Check for updates

Automated segmentation of 3D cine cardiovascular magnetic resonance imaging

Soroosh Tayebi Arasteh[1,2,3]*, Jennifer Romanowicz[1,4], Danielle F. Pace[5,6], Polina Golland[6], Andrew J. Powell[1], Andreas K. Maier[2], Daniel Truhn[3], Tom Brosch[7], Juergen Weese[7], Mahshad Lotfinia[8], Rob J. van der Geest[9] and Mehdi H. Moghari[10]

[1]Department of Cardiology, Boston Children's Hospital, and Department of Pediatrics, Harvard Medical School, Boston, MA, United States, [2]Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany, [3]Department of Diagnostic and Interventional Radiology, University Hospital RWTH Aachen, Aachen, Germany, [4]Department of Cardiology, Children's Hospital Colorado, and School of Medicine, University of Colorado, Aurora, CO, United States, [5]Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, MA, United States, [6]Computer Science & Artificial Intelligence Lab, Massachusetts Institute of Technology, Cambridge, MA, United States, [7]Philips Research Laboratories, Hamburg, Germany, [8]Institute of Heat and Mass Transfer, RWTH Aachen University, Aachen, Germany, [9]Department of Radiology, Leiden University Medical Center, Leiden, Netherlands, [10]Department of Radiology, Children's Hospital Colorado, and School of Medicine, University of Colorado, Aurora, CO, United States

Introduction: As the life expectancy of children with congenital heart disease (CHD) is rapidly increasing and the adult population with CHD is growing, there is an unmet need to improve clinical workflow and efficiency of analysis. Cardiovascular magnetic resonance (CMR) is a noninvasive imaging modality for monitoring patients with CHD. CMR exam is based on multiple breath-hold 2-dimensional (2D) cine acquisitions that should be precisely prescribed and is expert and institution dependent. Moreover, 2D cine images have relatively thick slices, which does not allow for isotropic delineation of ventricular structures. Thus, development of an isotropic 3D cine acquisition and automatic segmentation method is worthwhile to make CMR workflow straightforward and efficient, as the present work aims to establish.

Methods: Ninety-nine patients with many types of CHD were imaged using a non-angulated 3D cine CMR sequence covering the whole-heart and great vessels. Automatic supervised and semi-supervised deep-learning-based methods were developed for whole-heart segmentation of 3D cine images to separately delineate the cardiac structures, including both atria, both ventricles, aorta, pulmonary arteries, and superior and inferior vena cavae. The segmentation results derived from the two methods were compared with the manual segmentation in terms of Dice score, a degree of overlap agreement, and atrial and ventricular volume measurements.

Results: The semi-supervised method resulted in a better overlap agreement with the manual segmentation than the supervised method for all 8 structures (Dice score $83.23 \pm 16.76\%$ vs. $77.98 \pm 19.64\%$; P-value ≤ 0.001). The mean difference error in atrial and ventricular volumetric measurements between manual segmentation and semi-supervised method was lower (bias ≤ 5.2 ml) than the supervised method (bias ≤ 10.1 ml).

Discussion: The proposed semi-supervised method is capable of cardiac segmentation and chamber volume quantification in a CHD population with wide anatomical variability. It accurately delineates the heart chambers and great vessels and can be used to accurately calculate ventricular and atrial volumes throughout the cardiac cycle. Such a segmentation method can reduce inter- and intra- observer variability and make CMR exams more standardized and efficient.

1. Introduction

Congenital heart disease (CHD) affects nearly *1%* of live births in the United States and Europe (1, 2). It is the leading cause of perinatal and infant death, and accounts for about *4%* of all neonatal deaths in the United States (3). Luckily, *83%* of babies with CHD now survive infancy and reach adulthood in the United States, due to medical, surgical, and intensive care advances (4). Despite these advances, there is no cure for CHD, and thus, these patients must be monitored lifelong. In addition to echocardiography, cardiovascular magnetic resonance (CMR) is one of the primary imaging modalities for monitoring CHD patients since it is non-invasive and does not require ionizing radiation (5). Although CMR can provide high spatial and temporal resolution images to calculate cardiac function, these measurements are traditionally based on multiple 2-dimensional (2D) cine image acquisitions that are precisely prescribed. For instance, 2D cine images are planned in a short-axis view to calculate cardiac function. However, defining the short-axis view is expert and institution dependent, and serial examinations may prescribe slightly different planes resulting in measurement variability. Furthermore, 2D cine images have a relatively large slice thickness which precludes isotropic delineation of the left and right ventricle for a more accurate measurement of cardiac function. An isotropic 3-dimensional (3D) cine sequence can mitigate these problems. Because it is non-angulated, prescription is independent from the expert and institution. Isotropic 3D cine images also allow for isotropic segmentation of the heart.

Manual segmentation of 3D cine images is, however, time consuming and impractical. Therefore, an automatic segmentation method is necessary to delineate the heart chambers and great vessels and to evaluate cardiac function. At present, artificial intelligence (AI) and, specifically, deep learning (DL) is making strong progress in automatic segmentation of CMR images (6–11). AI-based methods have been successfully utilized to delineate adult heart disease and CHD using 2D cine images (8, 10, 12). However, they are not designed for segmenting 3D cine CMR images. Prior work has also addressed the physical distortions and appearance changes that arise when segmenting the right and left ventricles in adult patients (7, 13–15). Method development regarding automated image analysis for CHD is very limited and has mostly been aimed towards computer-aided diagnosis (16–18), ventricular function quantification (19–23), 2D cine (9, 18, 24), or 3D static imaging (20, 25–29). Recently, more novel AI-based segmentation methods are proposed in the literature (30–35) which are based on vision transformers (36) but their performance on whole-heart segmentation have not been investigated. To date, there is no automatic DL-based segmentation method available that works on 3D cine datasets for delineating whole-heart and great vessels, nor is there work addressing cardiac segmentation in a wide range of CHD lesions.

In this article, we sought to develop fully automatic supervised and semi-supervised DL-based segmentation methods for a wide range of CHD lesions that delineate the heart and great vessels including left atrium (LA), left ventricle (LV), aorta (AO), superior vena cava (SVC), inferior vena cava (IVC), right atrium (RA), right ventricle (RV), and pulmonary arteries (PA), from 3D cine images. We hope the proposed technique enables isotropic delineation of cardiac anatomy and great vessels throughout the cardiac cycle and can be used for the assessment of cardiac function and anatomy.

2. Materials and methods

We have developed two 3D DL-based segmentation methods: a fully automatic supervised method and a fully automatic semi-supervised method. The objective was to evaluate and compare their performances in delineating the whole-heart and great vessels with respect to each other and to manual segmentation. Both methods employ the same network architecture and prediction pipeline; the only difference lies in their training processes.

2.1. The 3D cine datasets

2.1.1. CMR imaging protocol and patient population

We retrospectively identified 99 patients who had undergone 3D cine CMR imaging at our institution. For each patient, clinically indicated 3D cine images were acquired about 2 min after injection of 0.15 mmol/kg of gadobutrol contrast agent using a 1.5 T Philips scanner (Philips, Best, The Netherlands). The imaging parameters for the 3D cine acquisition were as follows: field-of-view of 512 (frequency-encode) × 170 (phase-encode) × 170 (slice-encode) mm, Cartesian k-space trajectory, profile

ordering centra, isotropic resolution of $1.51 \times 1.48 \times 1.51$ mm reconstructed to $1.0 \times 1.0 \times 1.0$ mm, 20 acquired heart phases reconstructed to 30 phases, parallel imaging technique (SENSE) with reduction factor of 2 (phase-encode) $\times 2$ (slice-encode), TFE factor 11, and TFE shots 310, flip angle 60°, repetition time of 4.5 ms, echo time of 2.3 ms, nominal scan time of approximately 5:12 min assuming heartrate of 60 bpm and a 100% scan efficiency. A respiratory navigator was used for minimizing the respiratory motion of the heart. The images were reconstructed online on the scanner. After acquisition, the 3D cine images were de-identified for analysis. This retrospective study was reviewed and approved by the institutional review board of Boston Children's Hospital, Boston, MA, USA (IRB-P00011748).

The patient cohort had an age range of 0.8–72 years (median 16), heart rate range of 57–131 bpm (median 83), and weight range of 5.8–113.6 kg (median 55.3). The patient population was imaged to assess CHD ($n = 81$), acquired heart disease ($n = 2$), connective tissue disease or aortopathy ($n = 6$), cardiomyopathy ($n = 5$), arrhythmia ($n = 3$), or a family history of inheritable heart disease ($n = 2$). With consideration to what features of abnormal cardiac anatomy would make automatic segmentation of the heart more difficult, we devised a set of rules to categorize patients into 3 complexity classifications of normal, mild/moderate, and severe (Table 1).

Table 2 illustrates further details about final patient categorization into normal anatomy, mild/moderate, and severely complex groups. Supplementary Table S2 provides detailed information about all the subjects in this study.

2.1.2. Pre-processing

All the 3D frames of every 3D cine image were cropped around the heart so that all 30 frames were the same size for each patient. Per patient, a specific 3D bounding box containing the whole heart was visually chosen and the cropping occurred accordingly. This was done to lower computational costs and standardize the fields of view. Moreover, the frames corresponding to the end-diastole (ED) and end-systole (ES) were visually selected for each 3D cine dataset. Because intensity distributions vary across CMR images, intensity normalization is required. Therefore, we created an intensity normalization scheme based on the estimated blood pool and lung intensities in each image of the 3D cine dataset (24).

2.1.3. Manual annotation and ground truth labels

From the 99 total cases analyzed in this study, 74 cases were utilized for training, 13 cases were chosen as validation images for optimizing the networks during the deep learning training, and 12 cases were selected as test images. The test and validation patients were selected in a balanced manner from the 3 anatomy complexity categories (normal, mild/moderate, and severe) to maintain generalizability as much as possible. Supplementary Table S1 shows the distribution of the 3D cine datasets based on their severity classes. The training and validation images were manually annotated on the ED and ES frames, i.e., only two out of thirty 3D frames contained manual annotations for training the DL-based segmentation methods. However, the test images were manually annotated on all 30 frames to serve as a reference, and to evaluate the DL-based segmentation methods throughout the cardiac cycle. A 3D U-Net (37, 38) convolutional neural network trained on 3D static CMR images (25, 25, 39) was used to pre-segment the ED and ES phases for the training and validation datasets. The output of the network was then refined using 3DSlicer toolkit (40) via generic segmentation tools such island relabeling, further painting, or erasing. The segmentation results were supervised and, if necessary, modified by a pediatric cardiologist with expertise in CMR imaging. A similar procedure was performed to segment all 30 phases of test datasets. Specifically, each of the 30 frames underwent pre-segmentation using the 3D U-Net convolutional neural network (37, 38) to segment all 30 phases of the test datasets individually. The pre-segmentation results were then manually adjusted for each phase by a single observer using 3DSlicer and its generic segmentation modules. Finally, a clinician reviewed and supervised the final segmentation results.

Two DL-based segmentation methods were developed for fully automatic segmentation of 3D cine images: 1. Supervised DL-based segmentation method and 2. Semi-supervised segmentation method. In the supervised DL-based segmentation method only the manual annotation of ED and

TABLE 1 Method for classifying subjects' cardiac anatomies into 3 categories of complexity. The 3 categories of complexity include normal anatomy and mild/moderate and severely complexity.

Category	Segmental anatomy	Ventricles	Shunts	Valve abnormalities	Miscellaneous
Normal	Normal	Two	Patent foramen ovale	Bicuspid aortic valve (BAV)	Cardiomyopathy, coarctation, interrupted aortic arch, aortopathy, anomalous coronaries
Mild/ moderate	Double-outlet right ventricle (DORV), unrepaired D-TGA	Two/Two, but unequal	Arterial septal defect (ASD), ventricular septal defect (VSD), Patent Ductus Arteriosus (PDA), coronary fistula, aortopulmonary (AP) window, partial/ total anomalous pulmonary venous connection (PAPVC/TAPVC)	Common atrioventricular valve (AVV), any valvular stenosis, any prosthetic valve, AVV straddle, Tetralogy of fallot (TOF), truncus, Ebstein	Dextro transposition of the great arteries (D-TGA) status post (S/P) arterial switch, S/P Glenn of Fontan, dextrocardia, 2× mild
Severe	Levo Transposition of the great arteries (L-TGA), more complex arrangements	One		Any valvular atresia	S/P atrial switch, 2× moderate

TABLE 2 Details of the subjects according to cardiac diagnosis.

Normal	26
Cardiomyopathy	5
Hypertrophic cardiomyopathy	3
Arrhythmogenic right ventricular cardiomyopathy (ARVC)	1
Left ventricular non-compaction cardiomyopathy	1
Arrhythmia	3
Rule out (R/O) ARVC	2
Wolff-Parkinson-White	1
Family history	2
Ventricular fibrillation arrest	1
Aortopathy	1
Coarctation/BAV	8
Isolated coarctation of the aorta (CoA)	3
CoA + BAV	2
Isolated BAV	3
Connective tissue disease (CTD)	6
Marfan's syndrome	4
CTD, aortopathy nitric oxide synthesis (NOS)	2
Miscellaneous	2
ARCAPA	1
Lipomatous hypertrophy of the interatrial septum	1
Mild/moderate	39
Left-to-right shunt	8
Repaired VSD + arch repair	3
Repaired ASD	1
Repaired TAPVC	1
Repaired AP window + arch repair	1
Left coronary artery to RV fistula	1
Sinus venosus defect	1
Prosthetic valve	4
Congenital aortic stenosis, S/P aortic valve replacement (AVR)	3
Rheumatic heart disease, S/P mitral valve regurgitation & AVR	1
Congenital aortic valve disease	9
Congenital pulmonary valve disease	2
D-TGA, S/P arterial switch operation	1
Repaired TOF	7
Repaired truncus arteriosus	2
Ebstein anomaly	2
DORV	2
D-TGA, S/P Rastelli repair	2
Severe	34
Single ventricle	5
Double-inlet left ventricle	2
Hypoplastic left heart syndrome, S/P Fontan	1
Tricuspid atresia S/P Fontan	1
Double-inlet/double-outlet RV	1
Complex segmental anatomy	3
Superior-inferior ventricles, S/P Glenn	2
Dextrocardia, D-TGA, crisscrossed AV valves, S/P Glenn	1
L-TGA	5
valvular atresia	11
TOF/PA	4
DORV/PA	4
PA/interventricular septum	2
Ebstein/PA	1
Atrial switch	2
2× Moderate Lesions	8
DORV, right-dominant atrioventricular canal	3

(continued)

TABLE 2 Continued

DORV, S/P Glenn	2
TOF, S/P Fontan	1
Dextrocardia, D-TGA	1
D-TGA, S/P Glenn	1

ES frames were used for training. In the semi-supervised DL-based segmentation method, the manual annotation of ED and ES frames were automatically propagated to other 28 frames in the cardiac cycle, and manual annotation of ED and ES frames and propagated annotation of other 28 frames were used for training.
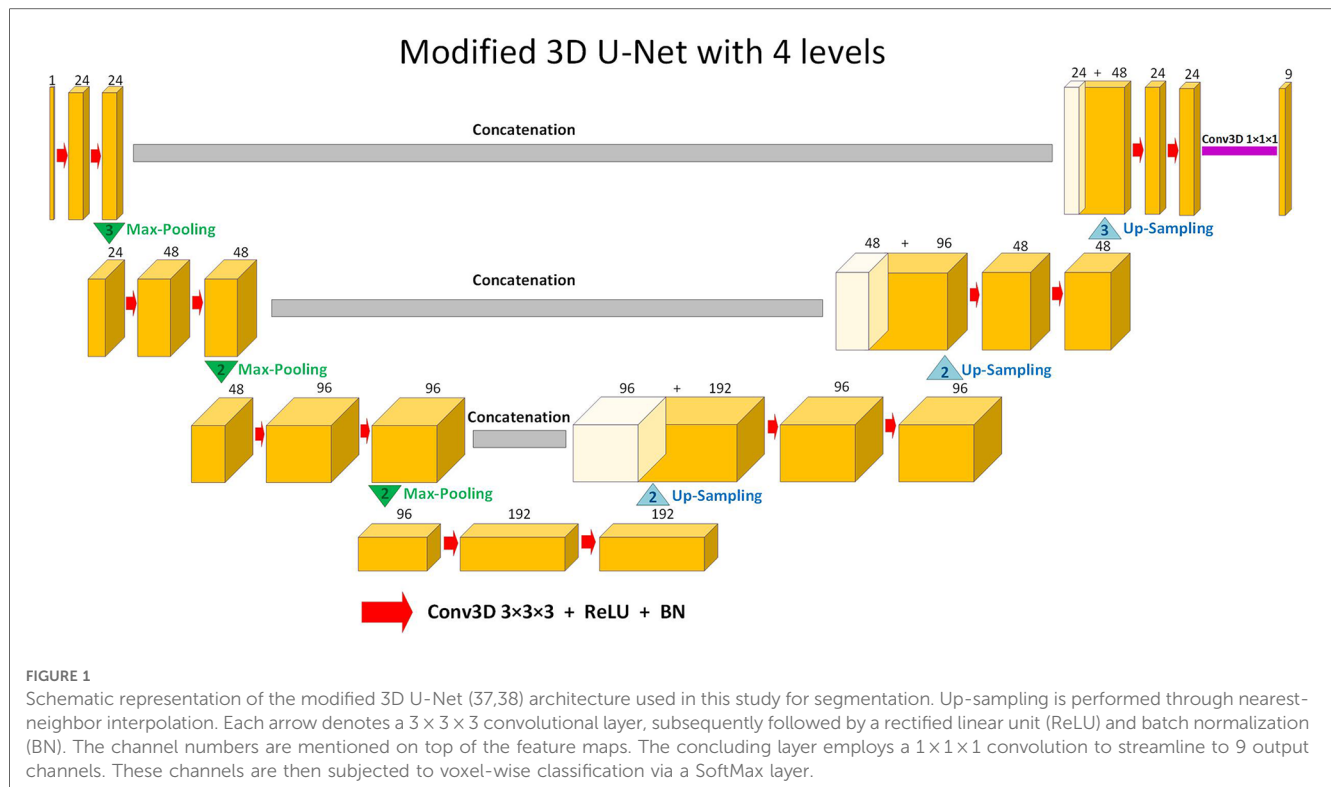
2.2. The supervised DL-based segmentation method

2.2.1. Approach

The supervised DL-based segmentation method has a 3D U-Net structure (Figure 1). Like the standard U-Net structure (37) and benefiting from nnU-Net (41) configurations, it has a contracting and an expanding path, each with four resolution levels (37, 38). In the contracting path, each layer contains two $3 \times 3 \times 3$ convolutions, each followed by a rectified linear unit (ReLU) (42), a batch normalization (43), and then a $2 \times 2 \times 2$ max pooling layer with strides of two in each dimension. In the expanding path, each layer consists of a nearest neighbor up-sampling of $2 \times 2 \times 2$ in each dimension, followed by two $3 \times 3 \times 3$ convolutions each followed by a ReLU and a batch normalization layer. Shortcut connections from layers of equal resolution in the contracting path provide high-resolution features to the expanding path. In the last layer, a $1 \times 1 \times 1$ convolution, which reduces the number of output channels to the number of labels, followed by a SoftMax layer, is used for the voxel-wise classification.

2.2.2. Supervised training

We followed a supervised approach and solely utilized the manually annotated ED and ES frames of the 3D cine datasets for training. The model was optimized using the AMSGrad (44) optimizer with a learning rate of 3×10^{-4}. As loss function, we chose the weighted cross-entropy with inverted class frequencies of the training data as loss weights to counteract the imbalanced class frequencies, with higher weights for smaller structures. To minimize the overhead and make maximum use of the graphics processing unit memory, we favored large input tiles over a large batch size and reduce the batch to a single 3D image (37). Consequently, the batch normalization acted like instance normalization in our implementation. We performed data augmentation by applying random affine transformations including rotation, shearing, translation, intensity scaling, and nonlinear transformations including left-right and anteroposterior flips (which are helpful for dextrocardia and other cardiac malposition in CHD), constant intensity shifts, and additive Gaussian noise on the datasets.

FIGURE 1
Schematic representation of the modified 3D U-Net (37,38) architecture used in this study for segmentation. Up-sampling is performed through nearest-neighbor interpolation. Each arrow denotes a $3 \times 3 \times 3$ convolutional layer, subsequently followed by a rectified linear unit (ReLU) and batch normalization (BN). The channel numbers are mentioned on top of the feature maps. The concluding layer employs a $1 \times 1 \times 1$ convolution to streamline to 9 output channels. These channels are then subjected to voxel-wise classification via a SoftMax layer.

2.3. The semi-supervised DL-based segmentation method

2.3.1. Approach

In each 3D cine dataset, only the ED and the ES phases of every 3D cine training image were manually annotated. Successive frames were not very different from each other. As such, inspired by Fan et al. (45), we utilized a label propagation algorithm to generate pseudo-labels for the intermediate frames which were not manually annotated to enhance our training. For each 3D cine image, we trained a separate DL-based segmentation neural network from scratch, using the ED and ES labels. We called this type of DL network the *propagator* network. This network was responsible for generating pseudo-labels for the remaining unlabeled 28 phases of the cine images.

The label propagation procedure is explained in more detail in **Algorithm 1**. Firstly, for each patient, we trained a propagator network, i.e., a supervised 3D U-Net (**Figure 1**) using only the ground truth labels, i.e., the ED and the ES traces. Consequently, in this stage, we had only two 3D images as the training data. Subsequently, this trained propagator network was used to automatically segment adjacent 3D volumes; specifically, we chose to segment three consecutive frames directly preceding and directly following both the ED and ES frames. In this way, we segmented 12 frames using the propagator network trained on only the ED and ES volumes (**Figure 2**). This process resulted in 1 manually annotated ED frame, 1 manually annotated ES frame, and 12 frames with pseudo-labels. As shown in **Figure 2**, we pooled the original ED

and ES frames with the newly segmented 12 frames and trained another propagator network using these images. This second propagator network followed the same procedure as the first, with the only exception that it segmented two consecutive frames of the training volumes. Likewise, we repeated this iterative procedure, segmenting two consecutive frames per iteration, until all the frames were segmented.

2.3.2. Weighted-probabilistic semi-supervised training

After performing the label propagation on all the available training images and generating all the pseudo-labels for the 28 intermediate non-annotated frames corresponding to each training patient, we trained the semi-supervised DL-based segmentation method. To be consistent with the supervised training method for comparison, we used the same U-Net structure and training parameters as the supervised DL-based segmentation method with the following modifications. Firstly, the loss weights, i.e., the inverted class frequencies, were updated according to the new distribution of data for each trained 3D U-Net. Secondly, we introduced a weighted-probabilistic training process to compensate for the accumulated error during the label propagations.

To reduce error propagation during training, we performed an iteration-based probabilistic training. We assigned a different weight to each 3D frame based on its distance from the ED and ES frames. The closer the frame was to either the ES or ED frame, the greater its weight. For each 3D cine image, we divided the 30 frames into 4 sets. The ED and ES frames, which had the

---

for *all 3D cine training images* **do**
    initialize the empty set *T*;
    initialize the empty set *S*;
    append all the 3D frames to *S*;
    append ED and ES volumes to *T*;
    remove the ED and ES volumes from *S*;
    train a supervised 3D U-Net using the elements of *T* (a propagator network);
    segment *k-nearest* neighbor 3D frames of ED and ES;
    append the newly-segmented volumes to *T*;
    remove the newly-segmented volumes from *S*;
    while *length(S) > 0* **do**
        train a propagator network using *T*;
        for all frames of *S* **do**
            if *frame in k − 1-nearest neighbors* **then**
                segment the frame;
            append the newly segmented volumes to *T*;
            remove the newly segmented volumes from *S*;

ALGORITHM 1
Label propagation strategy of the semi-supervised DL-based segmentation method. Here, *k* can be chosen according to the available dataset and the amount of motion between successive 3D frames of every 3D cine image. According to the 3D cine datasets and the average motion between successive 3D frames of every 3D cine image, we selected *k = 3* for all the 3D cine training images.

highest certainty of correctness for their labels, were assigned to set number 1 (D_1) and called *level-1* frames. The frames that were segmented using the first 3D U-Net of the label propagation process were assigned to set number 2 (D_2) and called *level-2* frames. We followed a similar strategy in dividing the remaining frames, until all 30 frames were assigned to a D_i set where, $i \in$ *(1,2,3,4)*. Consequently, we derived a training pool ordered according to the certainty of the annotation labels. To train the
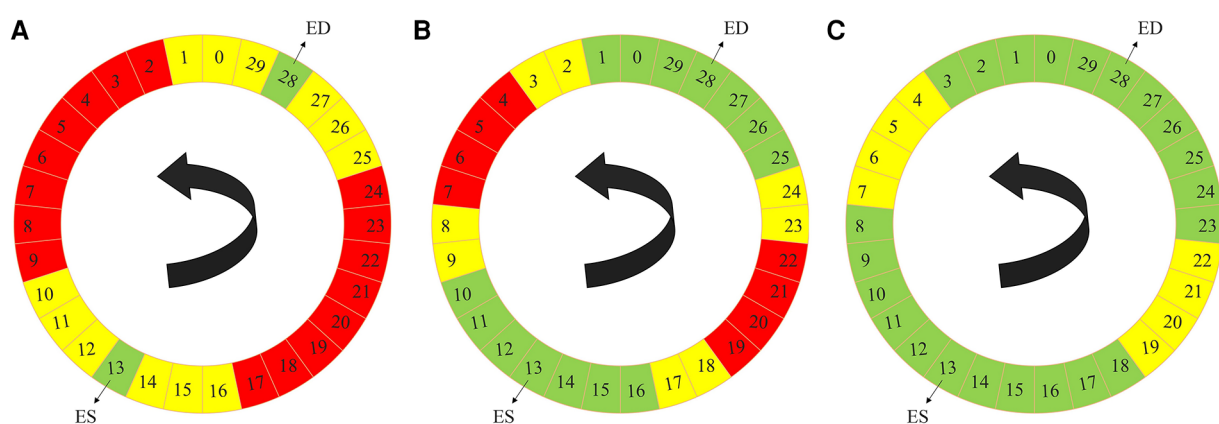

FIGURE 2
Label propagation process, to generate pseudo-labels for all the intermediate non-annotated frames to be used during weighted-probabilistic training, described by using one of the 3D cine datasets. (**A**) In the first step, the 3-nearest neighbors of the ED volume (29, 0, 1, 27, 26, and 25) and ES volume (14, 15, 16, 12, 11, and 10), shown in yellow, are segmented using the propagator network trained only on ED volume (28) and ES volume (13), shown in green. (**B**) In the second step, the 2-nearest neighbors of the segmented frames, from the ED volume side (2, 3, 24, and 23) and from the ES volume side (17, 18, 9, and 8), shown in yellow, are segmented using the propagator network trained on ED volume (28) and its 3-nearest neighbors (29, 0, 1, 27, 26, and 25) and ES volume (13) and its 3-nearest neighbors (14, 15, 16, 12, 11, and 10), shown in green. (**C**) In the last step, the 2-nearest neighbors of the newly-segmented frames, from the ED volume side (4, 5, 22, and 21) and from the ES volume side (19, 20, 7, and 6), shown in yellow, are segmented using the propagator network trained on the rest of the frames (3, 2, 1, 0, 29, 28, 27, 26, 25, 24, 23, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, and 18), shown in green, which concludes the segmentation of all the time frames.

network during each iteration, we randomly selected a 3D cine dataset from the 74 training datasets based on a uniform distribution. Then, we sampled two 3D volumes out of 30 frames of the chosen 3D cine dataset. The sampling occurred based on their certainty level (D_1, D_2, D_3, and D_4). The higher the level of a frame, the less its weight. Consequently, the highest weights were assigned to the D_1 set, i.e., the ED and ES frames had higher chance of getting chosen for a training batch during each training iteration.

2.4. The all-frames supervised DL-based segmentation method

2.4.1. Approach

To assess the effectiveness of the label propagation approach and weighted-probabilistic training, we devised a method involving a comparable 3D U-Net structure, as depicted in **Figure 1**. We trained this network using all the ED and ES frames from the training subjects. Afterward, we utilized the trained network to predict all unannotated frames within the training set, totaling 28 frames per training subject. The newly predicted 2,072 frames (28×74) were then combined with the ED and ES frames from the training subjects, resulting in a novel training set comprising 2,220 (30×74) 3D frames. Finally, we trained a new 3D U-Net structure using all frames from this augmented training set.

2.4.2. Supervised training

We adopted a supervised learning approach to train a new 3D U-Net, as depicted in **Figure 1**, using all 30 frames from the 74 training datasets. Each 3D frame, whether manually annotated (ED and ES) or automatically annotated, was assigned an equal probability of selection during each training iteration. This ensured that all frames in the training datasets were treated equally during the training process.

For optimization, we utilized the AMSGrad (44) optimizer with a learning rate of 3×10^{-4}, along with the weighted cross-entropy loss function. The class frequencies of the training data were inverted and used as weights during training. This approach considered all the 2,220 frames in the training data to ensure an effective learning process.

2.5. Quantitative evaluation of the method

As our evaluation metric, we used the "Dice score" to quantify segmentation accuracy. The Dice score evaluates the overlap between automated segmentation A and manual segmentation B and is defined as,

$$\text{DSC} = \frac{2|A \cap B|}{|A| + |B|} \tag{1}$$

It is a value between 0 and 1, with 0 denoting no overlap and 1 denoting 100% overlap between the segmentation results. The higher the Dice score, the better the agreement (9). We report the Dice score values in percentage. In addition to the Dice score calculation for individual structures, we used average Dice score over the 8 cardiac structures analyzed as,

$$\text{DSC}_{\text{total}} = \frac{1}{8} \sum_{i \in s} \text{DSC}_i, \tag{2}$$

where, $S = \{\text{AO, LV, PA, RA, SVC, IVC, LA, RV}\}$.

2.6. Ventricular and atrial volume measurement

We compared the ventricular and atrial volumes throughout the cardiac cycle among the manual, supervised, and semi-supervised segmentation results from the test datasets. We also reported the mean volume difference and mean volume difference expressed as a percentage over all 30 frames of the cardiac cycle.
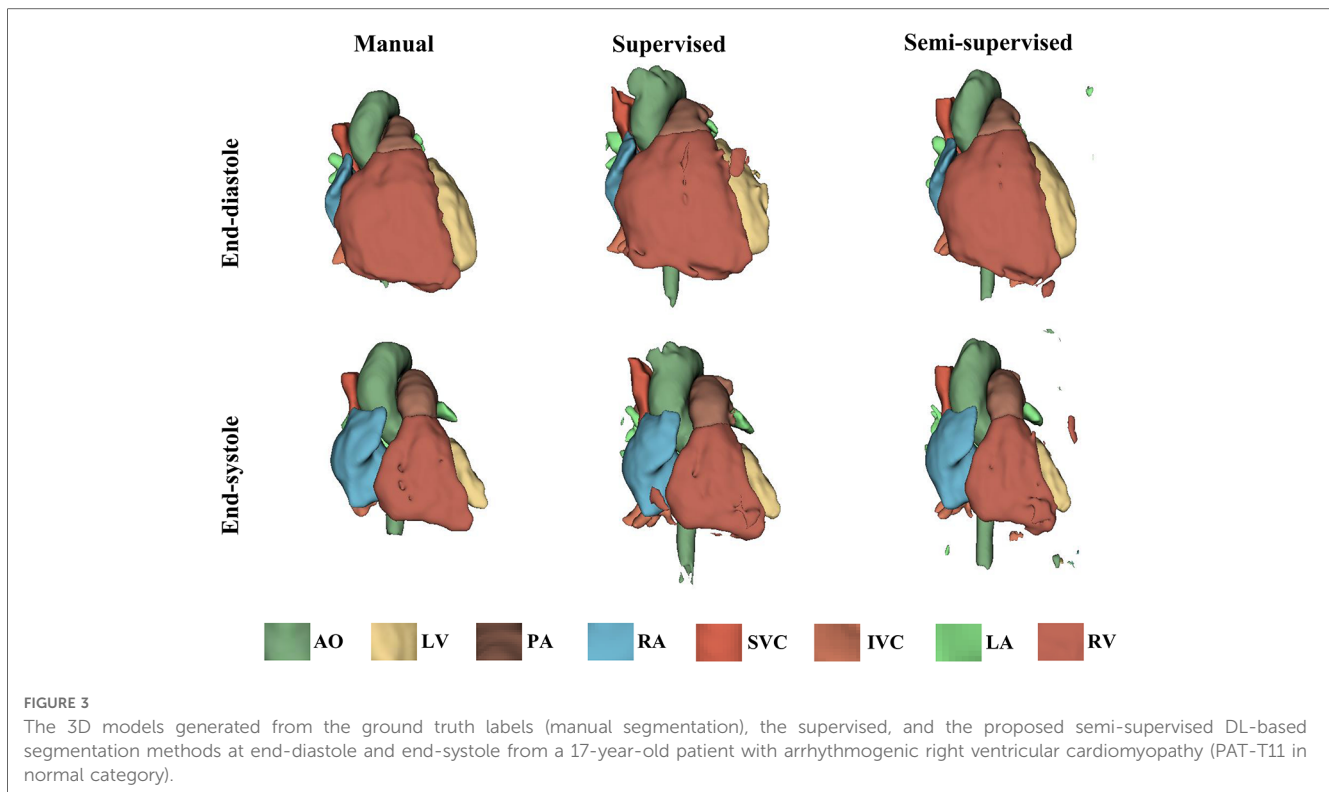
In addition, we reformatted the 3D cine images into a short-axis view, creating a series of multiple 2D cine images that covered the ventricles from base to apex. The slice thickness of the 2D images was set at 7 mm, with no slice gap. Subsequently, the 2D cine images were manually traced to accurately outline the end-systolic and end-diastolic volumes of both the LV and RV, allowing for the calculation of ejection fractions. These calculated volumes and ejection fractions were then compared [using intraclass correlation coefficient (ICC)] to those obtained from the manual segmentation of 3D cine images and the automatic segmentation of 3D cine images using supervised and semi-supervised DL-based methods.

2.7. Statistical analysis of the method

Descriptive statistics are reported as median and range, or mean ± standard deviation, as appropriate. Normality was tested using Shapiro-Wilk test (46). Either a two tailed *t*-test or a Wilcoxon signed-rank test (47) were used to compare two groups of paired data with Gaussian and non-Gaussian distributions, respectively. The two tailed *t*-test was used to compare the volumetric measurements and the Wilcoxon signed-rank test was used to compare the segmentation results in terms of Dice score. A *P*-value ≤0.05 was considered statistically significant.

2.8. Hardware

The utilized network architecture contained merely about 1.5 million parameters enabling fast and real-time segmentation of 3D volumetric images. Consequently, the hardware we used in our experiments for training and testing the networks were Intel

FIGURE 3
The 3D models generated from the ground truth labels (manual segmentation), the supervised, and the proposed semi-supervised DL-based segmentation methods at end-diastole and end-systole from a 17-year-old patient with arrhythmogenic right ventricular cardiomyopathy (PAT-T11 in normal category).

CPUs with 8 cores and 16 GB RAM and Nvidia GPUs of RTX 2060 with 6 GB memory.
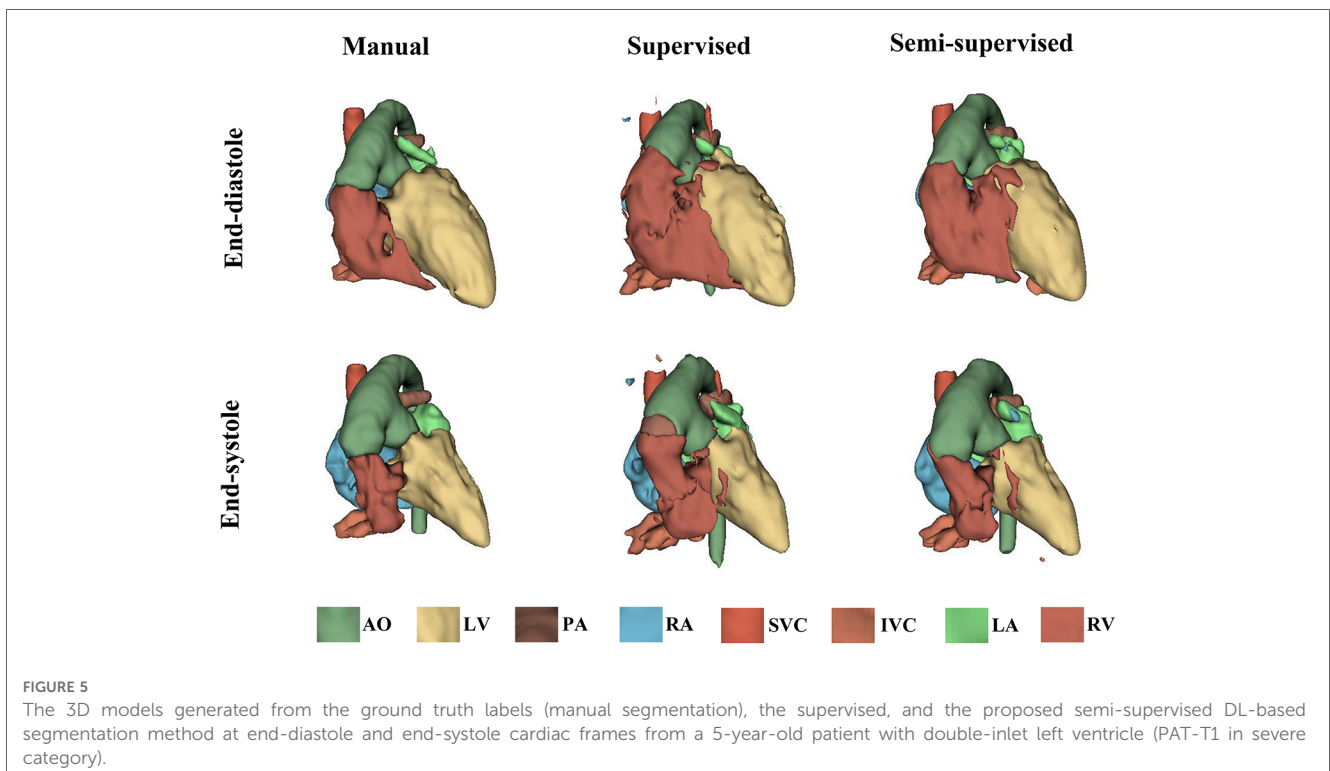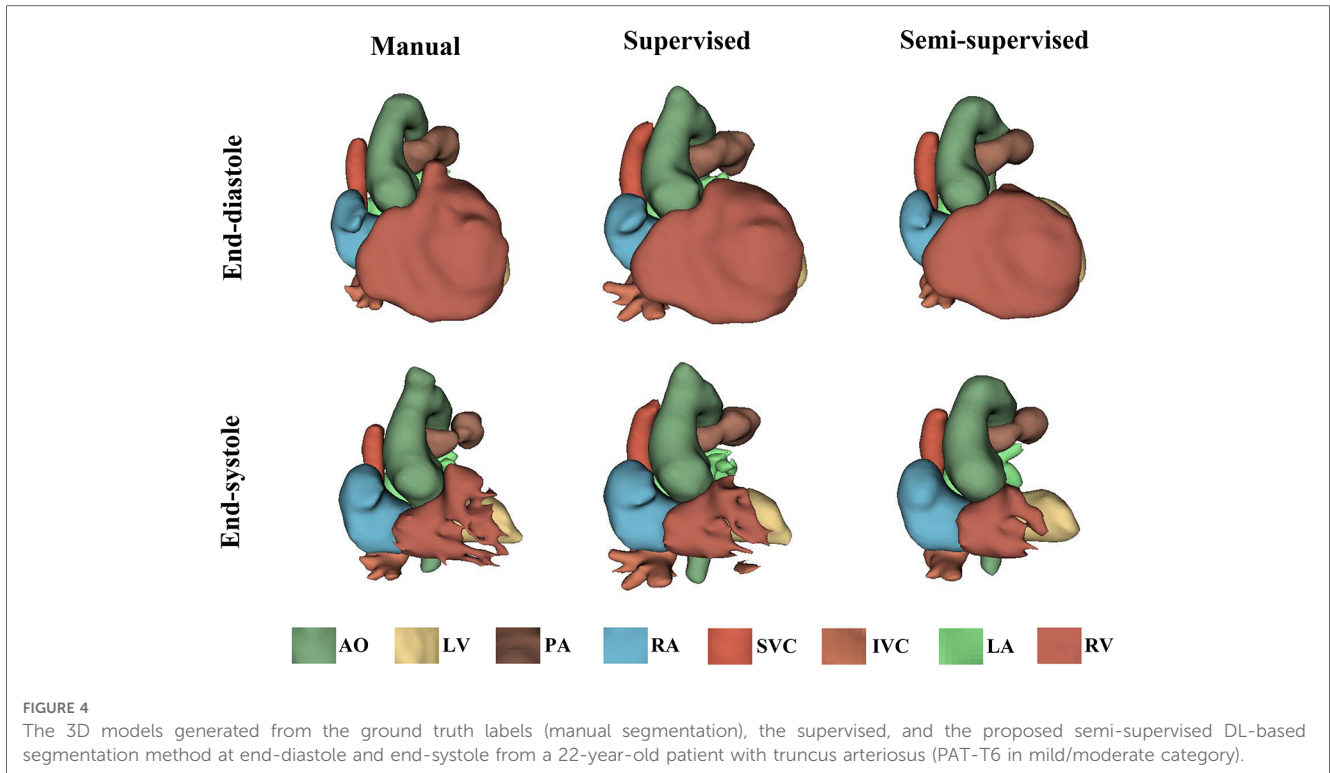
3. Results

It took approximately 10 h to train the supervised DL-based segmentation method and 24 h to train the semi-supervised method. **Supplementary Movie Files S1–S3** show the segmentation results of manual, supervised, and semi-supervised segmentation methods over all time frames in one test dataset with mild/moderate (subject PAT-T4 cf. **Supplementary Table S2**). **Supplementary Movie Files S4–S6** show the dynamic 3D models of the whole-heart and great vessels from the segmentation results of manual, supervised, and semi-supervised segmentation methods over all time frames in the same test dataset. **Figures 3–5**, compare the ground truth labels (manual segmentation), the supervised segmentation method, and the semi-supervised segmentation method results in three patients with normal anatomy (subject PAT-T11 cf. **Supplementary Table S2**), mild/moderate (subject PAT-T6 cf. **Supplementary Table S2**), and severe (subject PAT-T1 cf. **Supplementary Table S2**) complexity, at ED and ES frames.

Supplementary Figure S1 shows the Dice scores for 30 frames in all the test datasets evaluating the degree of overlap between the manual and supervised DL-based segmentation methods, and manual and semi-supervised DL-based segmentation methods. The semi-supervised DL-based segmentation method had a higher average Dice score with the manual method in the segmentation of all 8 cardiac structures across all cardiac anatomy complexities, compared to the supervised DL-based segmentation method (**Table 3**). **Supplementary Table S3** provides a comparative analysis between the proposed semi-supervised DL-based segmentation method and the all-frames supervised DL-based segmentation method. The results demonstrate that the semi-supervised DL-based segmentation method achieved significantly higher Dice scores for all structures (P-value ≤ 0.027) and across all three complexities (P-value ≤ 0.001). This indicates that the semi-supervised approach outperforms the all-frames supervised method in terms of segmentation accuracy.

Figure 6 compares the supervised and semi-supervised segmentation method in delineating different structures of the heart and great vessels over 30 frames in the cardiac cycle. As shown in **Table 3**, the semi-supervised method had a significantly higher Dice score compared to the supervised method for AO (P-value $= 0.034$), PA (P-value $= 0.027$), RA (P-value $= 0.002$), SVC (P-value $= 0.008$), and LA (P-value $= 0.009$); but not for LV (P-value $= 0.115$), IVC (P-value $= 0.176$), and RV (P-value $= 0.186$). On average over all structures, the performance of semi-supervised DL-based segmentation method was significantly better than that of the supervised method (P-value $= 0.003$).

Figures 7, **8** compare the LV and RV volumes computed from the manual segmentation and semi-supervised segmentation method in the test datasets. Similarly, **Supplementary Figures S2, S3** compare the LV and RV volumes computed from the manual segmentation and supervised segmentation method in the test datasets. As shown in **Table 4**, the overall mean volume difference error between the manual and semi-supervised segmentation was less than the mean difference error between the manual and supervised segmentation for LV and RV.

FIGURE 4
The 3D models generated from the ground truth labels (manual segmentation), the supervised, and the proposed semi-supervised DL-based segmentation method at end-diastole and end-systole from a 22-year-old patient with truncus arteriosus (PAT-T6 in mild/moderate category).

FIGURE 5
The 3D models generated from the ground truth labels (manual segmentation), the supervised, and the proposed semi-supervised DL-based segmentation method at end-diastole and end-systole cardiac frames from a 5-year-old patient with double-inlet left ventricle (PAT-T1 in severe category).

Figures 9, **10** compare the LA and RA volumes between the manual segmentation and semi-supervised segmentation method, and **Supplementary Figures S4, S5** compare the LA and RA volumes between the manual segmentation and supervised segmentation method in the test datasets. As shown in **Table 4**, the overall mean volume difference error between the manual

and semi-supervised segmentation was less than the mean difference error between the manual and the supervised segmentation for LA and RA.

Table 5 presents a comparison of ventricular volumes and ejection fractions obtained through various segmentation methods, including manual segmentation of reformatted 2D cine

TABLE 3 Statistical results comparing performance of the proposed semi-supervised segmentation method with the supervised method based on anatomic complexity category as well as structure in delineating heart chambers and great vessels in 12 3D cine test datasets.

	Dice score [%]		P-value
	Supervised vs. 3D manual	Semi-supervised vs. 3D manual	
Complexity			
Normal	89.71 ± 7.63	93.82 ± 4.11	**0.001**
Mild/moderate	87.52 ± 8.75	89.88 ± 6.10	**0.001**
Severe	67.71 ± 22.11	72.88 ± 20.95	**0.001**
Structure			
AO	87.19 ± 9.98	90.12 ± 6.40	**0.034**
LV	75.22 ± 31.50	79.22 ± 29.18	0.115
PA	76.88 ± 13.72	82.65 ± 12.49	**0.027**
RA	78.31 ± 26.67	83.04 ± 26.08	**0.002**
SVC	72.37 ± 15.32	83.92 ± 7.99	**0.008**
IVC	75.48 ± 9.01	78.28 ± 11.04	0.176
LA	76.84 ± 19.95	84.17 ± 11.04	**0.009**
RV	81.52 ± 15.31	84.48 ± 10.92	0.186
Over all structures	77.98 ± 19.64	83.23 ± 16.76	**0.001**

P-values corresponding to significant differences are indicated in bold.

images, manual segmentation of 3D cine images, and both automatic supervised and semi-supervised DL-based segmentation methods. The results demonstrate a strong agreement between ventricular volumes and ejection fractions obtained from the manual segmentation of 3D cine images and the automatic segmentation of 3D cine images using a semi-supervised DL-based algorithm (p-value > 0.087; ICC > 0.94). The manual 3D segmentation also exhibited good agreement with manual 2D segmentation in measuring ventricular volumes and ejection fractions (ICC > 0.96). However, the end-diastolic volumes of the LV and RV, calculated from manual segmentation of 3D cine images, were significantly larger than those measured from manual segmentation of reformatted 2D images (P-value < 0.05). Consequently, this difference resulted in distinct ejection fractions when comparing manual segmentation of 3D cine and reformatted 2D cine images (P-value < 0.05).
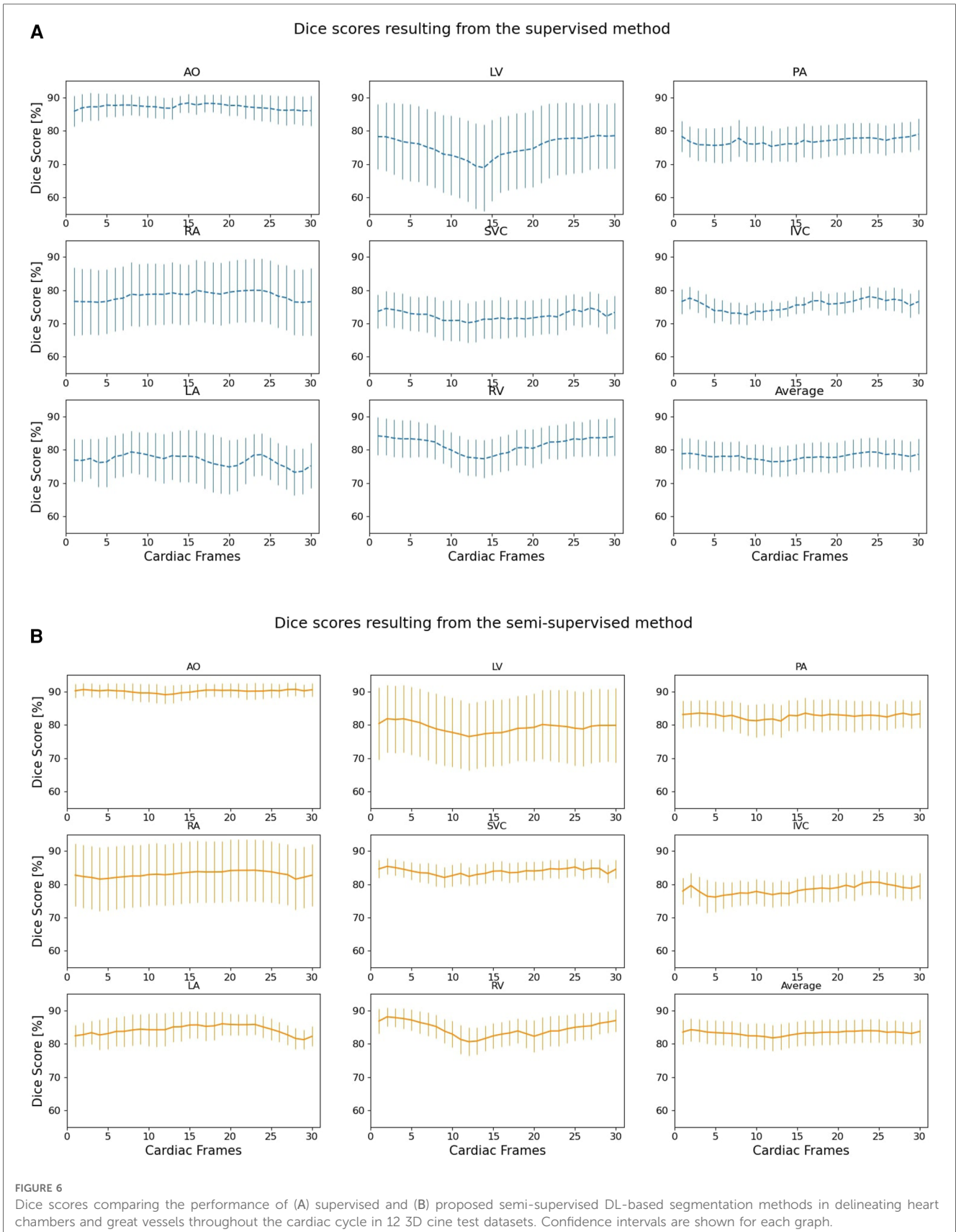
5. Discussion

We developed automatic supervised and semi-supervised DL-based segmentation methods for delineating great vessels (AO, PA, SVC, and IVC) and cardiac chambers (LV, RV, LA, and RA) from 3D cine whole-heart CMR images acquired from patients with a wide range of CHD lesions, based on a 3D U-Net structure (37, 38) and benefiting from the nnU-Net (41) configurations, one of the state-of-the-art frameworks for volumetric segmentation. We divided the patients into 3 categories based on complexity of cardiac anatomical variations (normal anatomy, and mild/moderate and severely complex) and investigated the performance of the automatic DL-based segmentation methods for each group. The semi-supervised segmentation method performed significantly better than the supervised method in each complexity category. However, the

performance of the supervised and semi-supervised segmentation methods significantly decreased as anatomical complexity increased.

On average, the semi-supervised DL-based segmentation method outperformed the supervised DL-based segmentation method in delineating the great vessels and heart chambers. This likely stems from the fact that the semi-supervised segmentation method was trained on all 30 phases in the cardiac cycle while the supervised method was trained only on the ES and ED cardiac phases. In order to further explore the implications of training on all 30 frames, we compared our proposed semi-supervised DL-based segmentation method against the all-frames supervised DL-based segmentation method. The all-frames supervised DL-based segmentation method was also trained on all 30 phases without incorporating our novel label propagation technique. As anticipated, the semi-supervised DL-based segmentation method demonstrated significantly superior segmentation results, thus validating its effectiveness. This outcome could potentially be attributed to the fact that errors would accumulate more significantly in the all-frames supervised DL-based segmentation method for frames that are further away from the ED and ES frames, i.e., the frames containing manual annotations.
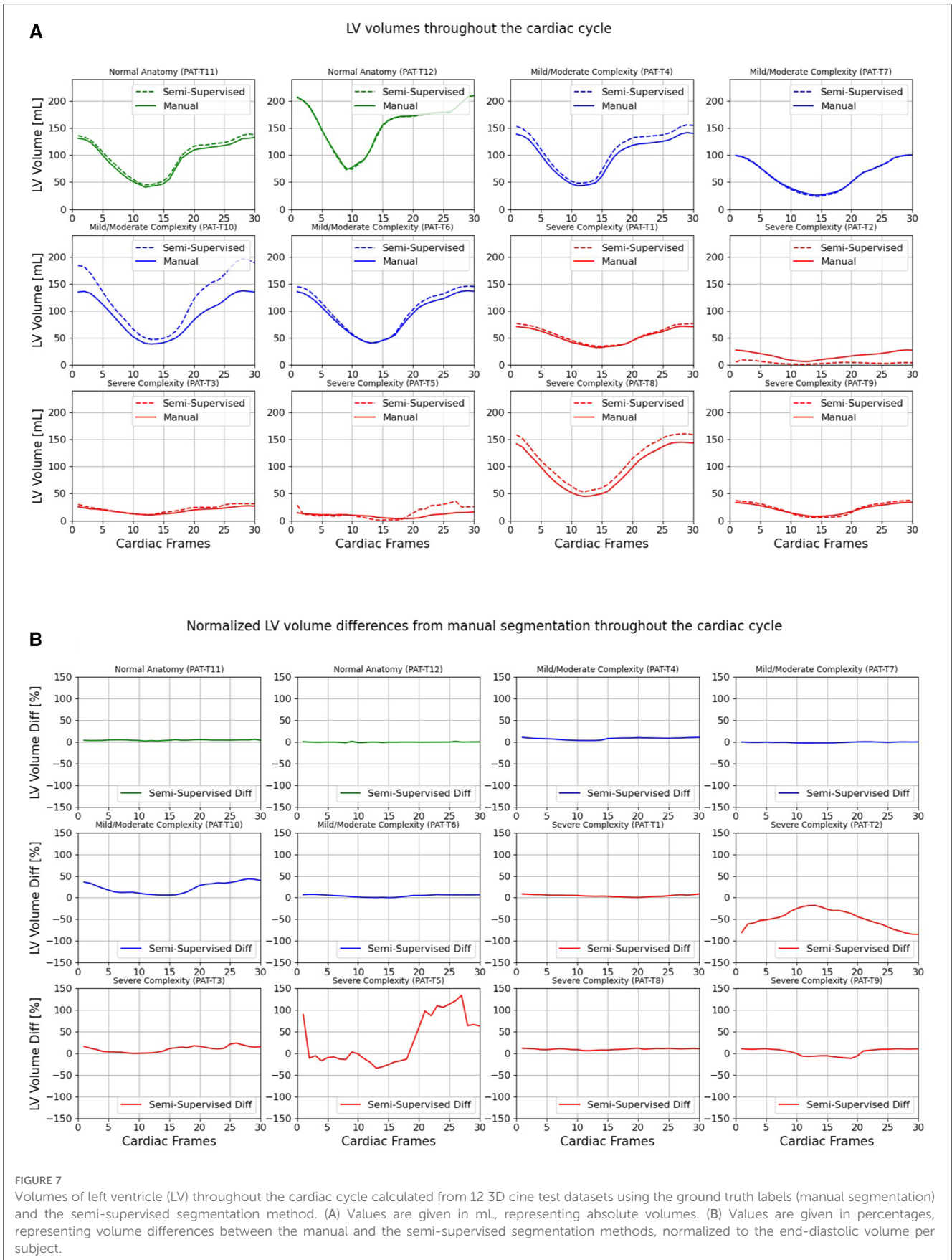
In subjects where there was only one ventricle, the network had difficulties determining whether the single ventricle was anatomically left or right, leading to a decrease in Dice scores for either the LV or RV. Consequently, the fact that the LV and RV Dice scores were not significantly different between the supervised and semi-supervised methods (**Table 4**) could be due to lower Dice scores for the aforementioned reason. In terms of volume measurements, there were no significant differences between the volumes of LV, RV, or LA calculated from manual segmentation (reference) and semi-supervised segmentation method. However, the volumes measured by the supervised method were significantly different compared to manual segmentation, except for the LA. This could be due to less dynamic motion of LA between ED and ES frames compared to the ventricles so that the supervised method (which was trained only on ED and ES frames) could perform as well as the semi-supervised method (which was trained on all 30 frames).

Like ours, previously described DL-based segmentation approaches can perform automatic delineation of great vessels and heart chambers. Pace et al. (25, 39) investigated the potential of active learning to automatically solicit user input in areas where segmentation error is likely to be high with an interactive algorithm. The objective was to segment the great vessels and heart chambers from static whole-heart 3D CMR images in patients with CHD. However, their algorithm was trained on one cardiac phase (i.e., ED phase). Since there is relatively large cardiac motion from ED to ES frames, it is not clear how their segmentation method would perform throughout the cardiac cycle. Qin et al. (24) presented a DL-based segmentation method for combined motion estimation and delineation of heart chambers from CMR images. However, their method is developed for 2D (not 3D) cine images. Bai et al. (48) proposed a CMR image segmentation method which combines fully

FIGURE 6
Dice scores comparing the performance of (A) supervised and (B) proposed semi-supervised DL-based segmentation methods in delineating heart chambers and great vessels throughout the cardiac cycle in 12 3D cine test datasets. Confidence intervals are shown for each graph.

convolutional and recurrent neural networks. They utilized a nonrigid registration algorithm for label propagation that accounts for motion of the heart in all cardiac frames. However, their method utilized a non-rigid registration algorithm for label propagation as opposed to our DL-based label propagation method. The performance of DL-based label propagation method

FIGURE 7
Volumes of left ventricle (LV) throughout the cardiac cycle calculated from 12 3D cine test datasets using the ground truth labels (manual segmentation) and the semi-supervised segmentation method. (A) Values are given in mL, representing absolute volumes. (B) Values are given in percentages, representing volume differences between the manual and the semi-supervised segmentation methods, normalized to the end-diastolic volume per subject.

FIGURE 8
Volumes of right ventricle (RV) throughout the cardiac cycle calculated from 12 3D cine test datasets using the ground truth labels (manual segmentation) and the semi-supervised segmentation method. (A) Values are given in mL, representing absolute volumes. (B) Values are given in percentages, representing volume differences between the manual and the semi-supervised segmentation methods, normalized to the end-diastolic volume per subject.
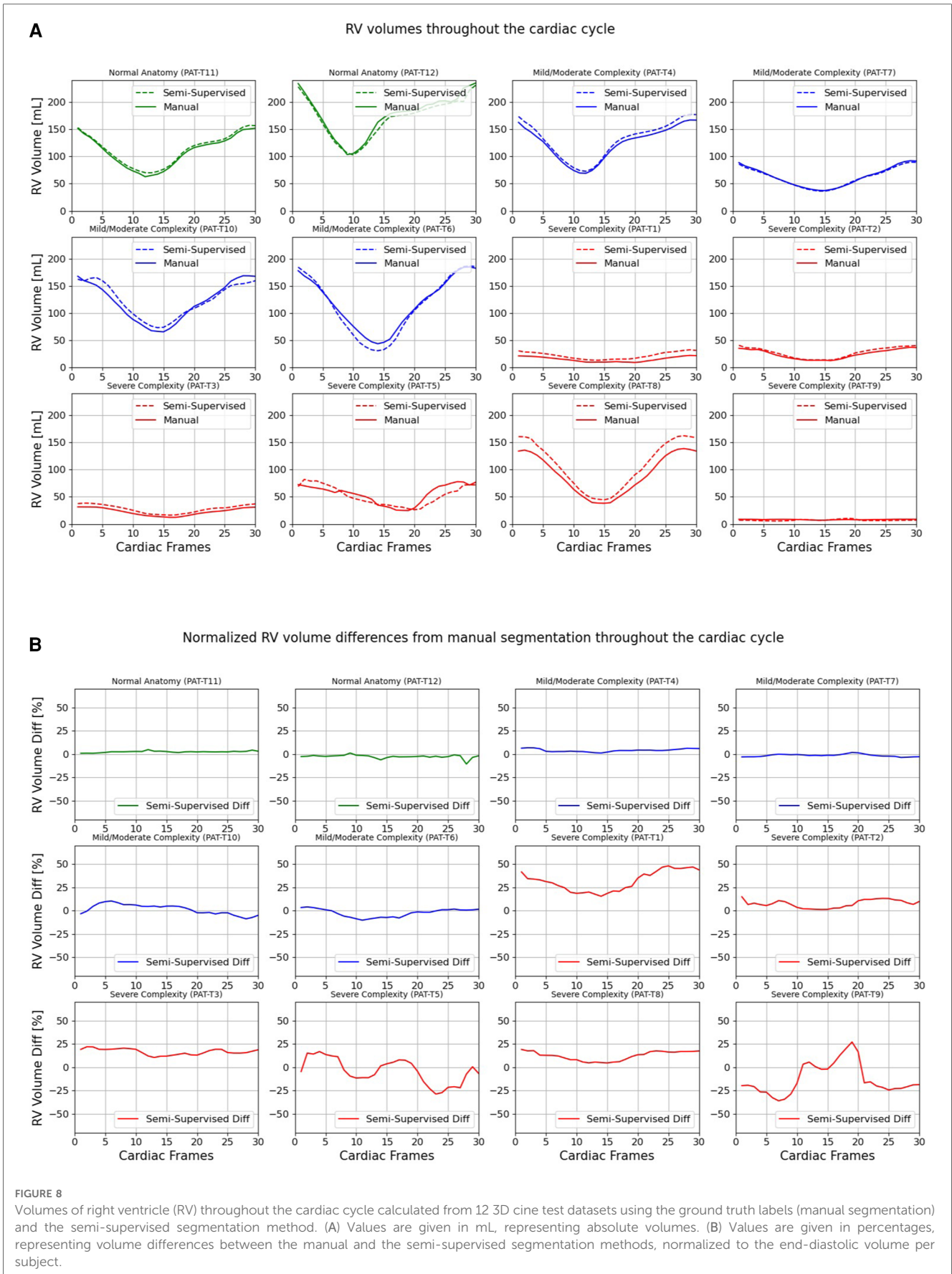
TABLE 4 Statistical results comparing LV, RV, LA, and RA volumes computed from the manual segmentation and supervised and semi-supervised segmentation methods in 12 3D cine test datasets using all 30 frames of the cardiac cycles.

Structure	Supervised vs. 3D manual			Semi-supervised vs. 3D manual		
	Difference [ml]	Difference [%]	*P*-value	Difference [ml]	Difference [%]	*P*-value
LV	10.0 ± 15.0	21.2 ± 76.8	0.030	5.2 ± 11.8	4.0 ± 40.8	**0.114**
RV	8.1 ± 11.1	24.1 ± 38.7	0.017	2.4 ± 8.1	7.1 ± 18.9	**0.212**
LA	3.9 ± 9.0	1.9 ± 32.6	**0.137**	1.9 ± 5.1	4.9 ± 16.3	**0.205**
RA	10.1 ± 10.2	32.5 ± 52.1	0.003	4.2 ± 4.0	13.1 ± 19.5	0.003

P-values corresponding to significant differences are indicated in bold.

needs to be compared with non-rigid label propagation methods. Furthermore, their method was developed for segmenting only AO. Recently, Cao et al. (30), Liu et al. (31), Hatamizadeh et al. (32), and Chen et al. (33) have proposed further general-purpose vision-transformer-based [Dosovitskiy et al. (36)] network architectures for 3D medical image segmentation. However, vision transformer architectures are known for their high computational resource requirements which would not suit our iterative algorithm (49). In addition, their performances have not yet been validated on the segmentation of CMR images.
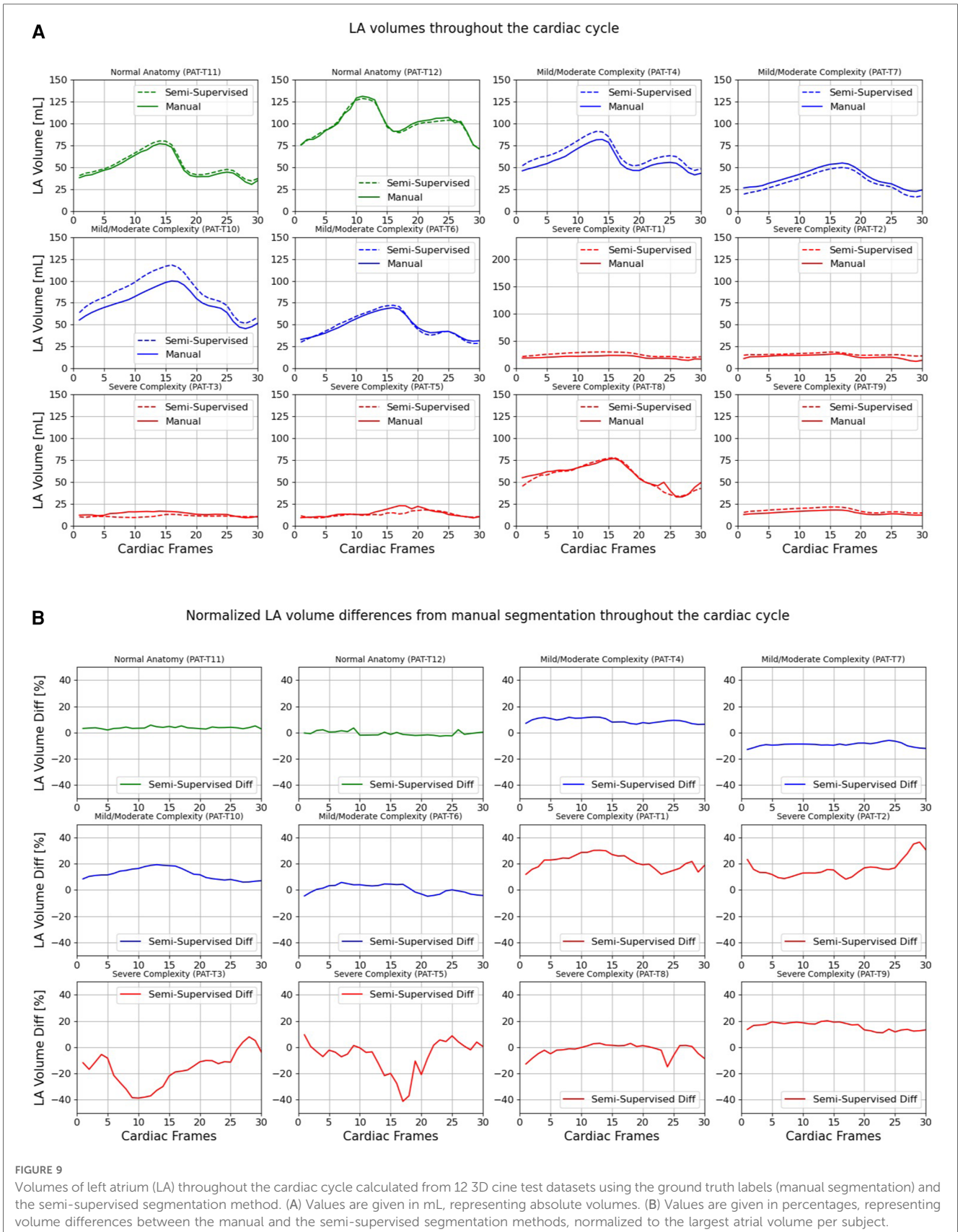
Our study has several limitations. A large number of patients had a coronary sinus close to their RA and it was not included in our segmentation labels due to its small size, significant motion, and difficulty in identification and manual segmentation. Furthermore, pulmonary veins were included in the segmentation of LA. Consequently, future work could consider the coronary sinus and pulmonary veins as additional structures that may require manual segmentation for training DL-based segmentation methods. In our manual segmentation of AO, we allowed the descending AO to have different lengths for each patient, as determined by the DL-based segmentation method. Similarly, SVC and IVC have different lengths for each patient in our manual segmentation, and for patients with an interrupted IVC, the hepatic veins were often segmented as IVC. We let the DL-based segmentation methods decide where the starting point of SVC and IVC would be. Hence, the lengths of AO, SVC, and IVC were different between the results of manual and automatic DL-based segmentation methods causing slightly lower Dice scores for the AO, SVC, and IVC. In addition, we did not have a uniform number of subjects assigned to each complexity category in the test. The test and training datasets could have been better balanced across complexity categories by increasing the number of patients. Furthermore, the number of test subjects ($n = 12$) could be increased. However, a major constraint in preparation of the data was the fact that manual annotation of the 3D cine datasets was extremely time-consuming, given that for each 3D cine image in the test dataset, all the 30 frames needed to be manually segmented. Cross-validation could be helpful in this regard; however, this would require all 30 frames of training and validation images to be manually segmented, which was infeasible. Future work will consider increasing the number of manually segmented datasets for training, validation, and test. The performance of recurrent neural networks (RNN) (50–53) in automatically segmenting the 30 time frames of cardiac cycle

in the 3D cine datasets should also be evaluated and compared with the performance of proposed supervised and semi-supervised DL-based segmentation methods.

Our study demonstrated the accurate segmentation of ventricular volumes using the automatic semi-supervised DL-based segmentation method compared to the manual segmentation of 3D cine images. However, a bias was observed between the manual segmentation of 3D cine images and reformatted 2D cine images when calculating the end-diastolic volumes of LV and RV. These biases, approximately 9 ml for LV and 19 ml for RV, may be attributed to the challenge of accurately delineating the atrial and ventricular cavity in the basal region during the manual segmentation of reformatted short-axis 2D cine images. In our future study, we plan to investigate the underlying reasons for this bias between manual 3D segmentation and manual segmentation of reformatted 2D images.

Finally, our DL-based segmentation methods did not take advantage of known patient diagnoses, which could provide significant information about the locations of great vessels and heart chambers. For instance, in cases where there was only one atrium (subject PAT-T9 cf. **Supplementary Table S2**), the networks had difficulty assigning left or right descriptors to the atrium present, e.g., subject PAT-T9 cf. **Supplementary Table S2** did not have RA. The absence of RA had affected the neighbouring structure, i.e., RV and the network assigned a very small part of the RV to RA as well. Many cases of single ventricle, dextrocardia, and common atrium resulted in the networks mistakenly assigning LV to RV or LA to RA, leading to a significant decrease in Dice score. For subject PAT-T2 cf. **Supplementary Table S2**, which had inverted ventricles, although the network was able to segment the ventricles almost perfectly, it had difficulties in differentiating between LV and RV and assigned a part of the LV to RV. In Fontan patients (subject PAT-T9 cf. **Supplementary Table S2**), the networks struggled to differentiate between the IVC and Fontan conduits, resulting in a low Dice score. Similarly, the networks failed to accurately identify where to stop the SVC in Glenn patients (subject PAT-T2 cf. **Supplementary Table S2**). By incorporating diagnosis information into the training and decision-making phases of the networks, the networks' performance in correctly assigning segment labels could be significantly improved.

Supervised and semi-supervised DL-based segmentation methods were presented for delineating the whole-heart and great vessels from 3D cine CMR images of patients with different

FIGURE 9

Volumes of left atrium (LA) throughout the cardiac cycle calculated from 12 3D cine test datasets using the ground truth labels (manual segmentation) and the semi-supervised segmentation method. (A) Values are given in mL, representing absolute volumes. (B) Values are given in percentages, representing volume differences between the manual and the semi-supervised segmentation methods, normalized to the largest atrial volume per subject.

types of CHD. The performance of the semi-supervised method was shown to be better than the supervised segmentation technique. The semi-supervised DL-based segmentation method

accurately delineates the heart chambers and great vessels and can be used to calculate ventricular and atrial volumes throughout the cardiac cycle. Such a segmentation technique can
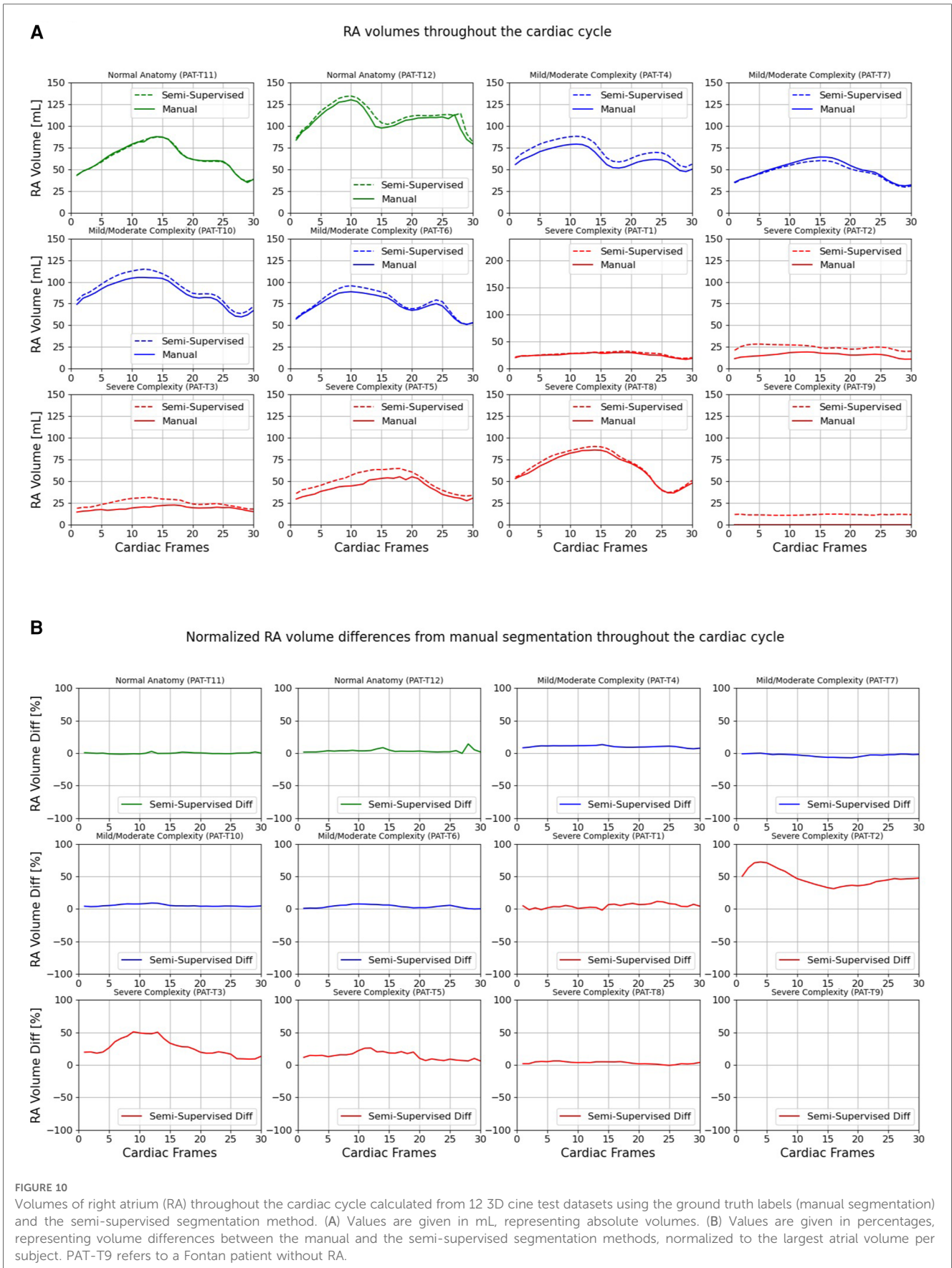
FIGURE 10
Volumes of right atrium (RA) throughout the cardiac cycle calculated from 12 3D cine test datasets using the ground truth labels (manual segmentation) and the semi-supervised segmentation method. (A) Values are given in mL, representing absolute volumes. (B) Values are given in percentages, representing volume differences between the manual and the semi-supervised segmentation methods, normalized to the largest atrial volume per subject. PAT-T9 refers to a Fontan patient without RA.

TABLE 5 Statistical results comparing ventricular volumes computed from the manual segmentation of reconstructed 2D cine images with the manual segmentation of 3D cine images and automatic segmentation of 3D cine images using supervised and semi-supervised methods in 12 3D cine test datasets.

	Left ventricle (LV)			Right ventricle (RV)		
	End-diastole [ml]	End-systole [ml]	Ejection fraction [%]	End-diastole [ml]	End-systole [ml]	Ejection fraction [%]
Manual (2D)	86.4 ± 51.9	31.8 ± 18.6	62.4 ± 5.6	87.0 ± 55.1	42.1 ± 27.5	47.4 ± 15.7
Manual (3D)	95.1 ± 55.1	32.0 ± 19.9	66.4 ± 7.4	106.5 ± 68.2	43.5 ± 31.8	56.4 ± 14.0
Supervised (3D)	109.2 ± 64.7	35.7 ± 26.0	71.2 ± 9.5	115.9 ± 72.6	48.8 ± 30.2	51.2 ± 15.6
Semi-supervised (3D)	104.7 ± 63.7	33.9 ± 22.7	68.4 ± 7.9	108.8 ± 67.2	43.4 ± 30.4	55.0 ± 18.3

reduce inter- and intra-observer variability and make CMR exams more standardized and efficient.

Data availability statement

The datasets presented in this article are not readily available because the datasets used and/or analyzed during the current study are internal data of Boston Children's Hospital and are available from the corresponding authors on reasonable request. Requests to access the datasets should be directed to AP, andrew.powell@cardio.chboston.org.

Ethics statement

This retrospective study involving humans was approved by the institutional review board of Boston Children's Hospital, Boston, MA, USA (IRB-P00011748). The retrospective study was conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin because this was a retrospective study.

Author contributions

STA designed, implemented, and trained the supervised and semi-supervised DL-based segmentation methods, pre-processed the datasets, performed the manual annotation, performed the statistical analyses, generated the movies files, implemented, and performed the study, and wrote the manuscript. JR categorized the patients based on their diagnosis into 3 groups, supervised and edited the manual annotations, performed the qualitative assessment of the results, and edited the manuscript. DFP provided support on the manual annotations and implementation of segmentation methods and reviewed the manuscript. ML contributed to designing the figures, read, approved, and edited the manuscript. RJG performed the 2D reconstruction of test datasets and edited the manuscript. PG, AJP, AKM, DT, TB, and JW, approved, and edited the

manuscript. MHM defined the project, supervised the manual annotations, performed the analysis, wrote, and edited the manuscript. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcvm.2023.1167500/full#supplementary-material

References

1. Best KE, Rankin J. Long-term survival of individuals born with congenital heart disease: a systematic review and meta-analysis. *J Am Heart Assoc*. (2016) 5(6): e002846. doi: 10.1161/JAHA.115.002846

2. Dolk H, Loane M, Garne E. European Surveillance of Congenital Anomalies (EUROCAT) Working Group. Congenital heart defects in Europe: prevalence and perinatal mortality, 2000 to 2005. *Circulation*. (2011) 123(8):841–9. doi: 10.1161/CIRCULATIONAHA.110.958405

3. Centers for Disease Control and Prevention (CDC). Racial differences by gestational age in neonatal deaths attributable to congenital heart defects — United States, 2003–2006. *MMWR Morb Mortal Wkly Rep*. (2010) 59(37):1208–11. https://pubmed.ncbi.nlm.nih.gov/20864921/

4. Oster ME, Lee KA, Honein MA, Riehle-Colarusso T, Shin M, Correa A. Temporal trends in survival among infants with critical congenital heart defects. *Pediatrics*. (2013) 131(5):e1502–1508. doi: 10.1542/peds.2012-3435

5. Lima JAC, Desai MY. Cardiovascular magnetic resonance imaging: current and emerging applications. *J Am Coll Cardiol*. (2004) 44(6):1164–71. doi: 10.1016/j.jacc.2004.06.033

6. Avendi MR, Kheradvar A, Jafarkhani H. A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI. *Med Image Anal*. (2016) 30:108–19. doi: 10.1016/j.media.2016.01.005

7. Avendi MR, Kheradvar A, Jafarkhani H. Automatic segmentation of the right ventricle from cardiac MRI using a learning-based approach. *Magn Reson Med*. (2017) 78(6):2439–48. doi: 10.1002/mrm.26631

8. Backhaus SJ, Staab W, Steinmetz M, Ritter CO, Lotz J, Hasenfuß G, et al. Fully automated quantification of biventricular volumes and function in cardiovascular magnetic resonance: applicability to clinical routine settings. *J Cardiovasc Magn Reson*. (2019) 21(1):24. doi: 10.1186/s12968-019-0532-9

9. Bai W, Sinclair M, Tarroni G, Oktay O, Rajchl M, Vaillant G, et al. Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. *J Cardiovasc Magn Reson*. (2018) 20(1):65. doi: 10.1186/s12968-018-0471-x

10. Karimi-Bidhendi S, Arafati A, Cheng AL, Wu Y, Kheradvar A, Jafarkhani H. Fully-automated deep-learning segmentation of pediatric cardiovascular magnetic resonance of patients with complex congenital heart diseases. *J Cardiovasc Magn Reson*. (2020) 22(1):80. doi: 10.1186/s12968-020-00678-0

11. Bernard O, Lalande A, Zotti C, Cervenansky F, Yang X, Heng PA, et al. Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved? *IEEE Trans Med Imaging*. (2018) 37(11):2514–25. doi: 10.1109/TMI.2018.2837502

12. Arafati A, Hu P, Finn JP, Rickers C, Cheng AL, Jafarkhani H, et al. Artificial intelligence in pediatric and adult congenital cardiac MRI: an unmet clinical need. *Cardiovasc Diagn Ther*. (2019) 9(Suppl 2):S310–25. doi: 10.21037/cdt.2019.06.09

13. Albà X, Pereañez M, Hoogendoorn C, Swift AJ, Wild JM, Frangi AF, et al. An algorithm for the segmentation of highly abnormal hearts using a generic statistical shape model. *IEEE Trans Med Imaging*. (2016) 35(3):845–59. doi: 10.1109/TMI.2015.2497906

14. Eslami A, Karamalis A, Katouzian A, Navab N. Segmentation by retrieval with guided random walks: application to left ventricle segmentation in MRI. *Med Image Anal*. (2013) 17(2):236–53. doi: 10.1016/j.media.2012.10.005

15. Shi W, Zhuang X, Wang H, Duckett S, Oregan D, Edwards P, et al. Automatic segmentation of different pathologies from cardiac cine MRI using registration and multiple component EM estimation. In: Metaxas DN, Axel L, editors. *Functional imaging and modeling of the heart*. Berlin, Heidelberg: Springer Berlin Heidelberg (2011). p. 163–70. (Lecture Notes in Computer Science; vol. 6666). Available at: http://link.springer.com/10.1007/978-3-642-21028-0_21 (Cited March 29, 2023).

16. Bruse JL, McLeod K, Biglino G, Ntsinjana HN, Capelli C, Hsia TY, et al. A non-parametric statistical shape model for assessment of the surgically repaired aortic arch in coarctation of the aorta: how normal is abnormal? In: Camara O, Mansi T, Pop M, Rhode K, Sermesant M, Young A, editors. *Statistical atlases and computational models of the heart imaging and modelling challenges*. Cham: Springer International Publishing (2016). p. 21–9. (Lecture Notes in Computer Science; vol. 9534). Available at: http://link.springer.com/10.1007/978-3-319-28712-6_3 (Cited March 29, 2023).

17. Ye DH, Desjardins B, Hamm J, Litt H, Pohl KM. Regional manifold learning for disease classification. *IEEE Trans Med Imaging*. (2014) 33(6):1236–47. doi: 10.1109/TMI.2014.2305751

18. Zhang H, Wahle A, Johnson RK, Scholz TD, Sonka M. 4-D Cardiac MR image analysis: left and right ventricular morphology and function. *IEEE Trans Med Imaging*. (2010) 29(2):350–64. doi: 10.1109/TMI.2009.2030799

19. Gilbert K, Cowan BR, Suinesiaputra A, Occleshaw C, Young AA. Rapid D-Affine biventricular cardiac function with polar prediction. In: Golland P, Hata N, Barillot C, Hornegger J, Howe R, editors. *Medical image computing and computer-assisted intervention—mICCAI 2014*. Cham: Springer International Publishing

(2014). p. 546–53. (Lecture Notes in Computer Science; vol. 8674). Available at: http://link.springer.com/10.1007/978-3-319-10470-6_68 (Cited March 29, 2023).

20. Mansi T, Voigt I, Leonardi B, Pennec X, Durrleman S, Sermesant M, et al. A statistical model for quantification and prediction of cardiac remodelling: application to tetralogy of fallot. *IEEE Trans Med Imaging*. (2011) 30(9):1605–16. doi: 10.1109/TMI.2011.2135375

21. Punithakumar K, Noga M, Ben Ayed I, Boulanger P. Right ventricular segmentation in cardiac MRI with moving mesh correspondences. *Comput Med Imaging Graph*. (2015) 43:15–25. doi: 10.1016/j.compmedimag.2015.01.004

22. Ralovich K, Itu L, Vitanovski D, Sharma P, Ionasec R, Mihalef V, et al. Noninvasive hemodynamic assessment, treatment outcome prediction and follow-up of aortic coarctation from MR imaging. *Med Phys*. (2015) 42(5):2143–56. doi: 10.1118/1.4914856

23. Zhang Y, Kwon D, Pohl KM. Computing group cardinality constraint solutions for logistic regression problems. *Med Image Anal*. (2017) 35:58–69. doi: 10.1016/j.media.2016.05.011

24. Qin C, Bai W, Schlemper J, Petersen SE, Piechnik SK, Neubauer S, et al. *Joint learning of motion estimation and segmentation for cardiac MR image sequences*. arXiv; (2018). Available at: http://arxiv.org/abs/1806.04066 (Cited March 29).

25. Pace DF, Dalca AV, Geva T, Powell AJ, Moghari MH, Golland P. Interactive whole-heart segmentation in congenital heart disease. *Med Image Comput Comput Assist Interv*. (2015) 9351:80–8. doi: 10.1007/978-3-319-24574-4_10

26. Yu L, Cheng JZ, Dou Q, Yang X, Chen H, Qin J, et al. Automatic 3D cardiovascular MR segmentation with densely-connected volumetric ConvNets. In: Descoteaux M, Maier-Hein L, Franz A, Jannin P, Collins DL, Duchesne S, editors. *Medical image computing and computer-assisted intervention−MICCAI 2017*. Cham: Springer International Publishing (2017). p. 287–95. (Lecture Notes in Computer Science; vol. 10434). Available at: https://link.springer.com/10.1007/978-3-319-66185-8_33 (Cited April 22, 2023).

27. Shi Z, Zeng G, Zhang L, Zhuang X, Li L, Yang G, et al. Bayesian VoxDRN: a probabilistic deep voxelwise dilated residual network for whole heart segmentation from 3D MR images. In: Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G, editors. *Medical image computing and computer assisted intervention—mICCAI 2018*. Cham: Springer International Publishing (2018). p. 569–77. (Lecture Notes in Computer Science; vol. 11073). Available at: http://link.springer.com/10.1007/978-3-030-00937-3_65 (Cited April 22, 2023).

28. Chen C, Qin C, Qiu H, Tarroni G, Duan J, Bai W, et al. Deep learning for cardiac image segmentation: a review. *Front Cardiovasc Med*. (2020) 7:25. doi: 10.3389/fcvm.2020.00025

29. Wolterink JM, Leiner T, Viergever MA, Išgum I. Dilated convolutional neural networks for cardiovascular MR segmentation in congenital heart disease. In: Zuluaga MA, Bhatia K, Kainz B, Moghari MH, Pace DF, editors. *Reconstruction, segmentation, and analysis of medical images*. Cham: Springer International Publishing (2017). p. 95–102. (Lecture Notes in Computer Science; vol. 10129). Available at: http://link.springer.com/10.1007/978-3-319-52280-7_9 (Cited March 29, 2023).

30. Cao H, Wang Y, Chen J, Jiang D, Zhang X, Tian Q, et al. Swin-Unet: unet-like pure transformer for medical image segmentation. In: Karlinsky L, Michaeli T, Nishino K, editors. *Computer vision—eCCV 2022 workshops*. Cham: Springer Nature Switzerland (2023). p. 205–18. (Lecture Notes in Computer Science; vol. 13803). Available at: https://link.springer.com/10.1007/978-3-031-25066-8_9 (Cited March 29, 2023).

31. Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, et al. *Swin transformer: hierarchical vision transformer using shifted windows*. arXiv; (2021). Available at: http://arxiv.org/abs/2103.14030 (Cited November 21, 2022).

32. Hatamizadeh A, Nath V, Tang Y, Yang D, Roth HR, Xu D. Swin UNETR: swin transformers for semantic segmentation of brain tumors in MRI images. In: Crimi A, Bakas S, editors. *Brainlesion: Glioma, multiple sclerosis, stroke and traumatic brain injuries*. Cham: Springer International Publishing (2022). p. 272–84. (Lecture Notes in Computer Science; vol. 12962). Available at: https://link.springer.com/10.1007/978-3-031-08999-2_22 (Cited March 29, 2023).

33. Chen J, Lu Y, Yu Q, Luo X, Adeli E, Wang Y, et al. *TransUNet: transformers make strong encoders for medical image segmentation*. arXiv; (2021). Available at: http://arxiv.org/abs/2102.04306 (Cited March 29, 2023).

34. Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, et al. *Segment anything*. arXiv; (2023). Available at: http://arxiv.org/abs/2304.02643 (Cited June 18, 2023).

35. Oquab M, Darcet T, Moutakanni T, Vo H, Szafraniec M, Khalidov V, et al. *DINOv2: learning robust visual features without supervision*. arXiv; (2023). Available at: http://arxiv.org/abs/2304.07193 (Cited June 13, 2023).

36. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. *An image is worth 16×16 words: transformers for image recognition at scale*. arXiv; (2021). Available at: http://arxiv.org/abs/2010.11929 (Cited November 21, 2022).

37. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF, editors. *Medical image computing and computer-assisted intervention—mICCAI 2015*. Cham: Springer International Publishing (2015). p. 234–41. (Lecture Notes in Computer Science; vol. 9351). Available at: http://link.springer.com/10.1007/978-3-319-24574-4_28 (Cited February 27, 2023).

38. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: Ourselin S, Joskowicz L, Sabuncu MR, Unal G, Wells W, editors. *Medical image computing and computer-assisted intervention—mICCAI 2016*. Cham: Springer International Publishing (2016). p. 424–32. (Lecture Notes in Computer Science; vol. 9901). Available at: https://link.springer.com/10.1007/978-3-319-46723-8_49 (Cited February 27, 2023).

39. Pace DF, Dalca AV, Brosch T, Geva T, Powell AJ, Weese J, et al. Learned iterative segmentation of highly variable anatomy from limited data: applications to whole heart segmentation for congenital heart disease. *Med Image Anal*. (2022) 80:102469. doi: 10.1016/j.media.2022.102469

40. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin JC, Pujol S, et al. 3D Slicer as an image computing platform for the quantitative imaging network. *Magn Reson Imaging*. (2012) 30(9):1323–41. doi: 10.1016/j.mri.2012.05.001

41. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods*. (2021) 18(2):203–11. doi: 10.1038/s41592-020-01008-z

42. Agarap AF. *Deep learning using rectified linear units (ReLU)*. arXiv; (2019). Available at: http://arxiv.org/abs/1803.08375 (Cited December 15, 2022).

43. Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. In: Bach F, Blei D, editors. *Proceedings of the 32nd international conference on international conference on machine learning*. Lille, France: JMLR.org (2015). p. 448–56.

44. Reddi SJ, Kale S, Kumar S. *On the convergence of Adam and beyond*. arXiv; (2019). Available at: http://arxiv.org/abs/1904.09237 (Cited March 29, 2023).

45. Fan DP, Zhou T, Ji GP, Zhou Y, Chen G, Fu H, et al. Inf-Net: automatic COVID-19 lung infection segmentation from CT images. *IEEE Trans Med Imaging*. (2020) 39 (8):2626–37. doi: 10.1109/TMI.2020.2996645

46. Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). *Biometrika*. (1965) 52(3/4):591. doi: 10.2307/2333709

47. Wilcoxon F. Individual comparisons by ranking methods. In: Kotz S, Johnson NL, editors. *Breakthroughs in statistics*. New York, NY: Springer New York (1992). p. 196–202. (Springer Series in Statistics). Available at: http://link.springer.com/10.1007/978-1-4612-4380-9_16 (Cited January 17, 2023).

48. Bai W, Suzuki H, Qin C, Tarroni G, Oktay O, Matthews PM, et al. Recurrent neural networks for aortic image sequence segmentation with sparse annotations. In: Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G, editors. *Medical image computing and computer assisted intervention—mICCAI 2018*. Cham: Springer International Publishing (2018). p. 586–94. (Lecture Notes in Computer Science; vol. 11073). Available at: http://link.springer.com/10.1007/978-3-030-00937-3_67 (Cited March 29, 2023).

49. Khader F, Mueller-Franzes G, Wang T, Han T, Arasteh ST, Haarburger C, et al. *Medical diagnosis with large scale multimodal transformers: leveraging diverse data for more accurate diagnosis*. arXiv; (2022). Available at: http://arxiv.org/abs/2212.09162 (Cited January 15, 2023).

50. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput*. (1997) 9(8):1735–80. doi: 10.1162/neco.1997.9.8.1735

51. Elman JL. Finding structure in time. *Cogn Sci*. (1990) 14(2):179–211. doi: 10.1207/s15516709cog1402_1

52. Chakravarty A, Sivaswamy J. RACE-Net: a recurrent neural network for biomedical image segmentation. *IEEE J Biomed Health Inform*. (2019) 23 (3):1151–62. doi: 10.1109/JBHI.2018.2852635

53. Wang Y, Zhang W. A dense RNN for sequential four-chamber view left ventricle wall segmentation and cardiac state estimation. *Front Bioeng Biotechnol*. (2021) 9:696227. doi: 10.3389/fbioe.2021.696227