Self-Supervised Learning for ECG-based Emotion Recognition

1 Implementation Details

1.1 Network Architecture

1.1.1 Signal Transformation Recognition Network

The network architecture consists of three convolutional blocks and two dense layers. The convolutional layers are shared among tasks, while the dense layers are task-specific. Each convolutional block includes:

- Two 1D convolutional layers with ReLU activation.
- A max-pooling layer (size: 8).
- The number of filters increases from 32 to 128 across blocks.
- Kernel sizes reduce from 32 to 8 after each block.

The final layers consist of global max pooling, followed by task-specific dense layers. The dense layers use dropout (60%) and L2 regularization (with $\beta = 0.0001$).

1.1.2 Emotion Recognition Network

This network reuses the convolutional layers of the signal transformation recognition network. Two dense layers with 64 hidden units and a sigmoid activation follow. The transferred convolutional layers are frozen, and only the dense layers are trained using the ECG signals and emotion labels.
A Self-supervised learning framework was employed, using ECG data to pre-train a convolutional neural network (CNN). The model consists of convolutional layers followed by batch normalization and ReLU activation. The architecture also includes a fully connected layer for emotion classification.

1.2 Pretext Task

For self-supervision, a contrastive learning approach was used. The pretext task was defined by generating augmented views of the ECG signals

1.3 Fine-tuning

After the pre-training, the model was fine-tuned on a labeled dataset for emotion recognition using categorical cross-entropy loss and stochastic gradient descent (SGD) optimization.

2 Dataset Description

The dataset used for this project is the *DREAMER* dataset which is a multi-modal database comprising of electrocardiogram (ECG) signals with **Sampling Rate:** 256 Hz from 23 participants recorded during audio-visual stimuli for affect elicitation. Participants self-assessed their emotions in terms of valence, arousal, and dominance.

3 Results