

3D U-Net-Based Deep Learning Approach for Pediatric Cardiovascular MRI Segmentation

Musab Ahmed Khan Umair
Artificial Intelligence
Istinye University
Istanbul, Turkey

musabkhan2003@gmail.com

Irem Yilmaz
Biomedical Engineering
Yildiz Technical University
Istanbul, Turkey
irem.yilmaz@std.yildiz.edu.tr

Omer Faruk Sahin
Bioengineering
Marmara University
Istanbul, Turkey
umarfaruksahin@outlook.com

Ahmed Boray
Computer Engineering
Istinye University
Istanbul, Turkey
ahmedegy263@gmail.com

Mohamad Humam Alzaeim
Biomedical Engineering &
Bioinformatics Istanbul
Medipol University
Istanbul, Turkey
mohamad.Alzaeim@std.medipol.edu.tr

Kevser Banu Kose
Biomedical Engineering Department
Istanbul Medipol University
Istanbul Turkey
kbkose@medipol.edu.tr

Fatma Zeynep Gungoren
Radiology
The Anadolu Medical Center
Istanbul, Turkey

fzgungoren@yahoo.com
Ruqaiyah Mirza
Software engineering
Istinye University
Istanbul, Turkey
ruqaiyah.mirza@stu.istinye.edu.tr

Beyza Nur Yazici
Computer Engineering
Istinye University
Istanbul, Turkey
beyzanur.yazici@stu.istinye.edu.tr

Senol Piskin*
Mechanical Engineering Department
Istinye University
Istanbul, Turkey

*Corresponding author:
senol.Piskin@istinye.edu.tr

JS Femilda Josephin
Computer Engineering Department
Istinye University
Istanbul, Turkey
femilda.bai@istinye.edu.tr

Ayham Pezri
Biomedical Engineering
Medipol University
Istanbul, Turkey
ayham.pezri@std.medipol.edu.tr

Nedim Yildirim
Sev American College
Istanbul, Turkey
ny.nedimyildirim@gmail.com

Ibrahim Faress
Software Engineering
Istinye University
Istanbul, Turkey
ibrahim.faress@stu.istinye.edu.tr

Abstract— Accurate segmentation of cardiovascular structures in pediatric patients is essential for clinical evaluation, but variability in imaging data and the labor-intensive nature of manual segmentation remain major challenges. The given paper provides a reproducible deep-learning model intended to counteract these issues. With our deep learning approach, cardiac MRI data is standardized by (i) converting anatomical labels into a four-class mask (involving background, aorta, chambers, pulmonary structures), (ii) aligning images and masks on coordinates using nearest-neighbor resampling, and (iii) automatically cropping the region-of-interest (ROI). To handle class imbalance, we trained a 3D U-Net with the Focal Tversky Loss function, using both public and private clinical datasets. To improve robustness and reduce outlier errors, we applied model ensembling by averaging predictions from independently trained networks. By automating preprocessing and segmentation, our approach provides a pediatric cardiac analysis tool that reduces manual workload and increases reproducibility.

Keywords— 3D U-Net, Cardiac Segmentation, Pediatric MRI, Deep Learning, Ensemble Methods, Focal Tversky Loss, Automated Preprocessing, ROI Cropping

I. INTRODUCTION

A. Problem Statement

variability, especially when dealing with small, complex pediatric anatomies. These challenges are further compounded by low-resolution or artifact-prone scans common in clinical datasets, especially in young children who may require sedation or who cannot remain still during imaging. Traditional

segmentation pipelines struggle with inconsistent image quality, label imbalance, and unreliable region delineation. Thus, there is a critical need for robust, automatic segmentation methods tailored to the anatomical and imaging-specific constraints of pediatric cardiac MRI.

To address these challenges, we present a fully automated segmentation approach using a 3D U-Net-based multi-class segmentation trained on both open and clinical datasets. This approach enables reliable extraction of critical cardiovascular structures.

B. Our Contribution

A novel deep learning model is proposed in this study to perform multi-class segmentation of pediatric cardiovascular structures in three-dimensional magnetic resonance imagery. The pipeline unifies the anatomical labels and resamples the scans and masks with nearest neighbor resampling in 3D Slicer to ensure consistency. A subsequent automatic ROI cropping step uses padding of relevant cardiac regions to make the input size smaller and speed up computation. When small and imbalanced structures need to be accurately segmented, training a 3D U-Net with the Focal Tversky Loss is effective. To improve robustness, an ensemble inference was applied where the predictions of a variety of individually trained models are averaged. The proposed approach generalizes well across diverse anatomies and imaging conditions. By automating preprocessing and segmentation, it significantly reduces manual workload and supports scalable deployment.



II. RELATED WORK

Recent developments in cardiovascular anatomical segmentation have been largely driven by hybrid atlas-learning pipelines and deep learning methods. Bruns et al. [1] provide a comprehensive study using CT data, detailing dataset cataloging, preprocessing (e.g., intensity normalization, cropping), and evaluation metrics such as Dice and Hausdorff scores. The authors emphasize persistent challenges like scanner-based domain shifts, class imbalance in small structures, and annotation limitations, proposing standardized benchmarks and external validation as remedies. Complementing this, Alnasser et al. [2] introduce a multi-class statistical atlas-based whole-heart segmentation (WHS) framework for CT and MRI, demonstrating the enduring value of atlas priors when combined with data-driven refinement to enhance boundary consistency across substructures.

REFERENCES

S.No	Journal/Authors	Combined Description
1	Alnasser TN, Abdulaal L, Maiter A, Sharkey M, Dwivedi K, Salehi M, Garg P, Swift AJ & Alabed S (Frontiers in Cardiovascular Medicine)	This 2024 review evaluated 18 CT-based studies using various deep learning models like ResNet50, U-Net, and nn-UNet. It reported strong segmentation performance (e.g., LA 0.88, LV 0.91) and assessed study quality using the CLAIM checklist.
2	Bruns S, Wolterink JM, van den Boogert TPW, Runge JH, Bouma BJ, Henriques JP, Baan J, Viergever MA, Planken RN & Išgum I (Computers in Biology and Medicine)	In 2022, this study used 3D CNNs to segment 4D cardiac CT images from 1,509 patients. The models achieved high Dice scores around 0.89 and demonstrated >90% clinical usability.
3	Ghosh TK, Hasan MK, Roy S, Alam MA, Hossain E & Ahmad M (IEEE Access)	This 2021 work proposed a probabilistic atlas-based segmentation method for CT and MRI heart images, using Bayesian EM and CNN variants, achieving improved accuracy and reduced volume error.

4	Chen C, Qin C, Qiu H, Tarroni G, Duan J, Bai W & Rueckert D (Frontiers in Cardiovascular Medicine)	This 2020 review covered over 100 studies on cardiac image segmentation, primarily using CNNs on datasets like ACDC and MM-WHS
5	Zhuang X (Journal of Healthcare Engineering)	This 2013 review discussed early automatic segmentation using atlas-based and deformable models, reporting modest Dice scores (~0.4–0.5) and highlighting challenges in generalization.
6	Karimi-Bidhendi et al.	A GAN-augmented fully convolutional network was developed for pediatric CHD CMR segmentation, addressing limited annotated datasets. The model achieved Dice scores up to 91% , outperforming U-Net and cvi42, and showed strong agreement with manual segmentation.
7	Yoshida A, Kondo Y, Yoshimura N, Kuramoto T et al. (Measurement Sensors)	The 2021 study applied a modified U-Net on pediatric CT scans, showing excellent whole-heart segmentation (DSC > 0.95) and reasonable chamber-level performance (0.78–0.88).

From a learning-centered viewpoint, Ghosh et al. [3] explore CNN-based heart segmentation, highlighting attention modules, multi-scale feature aggregation, and U-Net variants that effectively handle complex anatomy such as thin-walled ventricles and great veins. The existing study also emphasizes strategies to improve geometric fidelity, including 3D/2.5D contexts, patch-based training for memory efficiency, and post-processing techniques like connected-component filtering. Chen et al. [4] expand segmentation to cardiac radiography, where low soft-tissue contrast and overlapping projections present unique challenges. The researchers have also underlined the importance of robust augmentation, domain adaptation, and uncertainty estimation to minimize clinical errors. Song et al. [5] address segmentation in pediatric congenital heart disease, employing topology-aware losses, shape constraints, and tailored approaches for rare morphologies under limited training data. Across all studies [1–5], unresolved issues remain: achieving generalization across sites, consistent performance on small or atypical anatomies, and maintaining rigorous, externally validated reporting.

III. DATASETS

In the proposed study, two datasets have been utilized, namely, an open-source pediatric cardiac dataset and a combined dataset created by augmenting the public data with local clinical scans. The open-source data was used to develop early models and to serve as an external benchmark. Each of the cases was anonymized, transformed into NRRD, and named in a standardized scheme. The concatenated dataset includes local data gathered using institutional approvals, which cover different imaging conditions and pediatric anatomies. To make all scans consistent, the same preprocessing routines were employed, i.e., format conversion and label standardization (0: background, 1: aorta, 2: chambers, 3: pulmonary structures).

IV. METHODOLOGY

A. Data Preprocessing

Patients' cases consisted of a volumetric scan and its multi-label segmentation mask in NRRD format. All the data were preprocessed and cached in the form of NumPy arrays (.npy) to shorten training time and reduce I/O overhead. The trilinear interpolation was used to resize volumes of identical shapes of 128x128x64, and nearest neighbor interpolation was used out of a mask to maintain the accuracy of labels. The intensities were normalized to the range [0,1] using min-max normalization. The same scheme of classes was employed: 0 was adopted to describe background, 1 to describe the aorta, 2 to describe the chambers, and 3 to describe the pulmonary structures. Consistent preprocessing was applied across validation and test cases.

B. Data Augmentation

Training-time augmentation was used to increase model robustness to anatomical variation and the presence of acquisition artifacts, with augmentation implemented in TorchIO. This also involved random spatial changes, including flips, affine deformations (scaling, rotation, translation), and elastic deformations to model changes of enlistment. Scanner-related variations were modelled using such intensity augmentations as gamma correction, random bias field, and Gaussian noise. These augmentations were applied stochastically to every batch during training, improving the model's generalizability across varied datasets.

C. Model Training & Evaluation

The model utilizes a 3D U-Net architecture for multi-class cardiac segmentation, featuring a symmetric encoder-decoder structure with skip connections to combine semantic and detailed features. The encoder includes three downsampling blocks, each with two 3x3x3 convolutions, batch normalization, ReLU, and 2x2x2 max pooling. A bottleneck with two convolutional layers and 256 filters connects the encoder and decoder. The decoder mirrors the encoder's structure, using transposed convolutions for upsampling. At every stage,

encoder outputs are concatenated with the corresponding upsampled features from the decoder. The output layer applies a 1x1x1 convolution with four channels (background, aorta, chambers, and pulmonary vessels), and a Softmax layer transforms the logits into per-voxel class likelihoods.

$$\text{Softmax}(z_i) = \frac{e^{z_i}}{\sum_{j=1}^C e^{z_j}} \quad (1)$$

A softmax layer is used at the network output to convert logits into class probabilities. Here, $C=4$ is the total number of classes.

$$TI = \frac{TP + \epsilon}{TP + \alpha \cdot FN + \beta \cdot FP + \epsilon} \quad (2)$$

To address the challenges of class imbalance and indistinct boundaries in cardiac imaging, the model was trained using a **Focal Tversky Loss (FTL)** function.

$$L_{FTL} = (1 - TI)^\gamma \quad (3)$$

This loss function is built upon the Tversky Index (TI), which generalizes the Dice score by weighting false negatives and false positives differently ($\alpha = 0.7$, $\beta = 0.3$). The FTL includes a focusing parameter that places greater emphasis on difficult-to-classify areas, helping the model improve accuracy on underrepresented structures.

The 3D U-Net model was trained with a batch size of 2 for 1000 epochs, utilizing early stopping based on validation loss to prevent overfitting and optimize training time. Auto mixed-precision (AMP) was employed to reduce GPU memory consumption and accelerate convergence. Data loaders were multi-threaded to ensure efficient data processing during training and validation.

Model performance was assessed using three standard segmentation metrics: the **Dice Similarity Coefficient (DSC)** to measure spatial overlap, the **Jaccard Index (IoU)** to provide a stricter comparison of regions, and **Pixel Accuracy** to measure overall voxel-wise correctness. These metrics were calculated globally and on a per-class basis to provide a comprehensive evaluation.

D. Ensemble Approach

An ensemble learning strategy was chosen to enhance the stability of the segmentation and to decrease the variance. The 3D U-Net model of the specific architecture was trained multiple times with various sets of random initialization. When performing inference, consensus prediction was obtained by averaging, voxel-wise, the softmax probability maps output by each model. The ensemble solution minimized the possibility of outlier errors that could arise due to each model and enhanced the general reliability of segmentation.

V. EXPERIMENTS & RESULT

A. Qualitative Results

In order to analyze the visual quality of our test segmentation models, we provide two visual comparisons. The two figures, a and b, can prove together that data augmentation and longer training can positively affect segmentation performance

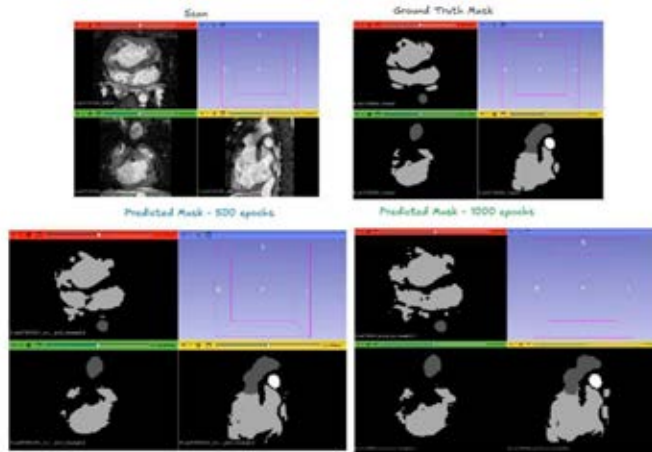


Figure 1: Multi-planar visual comparison of model predictions at 500 vs. 1000 training epochs.

Figure 1 shows a multi-planar view of Patient 55, with the results of the 500-epoch trained model and the 1000-epoch trained model. The latter model, trained for an increased amount of time, displays an accurate adherence to the boundaries and less noise. Overall, these two figures highlight how combining augmentation with long-term training enhances segmentation accuracy and reliability.

VI. DISCUSSION & CONCLUSION

This study shows that a reproducible, uniform preprocessing pipeline with a 3D U-Net as an ensemble yields good results on multi-class pediatric cardiovascular segmentation. The pipeline is easier to use and frees up far more processing resources, and the model reports good performance when applied to a standardized open-source dataset. It has high scores in the aorta and the chambers, which is indicative of good segmentation of well-defined structures.

There were two principal problems. First, the small number of cases and lack of diversity in both the combined and the open-source datasets were limited, constraining the model from learning different kinds of examples and impacting the performance across various clinical scans. Second, the pulmonary structures were the weakest group of structures. Although Focal Tversky Loss has been used to overcome the problem of class imbalance, segmentation of small, branching vessels is still challenging. An alternative direction is to investigate attention mechanisms or shape-aware priors in the future.

The most important contribution is the automatic pipeline of preprocessing that normalizes labels, resamples geometry, and crops ROIs, decreasing input size and training time, and resulting in consistency. This work introduces an end-to-end automated cardiovascular MRI segmentation pipeline, which uses standardized preprocessing to then use a 3D U-Net ensemble trained with Focal Tversky Loss to deliver high precision on published data, addressing the issues of label contradictions and inefficient computation.

VII. ACKNOWLEDGMENT

This work was supported by the Scientific and Technological Research Council of Türkiye (TUBITAK) under the 1001 Grant Program (Project No. 122E488).

VIII. AFFILIATIONS

This study was conducted in collaboration with Istinye University, Istanbul Medipol University, and Anadolu Medical Center.

REFERENCES

- [1] Bruns, S., Wolterink, J. M., van den Boogert, T. P., Runge, J. H., Bouma, B. J., Henriques, J. P., ... & Išgum, I. (2022). Deep learning-based whole-heart segmentation in 4D contrast-enhanced cardiac CT. *Computers in biology and medicine*, 142, 105191.
- [2] Alnasser, T. N., Abdulaal, L., Maiter, A., Sharkey, M., Dwivedi, K., Salehi, M., ... & Alabed, S. (2024). Advancements in cardiac structures segmentation: a comprehensive systematic review of deep learning in CT imaging. *Frontiers in cardiovascular medicine*, 11, 1323461.
- [3] Ghosh, T. K., Hasan, M. K., Roy, S., Alam, M. A., Hossain, E., & Ahmad, M. (2021). Multi-class probabilistic atlas-based whole heart segmentation method in cardiac CT and MRI. *IEEE Access*, 9, 66948-66964.
- [4] Chen, C., Qin, C., Qiu, H., Tarroni, G., Duan, J., Bai, W., & Rueckert, D. (2020). Deep learning for cardiac image segmentation: a review. *Frontiers in cardiovascular medicine*, 7, 25.
- [5] Song, Y., Ren, S., Lu, Y., Fu, X., & Wong, K. K. (2022). Deep learning-based automatic segmentation of images in cardiac radiography: a promising challenge. *Computer Methods and Programs in Biomedicine*, 220, 106821.
- [6] Karimi-Bidhendi, S., Arafati, A., Cheng, A. L., Wu, Y., Kheradvar, A., & Jafarkhani, H. (2020). Fully-automated deep-learning segmentation of pediatric cardiovascular magnetic resonance of patients with complex congenital heart diseases. *Journal of cardiovascular magnetic resonance*, 22(1), 80.
- [7] Yoshida, A., Lee, Y., Yoshimura, N., Kuramoto, T., Hasegawa, A., & Kanazawa, T. (2021). Automated heart segmentation using U-Net in pediatric cardiac CT. *Measurement: Sensors*, 18, 100127.