SIIM23 ANNUAL MEETING + InformaticsTECH Expo JUNE 14-16 | AUSTIN, TX
Learning. Innovating. Belonging.

Generalization Techniques and Solutions for Deep Learning in Echocardiography: Patients with Restrictive Cardiomyopathy

Jiwoong Jeong, MS, PhD Student, Arizona State University
Chieh-Ji Chao, MD; Reza Arsanjani, MD; Chadi Ayoub, MBBS, PhD; Juan Maria Farina, MD; Bhavik Patel, MD, MBA; Jae Oh, MD; Imon Banerjee, PhD

Introduction

The inherent characteristics of transthoracic echocardiography (TTE) images such as its specular nature and acquisition variations can limit the direct use of TTE images in the development and generalization of deep learning models. Constrictive pericarditis (CP) and cardiac amyloidosis (CA) were chosen as examples of restrictive cardiomyopathy that are important but challenging entities to differentiate clinically and even hemodynamically for model development.

Hypothesis

The purpose of this study was to propose an automated framework to address the common challenges of echocardiography deep learning model generalization.

Methods

Patients with apical 4 chamber (A4C) view from TTE studies and a confirmed diagnosis of CP, CA, and normal cases from Mayo Clinic Rochester and Arizona were identified to extract baseline demographics. We proposed a novel pre-processing and image generalization framework to process the images for the training and comparison of three deep learning architectures (ResNet50, ResNeXt101, and EfficientNetB2). Finally, ablation studies were conducted to justify the effect of each proposed processing step in the final classification performance.

Results

The models were initially trained and validated on 720 unique TTE studies from Mayo Rochester, and further validated on 225 studies from Mayo Arizona. With our proposed generalization framework, EfficientNetB2 generalized the best with an average AUC of 0.96 (±0.01) and 0.83 (±0.03) on the Rochester and Arizona test sets, respectively.

Conclusion

In applying the proposed generalization techniques, we successfully developed an echocardiography-based deep learning model, which can accurately differentiate CP from CA and normal cases and applied the model to images from two sites. The proposed framework can be further extended for the development and generalization of any echocardiography-based deep learning models.

Figure(s)

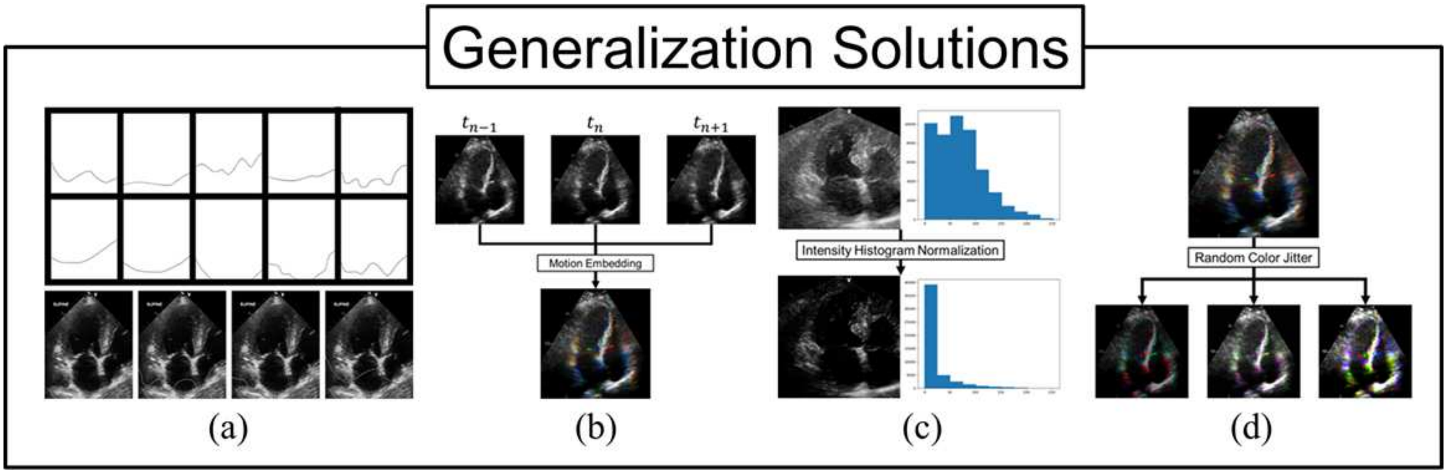

Figure 1. Four generalization solutions for the TTE challenges: (a) Ten hand drawn respiratory lines that were randomly sampled to augment the dataset with. The last row contains the original frame without respiratory lines (far left), and three randomly drawn respiratory lines applied on the original image. (b) Motion embedding using current, past and future frames and color in the image represents the motion in the current frame. (c) Preprocessing and augmentations. On the left is a figure of an image histogram matched to the training data. On the right are several examples of an image with color jitter.

Framewise Rochester Test									
	ResNet50			ResNeXt101			EfficientNet-B2		
Class	**Precision**	**Recall**	**F1 Score**	**Precision**	**Recall**	**F1 Score**	**Precision**	**Recall**	**F1 Score**
CA	**0.933±0.002**	**0.934±0.002**	**0.934±0.002**	0.922±0.002	0.922±0.002	0.922±0.002	**0.933±0.002**	**0.934±0.002**	**0.933±0.002**
CP	0.916±0.002	0.916±0.002	0.916±0.002	**0.951±0.002**	**0.951±0.002**	**0.951±0.002**	0.912±0.002	0.912±0.002	0.912±0.002
Normal	**0.931±0.002**	**0.931±0.002**	**0.931±0.002**	0.929±0.002	0.927±0.002	0.927±0.002	0.879±0.002	0.874±0.003	0.875±0.003
Studywise Rochester Test									
	ResNet50			**ResNeXt101**			**EfficientNet-B2**		
Class	**Precision**	**Recall**	**F1 Score**	**Precision**	**Recall**	**F1 Score**	**Precision**	**Recall**	**F1 Score**
CA	0.914±0.054	0.898±0.046	**0.905±0.038**	**0.918±0.051**	**0.902±0.043**	**0.909±0.034**	0.911±0.055	0.850±0.053	0.878±0.040
CP	**0.969±0.030**	**0.922±0.041**	**0.944±0.027**	0.940±0.042	0.921±0.042	0.930±0.031	0.911±0.056	0.894±0.048	0.901±0.040
Normal	**0.938±0.035**	**0.987±0.013**	**0.961±0.020**	0.920±0.043	0.957±0.026	0.938±0.026	0.906±0.049	0.971±0.020	0.937±0.029

Table 1. Tabular data showing the quantitative performance of the various model architectures on the Rochester test set in terms of precision, recall and F1-score. Optimal performance is highlighted in bold. 95% confidence interval added for both framewise and study wise results.

Keywords

Applications; Artificial Intelligence/Machine Learning; Imaging Research