

# Enhancing CVD Prediction through Multi-Modal SSL

Francesco Girlanda<sup>1</sup>, Olga Demler<sup>1,2</sup>, Bjoern Menze<sup>3</sup>, Neda Davoudi<sup>1,3,4</sup>

<sup>1</sup>Department of Computer Science, ETH Zurich, CH, <sup>2</sup>Brigham and Women's Hospital, Harvard Medical School, USA,

<sup>3</sup>Department of Quantitative Biomedicine, University of Zurich, CH, <sup>4</sup>ETH AI Center, ETH Zurich, CH

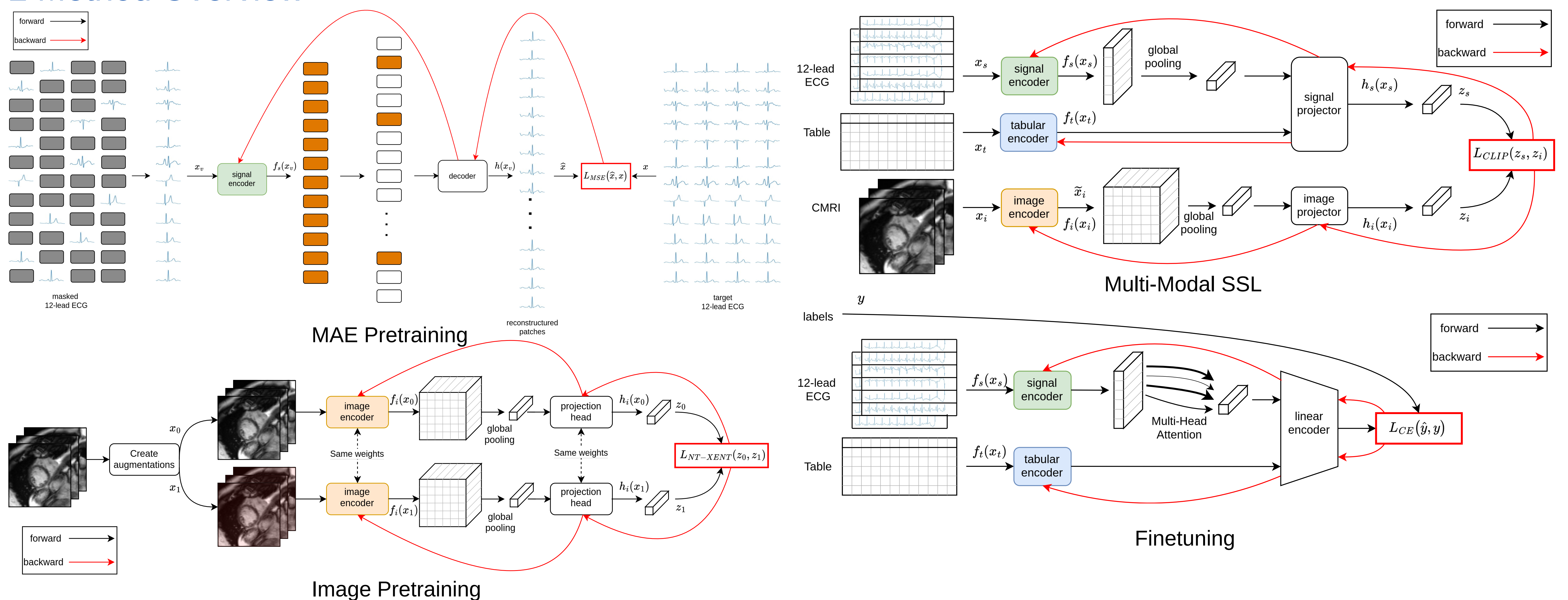
## 1 Introduction

Accurate prediction of cardiovascular diseases (CVD) is crucial for timely diagnosis and intervention. Traditional models focus on single data modalities or assume labeled data. This paper introduces:

- multi-modal approach that combines CMR images, ECG data, and medical records to provide a more comprehensive understanding of cardiovascular health;
- self-supervised pretraining make use of huge unlabeled data;
- a multi-modal contrastive learning objective is to transfer knowledge from expensive CMR to cheap and simple modalities such as ECG and tabular information;
- a finetuning step pre-trained encoders on specific predictive tasks using a small amount of labels.

Our proposed method outperformed the supervised approach by 7.6% in balanced accuracy with the same amount of labels.

## 2 Method Overview



## 3 Results on MI

SSL Modality	MAE pretr.	Image pretr.	ECG Train	Tab. Train	CMRI Train	AUC [%]	Bal. Acc [%]
Tabular	✗	✗	✗	✓	✗	0.62	0.5
ECG	✗	✗	✓	✗	✗	0.68	0.6
ECG	✓	✗	✓	✗	✗	0.76	0.69
CMRI	✗	✗	✗	✗	✓	0.74	0.67
ECG (MMCL [69])	✓	✓	✓	✗	✓	0.77	0.7
ECG + Tabular (Ours)	✓	✓	✓	✓	✓	<b>0.80</b>	<b>0.71</b>

Modality	ECG	Tabular	AUC [%]	Balanced Acc [%]
Supervised NN	✗	✓	0.55	0.54
Supervised NN	✓	✗	0.71	0.66
Supervised NN	✓	✓	0.70	0.65
Ours	✓	✓	<b>0.80</b>	<b>0.71</b>

## 5 Conclusion

In summary, this paper introduces:

- a novel framework for predicting CVD using limited labels by transferring knowledge from CMR images to ECG signals and clinical data on big unlabeled data;
- a model that learns informative representations that are robust and generalizable, ultimately leading to enhanced predictive accuracy for several downstream tasks;
- the usage of SSL techniques that mitigate the need for large annotated datasets, making our approach scalable and applicable to real-world clinical settings;
- a demonstrated improvement of 7.6% in balanced accuracy over supervised methods using only ECG and tabular data, proof of the fact that our pipeline progressively makes the model learn richer embeddings thanks to the transfer learning from expensive modalities.

## References

- [1] Chen et al. A simple framework for contrastive learning of visual representations.
- [2] He et al. Masked autoencoders are scalable vision learners.
- [3] Sudlow et al. Uk biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age.
- [4] Turgut et al. Unlocking the diagnostic potential of ecg through knowledge transfer from cardiac mri.