Improving Right Ventricle Segmentation in Cardiac Magnetic Resonance Imaging Through Transfer Learning

Abbas Rizvi
Radiology & Diagnostic Imaging
University of Alberta
Edmonton, Canada
aarizvi2@ualberta.ca

Sivalingam Ampatishan
Radiology & Diagnostic Imaging
University of Alberta
Edmonton, Canada
ampatish@ualberta.ca

Ramesh Mahdavifar
Radiology & Diagnostic Imaging
University of Alberta
Edmonton, Canada
ramesh1@ualberta.ca

Liang Zhong
Duke NUS Medical School and
Department of Biomedical Engineering
National University of Singapore
Singapore
zhong.liang@duke-nus.edu.sg

Michelle Noga
Radiology & Diagnostic Imaging
University of Alberta
Edmonton, Canada
mnoga@ualberta.ca

Kumaradevan Punithakumar
Radiology & Diagnostic Imaging
University of Alberta
Edmonton, Canada
punithak@ualberta.ca

***Abstract*—Segmentation of the right ventricle (RV) in magnetic resonance imaging (MRI) sequences is critical for assessing RV function. However, manual segmentation involves processing hundreds of images per patient, making it a tedious and time-consuming process. Recently, deep convolutional neural networks have emerged as an effective solution for automating RV segmentation in MRI sequences, substantially reducing manual workload. Accurate segmentation of the RV is crucial for reliable clinical applications. In this study, we demonstrate that transfer learning using a pre-trained segmentation model from the Medical Open Network for Artificial Intelligence (MONAI) Model Zoo significantly improves segmentation accuracy, as measured by the Dice similarity coefficient (DSC) and 95th percentile Hausdorff distance (HD95) scores, compared to manual annotations from medical experts. Our approach increased DSC-based segmentation accuracy from 74.93% (pre-trained MONAI Zoo model) and 83.15% (same architecture trained on our data) to 84.91% on 1,994 test images acquired from seven patients. Furthermore, it outperformed a state-of-the-art self-configuring network, nnU-Net, which achieved an accuracy of 81.98% on the same dataset. This study demonstrates the effectiveness of transfer learning in improving segmentation accuracy for the proposed task.**

***Index Terms*—Deep convolutional neural networks, right ventricle, magnetic resonance imaging, cardiac disease, transfer learning**

I. Introduction

Cardiovascular disease remains the leading cause of mortality worldwide [1], [2]. Magnetic resonance imaging (MRI) is widely regarded as the reference standard for evaluating cardiac structure and function [3]. In a typical cine cardiac MRI examination, both the left and right ventricles are imaged; however, for many years, clinical focus has been placed predominantly on the left ventricle. The right ventricle (RV) has often been referred to as the *forgotten chamber* [4]–[6], largely due to the limited clinical emphasis placed on its functional assessment. This historical oversight is now changing, as an increasing number of clinical studies highlight the critical role of RV function in diagnosing, prognosticating, and managing a variety of cardiovascular and pulmonary conditions.

Despite its growing clinical significance, assessing the RV remains challenging. Cine cardiac MRI typically generates hundreds of images per patient, and deriving functional parameters from these images through manual contouring is both time-consuming and labor-intensive for clinicians. Consequently, the automation of RV assessment has attracted significant research interest, offering the potential to dramatically reduce the time and effort required for quantitative ventricular analysis.

However, automated segmentation of the RV presents unique difficulties compared to the left ventricle. The RV 's thin myocardial walls, highly irregular geometry, and complex anatomical shape hinder accurate delineation. Traditional image processing techniques have been applied to RV segmentation [7], however, with limited success. In recent years, deep learning–based approaches have emerged as a promising alternative [8], [9], demonstrating improved accuracy and robustness in automated RV segmentation, and paving the way toward more efficient and reliable cardiac MRI analysis.

This study introduces a transfer learning–based approach aimed at enhancing the accuracy of RV segmentation in cardiac MRI. The proposed method leverages a pre-trained deep learning model obtained from the Medical Open Network for Artificial Intelligence (MONAI) Model Zoo, which serves

This work is in part supported by the Alliance International Catalyst grant by the NSERC. Dr L. Zhong is supported by the National Medical Research Council Clinician Innovator Award (MOH-001647-01), Individual Research Grant (NMRC/OFIRG/0018/2016 and NMRC/OFIRG25jan-0018).

as a robust initialization point for the segmentation task. This model is subsequently retrained using cine MRI image sequences acquired from a local hospital, allowing it to adapt to the specific imaging characteristics, scanner settings, and patient population of the target clinical environment.

Experimental results demonstrate that this transfer learning strategy substantially improves the segmentation accuracy of the neural network compared to the original pre-trained model. Furthermore, the proposed approach outperforms a leading state-of-the-art self-configuring segmentation framework, nnU-Net, achieving higher accuracy in delineating the RV's complex anatomical boundaries. These findings highlight the potential of transfer learning to bridge the gap between general pre-trained deep learning models and the nuanced requirements of institution-specific cardiac MRI datasets, ultimately advancing automated RV analysis in clinical practice.

II. Methodology

The proposed methodology employs transfer learning using a pre-trained model from the MONAI Zoo, specifically, the 3 Label Ventricular Segmentation model. This model retains the original architecture and training protocol described in [10]. The model was retrained using datasets acquired from the University of Alberta Hospital. The proposed study was approved by the Health Research Ethics Board at the University of Alberta. The details of the dataset are given below.

A. Dataset

The retrospective dataset included 10,566 short-axis cardiac MRI scans from 33 patients. Semi-automated RV segmentation was performed using proprietary in-house software, with all results manually verified and corrected by an experienced pediatric cardiac radiologist. From the total collection, 8,576 images were allocated for training and validation, while the remaining 1,994 short-axis MRI scans from 7 patients comprised the holdout test set. Table I shows the demographic details of our dataset.

TABLE I
Demographic Characteristics of Study Population

Set	Age (years)	Weight (kg)	Height (m)
Test	16.43 ± 3.33	62.04 ± 11.82	1.622 ± 0.092
Train	14.69 ± 5.75	48.20 ± 19.72	1.489 ± 0.195
Overall	15.06 ± 5.38	51.13 ± 19.19	1.517 ± 0.187

All scans were acquired using the SIEMENS Aera 1.5T scanner. Data presented as mean $\pm$ SD. Sex distribution: Test (2M:4F:1O), Train (13M:13F:0O), Overall (15M:17F:1O).

B. Data preprocessing and augmentation

Ground truth RV contours were rasterized into binary masks and paired with their source MRI images. All data were uniformly resized to 256×256 pixels to match network input requirements. The intensity values of the image were normalized to $[0, 1]$. We implemented comprehensive data augmentation including vertical and horizontal flipping, discrete 90° rotations (90°, 180°, or 270°), contrast adjustment through spatially smooth gamma fields ($\gamma \in [0.25, 3]$), random affine transformations incorporating rotation, translation, and scaling, and additive Gaussian noise with standard deviation $\sigma = 0.05$. This enhancement strategy was designed to improve the robustness of the model while maintaining the anatomical plausibility of cardiac MRI features.

C. Training configuration

We implemented a five-fold cross-validation strategy using the pretrained UNet architecture as our base model, with complete training/validation splits detailed in Fig. 1. The holdout test set, comprising data from 7 patients (see Fig. 1), was reserved for evaluating the final model. For each fold, we adapted the pretrained model to our RV-focused dataset by implementing a weighted Dice loss function with zero weights for the left ventricle and myocardium regions while maintaining the original architecture. The neural network was retrained over 1000 epochs, with 400 iterations per epoch and a batch size of 32, using the Adam optimizer with a reduced learning rate of 0.0005 and cosine decay scheduling for gradual rate reduction. The final model selection was based on optimal validation Dice similarity coefficient (DSC).

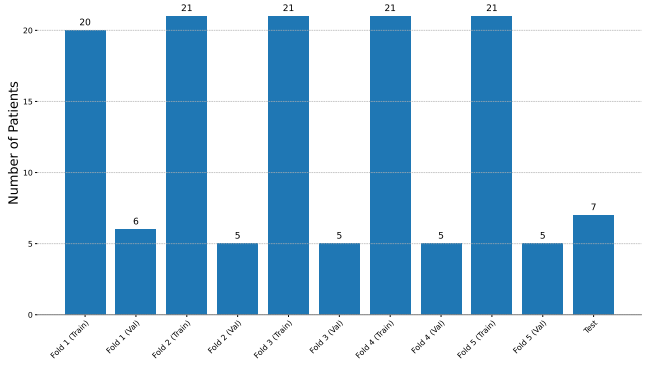

Fig. 1. Dataset partitioning scheme showing the five-fold cross-validation splits (training/validation sets) and independent holdout test set used for final model evaluation.

D. Inference configuration

Final model evaluation was conducted on the holdout test set using an ensemble of the five models generated during the cross-validation process. For each input slice, inference was performed in parallel across all five networks. The resulting softmax probability maps were then averaged to produce a single prediction. This ensemble map underwent a series of post-processing operations, including hole filling, retaining the largest connected component, and removing minor artifacts, to generate the final segmentation mask.

E. Baseline Methods

The performance of the proposed transfer learning approach was benchmarked against three baseline methods: the original

pre-trained model from the MONAI Model Zoo [10], the same architecture trained from scratch on our data, and the nnU-Net framework [11]. The pre-trained MONAI model was evaluated directly on the test set. The model trained from scratch followed the same training configuration as our proposed method, however, without initializing from pre-trained weights.

An nnUNet framework [11] was trained from scratch for comparison with the proposed approach. The same training and holdout validation sets used for the proposed method were applied to nnUNet to ensure a fair evaluation. As a state-of-the-art, self-configuring neural network, nnUNet automatically determines many hyperparameters based on the data itself, making it well-suited for benchmarking. Since nnUNet is primarily designed for three-dimensional (3D) medical images, the temporal frames of each short-axis MRI slice were stacked along the depth axis, allowing the network to process cine sequences as volumetric inputs and thereby capture motion information across the cardiac cycle, together with their corresponding RV segmentation labels. Each fold of the built-in five-fold cross-validation strategy was trained in parallel on NVIDIA V100 GPUs, with early stopping applied to prevent overfitting. The final model for evaluation was chosen automatically using nnUNet's `find_best_model` function.

F. Quantitative evaluation

The quantitative evaluation of the neural network predictions was conducted against expert manual annotations using the DSC, 95$^{\text{th}}$ percentile Hausdorff distance (HD95), and reliability analysis.

1) DSC: The spatial overlap between the predicted segmentation $\mathcal{S}_a$ and the expert annotation $\mathcal{S}_m$ was quantified using the DSC, defined as

$$DSC(\mathcal{S}_a, \mathcal{S}_m) = \frac{2\mathcal{S}_{am}}{\mathcal{S}_a + \mathcal{S}_m}, \tag{1}$$

where $\mathcal{S}_{am}$ denotes the intersection between the predicted and manual segmentations. A value of $DSC = 1$ indicates perfect agreement, whereas $DSC = 0$ indicates no overlap between the two segmentations.

2) HD95: The Hausdorff distance (HD) [12] is a symmetric measure of the spatial discrepancy between the automatic and manual boundaries of the RV region. Let C_a and C_m denote the automatic and manual contours, respectively. For each point p_a^i on C_a, the minimum Euclidean distance to any point p_m^j on C_m is computed, and vice versa. The HD is then defined as

$$HD(C_a, C_m) = \max\left(\max_i \min_j d(p_a^i, p_m^j), \max_j \min_i d(p_m^j, p_a^i) \right), \tag{2}$$

where $d(\cdot)$ denotes the Euclidean distance. Instead of the absolute maximum, this study reports the 95$^{\text{th}}$percentile of the Hausdorff Distance, denoted as HD95, to reduce sensitivity to outliers.

3) Reliability: The reliability quantifies the consistency or dependability of a method in achieving a desired performance. It reflects the likelihood that the method will meet or exceed a specified quality threshold, thus serving as an indicator of robustness and stability.

In the context of our segmentation task, reliability measures the probability that the DSC between the predicted and ground-truth segmentations exceeds a given threshold $d \in [0, 1]$. Formally, this is expressed as:

$$\mathcal{R}(d) = Pr(DSC > d) \tag{3}$$

where $Pr(\cdot)$ denotes probability over the test set. The reliability curve, obtained by computing $R(d)$ across a range of thresholds d, provides insight into how consistently a segmentation method achieves various accuracy levels. Higher curves indicate greater robustness and dependability in producing accurate segmentations.

III. Results

The performance of the proposed transfer learning approach was compared with nnU-Net [11] trained from scratch on the local dataset, the pre-trained model from the MONAI Model Zoo, and the same model trained from scratch on the local dataset. Predictions from all four neural networks were evaluated against expert manual annotations using the DSC and HD95. Table II reports the mean and standard deviation values of the DSC and HD95 for the RV segmentation for all four neural networks. The results show that the proposed approach outperforms the other three neural networks in terms of both DSC and HD95.

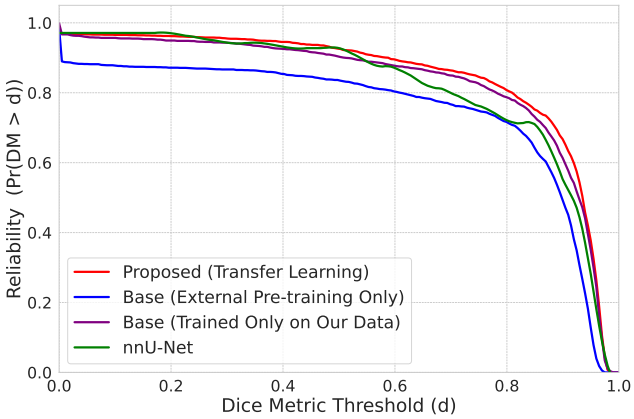

Fig. 2. The reliability ($\mathcal{R}(d) = Pr(Dice > d)$) for the proposed method and baseline methods. The proposed method yielded a higher reliability curve than both MONAI Zoo variants (pre-trained and trained on our dataset only) as well as nnU-Net, demonstrating the benefit of performing transfer learning for the proposed segmentation problem.

The reliability analysis in Fig. 2 offers additional insight into model consistency. The pre-trained MONAI Zoo model shows the lowest reliability, with a reduced likelihood of achieving higher Dice thresholds. Retraining the same architecture on our dataset markedly improves reliability, not only surpassing the pre-trained version but also outperforming nnU-Net across

TABLE II
QUANTITATIVE SEGMENTATION PERFORMANCE METRICS COMPARING AUTOMATED PREDICTIONS AGAINST EXPERT MANUAL ANNOTATIONS FOR RV DELINEATION. RESULTS SHOW MEAN ± STANDARD DEVIATION FOR DICE SIMILARITY COEFFICIENT AND HD95 ACROSS 1,994 TEST SET SHORT-AXIS MRI IMAGES FROM 7 PATIENTS.

Method	**DSC** (↑) **Mean (%) ± Std (%)**	**HD95** (↓) **Mean (mm) ± Std (mm)**
nnUNet [11]	81.98 ± 22.21	7.09 ± 14.49
MONAI Zoo Model (Pre-trained) [10]	74.93 ± 31.17	12.78 ± 17.48
MONAI Zoo Model (Trained on local data)	83.15 ± 23.32	7.71 ± 8.85
Proposed approach (Transfer learning)	84.90 ± 21.40	6.02 ± 6.49

much of the threshold range. The proposed transfer learning approach achieves the highest reliability overall, particularly at stricter thresholds, demonstrating its superior ability to deliver high-quality segmentations consistently.

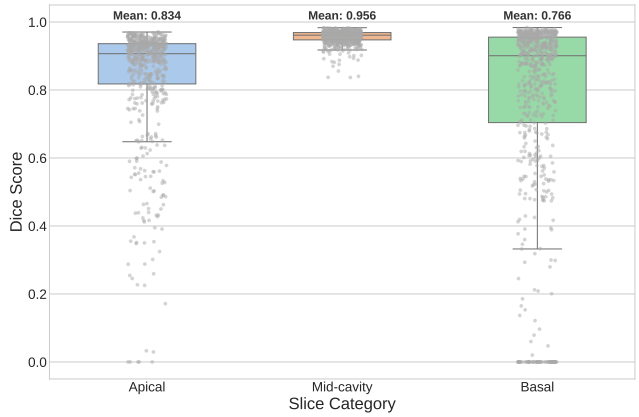

Fig. 3. The performance of the proposed algorithm for the RV segmentation at different short-axis slices, namely, the apical, mid-cavity, and basal slice levels.

We analyzed the proposed method's performance on different anatomical regions of the RV, namely, apical, mid-cavity, and basal regions. As shown by the box plot in Fig. 3, the performance is not uniform. The algorithm performed exceptionally well on mid-cavity slices, achieving a high mean Dice score of 95.58% with a standard deviation of just 1.98%. The performance on apical slices was weaker; however, it still yielded a respectable mean Dice score of 83.41% with a standard deviation of 17.73%.

The primary challenge for the model was the basal region, which yielded a considerably lower mean score of 76.57% with a standard deviation of 28.58%. It also exhibited a much wider distribution of results with numerous low-scoring outliers. The qualitative examples in Figure 4 visually confirm this trend. The mid-cavity (middle row) and apical (top row) segmentations are generally accurate, even at lower percentiles. In contrast, the basal region (bottom row) is prone to significant errors, particularly in lower-percentile cases, which explains the decreased and more variable performance. This indicates that while the model is robust for most of the ventricular anatomy, the complexity of the basal slices is the main source of segmentation inaccuracy in the proposed model.

A. Visual results

The predicted RV contours generated by the proposed method, alongside the expert manual contours, are illustrated on the MRI images at the apical, mid-cavity, and basal slices in Fig. 4. The columns correspond to cases with DSC scores at the 25^{th}, 50^{th}, 75^{th}, and 100^{th} percentiles. Overall, the results demonstrate strong agreement between the automated and manual contours across most images in the test set, indicating the method's robustness across different short-axis images.

IV. CONCLUSION

In this study, we present a transfer learning approach for RV segmentation from MRI images. Our method leverages a pre-trained neural network model from the MONAI Model Zoo, which is then fine-tuned using MRI datasets acquired from a local hospital. Evaluation on a holdout test set of 1,994 images, using the DSC and HD95 metrics, demonstrates that the proposed approach substantially enhances segmentation performance compared to both the pre-trained MONAI Zoo model and the same architecture trained on our dataset. The latter already outperformed nnU-Net, while the proposed method achieved the best overall performance. These results highlight the advantages of adapting pre-trained neural network models with local data, rather than applying them directly to domain-specific tasks without retraining.

The clinical impact of this work lies in its potential to automate the laborious task of RV segmentation, thereby enabling more routine and efficient functional assessment. The model is now being leveraged to generate a large-scale, annotated dataset from our patient cohort. This effort will directly enable our subsequent investigation into optimizing imaging protocols by determining the minimum number of short-axis slices required for accurate RV volume computation.

While the current results are promising, further work is needed to ensure the model's generalizability. A primary limitation is the training cohort, which could be expanded to ensure robustness across a broader spectrum of patient demographics and pathologies. Second, the model was developed on a single-vendor dataset, and establishing its performance on images from different manufacturers is a crucial step for broader applicability. Addressing both patient diversity and hardware variability is essential for developing a versatile tool. Beyond enhancing robustness, future work could also focus

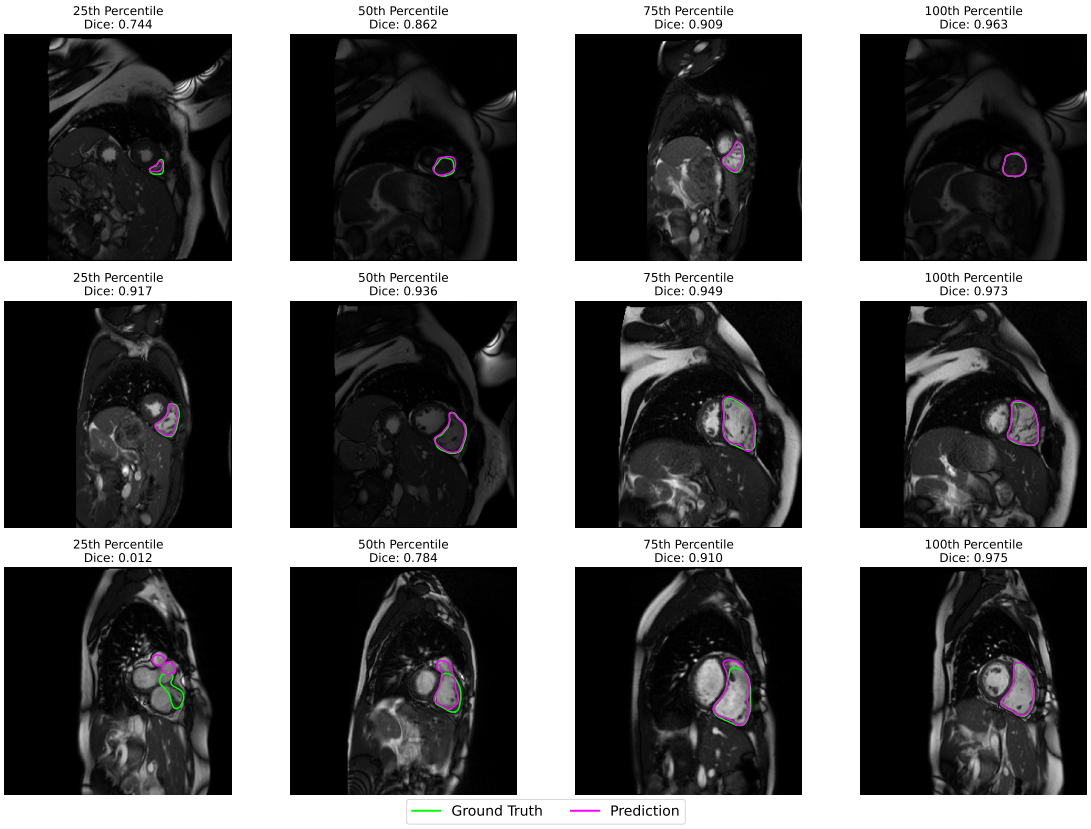

Fig. 4. Comparison of RV segmentation contours between the proposed method (purple) and manual annotations (green). The figure shows representative slices stratified by anatomical location, with the first, second, and third rows corresponding to apical, mid-cavity, and basal regions, respectively. For each category, images at the 25th, 50th, 75th, and 100th percentiles of the Dice score are displayed.

on expanding the model's scope by applying similar transfer learning techniques to segment long-axis views.

REFERENCES

[1] M. Vaduganathan, G. A. Mensah, J. V. Turco, V. Fuster, and G. A. Roth, "The Global Burden of Cardiovascular Diseases and Risk," *Journal of the American College of Cardiology*, vol. 80, no. 25, pp. 2361–2371, Dec. 2022.

[2] M. Di Cesare, P. Perel, S. Taylor, C. Kabudula, H. Bixby, T. A. Gaziano, D. V. McGhie, J. Mwangi, B. Pervan, J. Narula, D. Pineiro, and F. J. Pinto, "The Heart of the World," *Global Heart*, vol. 19, no. 1, p. 11, Jan. 2024.

[3] Y.-R. Wang, K. Yang, Y. Wen, P. Wang, Y. Hu, Y. Lai, Y. Wang, K. Zhao, S. Tang, A. Zhang, H. Zhan, M. Lu, X. Chen, S. Yang, Z. Dong, Y. Wang, H. Liu, L. Zhao, L. Huang, Y. Li, L. Wu, Z. Chen, Y. Luo, D. Liu, P. Zhao, K. Lin, J. C. Wu, and S. Zhao, "Screening and diagnosis of cardiovascular disease using artificial intelligence-enabled cardiac magnetic resonance imaging," *Nature Medicine*, vol. 30, no. 5, pp. 1471–1480, May 2024.

[4] S. R. Mehta, "Impact of right ventricular involvement on mortality and morbidity in patients with inferior myocardial infarction1," vol. 37, no. 1, 2001.

[5] N. Galea, I. Carbone, D. Cannata, G. Cannavale, B. Conti, R. Galea, A. Frustaci, C. Catalano, and M. Francone, "Right ventricular cardiovascular magnetic resonance imaging: Normal anatomy and spectrum of pathological findings," *Insights into Imaging*, vol. 4, no. 2, pp. 213–223, Apr. 2013.

[6] L. Cavigli, M. Focardi, M. Cameli, G. E. Mandoli, S. Mondillo, and F. D'Ascenzi, "The right ventricle in "Left-sided" cardiomyopathies: The dark side of the moon," *Trends in Cardiovascular Medicine*, vol. 31, no. 8, pp. 476–484, Nov. 2021.

[7] K. Punithakumar, P. Boulanger, and M. Noga, "A GPU-accelerated deformable image registration algorithm with applications to right ventricular segmentation," *IEEE Access*, vol. 5, pp. 20 374–20 382, 2017.

[8] H. Purmehdi, A. R. Hareendranathan, M. Noga, and K. Punithakumar, "Right ventricular segmentation from MRI using deep convolutional neural networks," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, July 2019, pp. 4020–4023.

[9] C. Martín-Isla, V. M. Campello, C. Izquierdo, K. Kushibar, C. Sendra-Balcells, P. Gkontra, A. Sojoudi, M. J. Fulton, T. W. Arega, K. Punithakumar, L. Li, X. Sun, Y. Al Khalil, D. Liu, S. Jabbar, S. Queirós, F. Galati, M. Mazher, Z. Gao, M. Beetz, L. Tautz, C. Galazis, M. Varela, M. Hüllebrand, V. Grau, X. Zhuang, D. Puig, M. A. Zuluaga, H. Mohy-ud-Din, D. Metaxas, M. Breeuwer, R. J. Van Der Geest, M. Noga, S. Bricq, M. E. Rentschler, A. Guala, S. E. Petersen, S. Escalera, J. F. R. Palomares, and K. Lekadir, "Deep learning segmentation of the right ventricle in cardiac MRI: The M&Ms challenge," *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 7, pp. 3302–3313, July 2023.

[10] E. Kerfoot, J. Clough, I. Oksuz, J. Lee, A. P. King, and J. A. Schnabel, "Left ventricle quantification using residual U-Net," in *Statistical Atlases and Computational Models of the Heart. Atrial Segmentation and LV Quantification Challenges*, M. Pop, M. Sermesant, J. Zhao, S. Li, K. McLeod, A. Young, K. Rhode, and T. Mansi, Eds. Cham: Springer International Publishing, 2019, vol. 11395, pp. 371–380.

[11] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, no. 2, pp. 203–211, Feb. 2021.

[12] D. Huttenlocher, G. Klanderman, and W. Rucklidge, "Comparing images using the Hausdorff distance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 9, pp. 850–863, Sept. 1993.