CSCI 5980/8980: Machine Learning for Healthcare
Final Project Presentation - Fall 2024

Self-Supervised Representation Learning on ECG Signals via Dual-View Transformers

Group 12; Shady Ali

Intro/abstract

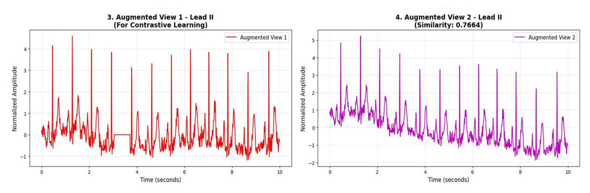
ECG labels are rare & imbalanced in datasets. We evaluated a *self-supervised Dual-View Transformer* on **PTB-XL** dataset, comparing three augmentation strategies. All methods displayed **limited & varying** embedding space separation and downstream classification performance. These results highlight the critical importance of preprocessing in self-supervised settings and the significant impact of augmentation choice on model convergence.

Main Contributions of the Project

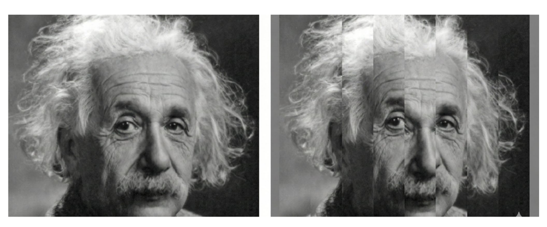
- Testing pure contrastive training with different types of augmentations and/or Random Lead Masking (RLM)[1].
- Testing Dual-View on channels (Channel Tokens) & Timesteps (Time Tokens).

Dataset & Preprocessing

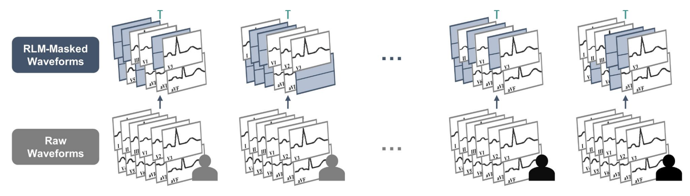
PTB-XL: 21,799 12-lead ECGs with 5 Superclasses: Normal, Myocardial Infarction, ST/T Change, Conduction Disturbance, Hypertrophy. Removed multi-label ECGs



1. Normalize.
2. Sample a pool of augmentations to apply sequentially (Noise, Crop, etc.)
3. Keep full signal length.

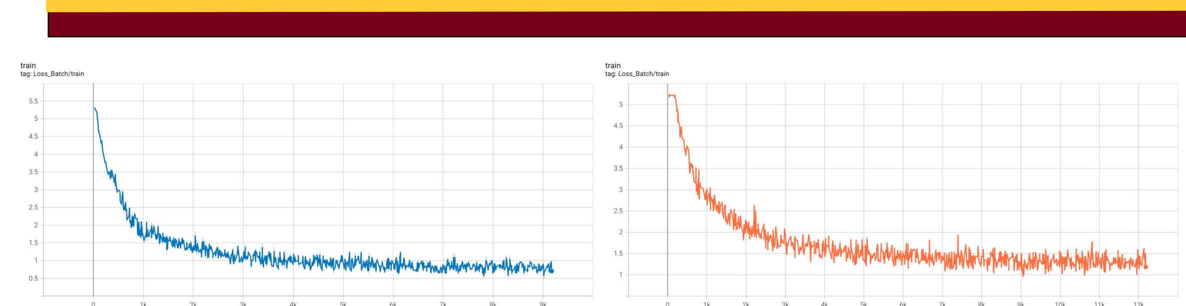


1. Constructing views by subsampling even vs. odd timesteps to preserve global context.
2. Normalize per sample.
3. Mask random channels[1].



1. Temporal Partitioning to two 5-second segments from each ECG[3].
2. Normalize per sample.
3. Mask random channels[1].

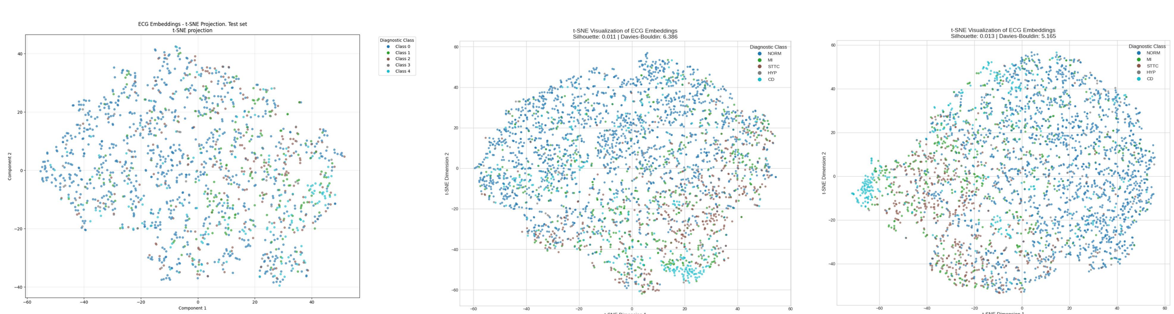
Results



Loss per batch:

Interleaving (**blue**) & non-overlapping (**orange**).

- Loss on test is similar.



Embedding space (L to R):

1. Sequential Augmentations.
2. Interleaving Segments.
3. Non-Overlapping Segments.

Cross-Attention Map:

Sequential Augmentations Interleaving Segments Non-Overlapping Segments

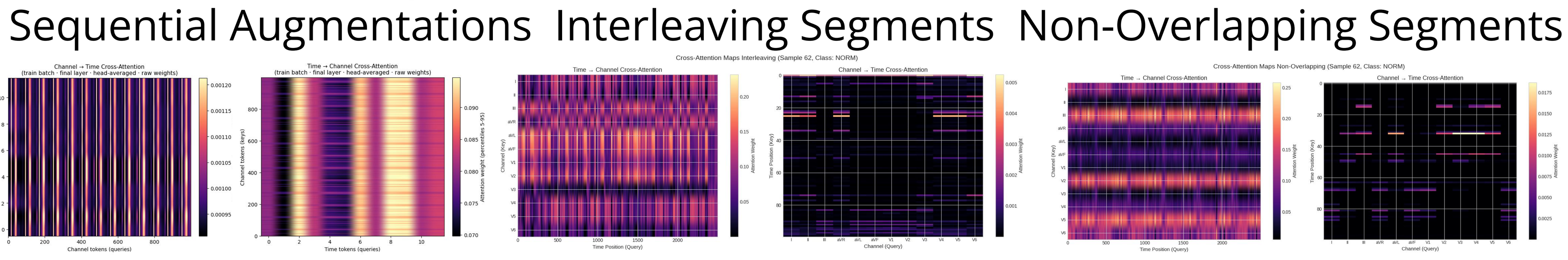


Table 1: Comparison of Model Performance (F1-Score)

Class / Metric	Sequential Augmentations	Interleaved Segments	Non-Overlapping Segments
Normal	0.78	0.82	**0.83**
Myocardial Infarction	0.07	0.38	**0.42**
ST/T Change	0.33	0.57	**0.63**
Hypertrophy	**0.11**	0.00	0.01
Conduction Disturbance	0.53	0.60	**0.66**
Overall Accuracy	0.63	0.70	**0.72**
Macro F1-Score	0.36	0.47	**0.51**

Linear-Probing F1-Score across the 3 methods.

- **Non-Overlapping is better**, but not by a lot from Interleaving.
- A **ResNet-18 Baseline** had similar/lower results as well.

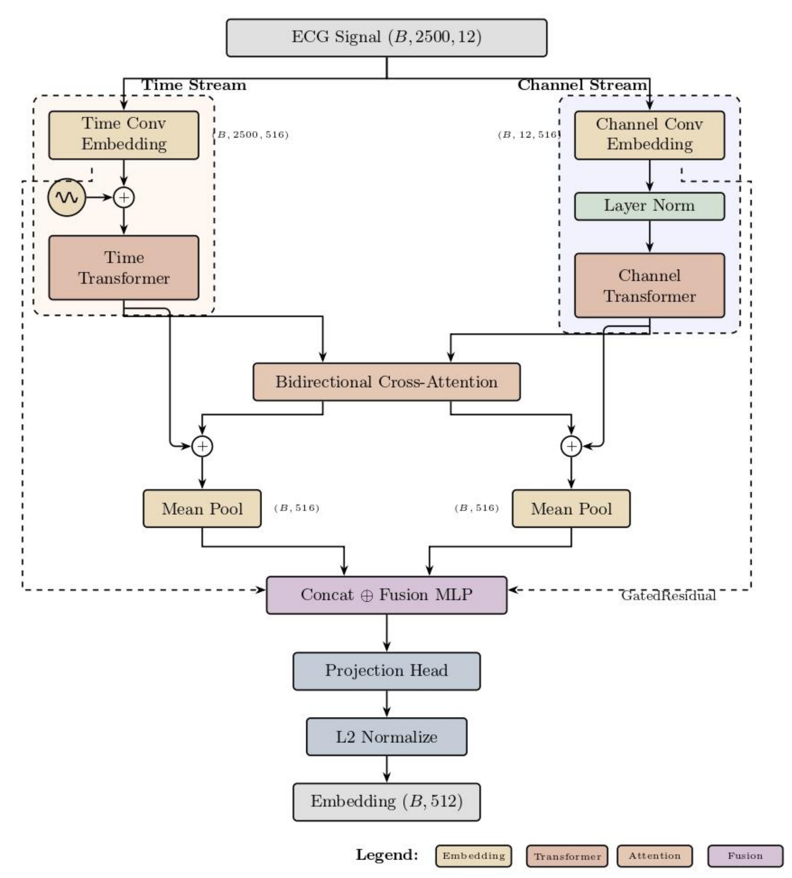
Background

- Electrocardiograms (ECGs) are commonly used for cardiovascular diagnosis, but manual annotation is time-consuming, error-prone, and requires specialized cardiologists.
- Supervised models require massive labeled datasets, & annotated medical data is a lot less and imbalanced.

Current unsupervised approaches include contrastive + generative objective (Transformer & CNN)[2], contrastive objective with an augmentation family[3], and a mixed contrastive objective (patient identification + cardiac arrhythmia classification.)[1]

Proposed Approach & Evaluation

Self-Supervised Contrastive Learning (SSCL) to create representations that are class-separable in the embedding space.



Training:

- Trained on 2x Nvidia GB10 (256GB VRAM) for 10 hours
- 5 epoch warm start.
- NT-Xent loss[4].

Evaluation:

- Clustering Tendency in Embedding space.
- Attention Map Analysis.
- Downstream classification (Linear Probing & Fine-Tuning).

Discussion & Future Directions

Discussion

- Random augmentions cancel a lot of learnable information
- Time tokens do not hold much **new** information for the Channel tokens to attend to.
- Channel tokens do have more information for Time to attend to.
- ECG segmentation with *RLM* more effective than random augmentations.
- Scaling the data further can help analyze if Interleaving holds potential to capture better global context.

Future Directions

- Deeper analysis on why types of augmentations work or don't.
- Scaling to distributed training to use larger batch sizes for contrastive training (**big limitation**).
- Seeing how different augmentations scale with data size and model size.

1. Oh, J., Chung, H., Kwon, J. M., Hong, D. G., & Choi, E. (2022, April). Lead-agnostic self-supervised learning for local and global representations of electrocardiogram. In Conference on Health, Inference, and Learning (pp. 338-353). PMLR.
2. McKeen, K., Masood, S., Toma, A., Rubin, B., & Wang, B. (2025). Ecg-fm: An open electrocardiogram foundation model. JAMIA open, 8(5), ooaf122.
3. Kiyasseh, D., Zhu, T., & Clifton, D. A. (2021, July). Clocs: Contrastive learning of cardiac signals across space, time, and patients. In International Conference on Machine Learning (pp. 5606-5615). PMLR.
4. Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020, November). A simple framework for contrastive learning of visual representations. In International conference on machine learning (pp. 1597-1607). PmLR.