ECG Signal Data Augmentation Strategy

Shared by Jennifer

J

I am working on training a deep learning model that will correctly classify ECG signals ouput from a prototype ECG device I am building. I have very small ouput data from this prototype ECG device. But I have 100,000 patient data from hospital 12-signal ECG machines that are publicly available, such as PTB-XL and SPH.

Let's say I want to use these public 12-lead ECG dataset, to create training and test data for a ResNet + SE model to classify ten different types of heart conditions. But instead of hospital-style 12-signal measurements with conductive gel, the data comes from my prototype that has leads embedded into a elastic shirt. It's likely that there will be some displacement of each lead in the shirt from where the lead would be if placed using each patient's fiducials. Also, there will be added noise due to body & shirt movement. How would you augment the public and test data to approximate the displacements and noise from the prototype. Would there be a much larger augmented dataset (using the public dataset as the basis)?

This is a fascinating domain adaptation problem where you need to bridge the gap between high-quality hospital ECG data and your prototype wearable device. Here's a comprehensive approach to data augmentation that should help your model generalize to your shirt-based ECG system:

Lead Displacement Augmentation

Spatial Transformations:

- Apply small rotations (±5-15°) to individual leads to simulate electrode misplacement
- Use time-domain stretching/compression (±5-10%) to simulate different chest geometries
- Add inter-lead timing delays (±10-50ms) to simulate asynchronous sampling from displaced electrodes
- Apply amplitude scaling per lead (0.8-1.2x) to account for varying contact quality

Anatomical Variation Simulation:

- Create lead-specific transformation matrices based on typical displacement patterns in shirt-based systems
- Use morphological transformations to simulate how signals change with electrode position relative to heart anatomy
- Apply different scaling factors for precordial vs limb leads since shirt-based systems primarily affect chest leads

Noise and Artifact Augmentation

Movement Artifacts:

- Add low-frequency drift (0.05-2 … o simulate baseline wander from breathing and movement