Multi-Modal Supervised Contrastive Learning for Myocardial Infarction Classification

Charlotte Richardson*, Thalia Seale, Vicente Grau, Abhirup Banerjee

Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford, UK

Introduction: Myocardial infarction (MI) classification typically relies on a single modality. ECG-based approaches miss cases with normal morphology, whereas imaging metrics fail to capture the full cardiac cycle dynamics. A multi-modal approach could better characterise dysfunction.

Aim: To develop a novel multi-modal supervised contrastive learning (SCL) framework for MI classification that integrates the 12-lead ECG with 3D motion of the heart from cardiac magnetic resonance imaging (MRI).

Methods: 719 subjects from the UK Biobank were selected, with 15% labelled incident MI (iMI) and 22% prevalent MI (pMI). A β-total correlation variational autoencoder (β-TCVAE), pretrained on 10,646 patients and fine-tuned on UK Biobank data, encoded 12-lead ECGs. A β-VAE with a point completion network encoder is followed by a temporal VAE to represent biventricular motion. We compared three fusion strategies: latent concatenation, ensemble classification, and cross-modal SCL with embeddings projected to a shared space. All methods were evaluated using 5-fold cross validation on the entire dataset and reported on the global confusion matrix.

Results: Across fusion strategies, pMI F1 improved over the ECG baseline (0.42) by 14% for concatenation, 28% for ensemble, and 33% for SCL, achieving 0.55. Cross-modal SCL achieved the highest accuracy (0.73) and precision (0.48), while ensemble achieved the highest recall (0.69) and AUROC (0.73). The iMI gains were smaller, with F1 improvements of 11%, 9.0% and 9.9% respectively, though recall improved substantially under SCL (0.60), a 23% increase over MRI alone and 44% increase over ECG alone.

Conclusion: Multi-modal fusion of ECG and cardiac MRI consistently outperforms single-modality methods. Our SCL achieved the strongest performance, demonstrating that jointly encoding electrophysiology and mechanical motion captured complementary discriminative features.

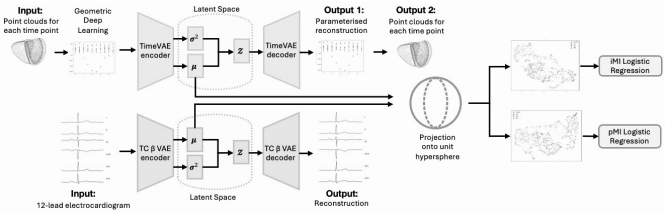

Figure 1. SCL model architecture.