Generalist deep learning for cross-modality landmark annotation in cardiovascular magnetic resonance

Ricardo A. Gonzales[1], **Carlos R. Suárez Gurruchaga**[1], **Ana L. Manrique**[1], **Daniel Atkinson**[1], **Vanessa M. Ferreira**[1], **Qiang Zhang**[1], **and Stefan K. Piechnik**[1]

[1]Oxford Centre for Clinical Magnetic Resonance Research (OCMR), Division of Cardiovascular Medicine, Radcliffe Department of Medicine, John Radcliffe Hospital, University of Oxford, Oxford, United Kingdom

Background

Cardiovascular magnetic resonance (CMR) enables detailed myocardial tissue characterisation through complementary modalities: quantitative T1 mapping for tissue properties, late gadolinium enhancement (LGE) for scar detection, and virtual native enhancement (VNE) as a contrast-free alternative[1]. Standardised analysis by the AHA 16-segment model requires precise annotation of landmarks such as the anterior right ventricular insertion and left ventricular centre, to ensure consistent segmental quantification. Manual annotation is time-consuming and subject to variability. Although recent deep learning approaches have advanced landmark detection, most have been developed for individual modalities and have not been extended to cover ShMOLLI T1 maps or VNE. This restricts their generalisability in clinical practice, where multiple modalities are routinely integrated. This study proposes a generalist deep learning framework that works robustly across CMR modalities, overcoming these limitations.

Methods

This retrospective study included 16,160 short-axis CMR images from 1,299 subjects across 20+ centres, spanning five modalities: native T1 (rest and stress), post-contrast T1, LGE (magnitude and PSIR), and VNE. A dual-stage ResNet-50v2 architecture[2] was developed: Stage 1 generated coarse predictions from preprocessed 384×384 images; Stage 2 refined predictions on spatially normalized 192×192 crops. The model was trained using data augmentation and evaluated on a held-out internal test set of 3,568 images from 257 subjects. Euclidean distance to manual annotations by four trained readers served as the performance metric. Model performance was compared with five specialist models (trained per modality) and one generalist model trained on the merged training data from all modalities. Inter-observer variability was assessed using an external test set of 42 native T1 images annotated by 11 independent observers.

Results

In the cross-modality testing (Figure 1A), the generalist model achieved an error of 2.3 mm (IQR: 1.3–4.5), outperforming all specialist models with errors smaller by a factor of 2.3 (2.0–3.7) overall. For in-distribution data, performance improved by a factor of 1.1 (1.0–1.1); for out-of-distribution data, generalist training improved results by 3.7 (2.4–4.9). In the external test set, inter-observer error was 1.3 mm (0.6–3.2), and the generalist model achieved 1.8 mm (1.1–3.0) on the same images. Accurate annotations (Figure 1B) were generated in 225 ms per sample in batch processing.

Conclusion

A generalist, dual-stage deep learning model can accurately annotate cardiac landmarks across multiple CMR modalities. It outperforms specialist models, especially in cross-modality settings, and approaches expert-level consistency. Its exposure to broader, more varied training data likely enhances generalisation and robustness, offering a scalable solution for standardised myocardial segmentation. This approach may extend to unseen or rare modalities with minimal tuning.

References

1. Zhang, Q. *et al.* Artificial intelligence for contrast-free MRI: Scar assessment in myocardial infarction using deep learning–based Virtual Native Enhancement. *Circulation* **146**, 1492–1503, DOI: https://doi.org/10.1161/CIRCULATIONAHA.122.060137 (2022).
2. Gonzales, R. A. *et al.* MVnet: automated time-resolved tracking of the mitral valve plane in CMR long-axis cine images with residual neural networks: a multi-center, multi-vendor study. *J. Cardiovasc. Magn. Reson.* **23**, 137, DOI: https://doi.org/10.1186/s12968-021-00824-2 (2021).

(A)

Test Train	**T1$_{\text{native}}$** (n=1,550)	**T1$_{\text{post}}$** (n=80)	**LGE$_{\text{mag}}$** (n=646)	**LGE$_{\text{PSIR}}$** (n=646)	**VNE** (n=646)	**All** (n=3,568)
T1$_{\text{native}}$ (n=4,966)	**2.0** (1.1-4.4)	9.6 (7.4-16.1)	11.4 (3.8-24.9)	9.7 (3.8-23.9)	10.8 (4.3-47.2)	4.7 (1.9-14.5)
T1$_{\text{post}}$ (n=740)	4.9 (3.0-8.7)	4.0 (2.8-6.7)	13.2 (5.7-24.8)	13.6 (5.4-26.3)	13.1 (7.0-25.7)	7.6 (4.0-17.9)
LGE$_{\text{mag}}$ (n=1,875)	6.0 (2.7-14.1)	6.3 (3.5-14.4)	3.0 (1.8-5.1)	6.3 (2.5-23.1)	4.8 (2.3-15.8)	4.8 (2.4-12.9)
LGE$_{\text{PSIR}}$ (n=1,990)	4.1 (2.1-10.7)	6.3 (3.7-17.7)	3.0 (1.8-5.4)	2.8 (1.7-4.7)	4.4 (2.0-13.8)	3.6 (2.0-8.6)
VNE (n=2,431)	13.7 (6.2-25.7)	15.1 (9.5-22.9)	10.5 (3.2-21.9)	10.5 (3.1-23.2)	2.1 (1.2-4.3)	9.0 (3.1-21.2)
All (n=12,002)	2.1 (1.2-4.5)	**2.5** (1.2-4.8)	**2.8** (1.6-4.8)	**2.6** (1.6-4.6)	**2.0** (1.1-3.9)	**2.3** (1.3-4.5)

(B)

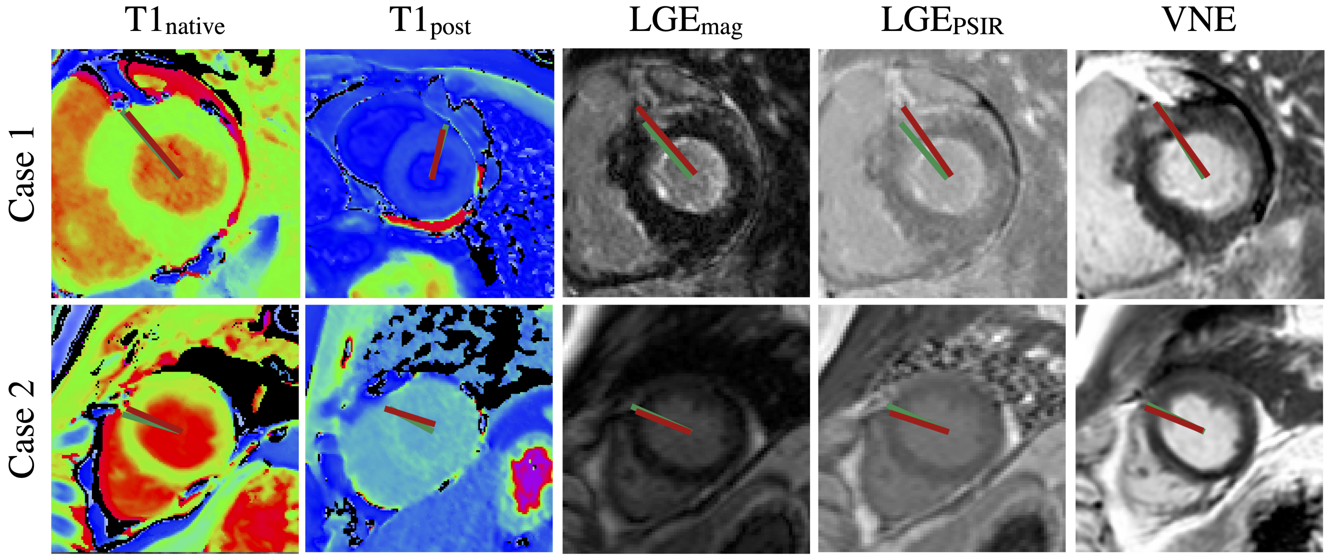

Figure 1. Generalist model performance and representative cross-modality predictions. (A) Euclidean error (mm, median [IQR]) and (B) examples showing predicted (red) vs reference (green) landmarks, in test set across five cardiovascular magnetic resonance modalities.