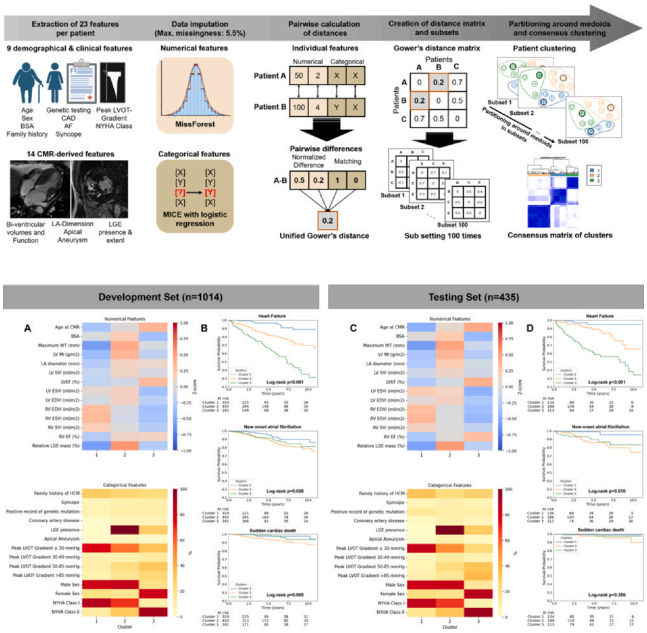
reported as frequencies with corresponding percentages and compared using the chi-squared test or Fisher's exact test, where applicable.

Abbreviations: BMI – body mass index; ICD – implantable cardioverter-defibrillator; NYHA – New York Heart Association; LVOT – left ventricular outflow tract; ACE – angiotensin-converting enzyme; ARB – angiotensin receptor blocker; LV/RV – left/right ventricle; LA – left atrium; WT – wall thickness; MI – mass index; EDVI/ESVI – end-diastolic/end-systolic volume index; SVI – stroke volume index; EF – ejection fraction; LGE – late gadolinium enhancement.



https://doi.org/10.1016/j.jocmr.2025.102263

ID#: 2211595

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

Ricardo A. Gonzales, PhD, FSCMR[1],
Carlos R. Suárez Gurruchaga, MSc[2], Ana L. Manrique, BSc[3],
Daniel Atkinson, MSc[4],
Vanessa M. Ferreira, MD, PhD, FSCMR[4],
Qiang Zhang, PhD, FSCMR[4],
Stefan K. Piechnik, PhD, FSCMR[4]

[1] *University of Oxford, Lima, Peru*
[2] *University of Oxford, Buenos Aires, Argentina*
[3] *University of Oxford, Piura, Peru*
[4] *University of Oxford, England, United Kingdom*

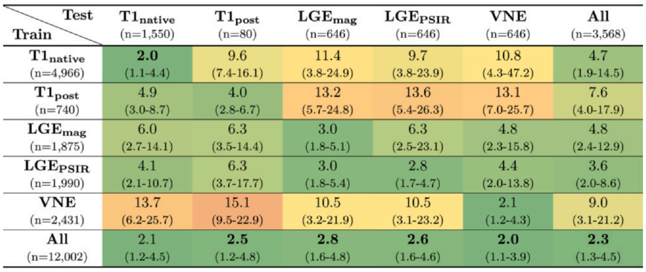
Background: Cardiovascular magnetic resonance (CMR) enables detailed myocardial tissue characterisation through complementary modalities: quantitative T1 mapping for tissue properties, late gadolinium enhancement (LGE) for scar detection, and virtual native enhancement (VNE) as a contrast-free alternative [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 1 A), 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.

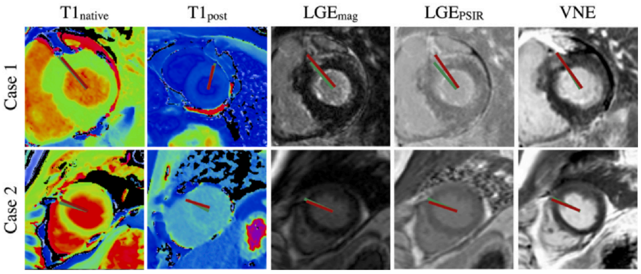
Conclusion: A generalist, dual-stage deep learning model can accurately annotate cardiac landmarks across multiple CMR modalities. It outperforms specialist models, especially in cross-modality settings, and approaches expert-level consistency. Its exposure to broader, more varied training data likely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

(A)



Test / Train	$T1_{native}$ (n=1,550)	$T1_{post}$ (n=80)	LGE_{mag} (n=646)	LGE_{PSIR} (n=646)	VNE (n=646)	All (n=3,568)
$T1_{native}$ (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)
$T1_{post}$ (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_{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_{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)



https://doi.org/10.1016/j.jocmr.2025.102264