

Deep-CodecG*: A Generalized Deep Autoencoder for Robust Segmentation of Left Atrium in Cardiac MRIs

AKHILESH RAWAT¹
RAJEEV KUMAR²

Data to Knowledge (D2K) Lab
School of Computer & Systems Sciences
Jawaharlal Nehru University, New Delhi, 110 067, India
¹akhile62_scs@jnu.ac.in, ORCID : 0000-0001-5138-4540
²rajeevkumar.cse@gmail.com, ORCID : 0000-0003-0233-6563

Abstract. The left atrium receives oxygenated blood from pulmonary veins and is a vital organ concerning congestive heart failure. Several deep learning-based architectures and learning methodologies have been proposed for left atrium semantic segmentation. These studies have shown good performance in learning *known* datasets. However, generalization remains challenging. In this work, we propose a deep auto-encoder architecture with generalization ability which we call *Deep-CodecG**. The proposed model utilizes a CNN-based auto-encoder, replacing the standard convolution with a two-convolution layer block. This proposed model is generalization enabled with a proper parameterization for (near-) optimal performance. The proposed Deep-CodecG* improves performance on *unseen* test data, with a dice score of 0.95, which is 6.3% higher than a standard auto-encoder, and 4% higher than V-Net. The proposed model gave superior sensitivity, specificity, Jaccard, structural similarity values, and Hausdorff distance, indicating improvement over an autoencoder with similar two-convolution layer blocks. Though these quantitative improvements seem marginal, they are shown to impact visualizing the segmented image significantly. The segmented left atrium closely matches the ground-truth data and yields additional info. Thus, the proposed Deep-CodecG* architecture for left atrium segmentation exhibits well-generalized and robust performance over various image datasets.

Keywords: Deep Learning, Auto-encoder, CNN, Left Atrium, Generalization

(Received April 1st, 2022 / Accepted May 1st, 2022)

1 Introduction

In medical imaging, CNN-based Deep Learning (DL) has been extensively used [17, 21]. DL models extract essential features from the images generated by the various medical modalities. Medical modalities are indispensable for diagnosing and treating a broad spectrum of diseases and conditions and play a crucial role in medical research and healthcare. Numerous researchers learn supervised, semi-supervised, and unsupervised DL architectures by analyzing cardiac MRI and CT images [5]. Using cardiac images, DL architectures segment the left atrium, left ventricle, right ventri-

cle, etc.

The left atrium receives oxygenated blood returning to the heart from the pulmonary vessels. During elevated blood pressure, the left atrium grows in size, an occurrence called left atrial enlargement [22]. Radiologists manually determine the left atrium size using cardiac MRI, CT imaging, and other modalities [30]. This helps physicians determine the risk of a heart attack in hypertensive patients. Left atrial enlargement may also indicate heart failure. Manual segmentation [6] is a laborious process. In medical imaging, Machine/Deep Learning (ML/DL) operates automatically

and reduces manual labor [27]. Numerous researchers used DL architectures to segment the left atrium automatically.

For medical image segmentation, DL techniques implement encoder-decoder using CNNs [13]. For left atrium segmentation, researchers rely significantly on U-Net architectures. Numerous studies have created various enhanced variants of a U-Net [25] architecture utilizing CNNs [27, 14, 26]. These architectures employed skip connections between the encoder and decoder to mitigate the vanishing gradient problem.

This paper proposes a CNN-based deep learning model called Deep-CodecG*, wherein deep stands for deep learning, codec for autoencoder, and G* for generalization with appropriate parameterization specific to the model and data. The proposed model employs additional convolutional layers to comprehensively understand the autonomously extracted features from the input by the encoder-decoder. We employ various normalization and regularization techniques with appropriate parameterization for their effective use across the datasets. We combine two 2D Convolutional layers with the same mask and generalization techniques for the auto-encoder in a single block, which we call 2CLB (2 Convolutional Layers Block).

For experiments, we utilize a publicly available 3D cardiac MRI dataset and create several 2D slices from the dataset for empirical experiments. We also configure a standard CNN-based auto-encoder with 2CLB, as in the proposed Deep-CodecG*; thus, we make another model based on a U-Net architecture configured with 2CLB for a fairer comparison with the proposed model; we call this *Config. AE*. We also consider a V-Net model for comparison. For a fairer comparison, we use state-of-the-art performance measures. Finally, we evaluate the proposed Deep-CodecG* model with two others – a standard *Config. AE* and the V-Net and assess their efficacy on *unseen* data not used during training and validation.

The proposed model segments the left atrium in cardiac MRI more precisely than the *Config. AE* and the V-Net. Thus, the primary research contribution of the work presented in this paper is in demonstrating that using generalization techniques with proper parameterization and fine-tuned optimization improves the proposed architecture’s performance on various unseen image datasets. Thus, the proposed Deep-CodecG* architecture for the left atrium in cardiac MRIs is shown to be robust. The rest of the paper is organized as follows. Section 2 reviews the related work. Section 3 demonstrates the methodological formulation of the proposed model. Section 4 discusses the datasets, implementa-

tion details, empirical results, and analysis. Finally, we conclude the work in Section 5.

Table 1: Abbreviations

Abbreviation	Description
LA	Left Atrium
LAS	Left Atrium Segmentation
2CLB	Two Convolutional Layer Block
Deep-CodecG*	Deep Auto-Codec with Generalization (Proposed)
Config. AE	Auto Encoder with 2CLB

2 Related Work

2.1 Left Atrium in Cardiac MRI

Atria (singularly known as atrium) are two of a heart’s four chambers responsible for facilitating circulation by rhythmically contracting and relaxing their walls, pushing blood that is returned to them into the ventricles via major venous vessels. The left atrium, located on the upper left-hand side, is one of the four chambers in the heart; the lower chambers are known as ventricles. The left circumflex coronary artery governs the blood supply to the left atrium, which is drained by the oblique vein of the left atrium arising during embryonic development from the left side of the fetal superior vena cava.

Several heart problems can cause a left atrium to swell, resulting in left atrial enlargement, associated with conditions including atrial fibrillation and heart failure. While health conditions associated with left atrial enlargement may not always cause symptoms, some people may experience breathlessness or shortness of breath, fatigue, and other symptoms associated with other heart conditions.

The size of the left atrium varies from person to person and may change as a person ages. High blood pressure is one of the conditions that may cause left atrial enlargement. Diagnosing left atrial enlargement can help physicians determine the risk of a heart attack in people with high blood pressure. Determining the left atrium dimensions is uphill from the cardiac MRI and CT scans. In this domain, a DL learning model acts as an automated segmentation tool that can assist radiologists in estimating the size and other pertinent information.

2.2 State-of-the-Art Technology

Convolutional Neural Networks (CNNs) are a popular tool in Deep Learning (DL) for analyzing and processing medical imaging data for various human organs. Thus, CNNs have been widely used to diagnose, detect, and segment abnormalities in various human body

parts. For instance, CNNs have been extensively applied for brain MRI tumor segmentation [19], liver tumor segmentation [9], breast mass segmentation [3], and left atrium [12]. In addition, CNNs have been utilized to classify different types of cancerous body cells [18]. Several pieces of research use DL techniques for left atrium segmentation. However, they have not been effective as the LA segmentation results need to be robust for estimating the LA enlargement precisely.

Liu et al. [16] proposed a Fully Convolutional Network (FCN) that utilized 3D CT scans for left atrium segmentation. Their experiments were conducted on the STACOM13 dataset, which was modest in size. To create the FCN, the authors used upsampling and deconvolution. Xiong et al. [30] organized a challenge for segmenting the left atrium using 3D LGE-MRIs. They found that CNNs used in a double-sequential manner outperformed conventional methods and single CNNs. Kazi et al. [13] used a 3D U-Net architecture for left atrium segmentation and trained the network with CT volumes. They used full volumes of CT scans and small patches of CT volumes to train their network. Kausar et al. [12] proposed a 3D shallow DNN for left atrium segmentation. They utilized both brief and extended skip connections with residual learning properties and utilized 20 MRI cases from the LASC dataset. They believed a shallow network was better than a data-hungry deep network. Degel et al. [7] used a V-Net architecture and trained it with 3D ultrasound images for left atrium segmentation. They also handled domain invariance with their proposed network. Du et al. [8] proposed a unified DL model for left atrium segmentation and visualization. They used 2D convolution with their network.

Yu et al. [31] employed a semi-supervised uncertainty-aware model for LA segmentation. Their framework depended on the student and teacher model. The student-teacher model is advantageous, When the teacher network is too complex or resource-intensive to be deployed in real-world applications or when the available data for training the student network is limited. The soft targets from the instructor network provide additional information that assists the student network in more effectively learning from the available data and generalizing to new data. Li et al. [15] also used a student-teacher model with some enhancement for semi-supervised 3D LA segmentation. For LA segmentation, a few studies employed non-DL approaches such as the level set method [20] and reverse structural remodeling [23].

The above models effectively learned the image datasets and segmented the LAs. However, their ap-

propriateness to *newer* image datasets is mostly unreported. Therefore, in this work, we design and propose an architecture that takes care of generalization ability *explicitly*.

3 The Proposed Deep Learning Architecture

This section details the design and implementation of the proposed architecture, Deep-CodecG*, and the workflow for segmentation of the left atrium (LA) from cardiac MRIs using this proposed architecture. First, we explain the block diagram of the proposed model; next, we explain the preliminaries with generalization techniques; last, we describe the workflow of the Deep-CodecG* with detailed architecture.

3.1 The Deep-CodecG* Architecture

Many researchers have used DL-driven autoencoder architectures, such as U-Net [28], V-Net [7], Seg-Net [2], etc., for medical image segmentation. These architectures preserve more information during the down-sampling process by leveraging skip connections. In this work, we propose a model that enhances the performance of such contemporary DL models by applying normalization techniques on skip connections and other generalization techniques in the whole network for robust left atrium segmentation.

Figure 1 shows the block schematics of the proposed architecture. The proposed DL-driven model named, Deep-CodecG* is a CNN-based autoencoder with three distinctions; (i) each convolution block is a two-convolution layer block (2CLB), wherein each 2CLB unit is equipped with normalization and dropout units (Figure 2), (ii) each block of an autoencoder/decoder is configured with 2CLB (Figure 1), and (iii) the CodecG* is generalization enabled with a proper parameterization for (near-) optimal performance.

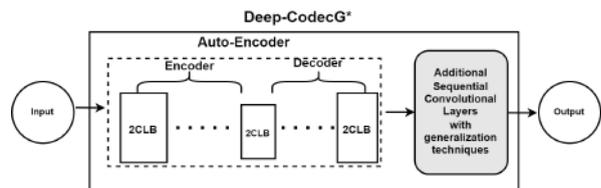


Figure 1: Block Diagram of the Proposed Deep-CodecG* Model

The Deep-CodecG* learns with the input of cardiac MRI and gives an output of a left atrium segmented image corresponding to the input image. The proposed model uses ten 2CLBs and one 2D convolution layer at the last phase; 21 convolutional layers are used in this architecture, as shown in Figure 4. We use the ReLU

activation function with 2CLBs and the sigmoid activation function after the last layer of 2D convolution.

Activation Function
Normalization
2DConvolution (kernel size 3 X 3, stride 1, padding 1)
Dropout 0.2%
Normalization
2DConvolution (kernel size 3 X 3, stride 1, padding 1)

Figure 2: Two Convolutional Layer Block (2CLB)

3.1.1 Two Convolution Layer Block (2CLB)

Instead of changing the convolution mask each time, we repeat the same mask twice. This has the advantage of reduced computational cost. In this work, we use 2CLB for every convolution operation. A 2CLB consisting of two 2D convolution layers, each with 3×3 filter size, stride size 1, and padding size 1, is shown in Figure 2. The number of filters or kernels increases during the encoder phase and decreases during the decoder phase; it depends on the channel size. In this block, the first 2D-convolution layer is followed by a normalization layer, a dropout layer, and a second 2D convolution layer, followed by a second normalization layer. Last, we use an activation function that helps to capture complex patterns in cardiac images.

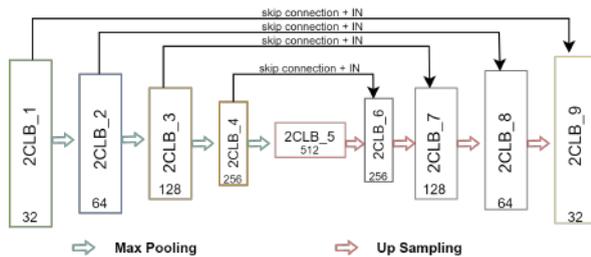


Figure 3: Autoencoder configured with 2CLB.

3.1.2 Autoencoders with 2CLB

Each encoder and decoder unit in the CodecG* (Figure 4) is configured with 2CLBs, as shown in Figure 2. For a fairer comparison of the performance of Deep-CodecG*, we use 2CLB for the customized autoencoder with proper parameterization of 2CLB, as in Figure 3; we call such customized autoencoder *Config. Auto-Encoder*. In the Config. AE, after the encoder-decoder, we add only one convolution layer followed by a Sigmoid activation function as used in U-Net architecture [28] to segment the LA. To determine the

empirical effect of the Deep-CodecG*, we compare the segmentation results of the proposed model with such Config. AE and state-of-the-art V-Net with 2D convolutional settings. All analyses of the empirical results of the three models are shown in Section 4.

3.1.3 Learning with Generalization

We employ various generalization techniques to accomplish normalization and regularization techniques in the Deep-CodecG* model. Normalization improves the flow of gradients during backpropagation [10, 29]. DL models update their parameters based on the error between the predicted and the ground truth masks. By standardizing the input data, normalization ensures that the gradients are neither too large nor too small; this helps to prevent the models from getting stuck in local optima. We use 20% dropout to prevent overfitting, Batch Normalization [10], and Instance normalization [29] techniques in the proposed model at various points. We apply these normalization methods to the parameters/weights of feature maps. In instance normalization (IN), we compute every channel's mean (μ) and standard deviation (σ) and normalize the particular channel weights concerning the computed μ and σ . In BN, we computed the μ and σ of identical mini-batches and normalized these weights with their computed μ and σ . After applying the normalization, we apply the parameterization again. Thus, we use these generalization techniques with proper parameterization and fine-tuning for (near-) optimal performance.

3.2 Workflow in Deep-CodecG* for LAS

The detailed workflow of LA segmentation using the above-described Deep-CodecG* architecture is depicted in Figure 4. As shown at the bottom of this figure, this has nine 2CLBs for making an encoder-decoder using skip connections. We use instance normalization (IN) with skip connection information; it helps improve the network's stability and convergence during training. Starting nine 2CLBs in the proposed model worked as auto-encoder (AE) (Figure 3). AE helps in extracting and learning the important features from the input. Four 2CLBs work as encoders; the first 2CLB starts with a channel size of 32. We double the channel size in the next 2CLB in the encoding phase. After each 2CLB in the encoder, we apply the max-pooling operation, which reduces the image size during the encoder phase and introduces some degree of translation invariance to the network. The middle 2CLB_5 is used as a bridge with a channel size 512. Next, four 2CLBs work as a decoder. We apply the up-sampling

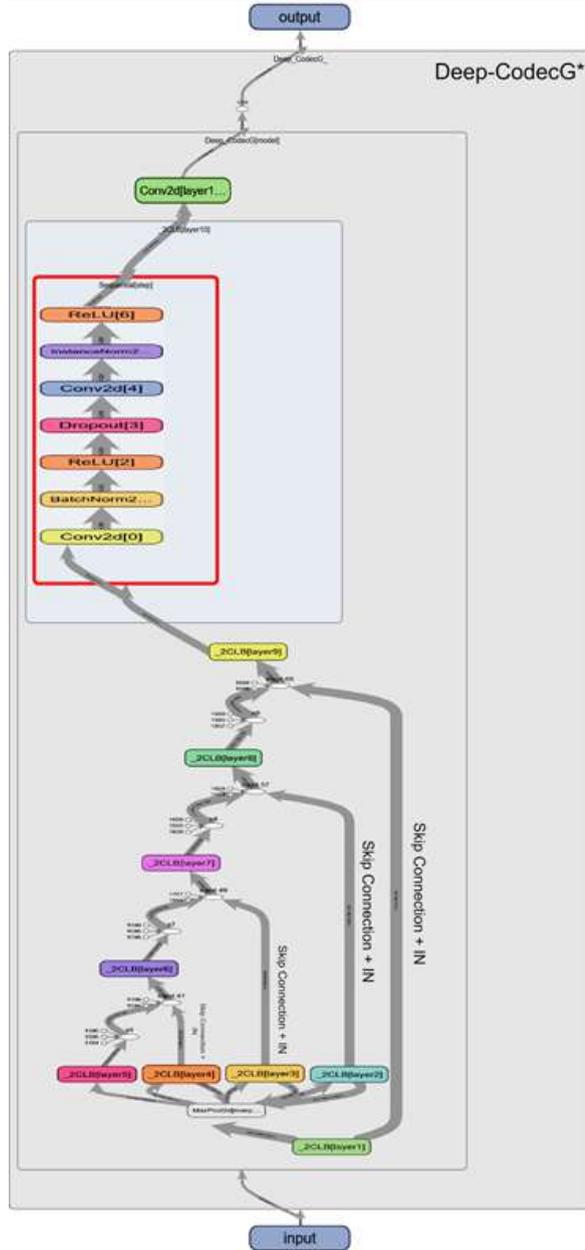


Figure 4: Workflow in Deep-CodecG* for Left Atrium Segmentation

operation and decrease the channel size during the decoder phase, allowing the network to capture more fine-grained details in the data, as mentioned in Figure 3.

After the encoder-decoder, we use one 2CLB to further understand these extracted features throughout the training process (shown with a grey box in Figure 1 and a red box in Figure 4). At last, we use one 2D Convolution layer with the filter size of 1×1 followed by the Sigmoid activation function and get the segmented left atrium image as output. There are a total of 7.9 Mil-

lion trainable parameters in the proposed model. We explain the dataset, preprocessing, and implementation details in Section 4.1.

4 Results and Analysis

4.1 Datasets & Preprocessing

We trained our model with a real dataset; the Medical Segmentation Decathlon (MSD)¹[1] has made the dataset publicly available. The corresponding dataset consists of 20 3D MRI scans of the hearts with corresponding ground truth masks of various patients with higher anatomical variability. Every MRI is stored in a 3D volume with *.nii* file format. The first two dimensions in the 3D volume represent height and width, while the third dimension represents the images' depth (D) in terms of number of slices.

Each image is $352 \times 352 \times D$, where D is depth, ranging from 100 to 180 approximately. We convert the 3D volume data into 2D slices. We crop the non-cardiac region from the given MRIs and masks during preprocessing. The dimensions of each cropped image are 256×256 . The augmentation techniques employed in 2D slices include: (i) Scaling (0.80, 1.15), (ii) Rotation ($-45, 45$), and (iii) Elastic Transformation. These augmentation techniques are applied to the corresponding ground truth mask of the image. Figure 5 shows a sample of the augmented data. The Z-Normalization technique is employed per-subject basis, followed by standardization of the normalized subject within the $[0, 1]$ range. This approach aids in mitigating the complexity of the training and validation processes. After preprocessing, we obtain a total of 4500 2D MRI-labeled slices and split them into a ratio of 70 : 20 : 10 for training, validation, and testing sets, respectively. We use 3150 2D MRI-labeled slices for training, 900 2D MRI-labeled slices for validation, and 450 MRIs used as testing. The red part in Figure 5 shows the left atrium of the cardiac MRI.

4.2 Experimental Setup and Evaluation Metrics

We implemented the proposed Deep-CodecG*, the Config. AE and V-Net with 2D convolutions on *Google Collaboratory Pro* utilizing the open-source Python library *PyTorchLightning*, which provides a high-level interface for *PyTorch*. Additional information can be found at². We experimented with an NVIDIA Tesla P100-PCI-E $\approx 16GB$ GPU and an Intel processor with 12GB of CPU RAM. Each model training is comprised

¹<https://www.medicaldecathlon.com/>

²<https://www.pytorchlightning.ai/>

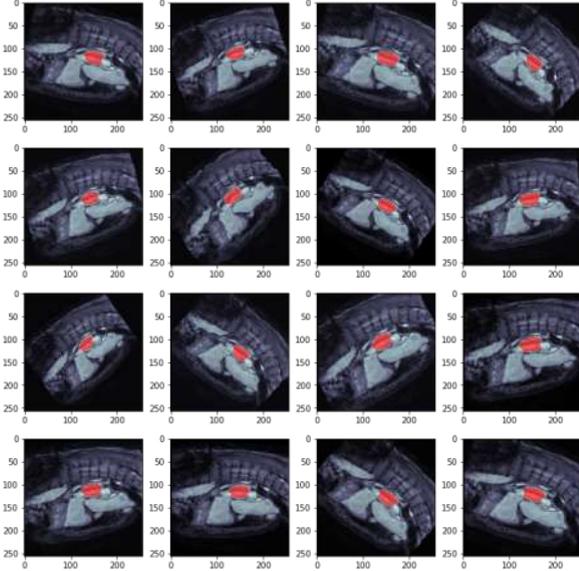


Figure 5: Dataset after the augmentation

of 100 epochs. Due to memory and computational limitations, we could complete all training with only six batches. We utilize Adam optimizer with a learning rate of $1e-4$. We use the dice-loss and dice similarity coefficient to calculate the training and validation error and segmentation accuracy, respectively. In our experiments, we have incorporated the dice coefficient as follows:-

$$DC(f, X, Y) = \frac{2 \times \sum_{i,j} f(X)_{ij} \times Y_{ij} + \epsilon}{\sum_{i,j} f(X)_{ij} + \sum_{i,j} Y_{ij} + \epsilon} \quad (1)$$

where variables X stands for the input image of cardiac MRI, Y for the ground truth, f for the model, and ϵ for a minimal integer to prevent division by zero. Our studies use Eq. 1 to assess the two overlapping regions. The Dice Loss function is defined as follows:

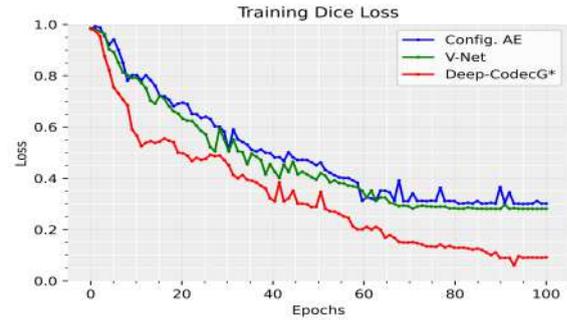
$$L_{Dice}(P, G) = 1 - \frac{2 \times \sum_{i,j} P_{ij} \times G_{ij} + \epsilon}{\sum_{i,j} (P_{ij}^2) + \sum_{i,j} (G_{ij}^2) + \epsilon} \quad (2)$$

where G represents the ground truth, P represents the model's prediction, and ϵ is a minimal number used to prevent division by zero. Additionally, we use Hausdorff Distance (HD) [11], Structural Similarity Index (SSIM) [4], Specificity, Sensitivity, Dice score, and Jaccard Index [8] to evaluate the LA segmentation outcomes. The minimum HD indicates the higher similarity between sets of points. The higher value of SSIM

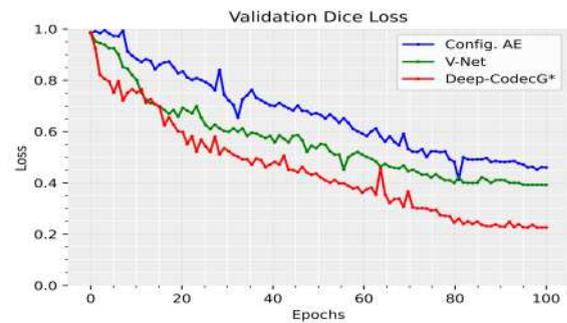
between the two images shows maximum similarity. The Jaccard Index is sensitive to both false positives and false negatives. In contrast, the Dice Score is more robust to the imbalance between the number of true negatives and false positives/negatives.

4.3 Comparison with State-of-the-Art Models

Figure 6 illustrates the loss graph during the training and validation of Config. AE, V-Net, and the proposed Deep-CodecG* model. During training and validation, the loss value progressively converged following numerous fluctuations. Config. AE halts the convergence of train and validation loss at 0.3012 and 0.4616, respectively, with a difference of 0.16. The V-Net halts the training and validation loss converged at 0.2812 and 0.3928, respectively. The proposed model stops the convergence of training and validation loss at 0.0919 and 0.211, respectively, with a difference of 0.11.



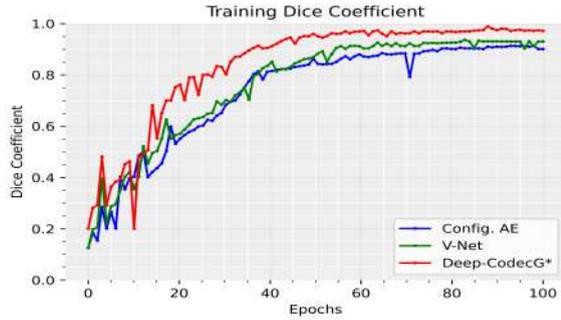
(a) Training Loss



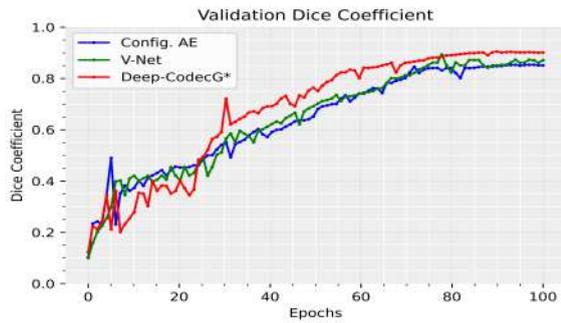
(b) Validation Loss

Figure 6: Comparative training and validation loss of Config. AE, V-Net, and the Proposed Model for left atrium segmentation

Figure 7 shows the dice coefficient score obtained during the training and validation processes. After twenty epochs, the proposed model has consistently improved DC, despite a lengthy learning process that included numerous up and down variations. When the



(a) Trainig Dice Coefficient



(b) Validation Dice Coefficient

Figure 7: Comparative training and validation dice coefficient of Config. AE, V-Net, and Deep-CodecG* for left atrium segmentation

90th epoch came along, it trained with more than 95% of the DC score. After the 100 epochs, the Config. AE stands at the training and validation DC of 0.9089 and 0.8515, respectively; the V-Net achieves the training DC of 0.9301 and validation DC of 0.8713, which is approx 2% higher than the DC of Config. AE during training and validation. In contrast, the proposed Deep-CodecG* architecture obtains DCs of 0.9720 and 0.9016 for training and validation, respectively. The CodecG* gains 4-6% higher training DC than V-Net and Config. AE, respectively, and 3-5% higher during the validation. We have listed all the numerical values achieved after the learning process in Table 2.

Figure 8 shows the variation in the mean dice score through the error bar graphs for all three models. The Config. AE model achieves the mean dice score of 0.89 with the standard deviation (std.) of ± 0.076 and standard error $SE = \frac{std.}{\sqrt{n}}$ of 0.013, where n is the number of test samples. The V-Net achieves a mean dice score of 0.90 with the std. of ± 0.064 and SE of 0.011. The proposed Deep-CodecG* model achieves a mean dice score of 0.95, which is 7% and 5% higher than Config. AE and V-Net, respectively, with the std. of ± 0.05 and

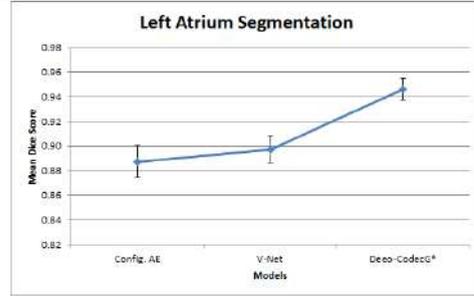


Figure 8: Error bar graph of Mean Dice score on test data set

SE of 0.008. The CodecG* gives a minimum standard error than both models on test data (see Table 3).

4.4 Visual Results

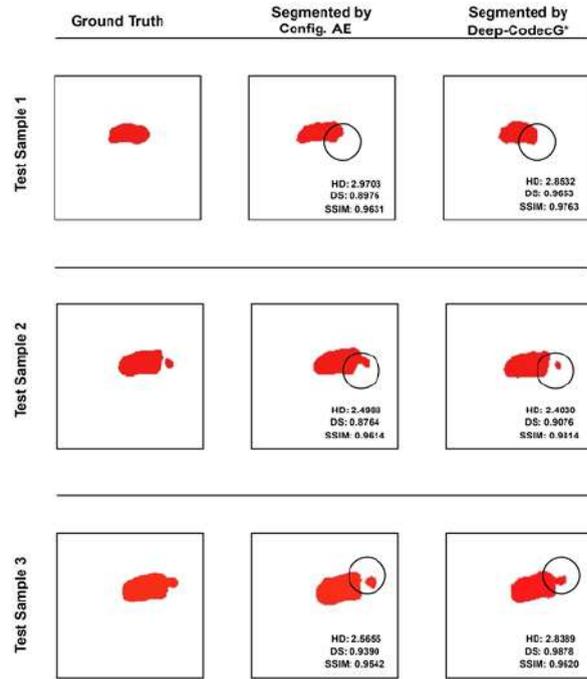


Figure 9: Test_Samples results explain with only left atrium mask

Next, we visually depict the ground truth masks and the segmented mask results of a few test samples. We show the results obtained by the proposed Deep-CodecG* architecture and its closest model Config. AE in Figure 9. In Test Sample 1, the proposed method segmented the left atrium similar to the ground truth with the SSIM of 97% and dice score of 96%, which is 7.5% higher than the Config. AE segmentation. The black circle represents the large discrepancy between the segmented results of the models. Similarly, we can see

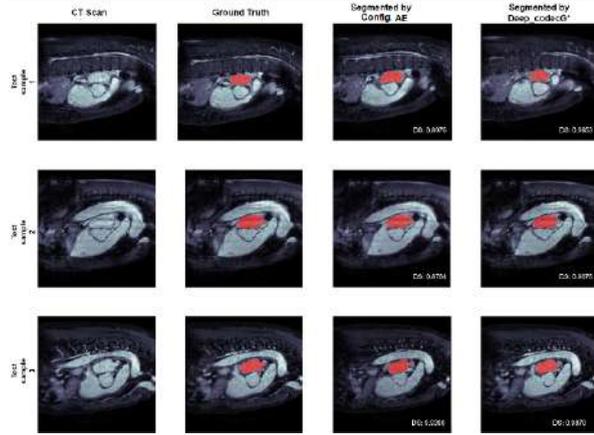


Figure 10: Sagittal view of left atrium segmentation results

that in Test Samples 2 and 3, the proposed model outperforms Config. AE in terms of Hausdorff Distance (HD), SSIM, and Dice Score (DS). The Deep-CodecG* model robustly segments the left atrium from the cardiac MRIs, which is close to the ground truth. In Figure 10, we represent the sagittal view of segmentation results in the cardiac MRIs, as we show the only left atrium of the same test samples in Figure 9.

Table 2 shows the quantitative outcomes of these three models’ training and validation phase. After 100 learning epochs, the proposed Deep-CodedG* has achieved 70% and 54% less training and validation loss than Config. AE, respectively. The V-Net performs better than Config. AE and lesser than the Proposed architecture. The CodecG* has achieved 67% and 46% less training and validation loss, respectively. Regarding accuracy, the proposed model gains 7.73% and 5.8% higher training and validation dice coefficients than the Config. AE, respectively. The CodecG* outperforms V-Net. Generalization techniques and additional convolution layers enabled the Deep-CodrecG* model to learn better than the Config. AE. and V-Net.

Table 2: Dice coefficient and loss of training and validation of three models after 100 epochs

[HTML]EJEFJEF	[HTML]EJEFJEF	[HTML]EJEFJEF Loss		[HTML]EJEFJEF Dice Coeff.	
[HTML]EJEFJEF-2	[HTML]EJEFJEF Model	[HTML]EJEFJEF Train	[HTML]EJEFJEF Val	[HTML]EJEFJEF Train	[HTML]EJEFJEF Val
Config. Auto Encoder (AE)		0.3012	0.4616	0.9089	0.8515
V-Net		0.2812	0.3928	0.9301	0.8713
Proposed Deep-CodecG*		0.0919	0.211	0.9720	0.9016

In Table 3, we have listed the average performance measures scores we achieved on test data of cardiac MRI by segmenting the left atrium using these models. The mean dice score of the proposed Deep-CodecG* is 7-5% higher than the Config. Autoencoder (AE) and V-Net; this implies the segmented left atrium by the pro-

posed Deep-CodecG* model is more accurate than both Config. AE and V-Net. Figure 9 shows the detailed segmentation results on the test samples.

Table 3: Average performance measures on test data for left atrium segmentation. Better values are shown in **Bold** face.

[HTML]EJEFJEF Performance Measures	[HTML]EJEFJEF Good	Configured Auto Encoder (AE)	V-Net	Proposed Deep-CodecG*
Dice Score	High	0.8962 ± 0.076	0.9002 ± 0.064	0.9532 ± 0.051
Standard Error	Low	0.0130	0.0110	0.0088
Sensitivity	High	0.9512	0.9554	0.9665
Specificity	High	0.9972	0.9936	0.9941
Jaccard	High	0.9146	0.9316	0.9743
Hausdorff Distance	Low	2.8916	2.7324	2.5123
SSIM	High	0.9662	0.9692	0.9787

The proposed model’s Hausdorff Distance (HD) is 13% and 8% less than the Config. AE and V-Net models, respectively; minimum HD shows a lesser difference between ground truth and predicted mask. Other performance measures, such as structural similarity index (SSIM) and Jaccard, are approx 1.2% and approx 7-5% higher than both Config. AE and V-Net, respectively. Regarding specificity and sensitivity, the proposed Deep-CodecG* model outperforms the Config. AE and V-Net model.

4.5 Discussion

We designed and implemented a DL-driven Deep-CodecG* model for the automatic left atrium segmentation. The G* stands for the generalization, which means our proposed model used different generalization techniques with a proper parameterization that increases the model’s ability to perform on unseen data and improves the robustness of the model; the numerous quantitative and visual results show the robustness of the proposed model thus achieved. We compared the empirical results of the proposed model with a Config. AE and V-Net models. The effectiveness of three models has been tested on unseen data presented in Table 3. We have used different performance metrics such as dice score, Hausdorff distance, and structural similarity index. The mean dice score of the proposed model is 6% higher than the state-of-the technique. Hausdorff distance in Figure 9 shows that the Deep-CodecG* has precisely segmented the left atrium than the Config. AE and V-Net models. During the training phase, Deep-CodecG* gains a 97% training dice coefficient score, which is approx 4% and 6% better than the Config. AE and V-Net, respectively. Thus, the Deep-CodecG* outperforms during validation and gives segmented LA results close to ground truth. Thus, the proposed model is shown to be robust, which may help radiologists and other experts diagnose heart-related problems.

Limitations The proposed Deep-CodecG* architecture uses the generalization techniques and additional convolutional layer to perform on 2D images made from 3D images. This architecture works better, specifically with the left atrium dataset, and may work with other medical images having variations in sensing.

5 Conclusion

In this paper, we performed left atrium (LA) segmentation in cardiac MRIs. Due to left atrial enlargement, the risk of a heart attack increases, and manual segmentation of the LA is a tedious task. Many researchers used U-Net architectures to perform medical image segmentation. We have proposed a deep learning-based Deep-CodecG* model for LA segmentation. We have used generalization techniques with proper parameterization in the whole network of the proposed model. In Deep-CodecG*, we used additional convolutional layers to understand the extracted features by the encoder-decoder further. For a fairer comparison of our proposed model, we customized an auto-encoder based on the U-Net framework called Config. AE, and also compared with a V-Net model. We used publicly available datasets for our experiments. Empirical results have shown that the proposed model outperforms Config. AE and V-Net on unseen data. Finally, this work shows that using convolutional layers with generalization techniques improves the efficacy of deep learning-based architectures. Even minor quantitative enhancements in performance by DL-based techniques on unseen data can play a significant role in medical imaging. They may help the radiologist and medical experts in the early detection and proper diagnosis. The proposed Deep-CodecG* is shown to match its segmented results with the ground truth closely.

In future work, we will perform with higher dimensional data and try to assess new generalization techniques with DL architectures in the medical domain.

Acknowledgement

The authors thank the editor and reviewer(s) for their valuable comments.

References

- [1] Antonelli, M., Reinke, A., Bakas, S., Farahani, K., Kopp-Schneider, A., Landman, B. A., Litjens, G., Menze, B., Ronneberger, O., and et al. The medical segmentation decathlon. *Nature Communications*, 13(1):4128, 2022.
- [2] Badrinarayanan, V., Kendall, A., and Cipolla, R. SEGNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Analysis & Machine Intelligence*, 39(12):2481–2495, 2017.
- [3] Cao, X., Chen, H., Li, Y., Peng, Y., Wang, S., and Cheng, L. Dilated densely connected U-Net with uncertainty focus loss for 3D ABUS mass segmentation. *Computer Methods & Programs in Biomedicine*, 209:106313, 2021.
- [4] Chang, Z., Cao, J., and Zhang, Y. A novel image segmentation approach for wood plate surface defect classification through convex optimization. *J. Forestry Research*, 29(6):1789–1795, 2018.
- [5] Chen, X., Wang, X., Zhang, K., Fung, K.-M., Thai, T. C., Moore, K., Mannel, R. S., Liu, H., Zheng, B., and Qiu, Y. Recent advances and clinical applications of DL in MIA. *Medical Image Analysis*, page 102444, 2022.
- [6] de Almeida, F. L., Rosa, R. L., and Rodríguez, D. Z. Voice quality assessment in communication services using deep learning. In *2018 15th International Symposium on Wireless Communication Systems (ISWCS)*, pages 1–6. IEEE, 2018.
- [7] Degel, M. A., Navab, N., and Albarqouni, S. Domain and geometry agnostic CNNs for LA segmentation in 3D ultrasound. In *Proc. 21st MICCAI, Part IV*, pages 630–637. Springer, 2018.
- [8] Du, X., Yin, S., Tang, R., Liu, Y., Song, Y., Zhang, Y., Liu, H., and Li, S. Segmentation and visualization of LA through a unified DL framework. *Int. J. Computer Assisted Radiology & Surgery*, 15:589–600, 2020.
- [9] Gao, Q. and Almekkawy, M. ASU-Net++: A nested U-Net with adaptive feature extractions for liver tumor segmentation. *Computers in Biology & Medicine*, 136:104688, 2021.
- [10] Ioffe, S. and Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proc. Int. Conf. Machine Learning (ICML)*, pages 448–456. PMLR, 2015.
- [11] Karimi, D. and Salcudean, S. E. Reducing the Hausdorff distance in medical image segmentation with CNNs. *IEEE Trans. Medical Imaging*, 39(2):499–513, 2019.

- [12] Kausar, A., Razzak, I., Shapiai, M. I., and Besheti, A. 3D shallow deep NN for fast and precise LA segmentation. *Multimedia Systems*, pages 1–11, 2021.
- [13] Kazi, A., Betko, S., Salvi, A., and Menon, P. G. Automatic LA segmentation from computed tomography angiography images. *Annals of Biomedical Engineering*, pages 1–10, 2023.
- [14] Lasmar, E. L., de Paula, F. O., Rosa, R. L., Abrahao, J. I., and Rodríguez, D. Z. Rsr: Ridesharing recommendation system based on social networks to improve the user's qoe. *IEEE Transactions on Intelligent Transportation Systems*, 20(12):4728–4740, 2019.
- [15] Li, S., Zhao, Z., Xu, K., Zeng, Z., and Guan, C. Hierarchical consistency regularized mean teacher for semi-supervised 3D LA segmentation. In *Proc. 43rd Annual Int. Conf. IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 3395–3398, 2021.
- [16] Liu, H., Feng, J., Feng, Z., Lu, J., and Zhou, J. LA segmentation in CT volumes with fully CNN. In *Proc. 3rd Int. Workshop DLMIA, and 7th Int. Workshop, Vol. 3: DL in MIA and Multimodal Learning for Clinical Decision Support*, pages 39–46. Springer, 2017.
- [17] Malhotra, P., Gupta, S., Koundal, D., Zaguia, A., and Enbeyle, W. Deep NNs for medical image segmentation. *J. Healthcare Engineering*, 2022, 2022.
- [18] Maron, R. C., Schlager, J. G., Hagenmüller, S., von Kalle, C., Utikal, J. S., Meier, F., Gellrich, F. F., Hobelsberger, S., Hauschild, A., French, L., et al. A benchmark for NN robustness in skin cancer classification. *European J. Cancer*, 155:191–199, 2021.
- [19] Naser, M. A. and Deen, M. J. Brain tumor segmentation and grading of lower-grade glioma using DL in MRI images. *Computers in Biology & Medicine*, 121:103758, 2020.
- [20] Ngo, T. A. and Carneiro, G. Left ventricle segmentation from cardiac MRI combining level set methods with deep belief networks. In *Proc. IEEE Int. Conf. Image Processing (ICIP)*, pages 695–699, 2013.
- [21] Nunes, R. D., Pereira, C. H., Rosa, R. L., and Rodríguez, D. Z. Real-time evaluation of speech quality in mobile communication services. In *2016 IEEE International Conference on Consumer Electronics (ICCE)*, pages 389–390. IEEE, 2016.
- [22] Peng, P., Lekadir, K., Gooya, A., Shao, L., Petersen, S. E., and Frangi, A. F. A review of heart chamber segmentation for structural and functional analysis using cardiac MRI. *Magnetic Resonance Materials in Physics, Biology & Medicine*, 29:155–195, 2016.
- [23] Rettmann, M. E., Holmes III, D. R., Breen, J. F., Ge, X., Karwoski, R. A., Monahan, K. H., Bahnsen, T. D., Packer, D. L., Robb, R. A., Investigators, C. P. I., et al. Measurements of the LA and pulmonary veins for analysis of reverse structural remodeling following cardiac ablation therapy. *Computer Methods & Programs in Biomedicine*, 118(2):198–206, 2015.
- [24] Rodríguez, D. Z., Rosa, R. L., and Bressan, G. A billing system model for voice call service in cellular networks based on voice quality. In *2013 IEEE International Symposium on Consumer Electronics (ISCE)*, pages 89–90. IEEE, 2013.
- [25] Rosa, R. L., Rodríguez, D. Z., and Bressan, G. Music recommendation system based on user's sentiments extracted from social networks. *IEEE Transactions on Consumer Electronics*, 61(3):359–367, 2015.
- [26] Shoaib, M. A., Chuah, J. H., Ali, R., Hasikin, K., Khalil, A., Hum, Y. C., Tee, Y. K., Dhanalakshmi, S., Lai, K. W., et al. An overview of DL methods for left ventricle segmentation. *Computational Intelligence & Neuroscience*, 2023, 2023.
- [27] Thamhilarasi, V. and Roselin, R. U-NET: CNN for lung image segmentation and classification in chest X-ray images. *Infocomp J. Computer Science*, 20(1), 2021.
- [28] Ulyanov, D., Vedaldi, A., and Lempitsky, V. Instance normalization: The missing ingredient for fast stylization. *arXiv 1607.08022*, 2016.
- [29] Xiong, Z., Fedorov, V. V., Fu, X., Cheng, E., Macleod, R., and Zhao, J. Fully automatic LA segmentation from late gadolinium enhanced mri using a dual fully CNN. *IEEE Trans. Medical Imaging*, 38(2):515–524, 2018.

-
- [30] Yu, L., Wang, S., Li, X., Fu, C.-W., and Heng, P.-A. Uncertainty-aware self-ensembling model for semi-supervised 3D LA segmentation. In *Proc. 22nd MICCAI, Part II 22*, pages 605–613. Springer, 2019.