Improving Robustness of Automatic Cardiac Function Quantification From Cine Magnetic Resonance Imaging Using Synthetic Image Data

Bogdan A. Gheorghiță (bogdan.gheorghita@siemens.com)
Siemens SRL
Lucian M. Itu
Siemens SRL
Puneet Sharma
Siemens Healthineers
Constantin Suciu
Siemens SRL
Jens Wetzl
Siemens Healthineers
Christian Geppert
Siemens Healthineers
Mohamed Ali Asik Ali
Siemens Healthineers
Aaron M. Lee
William Harvey Research Institute, NIHR Biomedical Research Centre at Barts, Queen Mary University of London
Stefan K Piechnik
University of Oxford
Stefan Neubauer
University of Oxford
Steffen E. Petersen
William Harvey Research Institute, NIHR Biomedical Research Centre at Barts, Queen Mary University of London
Jeanette Schulz-Menger
Charité – Universitätsmedizin Berlin, HELIOS Klinikum Berlin Buch, DZHK partnersite Berlin
Teodora Chitiboi
Siemens Healthineers

Research Article

Keywords: Robustness, Cardiac Function Quantification,cine Magnetic Resonance Imaging, Synthetic Image Data, Cardiovascular disease

Posted Date: July 20th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-715226/v1

- **Improving Robustness of Automatic Cardiac Function Quantification from cine Magnetic Resonance Imaging using Synthetic Image Data**

Bogdan A. Gheorghiță*[1, 2] , Lucian M. Itu[1, 2] , Puneet Sharma[3] , Constantin Suciu[1,2] , Jens Wetzl[4] , Christian Geppert[4] , Mohamed Ali Asik Ali[5] , Aaron M. Lee[7,8] , Stefan K Piechnik[6] , Stefan Neubauer[6] , Steffen E. Petersen[7,8,9,10] , Jeanette Schulz-Menger[11] , Teodora Chitiboi[3]

1 Advanta, Siemens SRL, Brașov, Romania

2 Systems Engineering, Transilvania University of Brașov, Brașov, Romania

3 Digital Technology and Innovation, Siemens Healthineers, Princeton, NJ, USA

4 Magnetic Resonance, Siemens Healthineers, Erlangen, Germany

5 Digital Technology and Innovation, Siemens Healthineers, Bangalore, India

6 Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford, UK.

7 William Harvey Research Institute, NIHR Biomedical Research Centre at Barts, Queen Mary University of London, London, United Kingdom

8 Barts Heart Centre, St Bartholomew's Hospital, Barts Health NHS Trust, West Smithfield, London, United Kingdom

9 Health Data Research UK, London, United Kingdom

10 The Alan Turing Institute, London, United Kingdom

11 Charité – Universitätsmedizin Berlin, Experimental and Clinical Research Center, Working Group on CMR and HELIOS Klinikum Berlin Buch, Cardiology Berlin, Germany. DZHK partnersite Berlin

Corresponding author name and email: Bogdan Gheorghita, bogdan.gheorghita@siemens.com

Acknowledgements: This research has been conducted using the UK Biobank Resource (access application 2964). The Data Science Bowl Cardiac Challenge Data was originally provided and publicly released by the National Heart, Lung, and Blood Institute (NHLBI). Special thanks to NHLBI Intramural Investigators Dr. Michael Hansen and Dr. Andrew Arai.

This work was partly funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No 825903 (euCanSHare project). SEP and AML acknowledges support from the National Institute for Health Research (NIHR) Biomedical Research Centre at Barts, from the SmartHeart EPSRC programme grant (EP/P001009/1) and the London Medical Imaging and AI Centre for Value-Based Healthcare. SEP acknowledges support from the CAP-AI programme, London's first AI enabling programme focused on stimulating growth in the capital's AI sector. SEP, SN and SKP acknowledge the British Heart Foundation for funding the manual analysis to create a cardiovascular magnetic resonance imaging reference standard for the UK Biobank imaging resource in 5000 CMR scans (PG/14/89/31194). This project was enabled through access to the Medical Research Council eMedLab Medical Bioinformatics infrastructure, supported by the Medical Research Council (MR/L016311/1).

Author contributions

BAG, LMI, PS, and TC made substantial contributions to the design of the machine learning approaches. PS, AML, CS made substantial contributions to the machine learning experiments. JSM, SN, SEP made substantial contributions to the data analysis and results interpretation. MAAA oversaw and validated data annotation. BAG, JW, CG, SKP, and TC drafted the manuscript. All authors were involved in critically reviewing and improving the manuscript and gave final approval of the version to be submitted.

Competing interests

BAG, LMI, PS, CS, JW, CG, MAAA, TC are employees of Siemens Healthineers (and affiliates). SEP acts as a paid consultant to Circle Cardiovascular Imaging Inc., Calgary, Canada and Servier. The other authors declare no competing financial interests.

Improving Robustness of Automatic Cardiac Function Quantification from cine Magnetic Resonance Imaging using Synthetic Image Data

Abstract:

Although having been the subject of intense research over the years, cardiac function quantification from MRI is still not a fully automatic process in the clinical practice. This is partly due to the shortage of training data covering all relevant cardiovascular disease phenotypes. We propose to synthetically generate short axis CINE MRI using a generative adversarial model to expand the available data sets that consist of predominantly healthy subjects to include more cases with reduced ejection fraction. We introduce a deep learning convolutional neural

network (CNN) to predict the end-diastolic volume, end-systolic volume, and implicitly the ejection fraction from cardiac MRI without explicit segmentation. The left ventricle volume predictions were compared to the ground truth values, showing superior accuracy compared to state-of-the-art segmentation methods. We show that using synthetic data generated for pre-training a CNN significantly improves the prediction compared to only using the limited amount of available data, when the training set is imbalanced.

Cardiovascular disease is the leading cause of death globally, according to the World Health Organization. Cardiovascular magnetic resonance imaging (MRI) is considered the gold standard for evaluating heart function. Estimating the ventricular end-systolic (ESV) and end-diastolic (EDV) volumes, stroke volume (SV) and ejection fraction (EF) from cardiac MRI is a prerequisite for assessing cardiovascular diseases, and typically requires careful and precise contouring of the ventricles.

Deep learning (DL) is predicted to bring substantial change to how cardiovascular MRI is acquired and analyzed (1). The gradual adoption of DL to solve medical image analysis tasks has spawned hundreds of articles addressing the automatic segmentation of cardiac chambers from MRI (2), including several segmentation challenges organized by societies such as MICCAI (3) and Kaggle (4). For example, Bai et al. (5) proposed a deep learning segmentation approach using a fully convolutional network (FCN). Liao et

al. (6) also proposed a deep learning segmentation approach using a modified FCN called Hypercolumns Fully Convolutional Neural Network (HFCN), where features from different levels are concatenated along channel axis. DL algorithms are increasing their performance thanks to the larger annotated datasets available such as the UK Biobank (7), but data is typically not sufficiently representative of cardiovascular disease phenotypes, which limits generalizability. While automatic segmentation approaches have the benefit of reproducibility, experts do not always agree on the precise contours location, as captured by the reduced inter-observer reproducibility of manual contours (8). In clinical practice this means that, while (semi-)automatic cardiac segmentation are commercially available, expert supervision, manual user input, and corrections are still routinely required (3). Other AI approaches focus on direct cardiac function quantification without producing an aggregated segmentation of the structure of interest, thus not requiring a manual contour correction step. Luo et al. (9) proposed a DL regression approach based on multi-scale atlas for the left ventricle (LV) location and a deep Convolutional Neural Network (CNN).

Data augmentation is routinely used in training DL models for medical imaging but is often limited to affine transformations and noise addition, which cannot generate cases with diverse clinical and scan parameters. In recent years, there has been a growing interest in DL for synthetic data generation, notably starting with Generative Adversarial Networks (GAN) (10) which can map a random noise vector to a synthetically generated image. A major disadvantage of GAN is the lack of control over the generated images, which was mitigated with the introduction of conditional GANs (11). Style transfer DL architectures (CycleGAN (12), Pix2Pix (13)) convert an input image from one domain to another, by modifying the style, while preserving the content.

Unsupervised style transfer has been applied from standard CINE MRI to LGE (14) and CT (15), but with limited application to cardiovascular pathologies. The main drawback of style transfer is the need for a large set of annotated images from at least one domain, that is representative of all cardiac anatomy phenotypes. Semantic image synthesis approaches (mask-to-image translation) map one or more segmentation masks to a corresponding image, i.e. the opposite of segmentation networks. GauGAN (16) (17) is a novel approach using a Spatially Adaptive Normalization (SPADE) technique which is a combination between batch normalization and instance normalization, implemented as a two-layer CNN. The network produces a realistic, completely new images, thus introducing more shape, texture, and background variations than conventional computer vision-based augmentation techniques. In one cardiac MRI application, Abbasi-Sureshjani et al. (18) have used a GauGAN network to synthesize labeled 3D+t CINE images.

Herein, we investigate the hypothesis that synthetically generated cardiac MRI can improve DL-based cardiac function quantification. We propose a DL approach to synthetically generate short axis (SAX) cardiac MRI stacks with a wide range of EF values, using the GauGAN (16) architecture, which we apply as a first step to generate a large synthetic subject dataset. We further propose a Residual Spatial Feature Encoding Recurrent network for Abstracting high-level patient features (SFERA) to predict left ventricle volumes (and implicitly the EF) without explicit segmentation. The network combines a fully convolutional feature encoder that learns the cardiac geometry with a recurrent network based on a bidirectional LSTM (19) that incorporates the volumetric information over a stack of variable number of short-axis slices. The network is pre-

trained on the previously generated large synthetic dataset, and then finetuned on real annotated cases. Our final EF prediction error is smaller than other state-of-the-art methods, most of them relying on cardiac chamber segmentation.

Results

Synthetic image generation

Figure 1 A, B shows the normal distribution of the EF parameter in the two large datasets. In the original datasets, the reported EF was reduced (<40%) in only 6.3% of the cases and high (>70%) in only 10.5% of the cases. This data imbalance can lead to suboptimal results for the pathological cases, i.e. an AI algorithm trained on such data distributions may perform poorly on the less represented low or high EF cases. Hence, by automatically enhancing the segmentation masks of our real training subjects, we synthetically generated 22653 new SAX stacks consisting of ED and ES masks for the left and right ventricles with a uniform distribution along the LV EF spectrum as shown in Figure 1C. Using a deep-learning network adversarial-trained for real patient data for mask-to-image generation, the synthetic masks were used to generate the same number of synthetic cardiac MR subject datasets. Figure 2 shows the entire workflow for generating new synthetic slices with a wide range of EF values, starting from a mid-ventricular slice of a real subject, as an example. For more details see the Methods section. The resulting synthetic cohort was approximately 32x larger than the real subject cohort.

Cardiac Function Prediction

The baseline results, obtained by training our proposed SFERA network for cardiac function prediction solely on real case data with a normal EF distribution are referred to as *Real Subjects Only* (RSO). The same network architecture trained entirely on synthetic data with a uniform EF distribution is referred to as *Synthetic Subjects Only* (SSO). The SSO model finetuned on real cases is referred to as *Real Subjects with Pretraining* (RSP).

Figure 3 shows the correlation between the manually annotated and the automatically predicted LV volumes and EF for the models with and without pretraining. The Pearson correlation values corresponding to RSO experiment (without pretraining) for EF, EDV and ESV are 80.2%, 90.3% and 93.7% (p < 0.001) for Dataset 1 and 81.5%, 94.8%, 92.1% (p < 0.001) for Dataset 2, as shown in Figure 3 top. In the RSP experiment (with pretraining), the Pearson correlation values for EF, EDV and ESV increased to 95.2%, 97.7% and 98.1% (p < 0.001) for Dataset 1, and 86.2%, 97.1%, 94.6% (p < 0.001) for Dataset 2, as shown in Figure 3 bottom.

Figures 4 shows the Bland-Altman analysis for the volumes and the EF predictions on our two test sets, for the experiments trained on real cases without and with pretraining. In both cases no bias was observed. The mean RMS error in the RSO experiment for the EF was 4.9% for Dataset 1 and 2.8% for Dataset 2. In the RSP experiment, the RMS error was significantly reduced to 2.7% for Dataset 1 and 2.5% for Dataset 2 (p < 0.005). Similarly, the RMS error was significantly reduced from 16.8 ml to 7.3 ml (p < 0.005) for Dataset 1 and from 8.0 ml to 6.2 ml for Dataset 2 for EDV. For ESV, the RMS error was reduced from 9.0 ml to 5.2 ml (p < 0.005) for Dataset 1 and from 5.6 ml to 4.7 ml for Dataset 2.

The 95% confidence intervals of the RMSE for EF, computed using bootstrapping, are [2.0, 2.1] in SSO experiment, [4.4, 5.3] and [2.7, 2.9] for RSO experiment Dataset 1 and Dataset 2, [2.4, 2.9] and [2.4, 2.6], for RSP experiment, Dataset 1 and Dataset 2.

Table 1 presents a comparison between the RMS error ± SD of EDV, ESV, and EF prediction, for the RSO and RSP experiments with our proposed approach, and the results of the winning team (20) of the Kaggle challenge (based on the mean CRPS (21) metric) and the results of the top 4 (6) team (which had the lowest RMS error for EF in the competition), previously reported results on the Dataset 2 (5). Namely, the winning team Luo et al. (20) obtained a 0.00948 CRPS (21) score, which is the equivalent of 12.0 ml RMS error for EDV, 10.2 ml for ESV and 4.9% ejection fraction. The smallest ejection fraction error, 4.7 was obtained by the top 4 team Liao et al. (6), even though the RMS error for volumes is a slightly bigger.

The inference time on a desktop computer with the following hardware configuration: Intel® Core™ i7-7700K CPU @ 4.20GHz, NVIDIA GeForce GTX 1080 Ti graphics card, 64GB RAM was around 5.5 ± 4.3 ms.

Discussion

The RSP model first pretrained on synthetic data outperforms the baseline RSO model trained only on real data. The EF prediction error decreases significantly when synthetic data is used for pretraining (RSP). Similarly, the Pearson correlation for the EDV, ESV, and EF is significantly higher for RSP compared to RSO. Pre-training has an

even higher impact for cases with low or very high EF values, which were less represented in the initial distribution.

Our final prediction model after pretraining on synthetic data (RSP) performs well compared to other state-of-the-art approaches. Since the original ground-truth of the Kaggle challenge test set is not publicly available, our results on Dataset 1 were based on our own manual segmentation of the CINE MRI data, so they are not directly comparable to the Kaggle challenge results. Nevertheless, our model shows very promising performance emphasized by a tight confidence interval.

A main advantage of our approach is that our image synthesis network is able to generate realistic-looking cardiac anatomy including papillary muscles and trabeculations inside the blood pool, which could then be used for pre-training. The synthetic data may also include small image artefacts, different image sharpness and varying contrast, similar to the original dataset used for training, which contribute to the realistic aspect. These synthetic cases thus reliably serve in the pre-training step for the ventricle volume and EF prediction task.

One limitation of our approach is that the synthetic image generator network was trained on individual 2D frames. This causes the image background to be somewhat inconsistent between ED and ES and for consecutive slices of the same case. In future work, we plan to extend the approach to generate fully consistent temporal volumes. Our data synthesis approach also requires an initial segmentation of the training data. In a novel approach, the need for manual segmentation could be circumvented by using an autoencoder (22), one direction which we will further investigate. Another limitation is that the ED-ES frames need to be preselected as input to the volume prediction network.

However, this task could also be performed by an independent neural network trained to automatically identify ED and ES timepoints from a CINE series such as (23).

Depending on how the basal slices are subjectively handled in manual vs. automatic contouring, segmentation-based approaches may introduce notable differenced in the EF in some cases. The segmentation-free EF prediction approach proposed has the advantage of not having to perform a pixel-wise binary decision. On the other hand, the disadvantage of any segmentation-free approach to EF prediction is that the final result is more difficult to confirm without the contours present. Nevertheless, hybrid approaches may employ an ensemble approach that combines different machine learning solutions, with or without segmentation (24). While manual or semi-supervised contouring of the myocardium is the preferred supervised clinical workflow for cardiac function evaluation, an AI algorithm that bypasses the segmentation step, and instead focuses on a more accurate image-based EF prediction, could serve as in independent second reader or as a first triage step before the MR exam is assigned to the main reader. In the future we plan to extend this approach to automatically detect normal cases, which could save reading time. We will also investigate the use of our EF prediction approach in combination with a fully automatic ventricle segmentation AI algorithm as independent AI readers who can self-verify results.

In conclusion, we showed that generating synthetic training data with machine learning can be a powerful tool for improving results of deep learning pipelines, especially when only unbalanced, scarce data is available. We have demonstrated that using synthetic

data for pre-training a network for a quantification task greatly improves the prediction compared to only using the limited amount of available data. To show this, we devised a two-step methodology: first, we generate synthetic data with a uniform distribution of EF values, by using a computer vision-based algorithm for generating binary masks and adopting a mask-to-image network. In the second phase, we pre-trained a neural network only on synthetic data, then finetuned it on the real cases. This methodology was demonstrated on the task of predicting the ventricle volumes and ejection fraction from CINE MRI, using two different datasets, with very accurate results in both cases. The same approach is generalizable to other medical image analysis tasks where the distribution of the available training data is insufficiently representative, or the amount of data is scarce.

Methods

Data

The Kaggle Data Science Bowl Cardiac Challenge Data (4) [Dataset 1] consists of CINE bSSFP cardiac MRI including a short-axis (SAX) stack which was used for ventricular volume quantification. This dataset is publically available (4). The data was acquired with 8-10 mm slice thickness, spatial resolution between 0.61-1.95 mm x 0.61-1.95 mm, and approximately 30 cardiac frames per slice, at 1.5 and 3 T (MAGNETOM Aera and Skyra, Siemens Healthcare, Erlangen, Germany). The average distance between consecutive SAX slices was 9.8+/- 0.6. Since the segmentation masks used to generate the EDV and ESV values used as ground truth in the competition were not made publicly available, the entire dataset was re-annotated by an expert observer. All individual ED and ES frames were manually identified, and the LV and right ventricle

(RV) were manually contoured. The annotations were used to compute ground truth values for the ED and ES LV volumes. The subjects with less than 5 consecutive SAX slices or with the presence of significant motion artefacts were excluded. 491 subject datasets were used for training, 187 for validation and the remaining 424 were reserved for testing

A second independent dataset was publically available from the UK Biobank Resource (7) [Dataset 2]. CINE bSSFP cardiac MR data was acquired using a standard protocol (25). The SAX stack spanning from the apex to the base of the left ventricle was acquired with 8 mm slice thickness, a spatial resolution ranging between 1.8-2.1 mm x 1.8-2.1 mm, and a 31 ms temporal resolution at 1.5 T (MAGNETOM Aera, Siemens Healthcare, Erlangen, Germany). The average slice distance was 8.89 ± 0.88 mm. A ground truth annotation of the LV and RV was obtained through manual segmentation of the end-systolic (ES) and end-diastolic (ED) phases by an expert observer. 2365 subjects were used for training, 224 for validation and the rest of 2084 were reserved for testing.

The data was resampled to 1x1 mm spatial resolution, cropped to 150x150 pixels around the image center and the image intensity values were normalized to the [3%, 97%] quantiles.

Synthetic Image generation

The right approach for synthetic data generation depends on several factors: availability of annotated data, desired quality of the synthetic data, reproducibility, and the amount of control over the characteristics of the generated data (e.g. class label, the

size and deformation of the structures). Herein, we describe a semantic image synthesis algorithm, capable of fully controlling the size and location of the resulting anatomical structures to obtain synthetic subjects with different EF values.

We adapted a state-of-the-art DL network architecture for mask-to-image translation GauGAN (16) (17) to the task of generating synthetic ED and ES image frames of a cardiac SAX image stack, while fully controlling the volume an ejection fraction of the LV. The generator consists of multiple SPADE blocks and the discriminator is a simple convolutional neural network. The loss function is computed from three weighted terms: Multiscale Adversarial Loss and two feature matching losses (one using the discriminator and the other one using a pretrained network).

We first trained the synthetic image generation network learns on a fixed set of CINE MR image and manually annotated segmentation masks consisting of three classes for the LV, RV, and myocardium, previously described in Dataset 1 and Dataset 2. The number of epochs used to train the image synthesis model was chosen empirically based on the subjective visual assessment of the generated images.

Next, we generated an extended dataset of synthetic masks to be used as input for the GauGAN model. For this, we used as starting point the segmentation masks in the Dataset 1 training subset. First, we perform for all slices an interpolation on (ED, ES) mask pairs, and return a number of $\mathrm{F} = 11$ intermediate interpolated masks computed as follows:

$$IM = (\frac{\alpha}{F} * SDT_1) + ((1 - \frac{F-\alpha}{F}) * SDT_2) \quad (1)$$

where IM represents the interpolated mask, $SDT_1\ and\ SDT_2$ represent signed distance transform masks of ED and ES and $\alpha \in (0, F)$. Pairs of (ED, interpolated ED) and (interpolated ED, ES) masks are used to create synthetic cases with reduced EF. In the second step, we use an affine transformation γ to rescale the ED and ES masks, such that anatomical structures become smaller at ES, and larger at ED. Thirteen uniformly distributed sample values of γ over the interval [0.7, 1) are used for rescaling the ES mask, leading to a smaller LV and implicitly a smaller volume. The same number of samples are used for the ED mask, but covering the interval [1, 1.2), resulting in a larger LV for ED mask and an increased EF for the case. The values of α and γ have been chosen empirically.

The synthetically generated masks contained the same number of slices as the real cases used as starting point. The EDV and ESV for the synthetic subjects were computed using Simpson's rule, assuming a constant slice thickness of 8 mm, and no gaps between slices. The EF is computed from the resulting volumes as:

$$EF = \frac{(EDV - ESV)}{EDV} \quad (2)$$

Finally, we applied the trained image synthesis model described above to the previously generated extended set of synthetic masks with a uniform EF distribution, in order to generate the synthetic CINE MR images. The resulted 22653 synthetic cases were split into 16491 synthetic datasets for training the SFERA network for cardiac function prediction, and 6162 for validation for the pretraining step in the RSP experiment.

Cardiac Function Prediction

We designed a custom deep neural network capable of processing a stack of CINE MR slices of variable number of slices and which outputs both EDV and ESV, further used to compute the EF. The architecture of the SFERA network is presented in Figure 5. The network input is a SAX stack of a varying number of slices, each consisting of one ED and one ES frame concatenated along the channel axis. A 2D residual CNN is employed in the first layers for every (ED, ES) pair. There are five residual blocks building the CNN. Every block consists of multiple 2D convolutional layers, ReLU activation functions, Batch Normalization (26) and Max Pooling layers. The first convolutional layer outputs 32 channels, and this parameter doubles in value at every convolutional block. Before feeding the resulting features to the LSTM (19), they are flattened and a linear network is used to reduce their dimensionality to 128 elements containing spatial information. Then, a bidirectional LSTM (19) network is applied to correlate the information between these feature vectors, resulting a vector containing both spatial and temporal information. As a final step, a Bayesian ridge regressor is employed to predict the final EDV and ESV volumes. The LSTM (19) approach enables the proposed model to process a variable length of slices. The training of the SFERA model is performed using the Rectified Adam optimizer and RMSE loss function.

Volume data is normalized by the distance between slices. We have employed a unity-based normalization to rescale the EDV and ESV values to the range [0, 1]. Only the slices between the basal and the apex planes were retained. After the inference step, the actual ventricular volume (ml) is obtained by scaling the voxel volume estimations output by the network by the original distance between slices.

We used root mean square (RMS) Error and Pearson correlation metrics to evaluate the performance of the trained model against the ground truth values for EDV, ESV and EF. The model prediction error was further investigated using Bland-Altman analysis where the confidence interval was defined as mean ± 1.96 SD. Kruskal-Wallis test was used to measure the statistical difference between the RMS errors obtain in the RSO and RSP experiments.

List of figures

Figure 1. Normal EF distributions of the 491 real cases from the Dataset 1 (left) and 2365 cases from Dataset 2 (middle), and our 22861 synthetically generated cases with a uniform EF distribution (right).

Figure 2. Workflow for the synthetic image generation step. The end-systolic and end-diastolic frames from every slice of the training data goes through this process to generate an extended set of masks with different EF values. Parameters α and γ are used to control the number of interpolated frames between [ES, ED], and the number of rescaled [ES, ED] pairs. In order to generate a new synthetic slice (ed, es pair) with smaller EF, new pairs of [ES original, Interpolated ED] and [Interpolated ES, ED original] are chosen. To generate a new synthetic slice with higher EF, a pair of [Smaller ESV, Bigger EDV] is chosen. An additional step is employed to filter only resulting subjects with EF between 10 and 80%.

Figure 3. Scatter plots of predicted and ground truth volumes and EF on the Dataset 1 test dataset (purple) and Dataset 2 test dataset (light blue), for the model trained on real

cases only, without pretraining (RSO) and the model finetuned on real cases after pretraining on synthetic data (RSP). RMS Error is computed for EDV, ESV and ejection fraction.

Figure 4. Bland-Altman (BA) plots of predicted and ground truth volumes and EF on the Dataset 1 test dataset (purple) and Dataset 2 test dataset (light blue), for the model trained on real cases only, without pretraining (RSO) and the model finetuned on real cases after pretraining on synthetic data (RSP). Bland-Altman (BA) analysis of the results, comparing the models trained on real cases without pretraining (top) and real cases with finetuning from the synthetic model (bottom).

Figure 5. Architecture of the Spatial Feature Encoding Recurrent network for Abstracting high-level patient features (SFERA) model. The model takes as input a SAX stack of variable number of slices which comprise individual ED and ES frames. The network outputs the ESV and ESV, which are subsequently used to compute the ejection fraction.

Table 1. Root mean squared error ± standard deviation for the EDV, ESV and EF prediction from top to bottom for our models trained on synthetic subjects only (SSO), real subjects only (RSO), and the SSO model finetuned on real cases is referred to as real subjects with pretraining (RSP). Below are the results of the winner of the Kaggle challenge (4) (based on the mean CRPS (21) metric), the results of the top 4 team (6) (which had the lowest RMSE for EF in the competition), and previously reported results on the Dataset 2 (5) for comparison.

	RMS Error ± STD					
Experiment	**EDV [ml] Dataset1**	**ESV [ml] Dataset1**	**EF [%] Dataset1**	**EDV [ml] Dataset2**	**ESV [ml] Dataset2**	**EF [%] Dataset2**
SSO (ours)	53.4 ± 16.3	34.3 ± 12.5	**6.9 ± 4.2**	-	-	-
RSO (ours)	16.8 ± 16.3	9.0 ± 8.3	**4.9 ± 4.6**	8.0 ± 7.6	5.6 ± 5.9	**2.8** ± **2.3**
RSP (ours)	7.3 ± 8.5	5.2 ± 4.9	**2.7 ± 2.4**	6.2 ± 5.7	4.7 ± 4.8	**2.5** ± **2.0**

Top1 Kaggle (4)	12.2	10.1	**4.8**	-	-	-
Top4 Kaggle (6)	13.2	9.3	**4.6**	-	-	-
Bai et al. (5)	-	-	**-**	6.1 ± 5.3	5.3 ± 4.9	**3.2** ± **2.9**

Bibliography

1. Litjens, G., Ciompi F, Wolterink JM, de Vos BD, Leiner T, Teuwen J, Išgum I. State-of-the-Art Deep Learning in Cardiovascular Image Analysis. JACC: Cardiovascular Imaging. , Vol. 12, pp. 1549-65 (2019). Litjens G8, Aug 1, 2019 , JACC: Cardiovascular Imaging. , Vol. 12, pp. 1549-65.

2. Chen, C, Qin C, Qiu H, Tarroni G, Duan J, Bai W, Rueckert D. Deep Learning for Cardiac Image Segmentation: A Review. Frontiers in Cardiovascular Medicine, Vol. 7 (2020).

3. Bernard, O., Lalande A, Zotti C, Cervenansky F, Yang X, Heng PA, Cetin I, Lekadir K, Camara O, Ballester MA, Sanroma G.Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved? *IEEE transactions on medical imaging*, Vol. 37, pp. 2514-25 (2018).

4. Second Annual Data Science Bowl - Transforming How We Diagnose Heart Disease. The National Heart, Lung, and Blood Institute (NHLBI). s.l. : Booz Allen Hamilton, 2016. Retrived from https://www.kaggle.com/c/second-annual-data-science-bowl.

5. Bai, W. , Sinclair M, Tarroni G, Oktay O, Rajchl M, Vaillant G, Lee AM, Aung N, Lukaschuk E, Sanghvi MM, Zemrak F, et al. Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. *Journal of Cardiovascular Magnetic Resonance*, Vol. 20, p. 65 (2018).

6. Liao, F., Chen, F., Hu, S. & Song, S. Estimation of the Volume of the Left Ventricle From MRI Images Using Deep Neural Networks. IEEE Transactions on Cybernetics, vol. 49, no. 2, pp. 495-504 (2019).

7. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P, Elliott P, Green J, Landray M, Liu B. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. 3, Mar 31, 2015, Plos med, Vol. 12.

8. Mikhail, G. D., Jos, J. M. W., Albert de Roos, Johan, H. C. R., Boudewijn, P. F. L. Operator induced variability in cardiovascular MR: left ventricular measurements and their reproducibility. *Journal of Cardiovascular Magnetic Resonance*, Vol. 7, pp. 447-57 (2005).

9. Gongning, L. et al. Multi-views Fusion CNN for Left Ventricular Volumes Estimation on Cardiac MR Images. *IEEE Trans Biomed Eng* (2018).

10. Goodfellow, J. I. et al. Generative Adversarial Nets. *NIPS'14. 2672--2680*. (2014).

11. Mehdi, M., Simon, O. Conditional Generative Adversarial Nets. *IEEE* (2014).

12. Jun-Yan, Z., Taesung, P., Phillip, I., Alexei, A. E. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. *IEEE Xplore* (2020).

13. Phillip, I., Jun-Yan, Z., Tinghui, Z., Alexei, A. E. Networks, Image-to-Image Translation with Conditional Adversarial. *CVPR* (2018).

14. Chen, C. et al. Unsupervised Multi-modal Style Transfer for Cardiac MR Segmentation. *International Workshop on Statistical Atlases and Computational Models of the Heart.* pp. 209-219.

15. Ouyang, C., Kamnitsas, K., Biffi, C., Duan, J., Rueckert, D. Data efficient unsupervised domain adaptation for cross-modality image segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention* (2019).

16. Taesung, P., Ming-Yu, L., Ting-Chun, W., Jun-Yan, Z. Semantic Image Synthesis with Spatially-Adaptive Normalization. *CVPR* (2019).

17. Github repository. NVlabs, Semantic Image Synthesis with SPADE. [Online] 2019. Retrived from https://github.com/NVlabs/SPADE.

18. Samaneh, A. S. et al. 4D Semantic Cardiac Magnetic Resonance Image Synthesis. *MIDL 2020 Conference Blind Submission* (2020).

19. Ralf, C. S., Eric, R. M.Understanding LSTM – a tutorial into Long Short-Term Memory. *Arxiv* (2019).

20. Gongning, L., Suyu, D., Kuanquan, W., Henggui, Z. Cardiac left ventricular volumes prediction method based on atlas location and deep learning.*IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, (2016).

21. Grimit, E. P., Gneiting T. V. J., Berrocal, N. A. J. The continuous ranked probability score for circular variables and its application to mesoscale forecast ensemble verification. *RMetS* (2007).

22. Kozerke, T. et al. Leveraging Anatomical Similarity for Unsupervised Model Learning and Synthetic MR Data. *ISMRM Annual Meeting* (2020).

23. Bin, K., Yiqiang, Z., Min, S., Thomas, D., Shaoting, Z. Recognizing End-diastole and End-systole Frames via Deep Temporal Regression Network. *International Conference on Medical Image Computing and Computer-Assisted Intervention* (2016) .

24. Hann, E. et al. Quality Control-Driven Image Segmentation Towards Reliable Automatic Image Analysis in Large-Scale Cardiovascular Magnetic Resonance Aortic Cine Imaging. International Conference on Medical Image Computing and Computer-Assisted Intervention pp. 750-758 (2019).

25. Petersen, S. E. et al. UK Biobank's cardiovascular magnetic resonance protocol. Journal of cardiovascular magnetic resonance Vol. 18, p. 8 (2015).

26. Sergey, I., Christian, S. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. Arxiv (2015).

27. Adam, P. et al. Automatic differentiation in PyTorch. NIPS 2017 Workshop on Autodiff

(2017).

28. Chitiboi, T. et al. Deep Learning-based Strain Quantification from CINE Cardiac MRI, ISMRM Annual Meeting 2020.

Figures

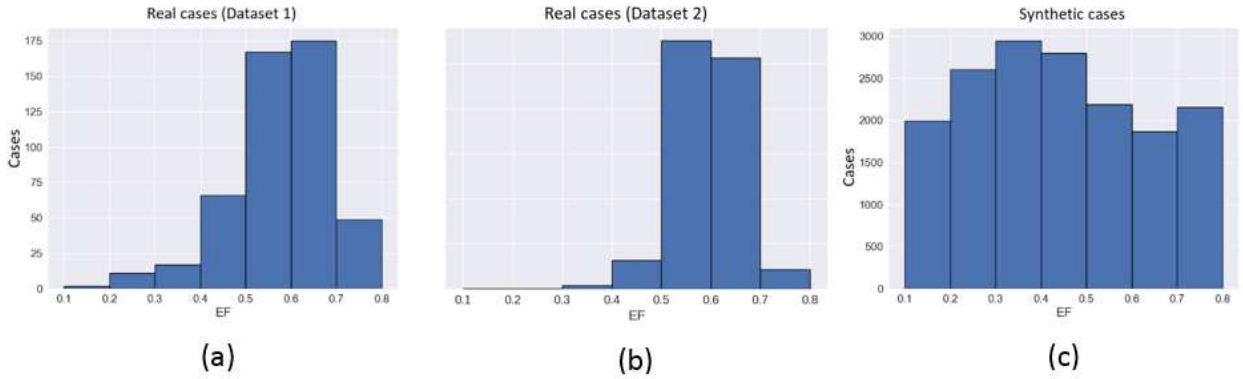

Figure 1

Normal EF distributions of the 491 real cases from the Dataset 1 (a) and 2365 cases from Dataset 2 (b), and our 22861 synthetically generated cases with a uniform EF distribution (c).

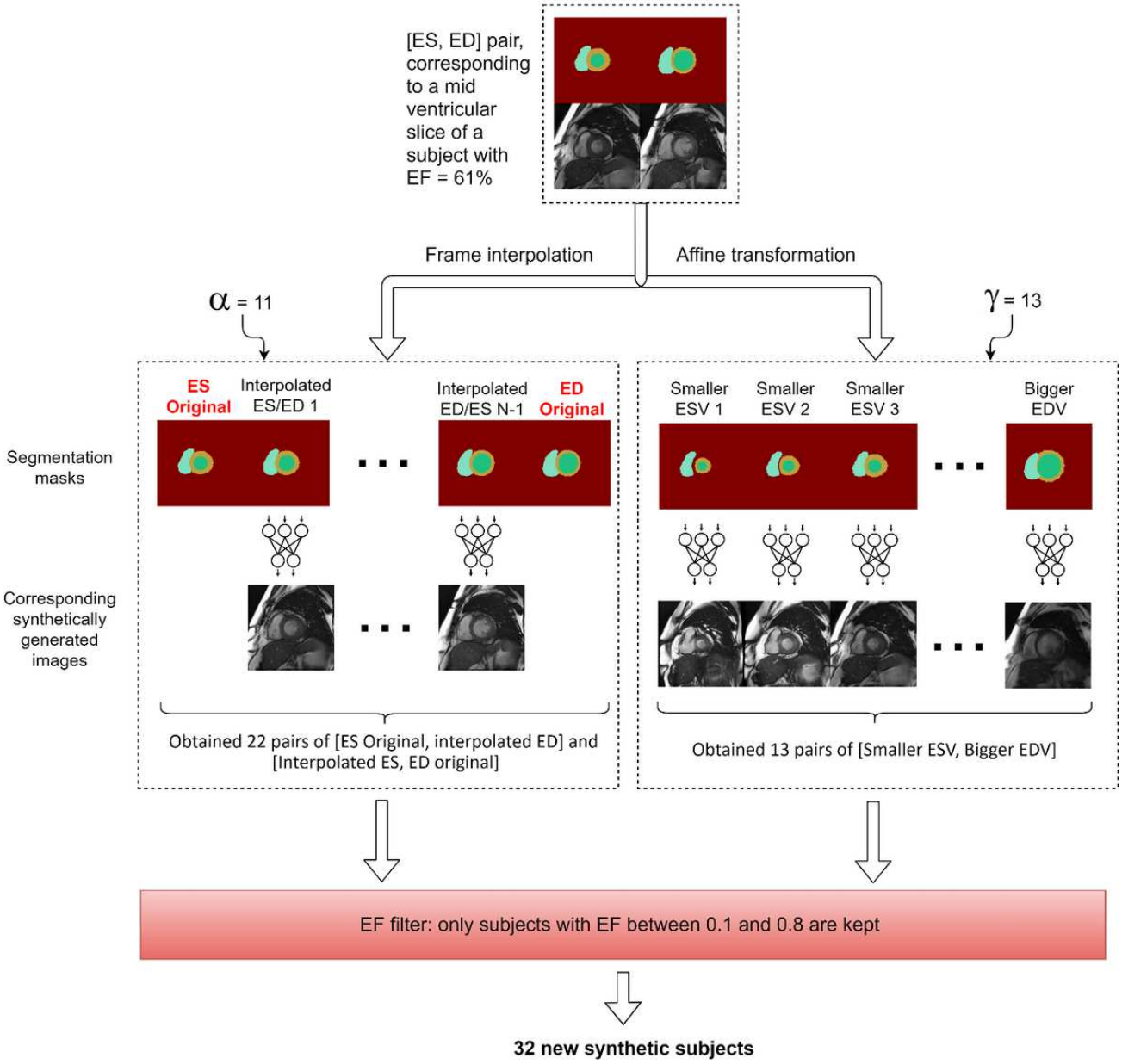

Figure 2

Workflow for the synthetic image generation step. The end-systolic and end-diastolic frames from every slice of the training data goes through this process to generate an extended set of masks with different EF values. Parameters α and γ are used to control the number of interpolated frames between [ES, ED], and the number of rescaled [ES, ED] pairs. In order to generate a new synthetic slice (ed, es pair) with smaller EF, new pairs of [ES original, Interpolated ED] and [Interpolated ES, ED original] are chosen. To generate a new synthetic slice with higher EF, a pair of [Smaller ESV, Bigger EDV] is chosen. An additional step is employed to filter only resulting subjects with EF between 10 and 80%.

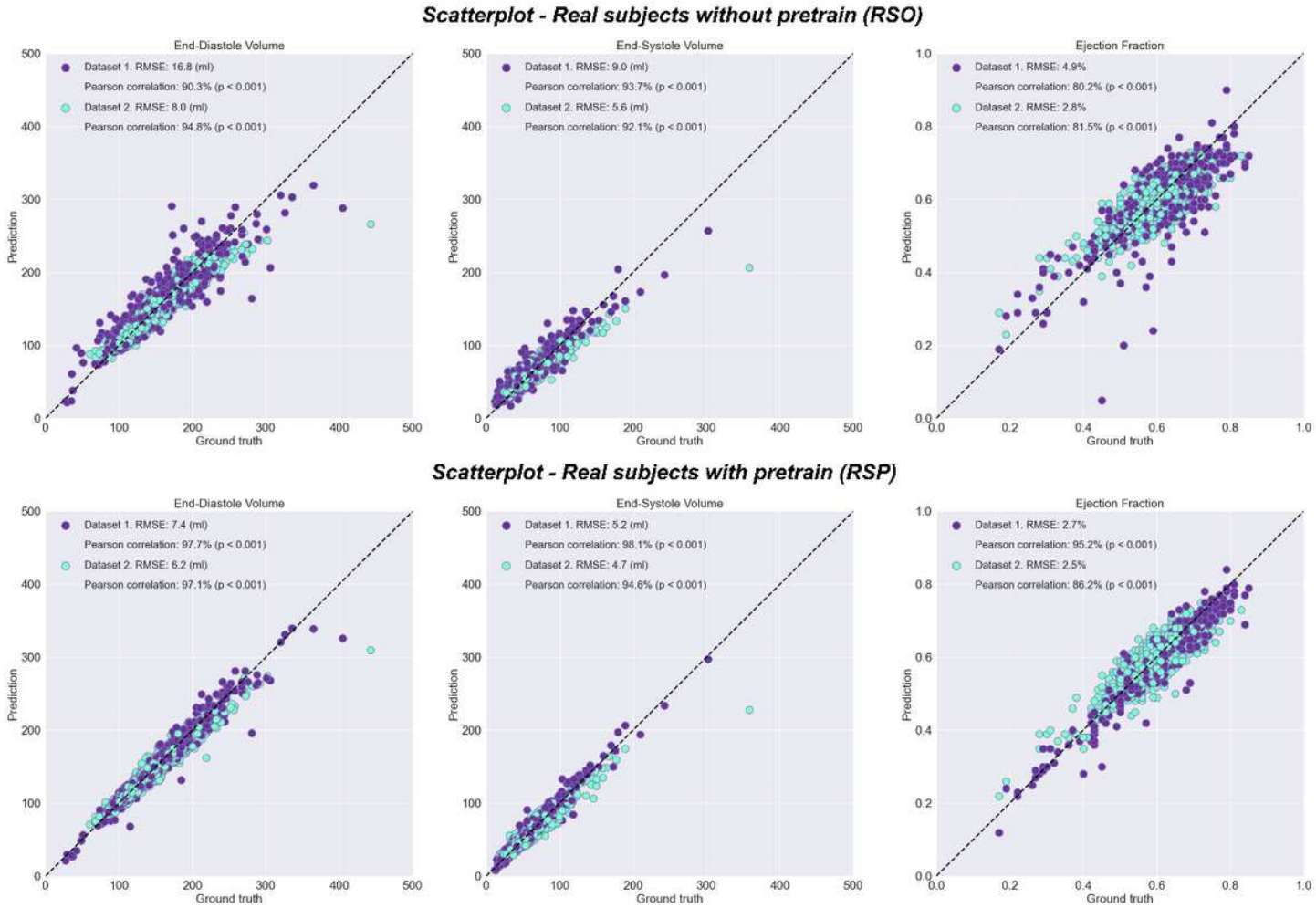

Figure 3

Scatter plots of predicted and ground truth volumes and EF on the Dataset 1 test dataset (purple) and Dataset 2 test dataset (light blue), for the model trained on real cases only, without pretraining (RSO) and the model finetuned on real cases after pretraining on synthetic data (RSP). RMS Error is computed for EDV, ESV and ejection fraction.

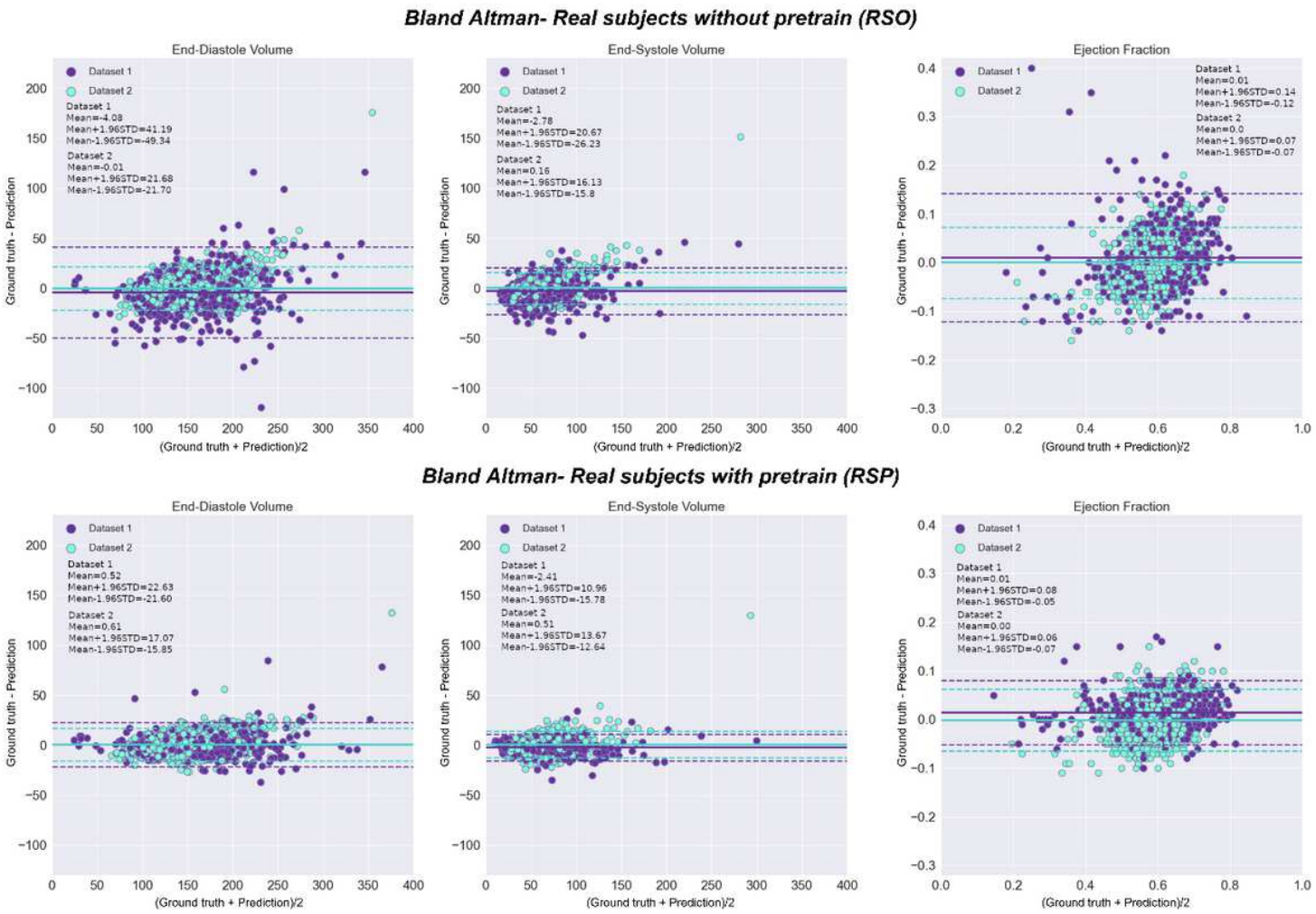

Figure 4

Bland-Altman (BA) plots of predicted and ground truth volumes and EF on the Dataset 1 test dataset (purple) and Dataset 2 test dataset (light blue), for the model trained on real cases only, without pretraining (RSO) and the model finetuned on real cases after pretraining on synthetic data (RSP). Bland-Altman (BA) analysis of the results, comparing the models trained on real cases without pretraining (top) and real cases with finetuning from the synthetic model (bottom).

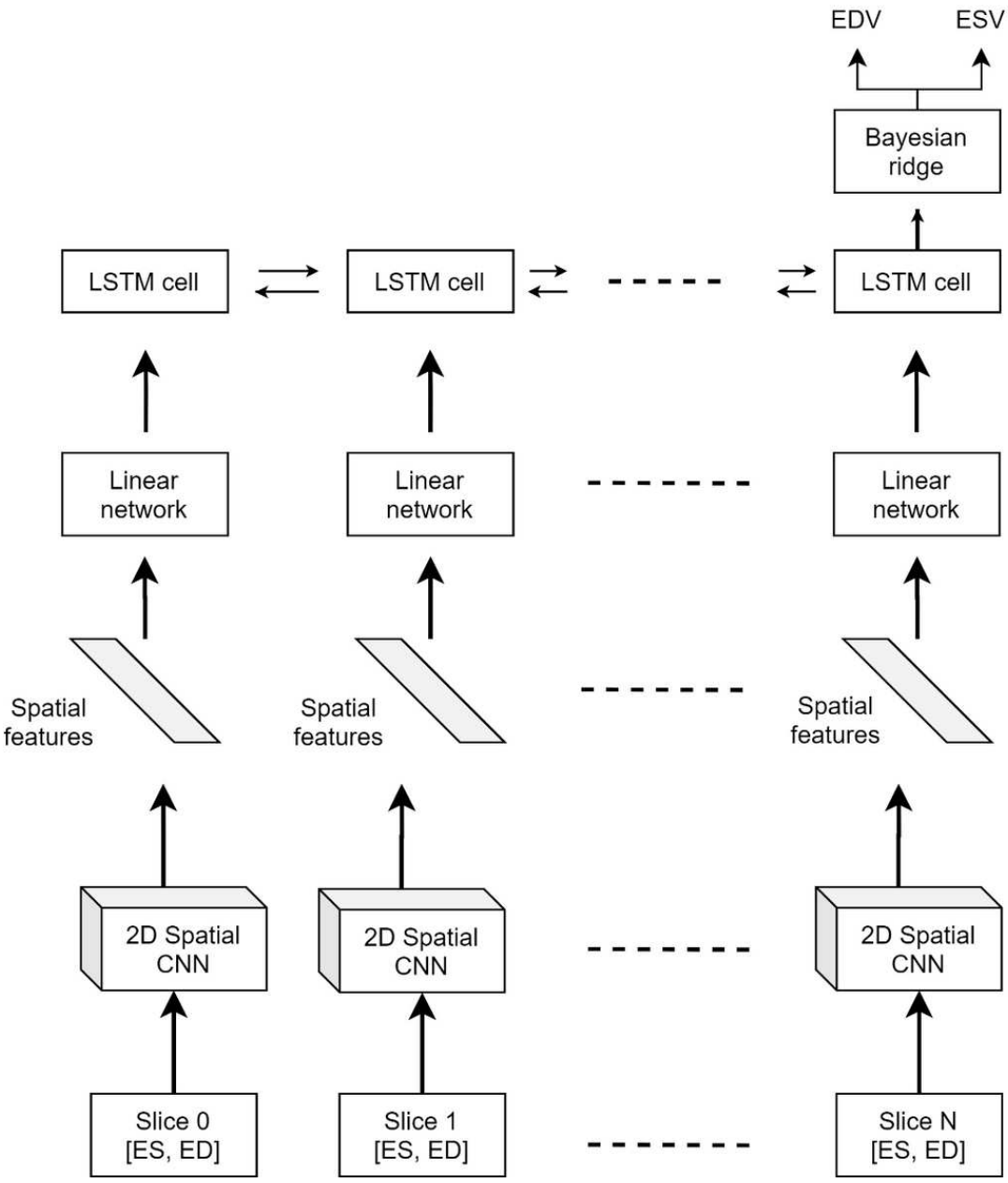

Figure 5

Architecture of the Spatial Feature Encoding Recurrent network for Abstracting high-level patient features (SFERA) model. The model takes as input a SAX stack of variable number of slices which comprise individual ED and ES frames. The network outputs the ESV and ESV, which are subsequently used to compute the ejection fraction.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- suplimentalNMIImprovingRobustnessofAutomaticCardiacFunctionQuantificationfromcineMRIusingSyntheticImageData.docx