 1/ncl in Eqs.(3) and (4).
(8) class I must be class i in the sentence just after Eq.(4)
(9) Add explanation for Mean IU more in detail.
(10) Definitions of pi and gi are missing in Eq.(5)
(11) Add explanation for Precision and Recall also.
(12) Authors should introduce full-term before using abbreviations. For example CNN, ASD, VSD, AVSD in Abstract, IU in page 11, HLHS and TGA in page 21.
(13) Page20: In Figure 9 we have also … -> In Figure 10 we have also…

Reviewer #2: The manuscript by Ria Nova, Siti Nurmaini, Radiyati Umi Partan, and Sukman Tulus Putra describes an automatic method for detecting cardiac septal defect. Using CNNs, authors created a methodology to segment

cardiac septal defects. The study is conceptually interesting, but there are some notes that could improve this valuable study.

1. What is the timing of capturing images during the process of transforming videos into frames? You can mention this value in the first paragraph of page 10. Whether you capture just a few shots (shown in figure 3) from each of the videos? In this case, please mention the exact time of capturing the image, and is that the same for all videos, or you choose it manually?
2. For ground truth, why did you use photoshop? Please mention your reason in the materials and methods section.
3. Regarding model architecture, why did you choose CNNs and U-Net? Are there any other studies about the septal defect in which different architectures were utilized (As you mentioned there is no other study utilizing CNN and U-Net for septal defect)? What are the plus points of CNN in comparison with other architectures in your application? Mentioning the answer to these questions in the materials and methods section is beneficial. (There is just a brief sentence in the Results section saying U-net have a good performance for the segmentation of heart images)
4. "The total of image about 4000, however in the segmentation process to obtain the good model of segmentation only good quality images are selected about 2609. From the selected image, ground truth was done for each of the 50 images", what was the reason for excluding a considerable amount of data? Please mention under which condition these "good" images were chosen?
5. As you mentioned, the possibility of wrong observations in clinical unautomated methods for detecting septal defects is very high, and your method is the first automated method, so it could be beneficial if the results of your method were compared with the clinical approach.

There are also some issues about the paper structure.

1. The hardware features of the system could be mentioned in the materials and methods rather than the abstract.
2. "Ventricular Septum Defect" in line 5 of introduction should be changed to "Ventricular Septal Defect"
3. Figure 3 could be removed or mentioned in the appendix since the step of extracting frames from a video could be described effectively without this figure.
4. Regarding Figure 4, the same as Figure3, it would be beneficial if authors show just one of their categories (for example, ASD) and show the changes more detailed and bigger in the material and methods section and then insert all other images into the appendix section.
5. Each reader must get the most information once read the caption of each Figure, so the captions should be completed.
6. Some of the values in Figure 6 should be typed in bigger size to be more readable.
7. Some of the parameters of formulas #2, #3, and #4 are not defined exactly under the formulas.

Reviewer #3: I have read the manuscript. It is an interesting work, but it is not well written. The structure of the paper is confusing and some methodological details are missing, making some results difficult to be interpret. Thus, I consider the manuscript not ready for publication.
Major Comments:
* The paper is not well written. Methods, results and discussion sections appears merged.
* The pre-processing section is not clear and authors didn't provide enough details to make the work repeatable. In order to clarify the pre-processing step, I suggest including a more detailed description of the used methods (not simply the used Phyton library) instead of reporting the results of this pre-processing.
* The paper presents no details about the data division in the method section. I suggest reporting the data division in the method section, and not in the results section. Moreover, the prevalence of different classes in the training dataset and in the testing dataset should be reported.
* Statistically, the gold standard for statistical measurements computation has to be described. Without information about the gold standard, the results are not clear.
* Results section and discussion section should be separated.
* Authors provided a comparison between their model and the V-Net in the results section, but the V-Net was not introduced in the method section. In order to make the comparison interpretable, authors should describe the V-Net (at least in the statistics section).
Minor Comments:
* Acronyms should be defined before their use, especially in the abstract.
* Figure 1 is not cited in the manuscript.

Reviewer #4: Dear authors I have read your paper entitled "Automated Image Segmentation for Cardiac Septal Defect based on Contour Region with Convolutional Neural Networks" with a lot of interest.

I have found your research very important and that gives us a taste of how diagnosis is going to be performed in the future. In your work physicians without special skills in echo studies can have diagnosis of congenital cardiac defects either analyzing the data they collected locally or even by transferring them to specialised centers. I was

impressed by the image quality and the process in order to reach specific standardized views that can be diagnostic. The outcome is successful.

Moreover I was impressed by the technical and mathematic explanation of the data process that my knowledge is not enough to fully appreciate. However being a cardiac surgeon I am looking forward to the use of Artificial Intelligence in diagnostic and therapeutic options and this work is a step forward to that. So this paper is worth to be published
My only comment as it is mentioned by you as well is that this is an initial try/observation with only 2 patients for each category : ASD, VSD, AVSD and Normal heart. So you have to return with a major number of patients with another paper in the future. I would also suggest to shorten the introduction and correct some grammatical errors in the discussion segment.

Apart from these minor corrections congratulations again for your work.

*****
Data in Brief (optional):
We invite you to convert your supplementary data (or a part of it) into an additional journal publication in Data in Brief, a multi-disciplinary open access journal. Data in Brief articles are a fantastic way to describe supplementary data and associated metadata, or full raw datasets deposited in an external repository, which are otherwise unnoticed. A Data in Brief article (which will be reviewed, formatted, indexed, and given a DOI) will make your data easier to find, reproduce, and cite.

You can submit to Data in Brief when you upload your revised manuscript. To do so, complete the template and follow the co-submission instructions found here: www.elsevier.com/dib-template. If your manuscript is accepted, your Data in Brief submission will automatically be transferred to Data in Brief for editorial review and publication.

Please note: an open access Article Publication Charge (APC) is payable by the author or research funder to cover the costs associated with publication in Data in Brief and ensure your data article is immediately and permanently free to access by all. For the current APC see: www.elsevier.com/journals/data-in-brief/2352-3409/open-access-journal

Please contact the Data in Brief editorial office at dib-me@elsevier.com or visit the Data in Brief homepage (www.journals.elsevier.com/data-in-brief/) if you have questions or need further information.

More information and support

FAQ: How do I revise my submission in Editorial Manager?
https://service.elsevier.com/app/answers/detail/a_id/28463/supporthub/publishing/

You will find information relevant for you as an author on Elsevier's Author Hub: https://www.elsevier.com/authors

FAQ: How can I reset a forgotten password?
https://service.elsevier.com/app/answers/detail/a_id/28452/supporthub/publishing/kw/editorial+manager/

For further assistance, please visit our customer service site:
https://service.elsevier.com/app/home/supporthub/publishing/. Here you can search for solutions on a range of topics, find answers to frequently asked questions, and learn more about Editorial Manager via interactive tutorials. You can also talk 24/7 to our customer support team by phone and 24/7 by live chat and email.

Automated image segmentation for cardiac septal defects based on contour region with convolutional neural networks: A preliminary study

by Ria Nova

Submission date: 03-Jan-2022 03:25PM (UTC+0700)
Submission ID: 1737051253
File name: 1-s2.0-S2352914821000915-main_1.pdf (11.5M)
Word count: 6285
Character count: 32103

Informatics in Medicine Unlocked 24 (2021) 100601

Contents lists available at ScienceDirect

Informatics in Medicine Unlocked

journal homepage: www.elsevier.com/locate/imu

Automated image segmentation for cardiac septal defects based on contour region with convolutional neural networks: A preliminary study

Ria Nova [a], Siti Nurmaini [b,*], Radiyati Umi Partan [c], Sukman Tulus Putra [d]

[a] Department of Child Health, Division of Pediatric Cardiology, Dr. Moh Hoesin Hospital, Faculty of Medicine, Universitas Sriwijaya, Palembang, 30126, Indonesia
[b] Intelligent System Research Group, Universitas Sriwijaya, Palembang, 30139, Indonesia
[c] Department of Medicine, Dr. MohHoesin Hospital, Faculty of Medicine, Universitas Sriwijaya, Palembang, 30126, Indonesia
[d] Department of Child Health, Division of Pediatric Cardiology, Dr. CiptoMangunkusumo Hospital, Faculty of Medicine, Universitas Indonesia, Jakarta, 10430, Indonesia

ARTICLE INFO

Keywords:
Cardiac septal defect
Contour segmentation
Deep learning
CNNs

ABSTRACT

Echocardiogram examination is important for diagnosing cardiac septal defects. With the development of AI-based technology, an echocardiogram examination previously performed manually by cardiologists can be done automatically. Automatic segmentation of cardiac septal defects can help a physician to make an initial diagnosis before referring a pediatric cardiologist for further treatment. In previous studies, automatic object segmentation using convolutional neural networks (CNNs) was one of the DL applications that have been developed for cardiac abnormalities. In this study, we propose a CNN-based U-Net architecture to automatically segment the cardiac chamber to detect abnormalities (holes) in the heart septum. In this study, echocardiogram examinations were performed on atrial septal defects (ASDs), ventricular septal defects (VSDs), atrioventricular septal defects (AVSDs), and normal hearts with patients undergoing echocardiogram examination at Moh Hoesin Hospital in Palembang. The results show that even for the relatively small number of datasets, the proposed technique can produce superior performance in the detection of the cardiac septal defects. Using the proposed segmentation model for four classes produces a pixel accuracy of 99.15%, mean intersection over union (IoU) of 94.69%, mean accuracy of 97.73%, sensitivity of 96.02%, and F1 score of 94.88%, respectively. The plots of the loss and accuracy curve show that all the errors were small, with accuracy rates reaching 99.05%, 98.62%, 99.39%, and 98.97% for ASD, VSD, AVSD, and normal heart, respectively. The comparison accuracy of contour prediction for U-Net was 99.01%, while V-Net was 93.70%. This shows that the U-Net has better accuracy than the V-Net model architecture. It can be proven that the architecture of CNNs has been successful in segmenting the cardiac chamber to detect defects in the heart septum and support the work of cardiologists.

1. Introduction

Congenital heart disease (CHD) is the most common congenital anomaly in new-born babies [1]. Anatomical abnormalities of the heart and blood vessels have even occured since the first trimester intra-uterine. There are many types of CHD varying from mild to severe, with both frequent and rare cases [2]. A Cardiac septal defect is one type of CHD that is marked by a hole in the atrial, ventricular, or both the atrial and ventricular septa, which correspond to atrial septal defect (ASD), ventricular septal defect (VSD), and atrioventricular septal defect (AVSD), respectively [1]. ASD and VSD are the most common CHD lesions, while AVSD is less common [2]. Even though AVSD is not as common as ASD and VSD cases, generally, the symptoms are more severe and overdue for detection [3,4]. Likewise, although there are many cases with ASD and VSD, they are still detected too late, so that treatment becomes delayed and ineffective [3].

Delay in early detection occurs because not all cases suspected with cardiac septal defects can be detected by an echocardiogram, whereas an echocardiogram is the gold standard examination to establish a diagnosis of CHD [5]. To resolve the problem of late detection, every case suspected of having a cardiac septal defect should be performed with an echocardiogram examination by a physician [5]. However, not all physicians can perform echocardiograms because this examination requires special skills to avoid misdiagnosis [6]. Moreover, cardiologists

* Corresponding author. Intelligent System Research Group, Faculty of Computer Science Universitas Sriwijaya, Palembang, 30137, Indonesia.
E-mail addresses: rialuthfan@yahoo.com (R. Nova), sitinurmaini@gmail.com (S. Nurmaini), radiandinadr@yahoo.co.id (R.U. Partan), sukmanputra@yahoo.com (S.T. Putra).

https://doi.org/10.1016/j.imu.2021.100601
Received 27 January 2021; Received in revised form 12 May 2021; Accepted 12 May 2021
Available online 23 May 2021

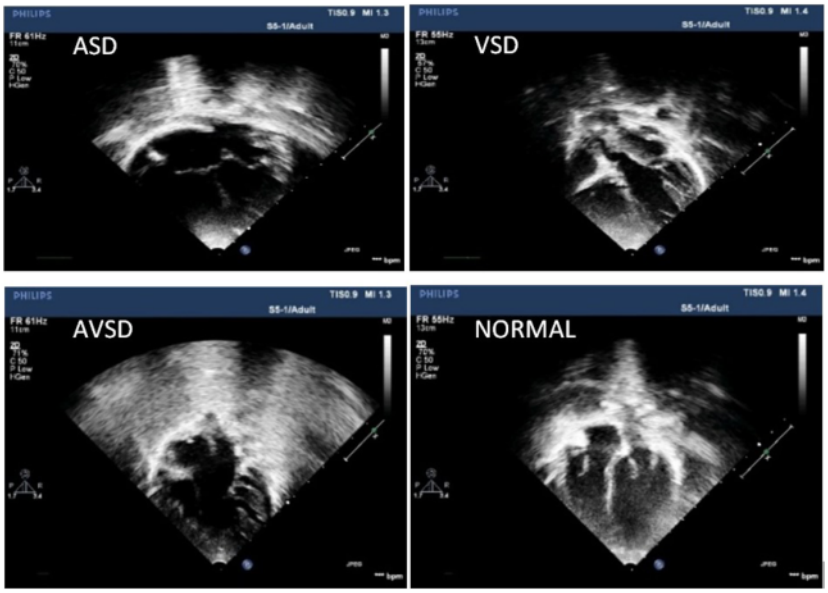


Fig. 1. Sample of raw data for ASD, VSD, AVSD, and normal heart from apical 4 and 5 chamber view in 2-D echocardiogram. Data source of RSMH medical records, November 7, 2019.

do not yet exist in every peripheral health service, as their numbers are still limited. For this reason, even though echocardiograms are available in peripheral health services, their use for the detection of CHD is still suboptimal [3]. Based on all these limitations, improving screening examinations with advanced technology to achieve accurate, automatic abnormal cardiac septum detection using echocardiograms has become a major issue.

To identify cardiac septal defects, the physician performs auscultation using a stethoscope to listen for heart sounds and murmurs [7]. Although the first heart sound is normal, the second heart sound is typical of a wide fixed split and a soft systolic ejection murmur is heard over the pulmonary area in the left upper sternal border [8], misdiagnosis of ASD is still common [3]. Misdiagnosis of VSD also frequently occurs because the sound of a typical holosystolic murmur in the mid-to lower-left sternal border varies depending on the type and size of VSD [9]. The same problem occurs with AVSD, and murmurs may often not be heard [1]. Therefore, echocardiogram is needed to confirm the diagnosis of cardiac septal defects.

These days, computer-based diagnosis systems have been developed. In other words, echocardiogram interpretation is done digitally, aided by a computer device (computer-aided diagnosis) using artificial intelligence (AI) [10–12]. With the development of AI-based technology, an echocardiogram examination for the detection of cardiac septal defects previously performed manually by cardiologists can be done automatically. An automatic echocardiogram examination can be used to assist physicians in early detection before referral to a cardiologist for further management. Deep learning (DL), as a part of AI, has demonstrated great potential in recent years for medical imaging. The most common applications of DL in medical imaging have been for image classification [13–15], detection [16], and segmentation [17–19].

Although DL has been widely applied to 2D cardiac images with high accuracy, to the best of our knowledge, there has been limited research until now that developed it for cardiac septal defects. Object segmentation is one of the DL applications that can be developed for cardiac septal defects. Contouring lesions can be identified through segmentation so that cardiac septal defects can be diagnosed accurately. Therefore, improving the 2D segmentation performance for cardiac septal defects using CNNs is important for a deep investigation. This study's novelty and contributions are as follows:

- To design a CNN model for segmenting cardiac septal defect conditions of the heart images with high accuracy;
- To develop a CNN-based U-Net architecture for segmenting the contour regions of ASD, VSD, AVSD, and normal condition; and
- To validate selected models with a V-Net architecture in terms of pixel accuracy, mean intersection union, mean accuracy, precision, recall and F1 score.

The rest of this paper is organized as follows: Section 2 explains the materials and methods, section 3 presents the results, and section 4 offers a discussion. Finally, the conclusions are presented in section 5.

2. Methods

2.1. Data acquisition

Echocardiogram examinations were performed on eight patients, consisting of ASD, VSD, AVSD, and normal heart patients. The age of subjects ranged from one to five years old, all of whom visited the children's heart clinic at Moh Hoesin Hospital between November 2019 and January 2020. All patients were examined for a 2-D echocardiogram with six standard views, namely, parasternal long and short, apical 4-(A4C) and 5-(A5C) chamber, subcostal, and suprasternal views. In this study, we have focused on A4C and A5C views, as can be seen in Fig. 1. This selection of focus is due to the atrial septum, ventricular septum and the four chambers of the heart being clearly visible in one view. We covered 200 images obtained from eight videos each of two videos of ASD, VSD, AVSD, and normal heart for training and validation.

2.2. Pre-processing

The pre-processing of infant video for segmentation consists of four

Fig. 2. The main steps of data pre-processing.

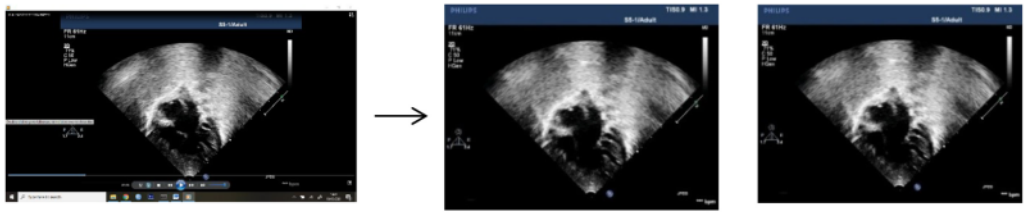


Fig. 3. Conversion of US video of AVSD to frames.

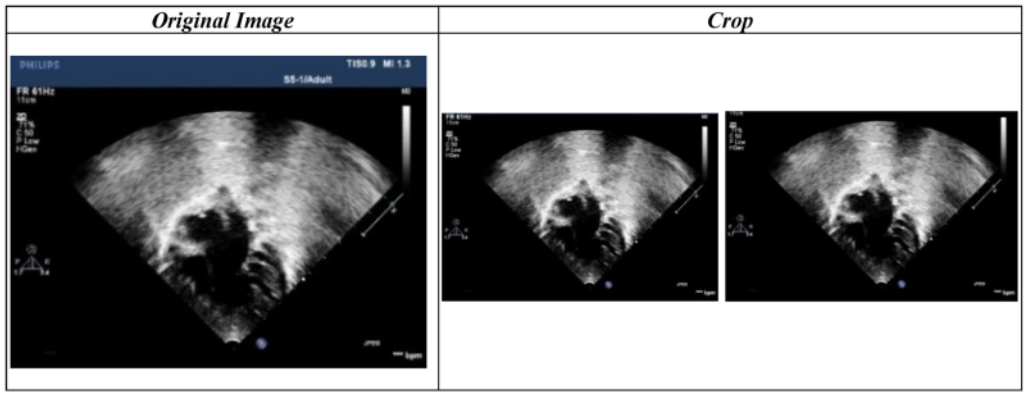


Fig. 4. Image cropping process for AVSD.

main steps, as shown in Fig. 2; (i) infant video to image framing. The type of file.avi and length is 5s. We used cv2.VideoCapture (); the infant video was then read frame by frame, where the frame will be stored in frame storage using cv2.imwrite () code to create a ground truth of infant images. Subsequently, we (ii) performed the data filtering process with a closed valve case; (iii) cropped all infant images from the frame based on an 800 × 600 pixel; (iv) and annotated the labels of infant images with a data annotation tool (Adobe Photoshop). The label consisted of a hole and a heart chamber. If there were only the chambers of the heart, it was identified as normal. The output of labels was saved in image thresholding.

For videos that have been obtained previously, the next step is to convert videos into frames or images. From the raw video data inserted into the Python library with OpenCV, the video will be converted into many frames. The data is recorded in the video in the.avi format and then converted into frames with the.jpg format. Fig. 3 shows the video being converted into frames of AVSD.

The results obtained in the process of converting the video to the frame will produce many frames based on the output obtained by the Python library. The data frame results obtained have a size of 800 × 600 pixels, and there is still much unnecessary information in the data frame. Thus, the next stage is to cut the image frame that has been performed before. This stage is performed the same as in the process of converting the video to the frame using library Python software. In the process of cropping the frame with the Python library, the crop range is adjusted to the right, left, top and bottom to remove unnecessary information, as shown in Fig. 4. Because the size of the pixel frame is maximal, it is enough to be used for the next process, so image cropping is not performed.

The final step taken in the pre-processing of this data is to label the data that has been cut before. The process of labeling images or ground truth uses the help of Adobe Photoshop and illustrator because PSD Photoshop software supports labels getting good results in the process of labeling image data. Fig. 5 shows the ground truth of the original ASD, VSD, AVSD and normal heart frames.

Eight echocardiogram videos were converted to become several frames (images)—about 100 to 500 images—which were used as the source of information. The total number of images was about 4000. However, in the process designed to obtain a good model of segmentation, only good quality images were selected, leaving 2609. From the selected images, ground truth was performed for each of the 50 images, as shown in Table 1.

The 200 ground truth dataset used for training and testing and the prevalence of different classes as can be seen in Table 2.

To ensure that the process of object detection was run in a good performance, an Intel i9-9900 k CPU with NVIDIA GPU RTX 2080ti 11 GB was used as the testing server. The processing time largely depended on the number of convolution layers in one image with Windows 10 OS.

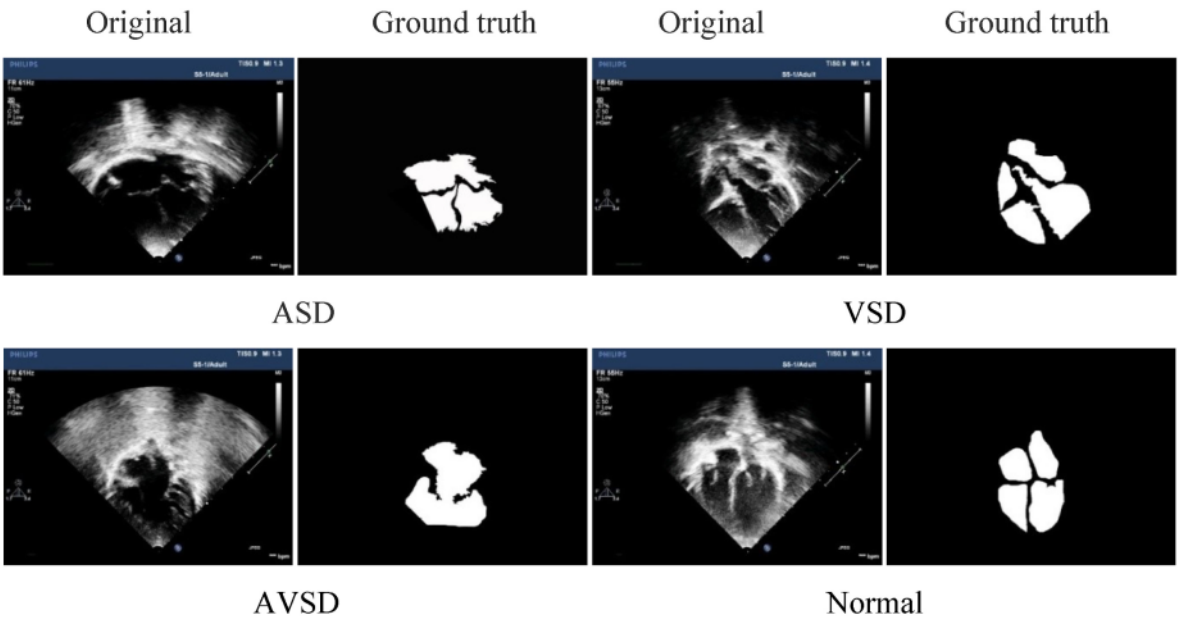


Fig. 5. Image labeling process for ASD, VSD, AVSD, and normal heart.

Table 1
Dataset of ASD, VSD, AVSD and normal heart.

No	Echocardiogram	Original Image		Ground Truth
		Patient 1	Patient 2	
1	ASD	302	557	50
2	VSD	166	115	50
3	AVSD	320	302	50
4	Normal	277	570	50
	Total	1065	1544	200

Table 2
Dataset for training and testing.

No	Echocardiogram	Training	Testing
1	ASD	34	10
2	VSD	30	11
3	AVSD	34	12
4	Normal	32	13
	Total	154	46

The validation process of CNN pre-trained models was carried out with different hyperparameters and network models. The learning rate process used was 10^{-5}, with an epoch 1000 and batch size of 64.

2.3. Model architecture

The deep learning method used in this study employs CNNs. The CNN-based U-Net architecture was chosen in this study because the architecture has been shown to exhibit good performance for the segmentation of heart images [20]. CNNs are designed to better utilize spatial and configural information by taking 2D images as input [21]. Structurally, CNNs have convolutional layers interspersed with pooling layers followed by fully connected layers, as in a standard multilayer neural network [21,22]. The role of a convolutional layer is to detect local features at different positions in the input feature maps with learnable kernels $k_{ij}^{(l)}$, namely, connection weights between the feature map I at layer $l-1$ and the feature map j at layer l. Specifically, the units of the convolutional layer l compute their activation $A_j^{(l)}$ on the basis of only a spatially contiguous subset of units in the feature maps $A_i^{(l-1)}$ of the preceding layer $l-1$ by convolving the kernels $k_{ij}^{(l)}$ as follows:

$$A_j^{(l)} = \left(\sum_{i=1}^{M(l-1)} A_i^{(l-1)} * k_{ij}^{(l)} + b_j^{(l)} \right), \quad (1)$$

where $M^{(l-1)}$ denotes the number of feature maps in layer $l-1$, the asterisk denotes a convolutional operator, *and* $b_j^{(l)}$ is a bias parameter. Due to the mechanisms of weight sharing and local receptive field, when the input feature map is slightly shifted, the activation of the units in the feature maps is shifted by the same amount. In this study, the architectural model of CNNs is U-Net for defect segmentation. Generally, the U- Net architecture is depicted in Fig. 6 as follow:

U-Net architecture consists of a contracting path (left side) and an expansive path (right side). The contracting path follows the typical architecture of a convolutional network. It consists of the repeated application of two 3 × 3 convolutions (unpadded convolutions), each followed by a rectified linear unit (ReLU) and a 2 × 2 max pooling operation with stride 2 for down-sampling. At each down-sampling step, we double the number of feature channels. Every step in the expansive path consists of an up-sampling of the feature map followed by a 2 × 2 convolution ("up-convolution") that halves the number of feature channels, a concatenation with the correspondingly cropped feature map from the contracting path, and two 3 × 3 convolutions, each followed by a ReLU. The cropping is necessary due to the loss of border pixels in every convolution. At the final layer, a 1 × 1 convolution is used to map each 64-component feature vector to the desired number of classes. In total, the network has 23 convolutional layers. All parameters of the architecture are defined in Table 3.

This research will also compare the architectural model with the V-Net. The V-Net approach comprises two main parts. The left section includes two features: the left area, which consists of a compression path, and the right area, which decompresses the input until its initial size is attained. The architecture V-Net is similar to the U-Net model, but

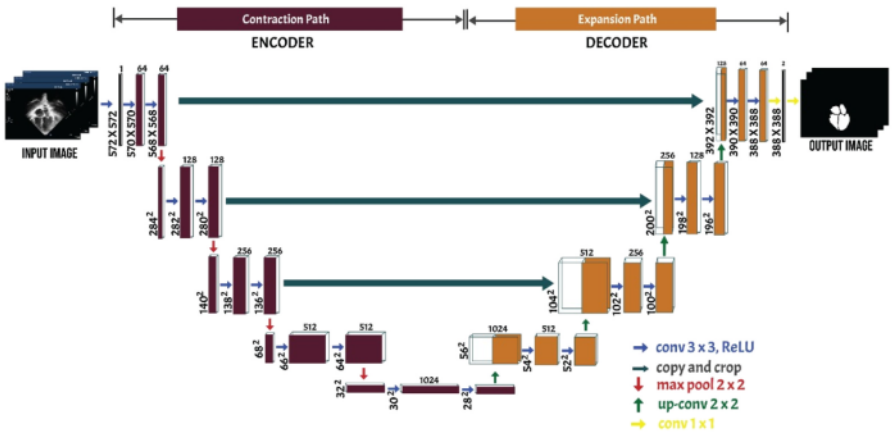


Fig. 6. U-Net architecture.

Table 3
U-net architecture.

Layer	Kernel Size, Feature Map	Stride	Activation Funtion	Output Shape
Input Layer	–	–	–	256 × 256 x 1
Convolution Layer 1	64 × 64 x 3	1	ReLu	128 × 128 x 3
Max Pooling 1	2 × 2	2	–	128 × 128 x 3
Convolution Layer 2	128 × 128 x 3	1	ReLu	256 × 256 x 3
Max Pooling 2	2 × 2	2	–	256 × 256 x 3
Convolution Layer 3	256 × 256 x 3	1	ReLu	512 × 512 x 3
Max Pooling 3	2 × 2	2	–	512 × 512 x 3
Convolution Layer 4	512 × 512 x 3	1	ReLu	1024 × 1024 x 3
Dropout	p = 0.5	–	–	1024
Max Pooling 4	2 × 2	2	–	1024 × 1024 x 3
Convolution 5	1024 × 1024 x 3	1	ReLu	512 × 512 x 2
Dropout	p = 0.5	–	–	512 × 512 x 2
Up	512 × 512 x 2	3(axis)	ReLu	512 × 512 x 3
Convolution Layer 6	512 × 512 x 3	1	ReLu	256 × 256 x 2
Up	256 × 256 x 2	3(axis)	ReLu	256 × 256 x 3
Convolution Layer 7	256 × 256 x 3	1	ReLu	128 × 128 x 2
Up	128 × 128 x 2	3(axis)	ReLu	128 × 128 x 3
Convolution Layer 8	128 × 128 x 3	1	ReLu	64 × 64 x 2
Up	64 × 64 x 2	3(axis)	ReLu	64 × 64 x 3
Convolution Layer 9	64 × 64 x 3	1	ReLu	2 × 2 x 3
Output Layer	–	–	Sigmoid	1

Table 4
Segmentation performance for four classes

Validation	Segmentation Prediction by U-Net Architecture (%)
Pixel Accuracy (PA)	99.15
Mean Intersection Union (MIU)	94.69
Mean Accuracy (MA)	97.73
Precision	93.83
Recall	96.02
F1 Score	94.88

Table 5
Segmentation performance for each class

Validation	Segmentation Prediction by U-Net Architecture (%)			
	ASD	VSD	AVSD	Normal
Pixel Accuracy	99.05	98.62	99.39	98.97
Mean Intersection Union	93.84	92.57	95.66	93.52
Mean Accuracy	98.21	95.32	97.56	96.18
Precision	91.06	94.58	96.27	94.67
Recall	97.22	91.21	95.41	92.83
F1 Score	93.99	92.83	95.81	93.66

with some differences. The left part of the network is divided into different stages that operate at various resolutions. Each step comprises one to three convolutional layers. At each stage, a residual function is learning. This architecture ensures convergence compared with non-residual learning network, such as U-Net. The convolutions use volumetric kernels. Resolution is reduced by convolution with 2 × 2 × 2 voxels wide seeds applied with stride 2. PreLU is used as a non-linearity activation function. The right network extracts features and expands the spatial support of the lower resolution feature maps to gather and assemble the necessary information to output a two-channel volumetric segmentation. Deconvolution operation is employed in order to increase the size of the inputs, followed by one to three convolutional layers. The residual function is learned. The last convolutional layer, having 1 × 1 × 1 kernel size, yields the same size as the input volume. The probabilistic segmentation of the foreground and background regions is achieved by applying softmax voxelwise. Similar to U-net, horizontal connections with location information are lost in the compression path (left). This can help to provide location information to the right part and improve the quality of the final contour prediction. Connection improves the convergence time of the model.

2.4. Performance metrics

To validate the cardiac segmentation performance of the proposed model, the statistical analysis used is pixel accuracy, mean accuracy, mean intersection union (mean IU), precision, recall, and F1 score, by comparing it with the ground truth, as defined below.

$$\text{Pixel Accuracy} = \frac{\sum_i n_i i}{\sum_i t_i} \tag{2}$$

$$\text{Mean Accuracy} = \frac{1}{n_c l} \sum \frac{n_c i}{t_i} \tag{3}$$

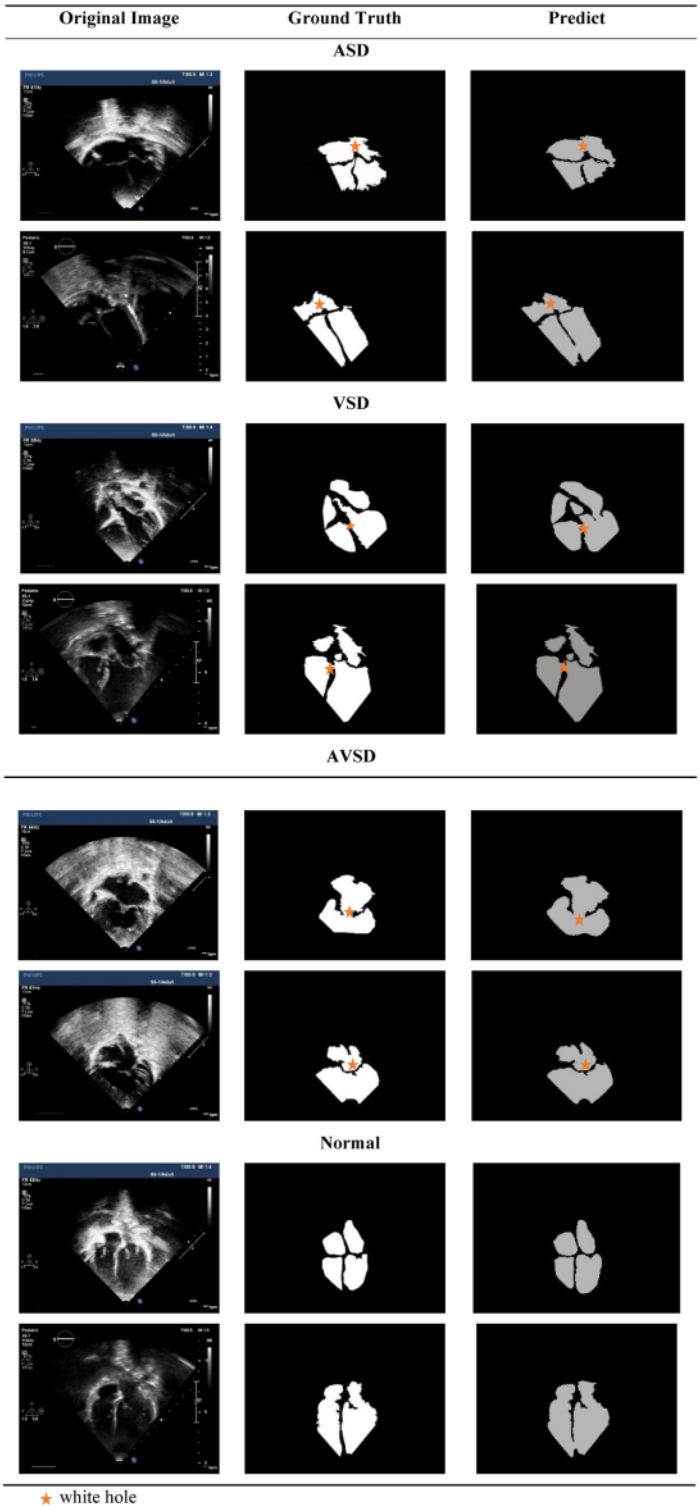


Fig. 7. Segmentation result using CNN-based U-Net architecture. Asterisk (★) indicates a hole in atrial and ventricular septal of ASD, VSD, and AVSD.

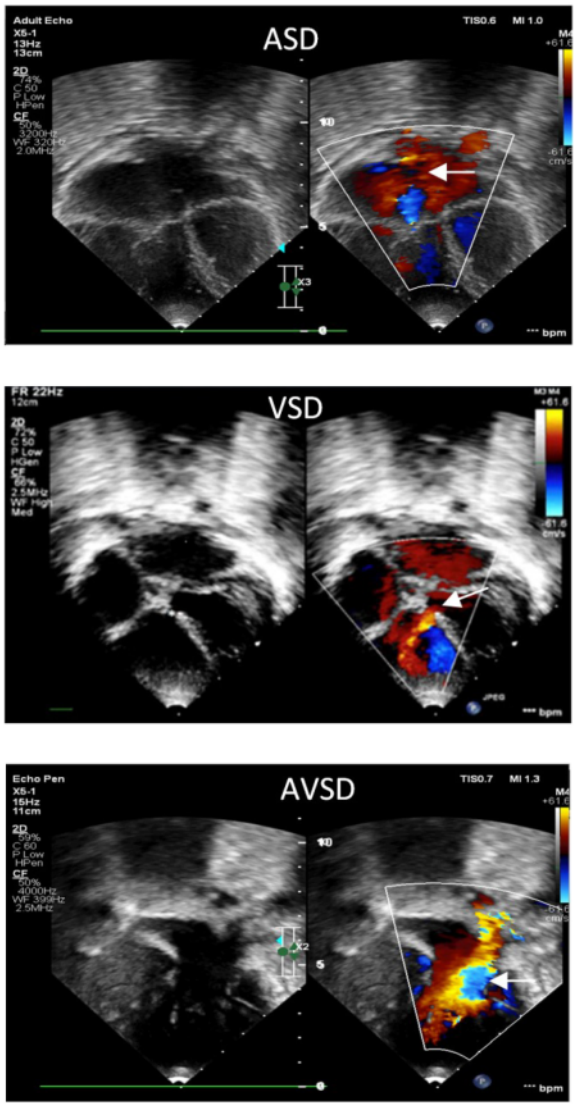


Fig. 8. Echo Doppler revealed flow from the left to the right of ASD, VSD and AVSD.

Table 6
Comparison of two architectures of segmentation performance model.

Validation	Performances (%)	
	U-Net	V-Net
Pixel Accuracy	99.15	97.27
Mean IU	94.69	84.88
Mean Accuracy	97.73	91.08
Precision	93.83	83.03
Recall	96.02	83.68
F1 Score	94.88	83.12

$$\text{Mean IU} = = \frac{1}{n_c l} \sum \frac{n_c i}{t_i + \sum j n_{ji} - n_{ii}} \tag{4}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{5}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{6}$$

where n_{ij} is the number of pixels of class i predicted to belong to class j, where there are n_{cl} different classes, and $t_i = \sum_j nij$ is the total number of pixels of class i. TP, FP and FN are true positives, false positives, and false negatives, respectively. For the F1 score, the dice coefficient equation is used as follows:

$$D = 2 * \frac{\sum_i^n pi\ gi}{\sum_i^n pi^2 + \sum_i^n gi^2} \tag{7}$$

where pi is prediction and gi is ground truth.

3. Results

In this study, the proposed model consists of nine convolution layers followed by the max-pooling layer, the drop-out layer, and the up layer. Using a proposed segmentation model for four classes produces a pixel accuracy of 99.15%, mean IU of 94.69%, mean accuracy of 97.73%, sensitivity of 96.02%, and F1 score of 94.88%, respectively, as shown in Table 4.

Table 5 reveals the results of the proposed segmentation models for ASD, VSD, AVSD, and normal heart, respectively. The U-Net architecture successfully predicted segmentation for all these groups. For each of these groups, performance for segmentation, especially pixel accuracy, reached above 95%—even for ASD and AVSD, at more than 99%.

In Fig. 7, we can see the segmentation results in terms of the original image, ground truth, and prediction image obtained using the CNN-based U-Net architecture. By looking at the picture, it can be seen that the presence of white holes connected in the atrial septum indicates the presence of ASD. Likewise, the presence of white holes connected in the ventricular septum indicates the presence of VSD, and their presence in both the atrial and ventricular septa shows AVSD. In this picture, we can also see that in a normal heart, there are no white holes connected in either the atrial or ventricular septum.

From the echo Doppler investigation, as can be seen in Fig. 8, the location of the defect in the atrium, ventricle or both septa is confirmed. This figure reveals Doppler flow (in red) from the left to the right in the atrial, ventricular or both septa, which demonstrates the location of ASD, VSD, and AVSD. If we compare the echo Doppler images with the proposed U-Net architecture segmentation, the location of the cardiac septum defect is very similar to the original image.

We have also compared the results of segmentation performance between U-Net and V-Net architectures. In Table 6, it can be seen that for segmentation performance, the U-Net model is better than the V-Net model. In U-Net pixel architecture, accuracy is 99.15, which is higher than V- Net. Likewise, mean IU, mean accuracy, precision, recall, and F1 score are higher than V-Net.

In addition to ground-truth predictions, we have also predicted the contours of the cardiac septal defects and compared the performance between the two architectural models. In Fig. 9, it is shown that the performance of ground truth and contour prediction of the cardiac septum defect is better in the U-Net architecture model compared to the V-Net.

We have revealed the graphs of accuracy and loss for each ASD, VSD, AVSD, and normal heart class in Fig. 10. The segmentation accuracy values obtained were 99.05%, 98.62%, 99.39%, and 98.97% for ASD, VSD, AVSD, and normal heart, respectively. In Fig. 10 we have also shown a loss model for each class. In the loss model, it can be seen that

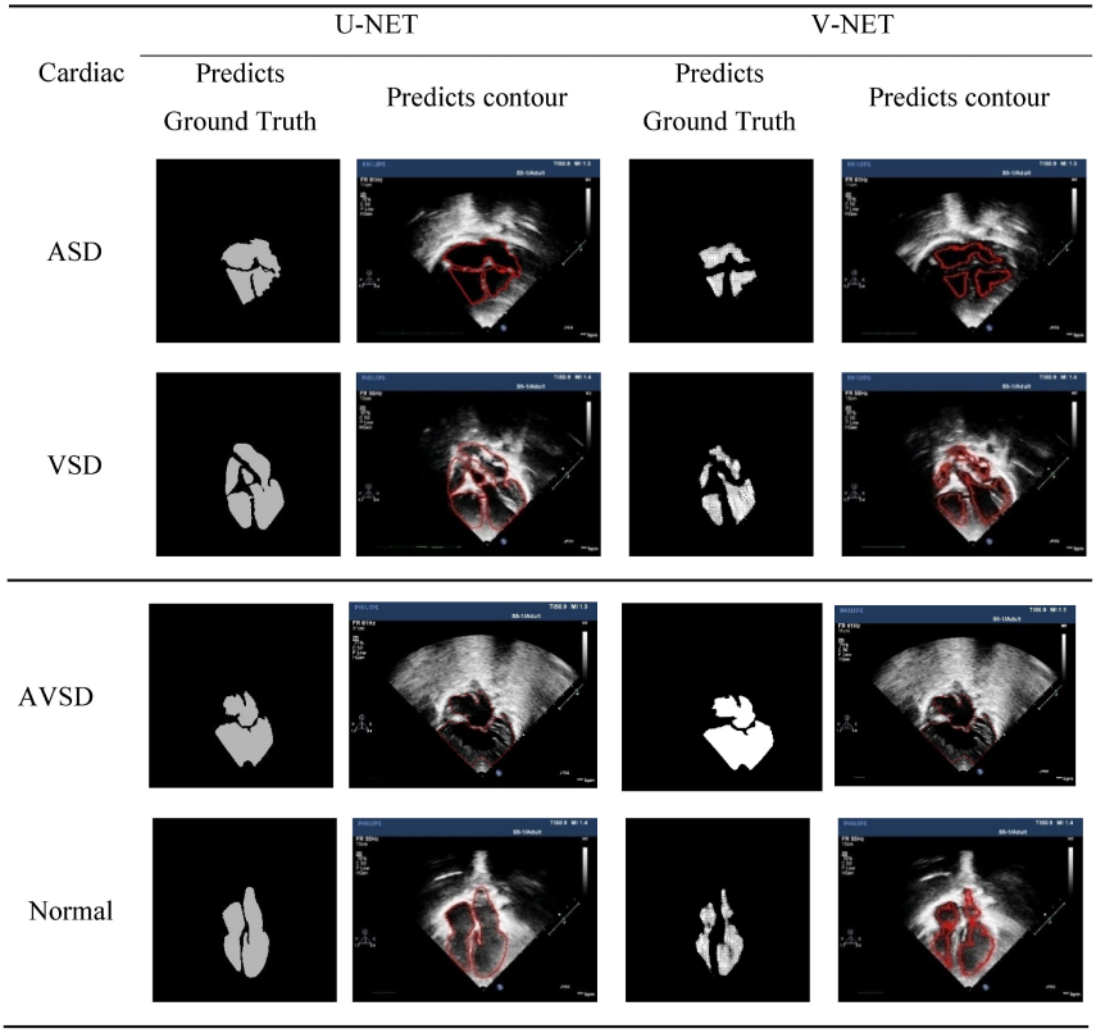


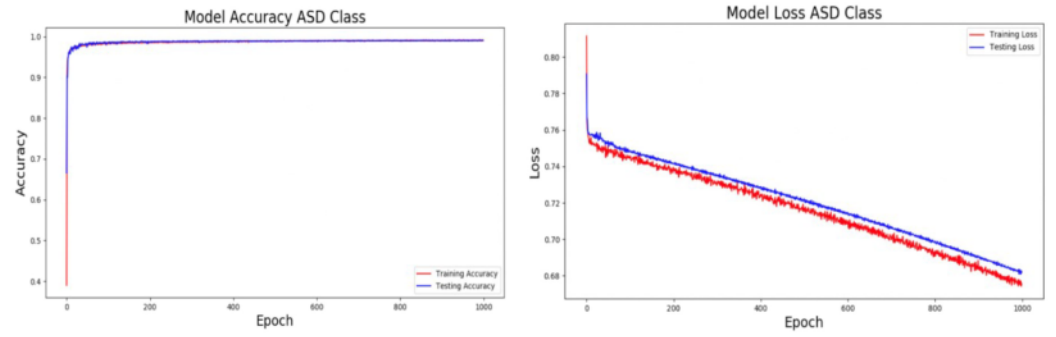
Fig. 9. Model comparison between U-Net and V-Net architecture.

from a small epoch to 1000 epoch, the loss value is low and stable, except for ASD.

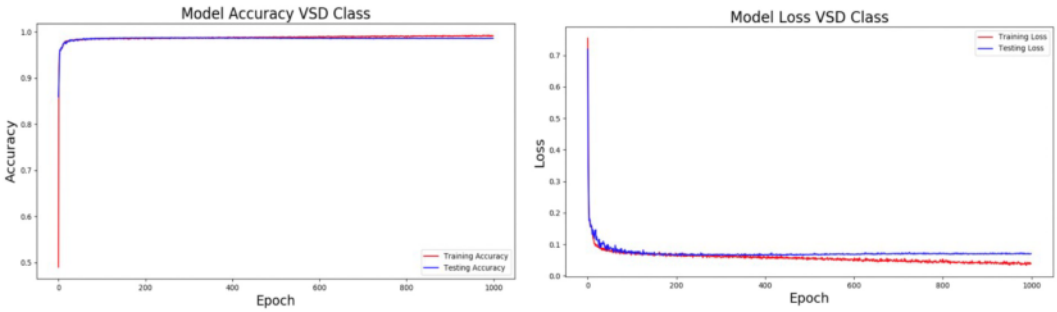
4. Discussion

To the best of our knowledge, this is the first report describing the segmentation of cardiac septal defect from a 2D echocardiogram image. In this study, we show that a CNN-based U-Net architecture can successfully account for segmentation of cardiac septal defects. The same was reported by Chen et al. although it was not with cardiac septal defects, regarding the success of CNN-based U-Net for cardiac segmentation [23].

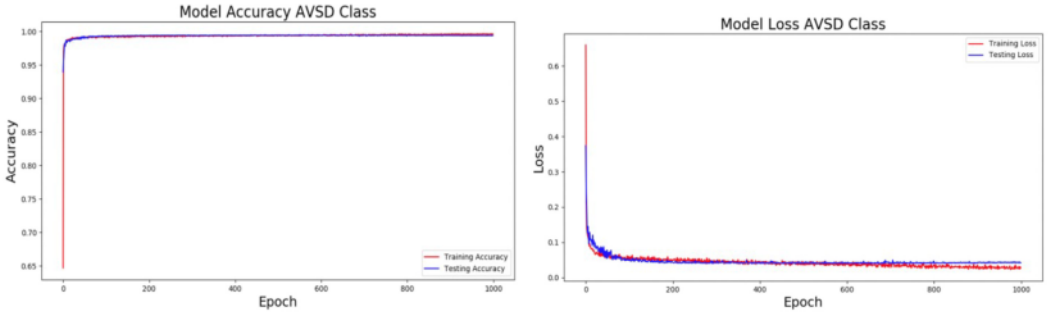
In this study, it has been shown that the proposed U-Net architecture can segment a normal heart and can also segment holes in the cardiac septum almost perfectly. The value of precision and recall is also high for each group, which is important to show us that there is no over or under segmentation of images. Likewise, with the dice score, the results were also high. A high dice score indicates that image segmentation with the proposed architectural model is almost similar to ground truth. Although there has not been a similar study, the performance results in our study are superior compared to research conducted by Perrin et al. In this study, we have described how CNNs were able to distinguish between hypoplastic left heart syndrome (HLHS) and transposition of the great arteries (TGA) with an accuracy of 92%; for aortic coarctation, however, the performance was still poor. The possibility of poor performance because the actual pathology of aortic coarctation is not in the field of view [24]. Several other studies have also proven the success of DL for 2D segmentation of cardiac ventricles, as shown in Table 7. Veni et al. and Smistad et al. have reported on the success of the DL method for carrying out the task of segmenting the left ventricle, but the results have not been as successful compared to studies conducted by recent researchers [25,27]. Research by Diller et al. has described the U-Net architecture model correctly in assessing patients with a systemic right ventricle and achieved high performance in segmenting the systemic right or left ventricle (with a dice metric between 0.79 and 0.88 depending on diagnosis) when compared with human experts [28]. In this study, they illustrate how appropriate DL models can be trained to recognize the systemic ventricle even in patients with complex cardiac anatomy and delineate the endocardial border in this setting [28]. Another study carried out by Jafari et al. has shown that U-Net



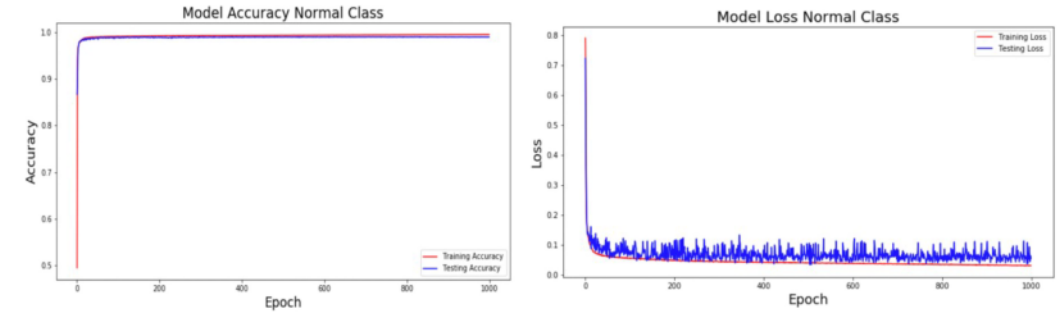
(a) ASD



(b) VSD



(c) AVSD



(d) Normal

Fig. 10. Loss and accuracy curve of ASD,VSD, AVSD, and normal heart.

Table 7
Summary review of deep learning methods for 2D ventricle segmentation.

Author	Method	Dice coefficient
Diller et al., 2019	U-Net architecture to segmenting the systemic right or left ventricle compared human experts	Normal heart 0.88 ± 0.06 TGA 0.86 ± 0.06 CcTGA 0.79 ± 0.08
Jafari et al., 2019	U-Net T-L net-based shape constraint on annotated frames	ED 94.1 ± 3.3 ES 93.0 ± 3.9
Leclerc et al., 2019	U-Net trained on a large heterogeneous dataset	ED 0.93 ± 0.04 ES 0.91 ± 0.06
Veni et al., 2018	FCN (U-Net) followed by level-set based deformable model	0.86 ± 0.06
Smistad et al., 2017	U net trained using labels generated by a Kalman filter-based method	0.86 ± 0.06
Our proposed model	U-Net for segmentation the infant heart	0.94 ± 0.05

architecture succeeded in segmenting the left ventricle during end-diastolic and end-systolic with semi-supervised learning methods [29]. Similarly, results were reported by Leclerc et al. about the success of the U-Net architecture for cardiac ventricular segmentation [30].

We have also compared the results of segmentation performance between U-Net and V-Net architectures. Both architecture models are used for medical image segmentation. The work of the V-Net architecture is the same as the U-Net, but the process in the architecture is slightly different. Moreover, V-Net is usually used for 3D images, while U-Net is used for 2D images [26,31]. In Table 6, it can be seen that for segmentation performance, the U-Net model is better than the V-Net model. In addition, using the U-Net architecture model proposed in this study can be more detailed and more accurately describe the atrial space, ventricular space, mitral valve, tricuspid valve, and aorta compared to V-Net. In U-Net pixel architecture, accuracy is 99.15 higher than V- Net. Likewise, mean IU, mean accuracy, precision, recall, and F1 score are higher than V-Net.

From the comparison result, we summarize as follows:

- The overall performance of the proposed model was better than its counterpart with limited datasets. This implies that it is more suitable for larger and more heterogeneous scale datasets.
- The overall performance of CNN segmentation-based U-Net architecture was better when assessed with four infant heart conditions: ASD, VSD, AVSD, and normal. This is an indication that the proposed model can be improved for other abnormalities in the heart.
- The performance result was compared with V-Net architecture, which produces higher performances in terms of pixel accuracy, mean IU, mean accuracy, precision, recall, and dice score.

Appendices.

Although the results look promising, there are some limitations of our study. (i) Only one echocardiogram view is segmented, and patient variation is still limited. To make the performance of cardiac septal defect detection more accurate, it is necessary to segment some echocardiogram views so that the type of cardiac septal defect can be determined. (ii) To expand this study to other abnormal conditions might add a great contribution to this line of research.

5. Conclusion

This study has been successful in establishing the automatic diagnosis of cardiac septal defects. A segmentation of the cardiac septal defect in 2D echocardiogram images was obtained using convolutional neural networks. The CNN-based U-Net architecture can successfully account for segmentation of cardiac septal defects. Using the proposed segmentation model for four classes, namely ASD, VSD, AVSD, and normal heart, produces a pixel accuracy of 99.15%, mean IU of 94.69%, mean accuracy of 97.73%, sensitivity of 96.02%, and F1 score of 94.88%, respectively. In this study, it was proven that the proposed U-Net architecture model has a very high degree of accuracy with a very small error rate for predicting contour lesions in cardiac septal defects. Through these findings, the diagnosis of a cardiac septal defect will be more precise and can be done automatically, so it can be utilized by all physicians when performing an echocardiogram examination. In the future, this research will be carried out with a greater number of patients and by combining several echocardiogram views.

Author contributions

S.N. Conceptualization, supervision, review & editing, data analyst; R.N. Data preparation, resources, and writing-original draft; R.U.P. Review & editing; S.T.P. review & editing.

Declaration of competing interest

The authors declare no conflict of interest.

Acknowledgment

We thank all our colleagues and all cardiac septal defect patients who contributed greatly to the current study. We greatly appreciate the support of all students in the Intelligent System Research Group (ISysRG), Faculty of Computer Science, Universitas Sriwijaya.

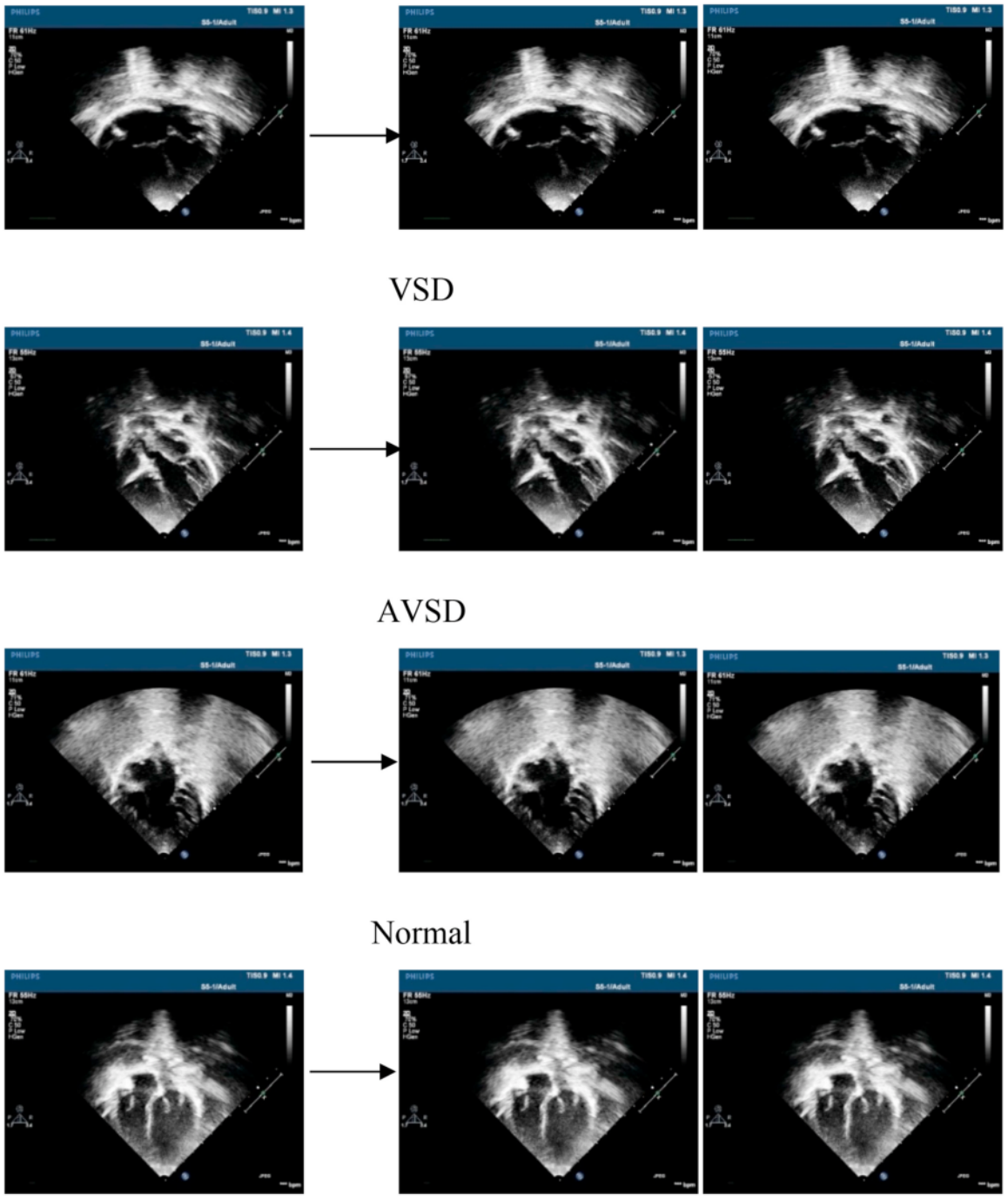


Fig. 3. Conversion of US video of ASD, VSD, AVSD, and normal heart to frames

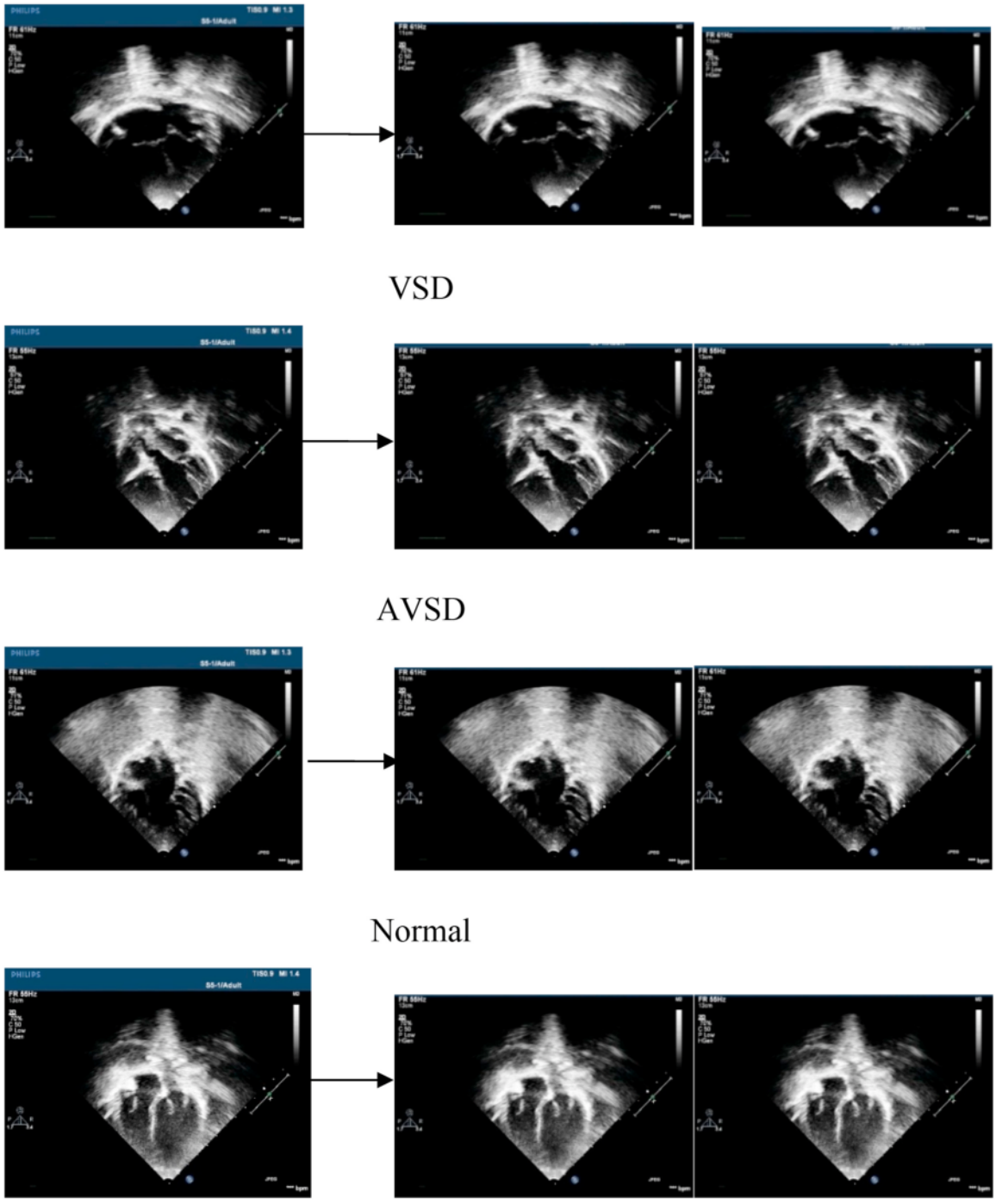


Fig. 4. Image-cropping process on ASD,VSD, AVSD, and normal heart

References

[1] Puri K, Allen HD, Qureshi AM. Congenital heart disease. Pediatr Rev 2017;38(10):471–86. https://doi.org/10.1542/pir.2017-0032.

[2] Liu Y, et al. Global birth prevalence of congenital heart defects 1970-2017: updated systematic review and meta-analysis of 260 studies. Int J Epidemiol 2019;48(2):455–63. https://doi.org/10.1093/ije/dyz009.

[3] Rashid U, Qureshi AU, Hyder SN, Sadiq M. "Pattern of congenital heart disease in a developing country tertiary care center : factors associated with delayed diagnosis. Ann Pediatr Cardiol 2016;9(3):210–5. https://doi.org/10.4103/0974-2069.189125.

[4] Backer CL, Eltayeb O, Mongé MC, Mazwi ML, Costello JM. Shunt lesions Part I: patent ductus arteriosus, atrial septal defect, ventricular septal defect, and atrioventricular septal defect. Pediatr Crit Care Med 2016;17(8):S302–9. https://doi.org/10.1097/PCC.0000000000000786.

[5] Mcleod G, et al. Echocardiography in congenital heart disease. Prog Cardiovasc Dis 2018. https://doi.org/10.1016/j.pcad.2018.11.004. 468–75.

[6] Saraf RP, V Suresh P, Maheshwari S, Shah SS. "Pediatric echocardiograms performed at primary centers : diagnostic errors and missing links ! Ann Pediatr Cardiol 2015;8(258):20–5. https://doi.org/10.4103/0974-2069.149514.

[7] Zhao QM, Niu C, Liu F, Wu L, Ma XJ, Huang GY. Accuracy of cardiac auscultation in detection of neonatal congenital heart disease by general paediatricians. Cardiol Young May 2019;29(5):679–83. https://doi.org/10.1017/S1047951119000799.

[8] Vick AGW, Bezold LI. Isolated atrial septal defects (ASDs) in children : classification , clinical features , and diagnosis. 2020. UptoDate.

[9] Fulton ADR, Saleeb S. Isolated ventricular septal defects in infants and children : anatomy , clinical features , and diagnosis. 2020. p. 1–26. UptoDate.

[10] Ghorbani A, et al. Deep learning interpretation of echocardiograms. npj Digit Med 2020:1–10. https://doi.org/10.1038/s41746-019-0216-8.

[11] Kusunose TSM, Haga K, Abe A. Utilization of artificial intelligence in echocardiography. Circ J 2019;83(8):1623–9. https://doi.org/10.1253/circj.CJ-19-0420.

[12] Gandhi S, Mosleh W, Shen J, Chow CM. Automation, machine learning, and artificial intelligence in echocardiography: a brave new world. Echocardiography 2018;35(9):1402–18. https://doi.org/10.1111/echo.14086.

[13] Litjen G TJ, Ciompi F, Wolterink JM, de Vos B, Leiner T. State-of-the-Art deep learning in cardiovascular image analysis. JACC (J Am Coll Cardiol): Cardiovascular Imaging 2019;12(8):1549–65. https://doi.org/10.1016/j.jcmg.2019.06.009.

[14] Madani A, Arnaout R, Mofrad M, Arnaout R. Fast and accurate view classification of echocardiograms using deep learning. npj Digital Medicine 2018;1(1):1–8. https://doi.org/10.1038/s41746-017-0013-1.

[15] Khamis H, Zurakhov G, Azar V, Raz A, Friedman Z, Adam D. Automatic apical view classification of echocardiograms using a discriminative learning dictionary. Med Image Anal 2017;36:15–21. https://doi.org/10.1016/j.media.2016.10.007.

[16] Dezaki FT, et al. Cardiac phase detection in echocardiograms with densely gated recurrent neural networks and global extrema loss. IEEE Trans Med Imag 2019;38(8):1821–32. https://doi.org/10.1109/TMI.2018.2888807.

[17] Smistad E, Ostvik A, Haugen BO, Lovstakken L. 2D left ventricle segmentation using deep learning. IEEE International Ultrasonics Symposium, IUS 2018. https://doi.org/10.1109/ULTSYM.2017.8092812. 1–4.

[18] Alcevska E. Segmentation of the left ventricle of the heart in 2D ultrasound images using convolutional neural networks. 2016.

[19] Leclerc S, et al. Deep learning for segmentation using an open large-scale dataset in 2D echocardiography. IEEE Trans Med Imag 2019;99:1–13. https://doi.org/10.1109/TMI.2019.2900516.

[20] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. 2015. p. 1–8. arXiv:1505.04597v1 [cs.CV].

[21] Shen D, Wu G, Suk H. Deep learning in medical image analysis. AnnuRev Biomed Eng 2017;19:221–48.

[22] Liu S, et al. Deep learning in medical ultrasound Analysis : a review. Engineering 2019;5(2):261–75. https://doi.org/10.1016/j.eng.2018.11.020.

[23] Chen C, et al. "Deep learning for cardiac image Segmentation : a review. Front Cardiovasc Med 2020;7:1–33. https://doi.org/10.3389/fcvm.2020.00025. March.

[24] Perrin DP, Bueno A, Rodriguez A, Marx GR, del Nido PJ. Application of convolutional artificial neural networks to echocardiograms for differentiating congenital heart diseases in a pediatric population. Med Imaging 2017. https://doi.org/10.1117/12.2254083. Computer-Aided Diagnosis, vol. 10134, p. 1013431, 2017.

[25] Veni G, Moradi M, Bulu H, Narayan G, Syeda-Mahmood T. Echocardiography segmentation based on a shape-guided deformable model driven by a fully convolutional network prior. Proceedings - International Symposium on Biomedical Imaging 2018:898–902. https://doi.org/10.1109/ISBI.2018.8363716. 2018-April, no. Isbi.

[26] Milletari F, Navab N, Ahmadi SA. V-Net: fully convolutional neural networks for volumetric medical image segmentation. In: Proceedings - 2016 4th international conference on 3D vision, 3DV 2016; 2016. p. 565–71. https://doi.org/10.1109/3DV.2016.79.

[27] Smistad E, Ostvik A, Haugen BO, Lovstakken L. 2D left ventricle segmentation using deep learning. In: IEEE international ultrasonics symposium. IUS; 2017. p. 1–5. https://doi.org/10.1109/ULTSYM.2017.8092573. July 2018.

[28] Diller G-P, et al. Utility of machine learning algorithms in assessing patients with a systemic right ventricle. Eur Heart J Cardiovasc Imaging 2019;20(8):925–31. https://doi.org/10.1093/ehjci/jey211.

[29] Jafari MH, et al. Semi-supervised learning for cardiac left ventricle segmentation using conditional deep generative models as prior. In: Proceedings - international symposium on biomedical imaging. vol. 2019; 2019. p. 649–52. https://doi.org/10.1109/ISBI.2019.8759292. April, no. Isbi.

[30] Leclerc S, et al. Deep learning for segmentation using an open large-scale dataset in 2D echocardiography. IEEE Trans Med Imag 2019;38(9):2198–210. https://doi.org/10.1109/TMI.2019.2900516.

[31] Zhao Z, Yang L, Zheng H, Guldner IH, Zhang S, Z D, B C. Segmentation in 3D biomedical images using weak annotation, vol. 1. Lv. Springer International Publishing; 2018.

Automated image segmentation for cardiac septal defects based on contour region with convolutional neural networks: A preliminary study

ORIGINALITY REPORT

15%	13%	12%	6%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.frontiersin.org Internet Source	2%
2	www.coursehero.com Internet Source	2%
3	Siti Nurmaini, Muhammad Naufal Rachmatullah, Ade Iriani Sapitri, Annisa Darmawahyuni et al. "Accurate Detection of Septal Defects with Fetal Ultrasonography Images Using Deep Learning-based Multiclass Instance Segmentation", IEEE Access, 2020 Publication	1%
4	Ria Nova, Sukman Tulus Putra, Siti Nurmaini, Radiyati Umi Partan. "Cardiac Septal Defects in Children: Hemodynamics, Clinical Manifestations and Detection", Bioscientia Medicina : Journal of Biomedicine and Translational Research, 2021 Publication	1%
	s3-eu-west-1.amazonaws.com	

5	Internet Source	1 %
6	academic.oup.com Internet Source	1 %
7	hal.archives-ouvertes.fr Internet Source	1 %
8	www.ncbi.nlm.nih.gov Internet Source	1 %
9	Submitted to University of North Carolina, Charlotte Student Paper	1 %
10	W. Zhang, H. Huang, M. Schmitz, X. Sun, H. Wang, H. Mayer. "A MULTI-RESOLUTION FUSION MODEL INCORPORATING COLOR AND ELEVATION FOR SEMANTIC SEGMENTATION", The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2017 Publication	1 %
11	export.arxiv.org Internet Source	1 %
12	annalspc.com Internet Source	1 %
13	www.spiedigitallibrary.org Internet Source	1 %

14 Siti Nurmaini, Alexander Edo Tondas, Annisa Darmawahyuni, Muhammad Naufal Rachmatullah et al. "Electrocardiogram Signal Classification for Automated Delineation using Bidirectional Long Short-Term Memory", Informatics in Medicine Unlocked, 2020
Publication

1 %

Exclude quotes	On	Exclude matches	< 1%
Exclude bibliography	On		