Summary

This paper introduces a novel Bimodal Masked Inverse Reconstruction-based Contrastive learning model (BMIRC) for self-supervised learning on electrocardiogram (ECG) signals. BMIRC enhances representation learning by jointly encoding both time and frequency modalities, achieving strong performance across multiple ECG datasets. The key contributions are:

1. **Introduction of a bimodal structure**: The paper proposes a dual-modal autoencoder framework that leverages both time and frequency modalities to improve feature extraction from ECG data.

2. **Enhanced masking strategy**: The model optimizes reconstruction and classification performance through a strategic combination of different masking rates for the time and frequency modalities.

3. **Combination of self-supervised and transfer learning**: By applying pre-training and fine-tuning across different datasets, the paper demonstrates the model's strong generalization ability in cross-domain tasks.

4. **Ablation experiments and visualization analysis**: Comprehensive ablation studies confirm the effectiveness of individual components, and visualizations illustrate the model's superior reconstruction capabilities and representation learning.

These contributions advance self-supervised learning approaches for ECG data and present significant potential for practical applications.

Positives

1. **Innovative Bimodal Approach**: The introduction of a bimodal structure that combines time and frequency modalities for self-supervised learning is novel and effective, resulting in improved representation learning for ECG data.

2. **Strong Performance**: The model consistently outperforms baselines across multiple datasets, especially in intra-domain settings, demonstrating its efficacy in both accuracy and generalization ability.

3. **Transfer Learning Success**: The use of pre-training on larger datasets and fine-tuning on smaller datasets highlights the model's ability to generalize well in cross-domain scenarios, making it valuable for real-world applications with limited labeled data.

4. **Detailed Ablation Studies**: The paper provides thorough ablation experiments that validate the contribution of each component, such as the importance of frequency modality and the Specific-Shared architecture.

5. **Comprehensive Visualization**: The visualization analysis, particularly using CKA similarity, offers insights into the model's internal workings and strengthens the argument for the effectiveness of its design, especially with the IRC module.

Comments

1. The author clearly describes the datasets used, but further details on the quality of the datasets would be beneficial, such as the sampling criteria, recording methods, and noise levels in each dataset. Additionally, information on data cleaning and exclusion criteria (e.g., excluding abnormal ECGs) could be expanded upon.

2. In the data preprocessing section, while steps such as denoising, down-sampling, and normalization are mentioned, it would be helpful to elaborate on how these steps impact model performance, especially regarding the choice of different filtering methods (e.g., Butterworth filter) and their effectiveness across different datasets.

3. Regarding the multi-label classification problem in ECG data, the author mentions label reconstruction but provides limited details on handling class imbalance. It is recommended to elaborate on how balancing strategies are implemented and include comparative experiments to demonstrate their effects.

4. It is recommended to further expand the discussion of the experimental results, particularly regarding the performance differences across different comparative methods. Could these differences be attributed to specific characteristics of the datasets (e.g., noise, class imbalance) or model architecture (e.g., the number of transformer blocks)?

5. The author mentions that all experiments were conducted on an RTX 2080-Ti GPU but does not provide information on training time or resource consumption. It is recommended to include details about the training time and computational efficiency of different models on the same hardware.

6. The experimental results include ACC, AUPRC, and other metrics across multiple tables. To enhance clarity, it is recommended to add graphical representations (e.g., bar charts or line graphs) alongside the tables. Furthermore, the table captions or footnotes could more clearly indicate the specific experimental conditions (e.g., pre-training status, dataset size).

7. In the section on varying the training set proportions, it would be beneficial to further explore how different training set sizes impact the model's generalization ability. The observation that BMIRC performs better on 50% of the dataset than a randomly initialized model trained on 100% of the dataset is particularly noteworthy and warrants a more detailed discussion.

8. While BMIRC outperforms other models in most experiments, it still exhibits limitations in certain scenarios, such as underperforming TS-TCCR on the Chapman dataset. It is recommended to elaborate on these limitations in the conclusion and discuss potential directions for improving the model.

9. BMIRC's performance on ECG datasets is impressive, but its ability to generalize to other time-series tasks is not explicitly discussed. The authors are encouraged to briefly discuss the potential applications of BMIRC to other time-series tasks, such as speech signals or stock market data, in either the experimental or conclusion sections.