Self-Supervised Representation Learning for ECG-based CTRCD Detection: A Comparative Study of Generative and Discriminative Paradigms

Natsu Suyama, Kazuko Tajiri, Takio Kurita, Yuki Ishizuka, Akira Furui

Hiroshima University
Higashi-Hiroshima, Japan

Backgrounds and Aims: Cancer therapy-related cardiac dysfunction (CTRCD) is conventionally diagnosed based on left ventricular ejection fraction measured by echocardiography. However, the accuracy of this approach depends heavily on operator expertise. ECG offers a more accessible alternative, but the scarcity of labeled CTRCD data limits the applicability of supervised deep learning. We aimed to evaluate the effectiveness of self-supervised learning (SSL) for ECG-based CTRCD detection, leveraging large-scale unlabeled ECGs to overcome label scarcity.

Methods: We compared two representative SSL paradigms: a generative method (masked autoencoders, MAE) and a discriminative method (momentum contrast, MoCo). Both methods were pre-trained on approximately $120{,}000$ unlabeled 12-lead ECG recordings. For the downstream CTRCD detection task, we used a separate labeled dataset of $1{,}262$ ECGs (107 positive, 1,155 negative), evaluating each method under fine-tuning and linear probing. A supervised baseline without pre-training was included for comparison. We further analyzed the learned representations via principal component analysis (PCA) before and after fine-tuning.

Results: SSL pre-training consistently improved detection performance: the AUROC increased from 0.80 (supervised baseline) to 0.88 (MAE with linear probing), and sensitivity improved from 0.27 to 0.55 (MAE with fine-tuning). PCA revealed qualitatively distinct representation trajectories: MAE acquired a structured representation of general ECG characteristics during pre-training, which was subsequently re-aligned toward the CTRCD axis through fine-tuning. In contrast, MoCo acquired task-relevant structure primarily during downstream adaptation. Despite these differences, both paradigms yielded well-separated CTRCD-positive and -negative clusters after fine-tuning.

Conclusion: SSL pre-training is an effective strategy for ECG-based CTRCD detection under label scarcity. Generative and discriminative paradigms reach comparable performance through qualitatively distinct representation trajectories, suggesting potential complementarity for future integration.