Improving the Generalizability of Cardiac MRI Segmentation Models Across Multi-Center and Multi-Vendor Datasets

부산대학교 정보컴퓨터공학부

지도교수: 감진규

분과명: 소프트웨어/인공지능

팀명: Cyber

202055633 이슬람 살리흐

202155631 케네스 예라슬

202155629 누가예바 알트나이

Table of Contents

1. Project Overview

1.1. Introduction and Problem Background

Cardiac magnetic resonance (cMRI) is known as the most acknowledged standard for the clinical evaluation and assessment of various cardiac diseases. When observing cMRIs, it is very important to gather information about the left ventricle (LV), myocardium, and right ventricle (RV). This is needed to identify the type of disease and its precise location. Although traditional manual segmentation methods are still widely used, it is very likely that the results of this assessment will greatly depend on the experience of each professional and the conditions in the hospital. Moreover, manual segmentation requires careful examination of each scan and often needs a lot of time, which probably will lead to errors and inconsistencies. The process requires extensive manual input to annotate cardiac boundaries across all image slices and phases, potentially compromising the consistency and accuracy of the result (Campello et al., 2021). This further proves the idea of improving and integrating automated or semi-automated assessment techniques to enhance the efficiency and reliability of cardiac assessment in clinical practice.

As a solution to the limitations of manual segmentation, deep learning-based methods are becoming increasingly popular as they provide more accurate and effective solutions to the problem. For example, the usage of CNN by Tran et al. shows a significant advantage in the segmentation of the left and right ventricles in short-axis (SA) MRI scans (Tran, 2017). Following this trend, Poudel introduced a recurrent FCN network that works with spatial information to improve the performance of left ventricle (LV) segmentation (Poudel et al., 2017). Although this method is developing very fast, we are still lacking precision and accuracy in the segmentation of the right ventricle (RV) as it is challenging because of its heterogeneous intensity, variable morphology, and indistinct boundaries in MRI scans. These challenges are further complicated in the basal and apical slices, where existing methods frequently suffer from reduced segmentation accuracy (Li et al., 2022).

Nonetheless, the main limitation of current deep learning models is their generalization. It often struggles when facing previously unseen datasets from different clinical centers or imaging vendors. As this kind of variability in imaging protocols and patient demographics or disease characteristics, and scanner-specific biases will occur constantly in real-life clinical practices, it can significantly degrade the overall performance and reliability of these models (Glocker et al., 2019).

Automating cardiac segmentation is very important for the efficient and reliable diagnosis of cardiovascular diseases. Even though deep learning methods have shown big potential, their limitations require further research in this field. There is an urgent need to develop models with enhanced robustness and generalization capabilities, which will consistently handle the variability of multi-center and multi-vendor datasets. More research is required to solve current challenges of deep learning models to ensure that they can be used effectively in the routine of clinical practices from cardiac assessment of patients.

1.2. Objectives

The primary objectives for our project in its current stage are as follows:

1. To conduct an extensive literature review on the cardiac MRI segmentation methodologies, by assessing their strength and limitations. We will particularly assess their generalization across diverse datasets. This objective aims to establish a comprehensive understanding of the problems that need to be solved in this domain.
2. To perform efficient and accurate analysis of the multi-center and multi-vendor datasets, we must implement different data preprocessing techniques such as normalization, and standardization to prepare the dataset for subsequent model training.
3. To visualize the cardiac MRI images and given segmentation labels to gain deeper insights into the dataset’s characteristics, thereby preparing for model development and evaluation steps.
4. To develop a new and reliable deep learning-based segmentation model with the objective of improving generalization across multi-center and multi-vendor datasets, addressing the problem of constant variability of these datasets.
5. To train the deep learning model using the preprocessed dataset, so that the training process includes strategies aimed at enhancing the model’s adaptability to diverse image conditions.
6. To evaluate the performance of our model developed on a chosen dataset for its ability to generalize across different clinical centers and vendors. This evaluation will include a deeper analysis of initial results and further improvements, mainly focusing on the accuracy and reliability of the model.
7. To solve the generalizability issues of deep learning models on the segmentation process, different methods like domain adaptation and data augmentation will be used. To assess the effectiveness of these methods, we will conduct comprehensive testing and analysis.
8. To provide practical usage of created models in clinical practices aimed at increasing the efficiency and reliability of deep learning models in diagnostic image workflow in cardiac imaging.

2. Modifications to Requirements and Constraints Analysis

2.1. Existing Requirements and Modifications

2.1.1. Initial Requirements

The initial requirements for the project were focused on the development of a reliable and robust deep learning model on the cardiac MRI segmentation model. These requirements included:

- **Data Requirements**: the usage of diverse datasets is essential to build reliable models. For this reason, we decided to utilize the M&M Challenge dataset, which contains 375 heterogeneous CMR datasets collected from 4 different scanner vendors across six hospitals in Spain, Canada, and Germany (Campello et al., 2021). Generalization is very important for developing models capable of working with images of different conditions and protocols. Other than M&M, the ACDC dataset was considered, but due to its limited size and variability, we deemed it less valuable.
- **Training Requirements**: to build a model that is reliable and capable of generation, it is required to incorporate data augmentation techniques. Data augmentation is used to artificially create more training materials by applying various geometric and photometric transformations to the images (Perez & Wang, 2017). We planned to implement domain adaptation using the Domain Adversarial Training of Neural Networks (DAAN) model, which helps to minimize domain shifts while keeping the model's effectiveness in segmenting cardiac structures (Corral Acero et al., 2021).

2.1.2. Modified Requirements

As we progress with the project, existing requirements get modified and new requirements are introduced to address emerging problems:

- **Data Requirements**: the ACDC dataset was ultimately excluded as an option for our model because of its limited diversity. This is due to its derivation from a single clinic with a single imaging protocol, which will not be enough to create a robust model. Then we conducted a deep review of the M&Ms and M&Ms-2 datasets. After that, it was decided to focus on the M&Ms-2 dataset because of its recent update, which presents a unique opportunity for novel contributions.
- **Training Requirements**: initially, we planned to implement standard data augmentation techniques, such as geometric and photometric transformations. Focusing on altering the data to better handle variability in cardiac MRI images across different clinical centers and imaging protocols will help make

an adaptive model. Additionally, the augmentation process also included transformations such as rotations, flipping, scaling, and intensity variations. Future work will implement the usage of Domain Adversarial Training of Neural Networks (DAAN) to mitigate performance drops in new domains or unseen images.

2.2. Existing Constraints and Modifications

2.2.1. Initial Constraints and Measures

The original constraints for the project included several important challenges:

Data Constraints:

- **Limited Availability of Diverse Data**: it was recognized that the lack of sufficient data for heart MRI scans potentially risked the possibility of training a reliable model.

Technical Constraints:

- **Computational Limitations**: it was expected that high computational power would be demanded for training deep learning models**.** The project required a sufficient server with the availability of GPU clusters.
- **Software Compatibility**: another challenge is the compatibility and integration of various software tools and libraries. Given the complexity of the deep learning framework, it was expected to work with different platforms and specialized imaging software.

2.2.2. Evolving Constraints and Measures

As the project progressed, more constraints were identified. To address the given challenges following adjustments were made:

Data Constraints:

- **Limited Availability of Diverse Data**: to mitigate this constraint, the team decided to leverage the M&Ms-2 dataset, which offers a broader range of imaging protocols and patient demographics. Also, preprocessing techniques like data cleaning, normalization, and standardization, were used to ensure that the data was ready for training. This approach aims to solve any initial data complication by making the available datasets more robust and suitable for further analysis.

Technical Constraints:

- **Computational Limitations**: to address the computational demands of training the deep learning models, the team members have optimized the training process by implementing mixed precision training and utilizing cloud-based GPU resources. Additionally, the support from institutional high-performance computing resources allowed us to reduce training time and manage computational load effectively.
- **Software Compatibility**: as the project progressed, the complexity of the software stack necessitated additional focus on ensuring compatibility between various tools and libraries. This led to modification in the original software environment setup, including the adoption of containerization technologies, such as Docker, to streamline development and deployment processes.

3. Team Member Progress

Name	Role
이슬람 살리흐	**Data Preprocessing and Image Standardization**: ● Develop and implement the *'preprocess_image'* function to preprocess MRI images for subsequent model development. ● Execute cropping operations using the *'crop_image'* function to eliminate unnecessary regions, which will help with the retention of only relevant image portions. ● Manage the transposition of images used to optimize spatial orientation. This will ensure that images are processed consistently across subsequent analyses. **Model Training and Checkpoint Management**: ● Implement checkpoint loading and fallback initialization mechanisms. This enables our model to resume training or depending on the data availability it can start new sessions.

	● Manage the supervised training loop by focusing on validation processes, checkpointing, and learning rate adjustments.
케네스 예라슬	**Segmentation Model Architecture Design**: ● Design the segmentation model using a U-Net-based architecture, including the implementation of the Baseline_2 class and the Generic_UNet structure. ● Structure the model with various modules, layers, and outputs. Use activation functions such as Leaky ReLU and normalization techniques like BatchNorm to improve model performance. **Model Testing and Inference Management**: ● Initialize and load models for inference, with a focus on processing test data and generating precise predictions. ● Handle the inference outputs, which include the storage of ground truth data, predictions, and reconstructed images.
누가예바 알트나이	**Data Rescaling and Normalization**: ● Perform image resizing to ensure a standardized resolution across all scans, maintaining uniform spacing. ● Implement the normalization of image data. This includes adjusting pixel values to achieve a mean of 0 and a standard deviation of 1 for given images. Also, apply specialized resizing techniques for segmentation maps to save the integrity of segmented regions. **Postprocessing and Model Evaluation**: ● Perform the postprocessing of predictions using the *'postprocess_predictions'* and *'postprocess_image'* functions by adjusting image spacings and shapes while applying transformations.

	● Evaluate the model's performance using the Dice Similarity Coefficient (DSC) and Hausdorff Distance (HD) metrics, ensuring the accuracy and robustness of the segmentation results.
Common	**Background Study and Information Gathering**: engage in comprehensive research and information gathering, focusing on current techniques and challenges in cardiac MRI segmentation. **Documentation**: contribute to report writing, presentation design and demonstration, and project defense preparations.

4. Design Details and Changes

4.1. Dataset Exploration and Analysis

We have chosen the M&Ms-2 Challenge dataset for our project, which contains the results of heart MRI of 360 patients. The number of patients includes individuals who suffer diseases that affect the right ventricle and left ventricle. There are also healthy individuals, which will help the model to be more generalized. This dataset is an ideal match for model creation due to its multi-center and multi-vendor nature. All the subjects were scanned at 3 clinical centers in Spain using MRI scanners from three different vendors: Siemens, General Electric, and Philips Medical Systems (Martín-Isla et al., 2023).

The MRI data were acquired using 9 different scanner models across the three vendors:

1. **General Electric (GE Healthcare)**: SIGNA EXCITE, Signa HDxt, Signa Explorer.
2. **Philips**: Achieva.
3. **Siemens**: SymphonyTim, Symphony, TrioTim, Avanto Fit Avanto.

These scanners vary in their imaging capabilities, technologies, and specific features used, which will create the range of imaging conditions that may be encountered in clinical practice. This is important to us as it creates multi-vendor variability, which is a significant factor in developing robust segmentation models that can generalize well across different image settings.

In total, the dataset includes 160 training and 40 validation instances covering 8 types of pathologies. The training set consists of 200 annotated images from four distinct centers. Experienced clinicians segmented the cardiac MRI (CMR) images, delineating the contours for the left ventricle (LV) and right ventricle (RV) blood pools, as well as the left ventricular myocardium (MYO). The labels used in the annotations are as follows: 1 for LV, 2 for MYO, and 3 for RV. These annotations are provided in both short-axis and long-axis views, encompassing a range of challenging RV pathologies and LV remodeling.

Pathology	**Num. studies training**	**Num. studies validation**
Normal subjects	40	5
Dilated Left Ventricle	30	5
Hypertrophic Cardiomyopathy	30	5
Congenital Arrhythmogenesis	20	5
Tetralogy of Fallot	20	5
Interatrial Communication	20	5
Dilated Right Ventricle	0	5
Tricuspid Regurgitation	0	5

Figure 1. Training and Validation Data Distribution

Two pathologies (Tricuspid Regurgitation and Congenital Arrhythmogenesis) are absent in the training set but present in the validation and testing sets to evaluate the model's generalization to unseen pathologies. The testing set consists of 160 studies, with the following pathology distribution:

Pathology	**Num. studies test**
Normal subjects	30
Dilated Left Ventricle	25
Hypertrophic Cardiomyopathy	25
Congenital Arrhythmogenesis	10
Tetralogy of Fallot	10
Interatrial Communication	10
Dilated Right Ventricle	25
Tricuspidal Regurgitation	25

Figure 2. Testing Data Distribution

Dataset Characteristics

The dataset encompasses a wide range of cardiovascular diseases, including Hypertrophic Cardiomyopathy (HCM), Dilated Cardiomyopathy (DCM), Hypertensive Heart Disease (HHD), Abnormal Right Ventricle (ARV), Athlete Heart Syndrome (AHS), Ischemic Heart Disease (IHD), and Left Ventricle Non-Compaction (LVNC). Additionally, we are given field strength in Teslas (T), which affects the quality, detail, and scan time of the MRI images produced, as well as patient comfort.

The subjects in this multi-disease study were carefully selected to cover a comprehensive spectrum of cardiovascular conditions, providing a robust basis for model training and validation.

Centre / Pathology	1	2	3	4	5	6
Healthy vol.	22	33	32	21	14	3
HCM	25	37	14	8	15	4
DCM	37	-	5	-	9	-
HHD	-	4	-	19	1	1
ARV	12	-	-	2	1	1
AHS	-	-	-	3	-	-
IHD	-	-	-	4	1	3
LVNC	-	-	-	-	2	2
Other	-	-	-	18	7	15

Figure 3. Distribution of Frequent Pathologies and Healthy Volunteers between Centers

- **Annotation Process**

Each CMR study was manually annotated by expert clinicians with experience ranging from 3 to over 10 years. Annotations were performed following clinical protocols, focusing on the end-diastolic (ED) and end-systolic (ES) phases in short-axis views. The annotated regions include the left and right ventricles (LV and RV) cavities and the left ventricle myocardium (MYO).

- **Standard Operating Procedure (SOP)**

To ensure consistency and accuracy in annotations, the Standard Operating Procedure (SOP) from the ACDC challenge was adopted. The SOP provides a standardized set of instructions for executing specific tasks, ensuring that all annotations meet high standards of quality and consistency.

4.2 Data Preprocessing and Visualization

First, to get a deeper understanding of datasets, we did an in-depth exploration of the cardiac MRI dataset. This helped us to better understand the anatomical structures and functions of the heart. There are cine MRI series captured in both long-axis (LA) and short-axis (SA) orientations in our dataset. Another thing to keep in mind is the presence of segmentation labels corresponding to different cardiac phases.

Data Loading and Initial Inspection

First, we loaded MRI data using the *'nibabel'* library. This library was used in our project to work with NIfTI format files. After loading the dataset, first, we performed dimensional analysis on the spatial and temporal parameters of an image. After examining the characteristics of the MRI data cubes, the following result was obtained: (256, 256, 1, 25)

It shows that the image has a spatial resolution of 256x256 pixels. The third '1' dimension indicates a single frame of the heart in one spatial dimension. Then, '25' represents the temporal resolution, meaning there are 25 frames captured over a specific period, corresponding to different phases of the cardiac cycle.

Visualization of Cine MRI Series

Visualization of the cine MRI is an important step for understanding the dynamic movement of the heart in the time of a cardiac cycle. We used *'matplotlib'* library to visualize frames of the heart, captured over time, in both LA and SA views. These visualizations are helpful for accessing the functional dynamics of the heart. It allows us to observe how the heart chambers and myocardial tissue behave during systole and diastole. Creed animations were saved as .gif files for later use.

Here are the visualized images of the LA and SA views:

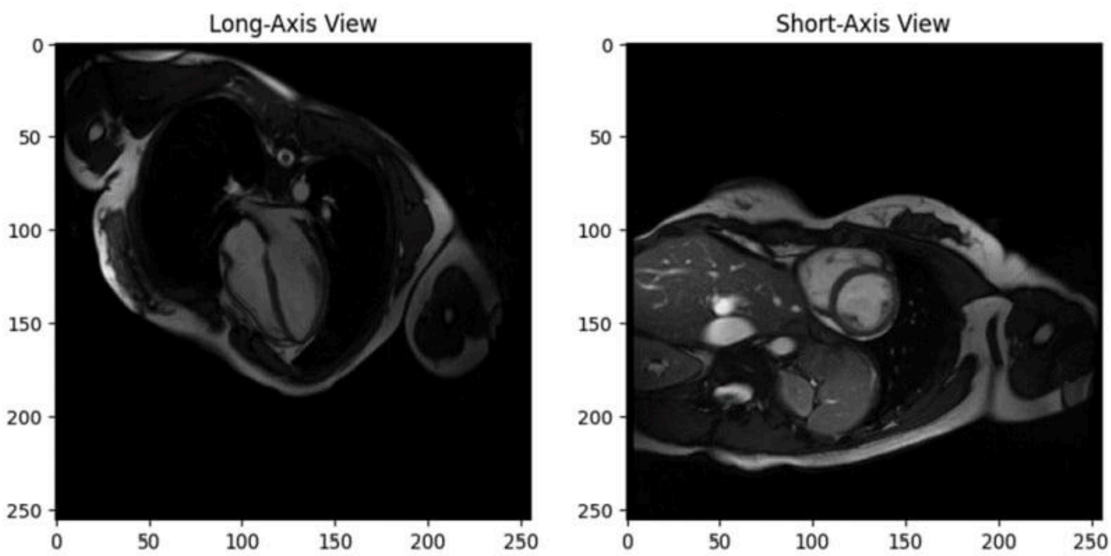

Figure 4. Resulting Frame of Cine MRI Data

By visualizing the cine MRI series, we can capture the movement of the heart across the entire cardiac cycle, providing necessary insights into the underlying cardiac anatomy and physiology. The LA view offers a perspective on the longitudinal movement of the heart, while the SA view provides a cross-sectional view. This gives us the ability to see the contraction and relaxation of the ventricles.

4.3 Analysis of the Methods

- **Data Augmentation**

Data augmentation is needed to create diversity in the training set. This will help the model to avoid the risk of overfitting and improve its generalizability. As imaging protocols and scanner vendors may vary greatly from case to case, it is important to apply various transformations like histogram matching, contrast modification, noise addition, and image synthesis. It makes the dataset more varied, which helps the model to be robust.

- **Data Augmentation Challenges**

Although data augmentation helps to create intensity-driven alteration to the dataset, it still does not change the fact that the segmentations obtained from the model tend to be more stable for trained vendors but will struggle to generalize to unseen vendors. This limitation requires more comprehensive strategies to address the full range of imaging protocols and vendors that may be encountered in clinical practice.

- **Domain Adaptation**

Domain adaptation theoretically should improve the model on the multi-vendor image segmentation by adapting to the unseen images. However, its application does not guarantee consistent improvements. Thus, it indicates that there is a dependency on the choice of baseline model and dataset.

- **Segmentation Accuracy Challenges**

There is a big difference in segmentation accuracy when a model is tested on unlabeled or unseen data samples compared to labeled ones. This shows that training and testing on datasets from the same vendor do not provide robust generalizability. This matter requires further advancements in model adaptation.

- **Robustness of Techniques**

Data augmentation and GANs are commonly used techniques in medical image segmentation. Even though they generally enhance model performance, they do not automatically guarantee robust results. It is apparent that if we don't calibrate it properly, the techniques may introduce noise or overfitting. Accurate validation is necessary to ensure the segmentation of our model for clinical applications.

- **Computational Considerations**

In general, it is better to have capable hardware. However, increased computational power does not always mean better accuracy. In deep learning models for medical image segmentation, it is important to balance the computational demand with model accuracy. It is needed to ensure that the model is practical and beneficial for effective work in clinical settings.

5. Project Activities and Interim Results

5.1. Project Activities

5.1.1. Model Selection

One of the most fundamentals of developing robust deep learning models for cardiac MRI segmentation and labeling is a selection of an appropriate model architecture. Our project emphasizes the importance of generalization of the model, which ensures the consistent performance of a model across diverse datasets from multiple clinics and scanner vendors. The model selection process is aimed at finding architecture that is capable of managing all the complexity and variability that come with medical imaging.

Segmentation Model: U-Net

For segmentation tasks, the U-Net architecture has been selected as the primary model. U-Net is known for its wide use in medical image segmentation. This wide recognition is due to the effectiveness of the model in handling various challenges. These are the key aspects of U-Net architecture:

- **Encoder-Decoder Structure**: The U-Net model employs a symmetric encoder-decoder framework. The encoder progressively captures contextual information through downsampling, while the decoder reconstructs the image through upsampling. This structure is particularly effective in retaining essential spatial features critical for accurate segmentation.
- **Symmetric encoder-decoder Architecture**: The U-Net model follows a symmetric and so-called medical image processing format where it uses both encoding and decoding. On the encoder side, context is accumulated throughout down-sampling, and on all levels, dense information maps are

obtained. This kind of architecture can help in preserving key spatial features that are important for accurate segmentation.

- **Skip Connections Symmetric**: These are same-level connections between the encoder and decoder paths that help in retaining fine details from both high-level and low-level features. This dual incorporation is required for proper segmentation of complex anatomical structures in cardiac MRI.
- **Precise Localization**: The proposed U-Net design allows for very accurate localization, which is critical in ascertaining the exact location of anatomical structures on cardiac MRI images.
- **Efficiency**: U-net is computationally efficient compared to more complex models. This high efficiency makes it ideal for training the initial model and deployment, under constraints in terms of computational resources available.

Model Flexibility and Scalability

Although U-Net acts as a base architecture for the first phase of development, this allows flexibility to try out other architectures later in future iterations:

- **U-Net Variants**: Further work could focus on more advanced variants like U-Net++ among other improvements that aim to increase performance through deeper networks or improved skip connections.
- **Complex Models**: Depending on the observed performance and specific challenges encountered, more complex architectures, such as attention-based models or transformers adapted for medical image segmentation, could be considered.
- **Adaption to Performance**: It is intended that the project will also be adaptable and that alternative architectures based on empirical results, as well as changes in segmentation tasks, may be integrated. In this way, it makes the solution state-of-the-art and efficient.

The focus remains on U-Net for its established efficacy in medical imaging, balanced with considerations for potential future enhancements to address evolving project requirements and performance needs.

5.1.2. Model Training

The training is a very important step in model creation. It ensures that the model will be able to adapt to new unseen data, thus making it more robust. Our project includes cross-domain and multi-center training strategies. Moreover, we utilize the full range of datasets to improve the model’s ability to work with images of different conditions and sources.

Cross-Domain and Multi-Center Training

Training on a wide range of data from multiple clinical centers, our model learns to handle a wide range of imaging conditions and variations. This will enhance robustness of the model, which is essential for medical image examinations. Our approach aims to improve the model's reliability in real-world clinical applications, where data variability is expected. That is why our model will be able to work freely with any clinical image from many different MRI results, thus improving the reliability of our model in medical practice.

Training Process

Our model's training process requires high-performance computer hardware equipment with GPUs. This is because computers must handle huge datasets and demanding model creation processes for deep learning. Training process include:

Iterative Optimization - current model is fine-tuned using backpropagation and stochastic gradient descent (SGD). This ensures that our model will create the most optimal results with the given dataset.

Hyperparameter Tuning - careful adjustment of hyperparameters such as learning rate, batch size, and number of epochs is applied to achieve the best possible performance.

This comprehensive training approach is designed to ensure that the model performs well not only on the training data but also on new, unseen datasets, making it a robust solution for cardiac MRI segmentation and labeling.

5.1.3. Model Testing

After the training phase, an in-depth testing process is essential to evaluate the model's performance and generalization capabilities. This testing ensures that the model not only performs well on the training data but also generalizes effectively to new, unseen examples.

Testing Procedure

Testing is carried out on a separate validation set comprising data from different clinical centers and vendors, which were not used during training. This approach is crucial for assessing the model's ability to generalize across various conditions and sources, reflecting its performance in real-world clinical settings.

Cross-Validation

To thoroughly evaluate the model's performance, cross-validation techniques are employed:

Folding Strategy - the dataset is divided into multiple folds, typically using k-fold cross-validation. In this approach, the model is trained on k−1 folds and tested on the remaining fold. This process is repeated for each fold, ensuring that every data point is used for training and testing.

Benefits - cross-validation helps identify overfitting or underfitting by evaluating the model's performance across different subsets of the data. It provides a more reliable and generalizable estimate of the model's performance.

Metrics and Performance Analysis

The evaluation of the model's performance involves several key metrics that are standard in medical image segmentation. These metrics offer a comprehensive view of the model's accuracy, sensitivity, and overall robustness:

Dice Coefficient (Dice score) - measures the overlap between the predicted segmentation and the ground truth. Providing a clear indication of how well the model captures the true positive areas in the segmentation.

Hausdorff Distance (HD) - evaluates the maximum distance between the boundary points of the predicted segmentation and the ground truth, offering insight into the worst-case boundary mismatches.

Jaccard Index (Intersection over Union, IoU) - quantifies the similarity between the predicted segmentation and the ground truth, providing an assessment of the model's balance between precision and recall.

Precision - indicates the proportion of true positive predictions among all positive predictions made by the model, reflecting its accuracy in identifying the target regions.

Recall - measures the proportion of true positive predictions among all positive instances made by the model, reflecting its accuracy in identifying the target regions.

F1-Score - the harmonic mean of precision and recall, offering a single metric that balances both aspects of performance.

These metrics are critical for assessing the model's performance comprehensively and ensuring it meets the required standards for clinical applicability. The analysis of these metrics helps in identifying areas for improvement and ensuring that the model is robust and reliable for real-world use.

5.2. Interim Results

Metrics and Performance Evaluation

The interim evaluation of the model's performance highlights its effectiveness in segmenting key anatomical structures, namely the Right Ventricle (RV), Left Ventricle (LV), and Myocardium (MYO). The evaluation employs the Dice Similarity Coefficient (DSC) and Hausdorff Distance (HD) as the primary metrics, assessed across different phases of the heart cycle under two acquisition conditions: Standard Acquisition (SA) and Long Acquisition (LA).

Evaluating Metrics

The '*evaluate_metrics*' function computes the performance metrics for the segmentation results across the three anatomical classes (LV, MYO, RV). For each class, the Dice Similarity Coefficient (DSC) and Hausdorff Distance (HD) are calculated by comparing the model's predictions against the ground truth data.

Right Ventricle (RV) Segmentation:

axis	RV_ED_DC	RV_ED_HD	RV_ES_DC	RV_ES_HD	RV_DC	RV_HD
SA	0.88142806	12.3269659	0.84282098	12.4128145	0.86212452	12.3698902
LA	0.91547298	6.85492246	0.89597437	7.07060923	0.90572367	6.96276585

Figure 5. Dice score and Hausdorff distance for RV

The segmentation of the Right Ventricle (RV) demonstrates robust performance, particularly under the Long Acquisition (LA) condition. The Dice scores are consistently high, with value exceeding 0.90 in LA, indicating accurate segmentation of the RV. The Hausdorff Distance is notably lower under LA conditions, suggesting improved boundary precision and reduced segmentation errors. This performance underscores the model's capability to adapt effectively to different acquisition settings.

Left Ventricle (LV) Segmentation:

axis	LV_ED_DC	LV_ED_HD	LV_ES_DC	LV_ES_HD	LV_DC	LV_HD
SA	0.93189517	4.33023297	0.88833767	6.24810182	0.91011642	5.28916739
LA	0.95587161	4.41682648	0.9353753	4.12913846	0.94562345	4.27298247

Figure 6. Dice score and Hausdorff distance for LV

The segmentation of the Left Ventricle (LV) represents strong performance, with Dice scores consistently above 0.99 across both SA and LA conditions. The LV segmentation achieves the highest Dice scores among all structures, particularly in the Long Acquisition phase, where the score reaches 0.9559. The Hausdorff Distance values are low, especially in the End-Diastole (ED) and End-Systole (ES) phases, indicating precise and accurate boundary delineation. These results confirm the model's proficiency in accurately segmenting the LV.

Myocardium (MYO) Segmentation:

axis	MYO_ED_DC	MYO_ED_HD	MYO_ES_DC	MYO_ES_HD	MYO_DC	MYO_HD
SA	0.79767209	7.39355115	0.82506684	7.76569478	0.81136947	7.57962297
LA	0.84370868	6.26822463	0.87066239	5.20997468	0.85718554	5.73909966

Figure 7. Dice score and Hausdorff distance for MYO

The performance of Myocardium (MYO) segmentation is competent, though slightly lower than the observed for the RV and LV. The Dice scores, while satisfactory, show room for improvement, with values ranging from 0.7977 to 0.8708. The Hausdorff Distances, particularly under the Standard Acquisition (SA) condition, are higher, indicating a need for refinement in boundary precision. The Long Acquisition (LA) condition shows improved performance, suggesting that the model benefits from the extended acquisition phase. These results indicate that while MYO segmentation is effective, targeted optimizations could further enhance accuracy and boundary precision.

Visualization of Preprocessing and Segmentation Results

To illustrate the effectiveness of preprocessing and segmentation, the following images are provided:

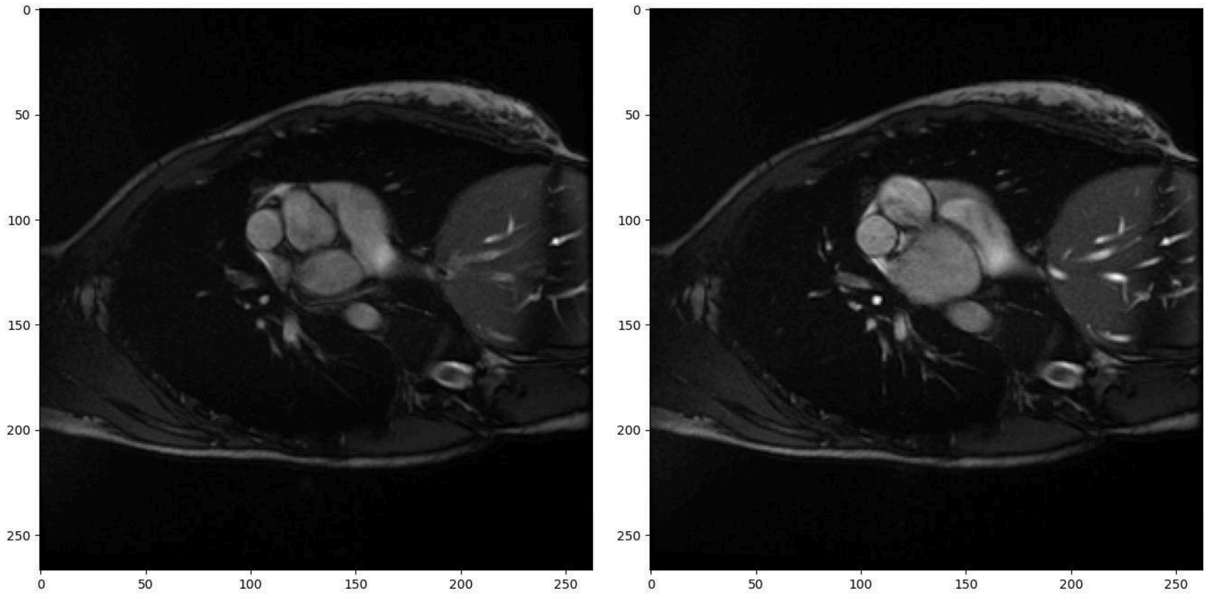

Figure 8. Short Axis ED and ES, After Soft Preprocessing

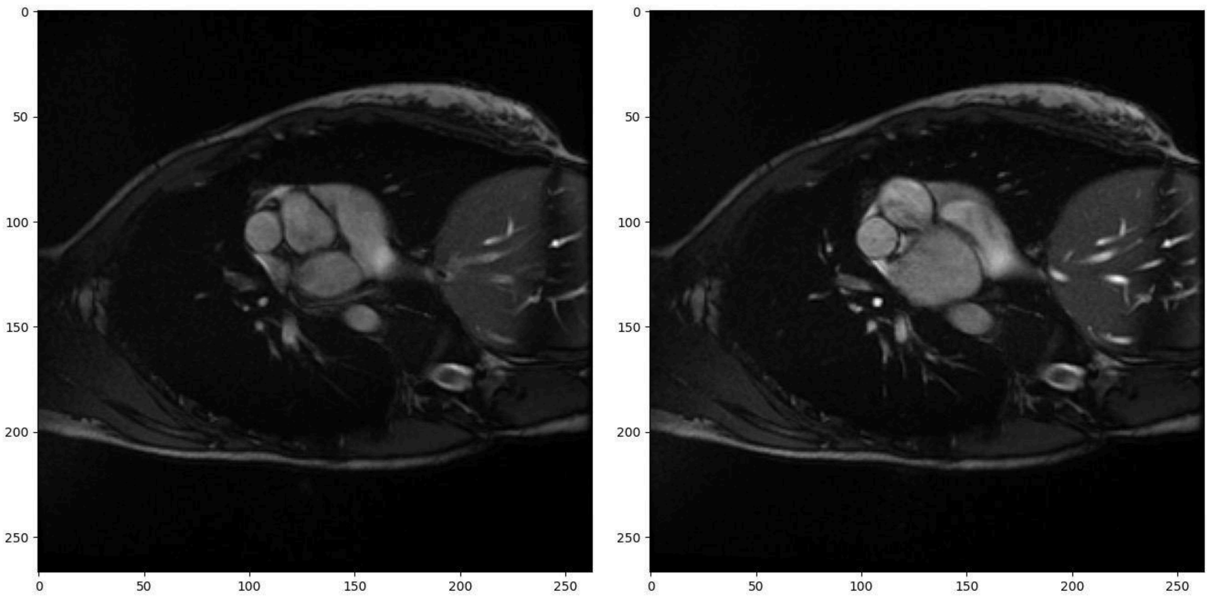

Figure 9. Short Axis ED and ES, After Validation Preprocessing

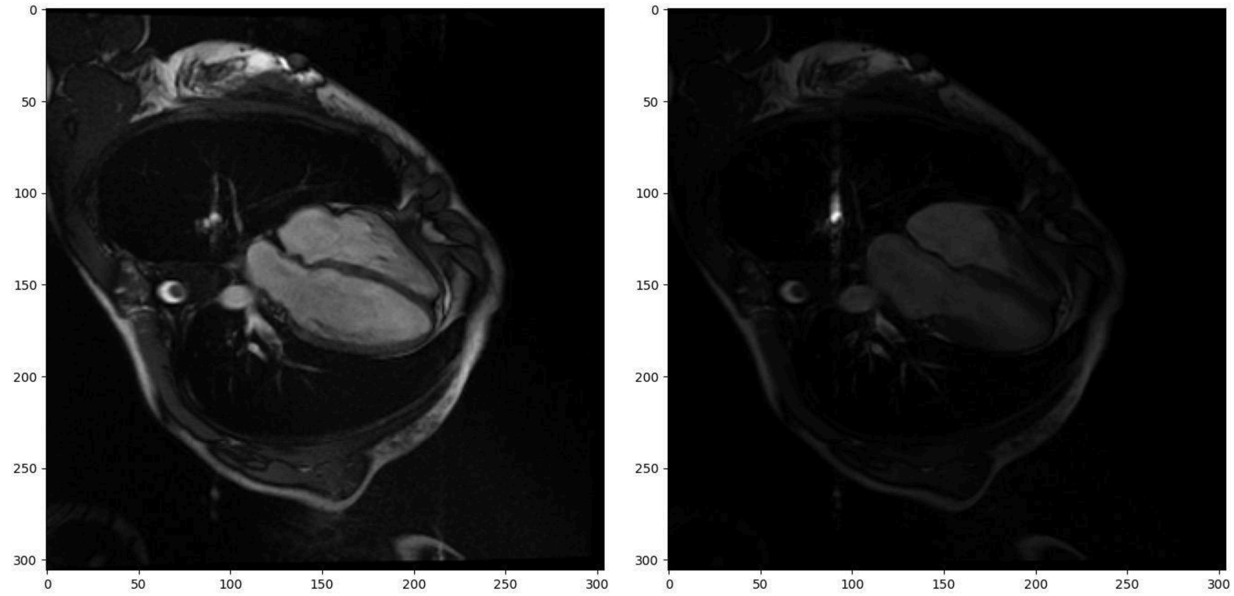

Figure 10. Long Axis ED and ES, After Soft Preprocessing

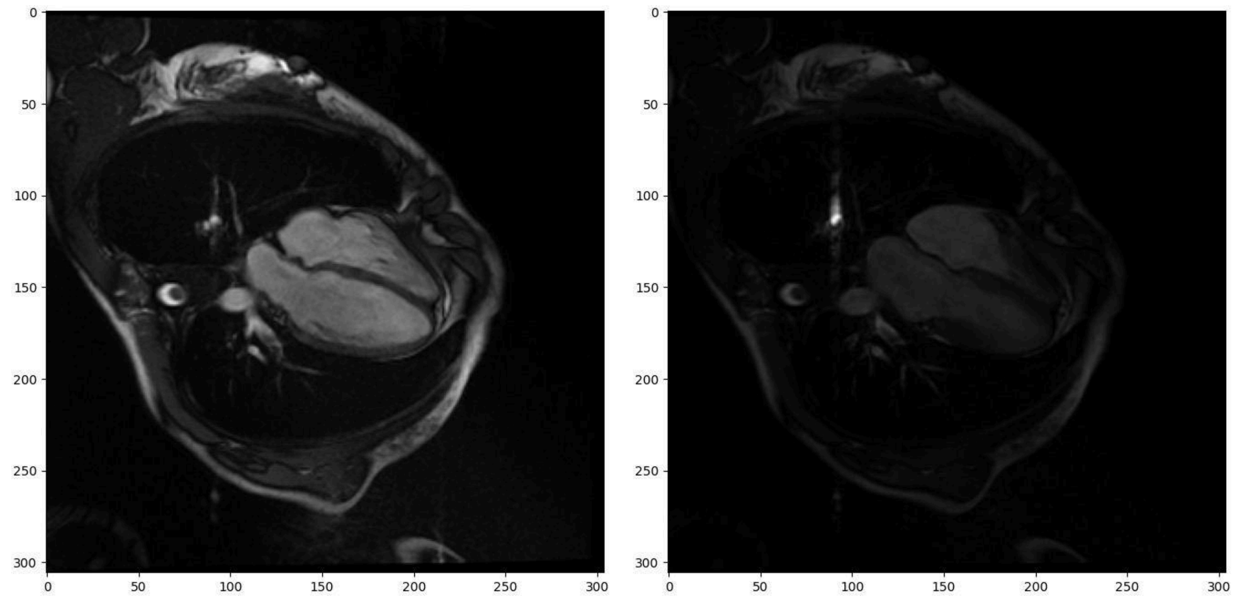

Figure 11. Long Axis ED and ES, After Validation Preprocessing

Result Interpretation

The model shows strong segmentation capabilities, especially for the Left Ventricle (LV), where DS is high and HD is low. This shows that the model is precise and accurate in segmentation of that part. The Right Ventricle (RV) also has shown good results, in particular the Long Anguisition (LA) is showing notable improvement in both accuracy and boundary precision. For the Myocardium (MYO), the model exhibits competent performance, although there is clear potential for further refinement to improve segmentation accuracy and reduce boundary errors.

Model Improvement and Optimization

To enhance the model's performance, particularly for the Myocardium (MYO) segmentation, several strategies can be implemented:

Data Augmentation

Increasing the diversity and quantity of training data using techniques such as rotation, scaling, and noise can improve the model’s ability to generalize and capture a broader range of anatomical variations, thereby enhancing overall accuracy.

Model Refinement

Exploring advanced model architectures or integrating additional layers and regularization techniques may improve segmentation precision, particularly in complex regions where the boundary detection is challenging. Thus the model will have more layers to detect and learn minor changes in cardiac MRI dataset.

Training Strategies

Employing advanced training methods, such as transfer learning or fine-tuning models pre-trained on similar tasks, could further boost performance, especially in more complex regions like the Myocardium. Fine tuning the parameters of the model might also increase the performance of this model.

Post-processing Enhancements

Refining post-processing steps, including improvements in image resizing, encoding, and smoothing techniques, could enhance the quality of segmentation results.

Additional Metrics

Incorporating more evaluation metrics, such as precision, recall, or Intersection over Union (IoU), would provide a more comprehensive assessment of the model’s performance, identifying specific areas where improvements may be needed.

Overall, the model demonstrates strong capabilities in segmenting the LV and RV, with promising performance for the MYO. Focused improvements and optimizations are likely to yield even better results across all anatomical regions, ensuring the model's readiness for clinical application.

5.3. Postprocessing and Evaluating Predictions

5.3.1 Post Processing Predictions

Post-processing represents an essential stage in the workflow of model-based image segmentation. It aimed at refining the initial predictions generated by the model to improve their accuracy and clinical relevance. This phase includes steps to correct and optimize the segmentation results generated during the model's inference stage. Assessment of post-processed data will help us to create or improve existing models.

1. **Purpose of Post-processing**

Enhancement of Prediction Accuracy

Postprocessing techniques are employed to adjust and fine-tune the model's initial predictions. These refinements correct any inaccuracies or artifacts in the initial output, resulting in a more precise and clinically relevant segmentation.

Preparation for Evaluation

Effective post-processing ensures that predictions are optimally prepared for comparison with the ground truth data. This step is crucial for an accurate assessment of the model's performance, ensuring that the predictions are in their best possible form before evaluation.

2. **Steps Involved in Post Processing**

Refinement and Correction - this involves smoothing edges, removing noise, and filling gaps within the segmented regions. Techniques such as morphological operations, spatial filtering, and boundary correction are commonly applied to enhance the segmentation's quality.

Resizing and Alignment - adjustments are made to the size and alignment of the segmented images to ensure that they match the dimensions and spatial orientation of the original ground truth images. This step is essential to ensure consistency and accuracy in the subsequent evaluation.

Validation of Results - the final post-processed predictions are validated against predefined quality standards before evaluation. This involves checking for consistency, accuracy, and alignment with predefined criteria to ensure the predictions are reliable.

3. **Comparison with Ground Truth**

Purpose of Comparison - after post processing, the refined predictions are compared to the ground truth annotations to assess how well the model's output aligns with the actual data. This comparison is critical for understanding the model's effectiveness and identifying areas for further improvements.

Evaluation Metrics - metrics such as Dice Similarity Coefficient (DSC) and Hausdorff Distance (HD) are employed to quantify the agreement between the post-processed predictions and the ground truth. These metrics provide valuable insights into the precision and accuracy of the segmentation.

4. **Impact on Model Performance**

Assessment of Quality - the comparison of post-processed predictions with the ground truth enables a thorough evaluation of the model's performance. This step is essential to determine whether the model's predictions are sufficiently accurate and reliable for practical use.

Identification of Improvement Areas - by analyzing discrepancies between the predictions and ground truth, specific areas where the model may require further training or adjustment can be identified. This feedback is crucial for iterative model improvement and optimization.

5.3.2 Inference Stage – Prediction vs. Ground Truth

To evaluate the model's performance, we present visual comparisons between the model's predictions and ground truth data across different phases and axes. These comparisons provide clear insights into the accuracy and reliability of the model's segmentation results.

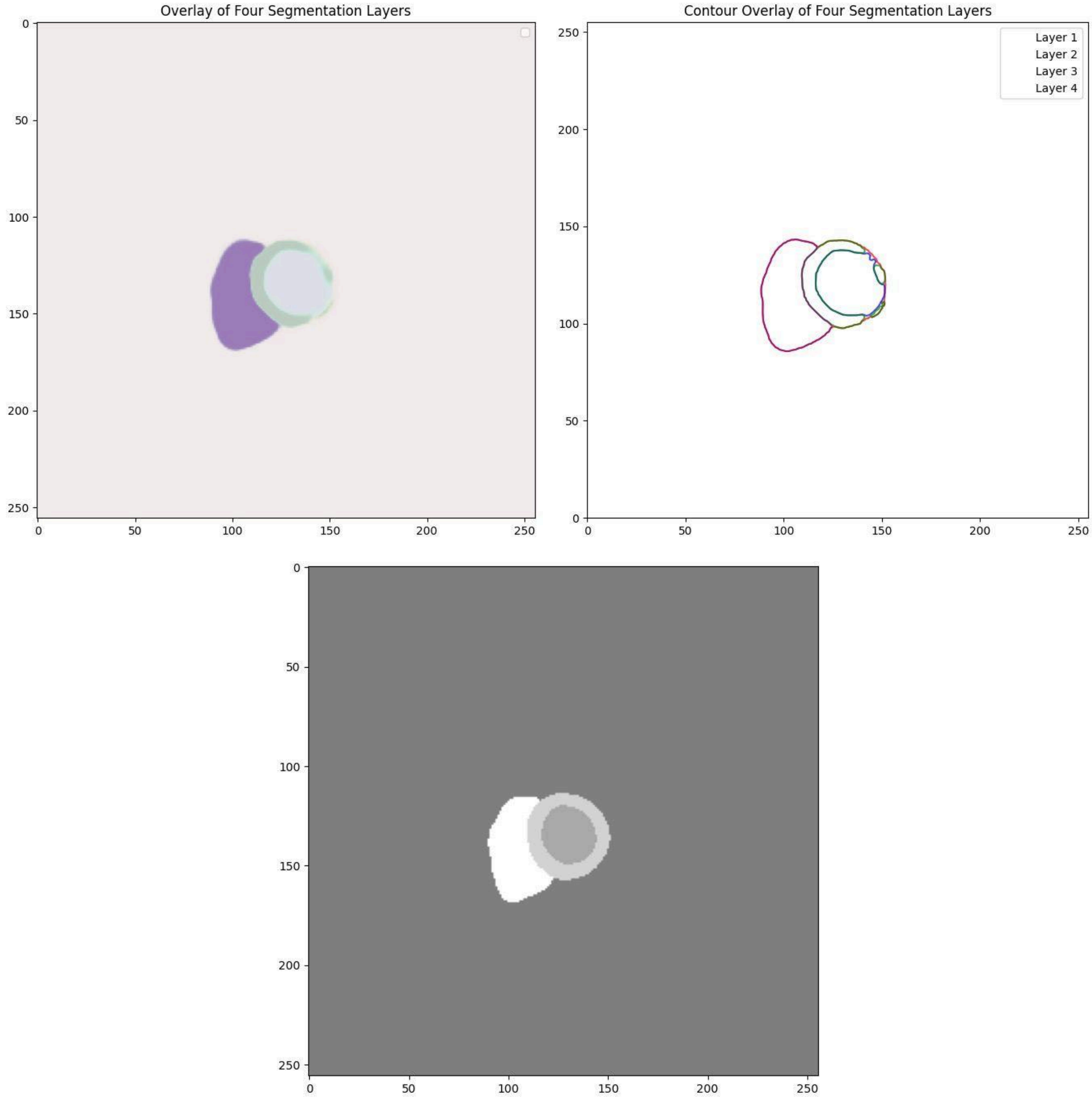

Figure 12. Short Axis, End Systole (ES)
Inference Stage - Prediction (Upper Two Panels), Ground Truth (Lower Panel)

The upper two panels display the model's predictions for the Short Axis during the End-Systole (ES) phase, while the lower panel shows the corresponding ground truth. This visual comparison allows us to evaluate how closely the model’s predictions align with the actual segmentation.

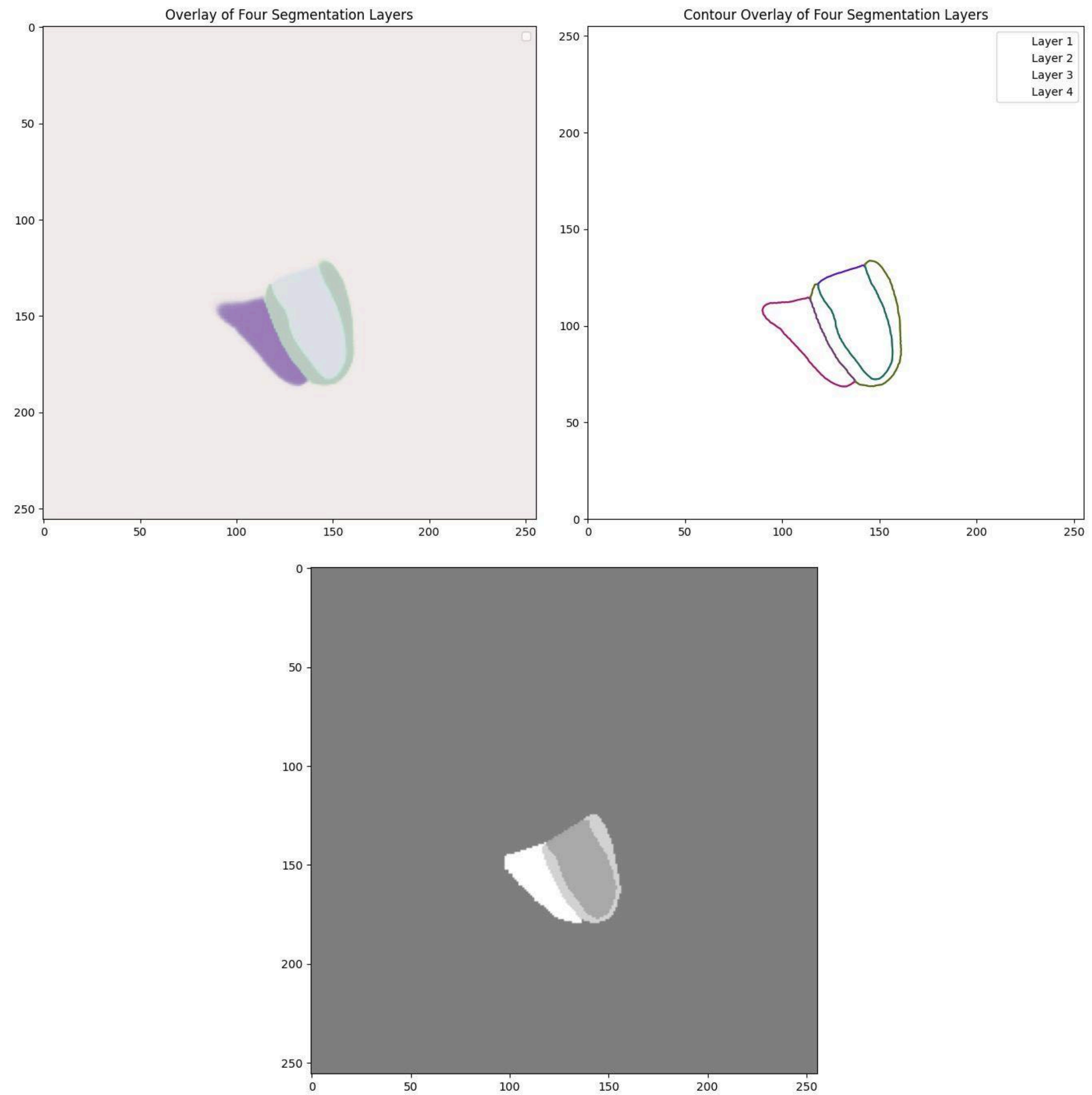

Figure 13. Long Axis, End Systole (ES)
Inference Stage - Prediction (Upper Two Panels), Ground Truth (Lower Panel)

The visual comparison includes the upper two panels displaying the model's predictions for the Long Axis during the End-Systole (ES) phase. At the same time, the lower panel shows the corresponding ground truth segmentation. This comparison is crucial for evaluating how closely the predictions match the annotated data.

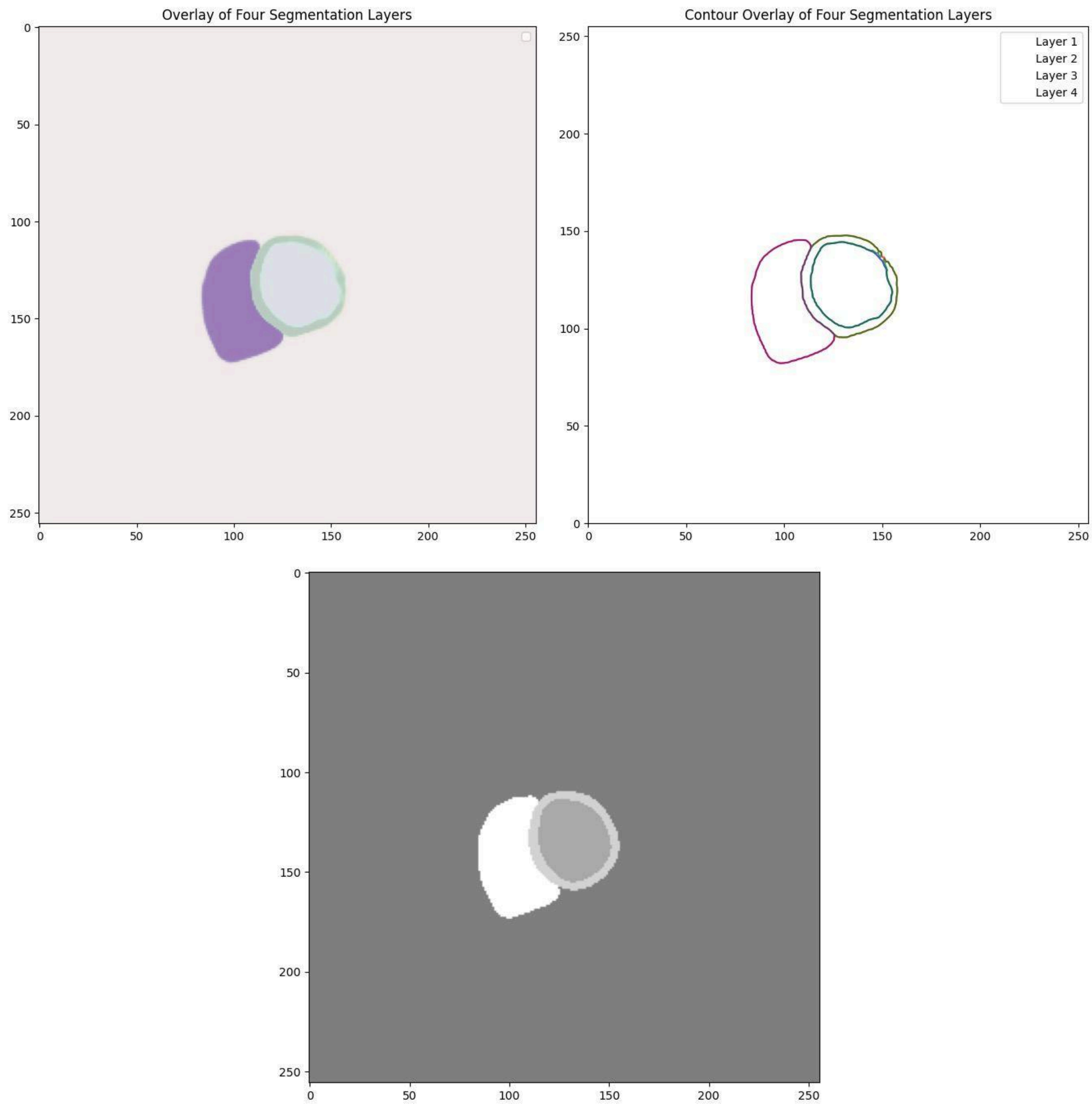

Figure 14. Short Axis, End Diastole (ED)
Inference Stage - Prediction (Upper Two Panels), Ground Truth (Lower Panel)

This image compares the model's predictions with the ground truth for the Short Axis during the End-Diastole (ED) phase. The upper part of the image shows the predicted segmentations, while the lower part presents the ground truth. The visual assessment is vital for determining the accuracy and reliability of the model's segmentation results in this phase.

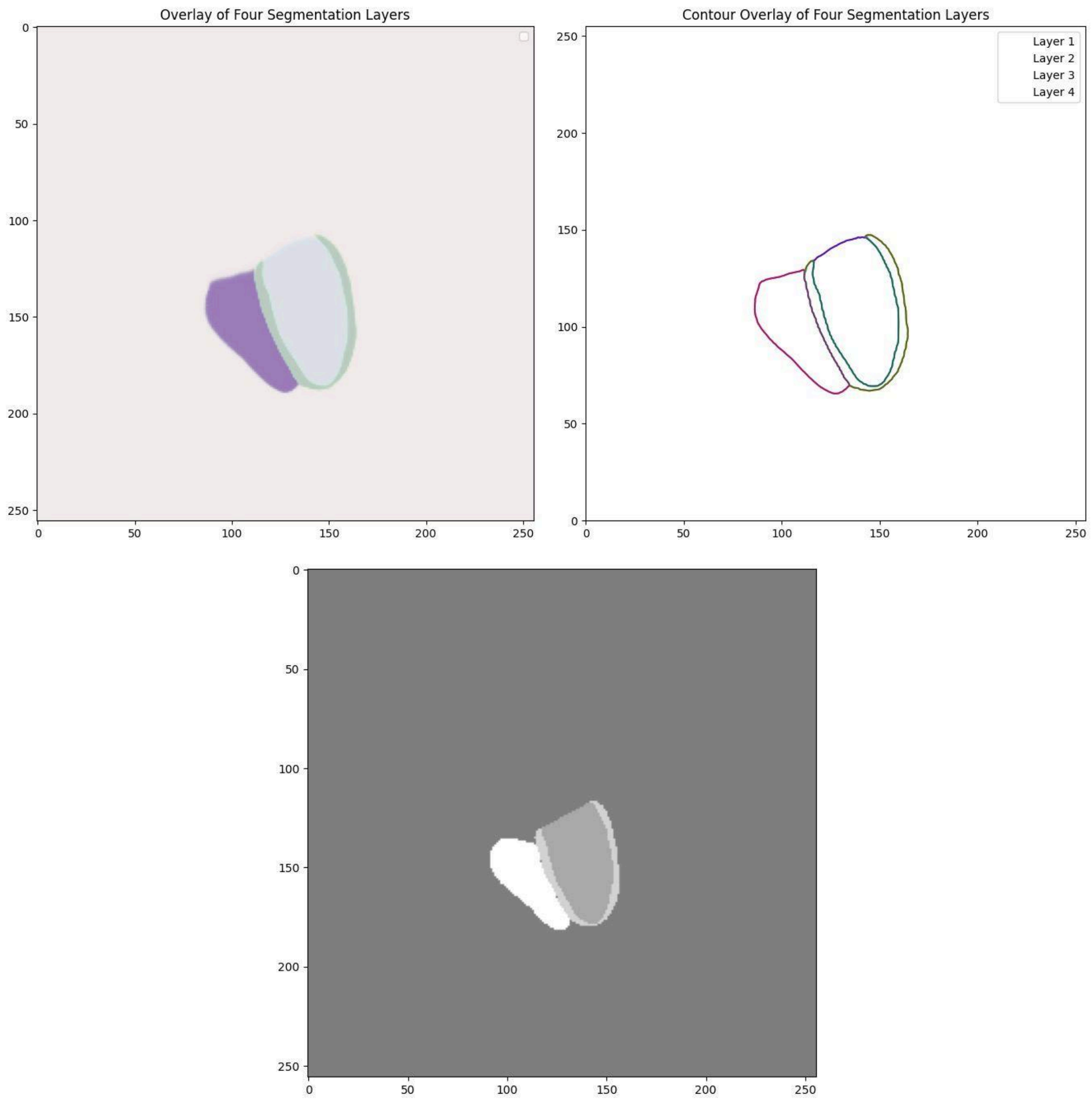

Figure 15. Long Axis, End Diastole (ED)
Inference Stage - Prediction (Upper Two Panels), Ground Truth (Lower Panel)

This image includes the model's predictions for the Long Axis during End-Diastole (ED) phase in the upper two panels, with the ground truth displayed in the lower panel. This visual comparison is vital for assessing the model's effectiveness in accurately segmenting the Long Axis during the ED phase.

6. Updated Project Implementation Plan

6.1. Project Progress Plan

6.1.1. Completed Work

Step 1: Literature Review and Research	
Objective:	Acquire a comprehensive understanding of the current state of cardiac MRI segmentation. Covering methodologies, recent advancements, and existing challenges.
Accomplishments:	● Conducted an extensive review of relevant academic literature and research publications in cardiac MRI segmentation. ● Studied major insights from the reviewed literature, showing great advancements in the field of deep learning models like U-Net for segmentation models. ● Identified crucial issues in existing approaches, like difficulties with multi-center datasets and domain variability, which informs about the necessity to strategize a project to overcome such obstacles.
Step 2: Dataset Exploration and Analysis	
Objective:	Understand the structure and characteristics of the multi-center and multi-vendor datasets and augment the dataset to enhance model robustness.
Accomplishments:	● Conducted a thorough examination of the datasets, focusing on structural characteristics and variations arising from different clinical centers and scanner vendors. ● Implemented data augmentation techniques, including random cropping, spatial and mirror transformations, probabilistic random transformations, padding and

	cropping, one-hot encoding, tensor conversion, and downsampling (increases dataset variability), improving the ability of the model to generalize across diverse imaging conditions.
Step 3: Data Preprocessing and Visualization	
Objective:	Prepare and standardize the dataset to ensure uniformity, facilitating effective model training and evaluation across all images.
Accomplishments:	● Developed and executed the *'preprocess_image'* function, which standardizes the dataset through different processes such as cropping, transposition, rescaling, and normalization. ● Ensured consistency in image dimensions, scale, and orientation across the entire dataset, thereby enhancing the reliability of the model training process. ● Leverage visualization tools to examine and validate the dimensions of cardiac MRI images and corresponding segmentation labels, confirming the efficiency of the preprocessing steps.
Step 4: Model Development and Training	
Objective:	Develop a robust segmentation model tailored for cardiac MRI and train it effectively on the preprocessed dataset.
Accomplishments:	● Designed and implemented a new segmentation model based on the U-net architecture, utilizing the Baseline_2 class and Generic_UNet framework. ● Structured the model with specialized submodules to handle short-axis and long-axis segmentation tasks. ● Integrated appropriate activation functions (Leaky ReLU) and normalization techniques (BatchNorm) to enhance model performance.

	● Trained the model using the preprocessed dataset, optimizing it through Stochastic Gradient Descent (SGD) with momentum. ● Established checkpointing and fallback mechanisms to ensure the robustness of the training process and mitigate overfitting. ● Performed an initial evaluation of the model's performance, analyzing key metrics such as Dice Similarity Coefficient (DSC) and Hausdorff Distance (HD), and conducted post-processing to refine the segmentation results.
Step 5: Documentation and Presentation	
Objective:	Document the progress of the project, methodologies, and preliminary findings in a clear and organized manner.
Accomplishments:	● Completed and submitted an intermediate report that provides a detailed account of the progress, methodologies applied, and preliminary results achieved in the early stages of the project.

6.1.2. Future Work

Step 1: Domain Adaptation for Generalizability	
Objective:	Improve the model's robustness and ability to generalize across diverse datasets by applying domain adaptation techniques.
Upcoming Tasks:	● Explore and implement domain adaptation strategies to address the variability inherent in multi-center and multi-vendor datasets. ● Apply Domain Adversarial Training of Neural Networks (DANN) to mitigate domain-specific biases and enhance the model's performance across different data sources.

Step 2: Model Evaluation and Analysis	
Objective:	Conduct a comprehensive evaluation of the model's performance using both quantitative metrics and qualitative analysis.
Upcoming Tasks:	• Perform an extensive quantitative assessment using evaluation metrics such as Dice Similarity Coefficient (DSC), Hausdorff Distance (HD), Precision, Recall, and F1-Score. • Conduct qualitative analysis by visually inspecting the segmentation results and comparing them with ground truth to validate the model's accuracy and reliability.
Step 3: Application Integration	
Objective:	Develop a user-friendly application interface and integrate the trained model into it.
Upcoming Tasks:	• Design and develop a user interface that facilitates intuitive interaction with the application. • Integrate the trained segmentation model into the application, ensuring it meets the functionality and real-time performance criteria required for clinical use.
Step 4: Application Testing and Debugging	
Objective:	Verify the application's performance and reliability through rigorous testing and debugging procedures.

Upcoming Tasks:	• Conduct comprehensive testing procedures to verify the application's performance under various conditions. • Implement necessary adjustments based on testing outcomes to ensure the application's robustness and efficiency. • Perform performance testing to evaluate the application's speed and reliability in processing real-time data.
Step 5: Final Documentation and Presentation	
Objective:	Summarize the entire project and present the outcomes and conclusions.
Upcoming Tasks:	• Write and submit the final project report detailing the methodologies employed, results obtained, and conclusions drawn throughout the project. • Prepare and deliver a final presentation that showcases the developed application, the model's performance, and the impact of domain adaptation and other strategies on the overall project success.

6.2. Updated Schedule

Division	Development Schedule																	
	May		June				July				August				September			
	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1. **Literature Review and Research**																		

● Conduct an extensive review of research papers focused on cardiac MRI segmentation. ● Summarize key findings from the literature, emphasizing the main advancements in the field. ● Identify and analyze the limitations inherent in current methodologies.																		
2. **Dataset Exploration and Analysis** ● Perform a comprehensive exploration of the multi-center and multi-vendor datasets, examining their structural characteristics and content. ● Apply data augmentation techniques to enhance the dataset, thereby improving the robustness and generalizability of the model.																		
3. **Data Preprocessing and Visualization** ● Preprocessing: o Design and implement the *'preprocess_image'* function to facilitate standardization. o Execute image preprocessing steps including cropping, transposition, rescaling, and normalization. o Ensure consistency in image size, scale, and orientation across the dataset.																		

● Visualization: o Employ visualization tools to explore and analyze the dimensions of cardiac MRI and the corresponding annotated segmentation labels.																
4. **Model Development and Training** ● Develop and implement a new segmentation model based on a U-net architecture. ○ Structure the model using the Baseline_2 class and Generic_UNet framework. ○ Incorporate distinct submodules for short-axis and long-axis segmentation. ○ Apply appropriate activation functions (e.g., Leaky ReLU) and normalization techniques (e.g., BatchNorm). ● Train the proposed model on the preprocessed dataset. ○ Implement checkpoint loading and fallback mechanisms to ensure robust training. ○ Execute supervised training loops for short-axis and long-axis submodules, managing validation,																

checkpointing, and learning rate adjustment. ○ Utilize Stochastic Gradient Descent (SGD) with momentum for optimization. ● Evaluate the model's performance and analyze initial results and observations. ○ Implement the inference process and handle model evaluation during testing. ○ Perform post-processing of results and evaluate model performance using relevant metrics such as Dice Similarity Coefficient (DSC) and Hausdorff Distance (HD).																		
5. **Documentation and Presentation** ● Complete and submit an intermediate report detailing progress, methodologies employed, and preliminary findings.																		
6. **Domain Adaptation for Generalizability** ● Investigate and apply domain adaptation techniques to enhance the model's robustness and generalizability.																		

● Implement learning methods to improve the model's performance in varied contexts and environments.																		
7. **Model Evaluation and Analysis** ● Conduct a quantitative assessment of the model's performance using various evaluation metrics. ● Perform qualitative analysis through visual inspection of segmentation results to validate model accuracy.																		
8. **Application Integration** ● Develop a user interface for the application, encompassing both front-end and back-end components to facilitate user interaction. ● Integrate the trained model into the application, ensuring it meets functionality and real-time performance requirements.																		
9. **Application Testing and Debugging** ● Execute comprehensive testing procedures to verify the application's performance and reliability. ● Implement necessary adjustments based on testing outcomes.																		

● Conduct performance testing to ensure the application operates efficiently under various conditions.																		
10. **Documentation and Presentation** ● Write and submit the final report, summarizing the entire project, including methodologies, results, and conclusions. ● Prepare and deliver a presentation demonstrating the developed application and the results of the model evaluations.																		

	Completed
	In Progress
	Planned

7. References

Campello, V. M., Gkontra, P., Izquierdo, C., Martin-Isla, C., Sojoudi, A., Full, P. M., Maier-Hein, K., Zhang, Y., He, Z., Ma, J., Parreno, M., Albiol, A., Kong, F., Shadden, S. C., Acero, J. C., Sundaresan, V., Saber, M., Elattar, M., Li, H., … Lekadir, K. (2021). Multi-Centre, multi-vendor and multi-disease cardiac segmentation: The M&MS Challenge. *IEEE Transactions on Medical Imaging*, *40*(12), 3543–3554. https://doi.org/10.1109/tmi.2021.3090082

Corral Acero, J., Sundaresan, V., Dinsdale, N., Grau, V., & Jenkinson, M. (2021). A 2-step deep learning method with domain adaptation for multi-centre, multi-vendor and multi-disease cardiac magnetic resonance segmentation. *Lecture Notes in Computer Science*, *12592*, 196–207. https://doi.org/10.1007/978-3-030-68107-4_20

Glocker, B., Robinson, R., Castro, D. C., Dou, Q., & Konukoglu, E. (2019, October 10). *Machine learning with multi-site imaging data: An empirical study on the impact of scanner effects*. arXiv.org. https://arxiv.org/abs/1910.04597

Li, L., Ding, W., Huang, L., & Zhuang, X. (2022). Right ventricular segmentation from short- and long-axis mris via information transition. *Lecture Notes in Computer Science*, 259–267. https://doi.org/10.1007/978-3-030-93722-5_28

Martín-Isla, C., Campello, V. M., Izquierdo, C., Kushibar, K., Sendra-Balcells, C., Gkontra, P., Sojoudi, A., Fulton, M. J., Arega, T. W., Punithakumar, K., Li, L., Sun, X., Al Khalil, Y., Liu, D., Jabbar, S., Queirós, S., Galati, F., Mazher, M., Gao, Z., … Lekadir, K. (2023). Deep learning segmentation of the right ventricle in cardiac MRI: The M&MS Challenge. *IEEE Journal of Biomedical and Health Informatics*, *27*(7), 3302–3313. https://doi.org/10.1109/jbhi.2023.3267857

Perez, L., & Wang, J. (2017, December 13). *The effectiveness of data augmentation in image classification using Deep Learning*. arXiv.org. https://arxiv.org/abs/1712.04621

Poudel, R. P., Lamata, P., & Montana, G. (2017). Recurrent fully convolutional neural networks for multi-slice MRI cardiac segmentation. *Lecture Notes in Computer Science*, 83–94. https://doi.org/10.1007/978-3-319-52280-7_8

Tran, P. V. (2017, April 27). *A fully convolutional neural network for cardiac segmentation in short-axis MRI*. arXiv.org. https://arxiv.org/abs/1604.00494