A Large Language Model-based Framework for Generalizable and Unified Physiological Signal Analysis

Yuer Ma[1,2], Haoxuan Qu[1], Hongbin Xu[3], Wenchang Chai[4], Zhiyuan Luo[2], Yufeng Zhang[2], Wenxiong Kang[2*], Jun Liu[1*]

[1*]School of Computing and Communications, Lancaster University, Bailrigg, Lancaster, LA1 4YW, United Kingdom.
[2]School of Automation Science and Engineering, South China University of Technology, Wushan Road, Guangzhou, 510641, Guangdong, China.
[3]ByteDance Seed, Zhongguancun Street, 100190, Beijing, China.
[4]Department of Computing, The Hong Kong Polytechnic University, Yuk Choi Road, 999077, Hong Kong, China.

*Corresponding author(s). E-mail(s): auwxkang@scut.edu.cn; j.liu81@lancaster.ac.uk;
Contributing authors: y.ma39@lancaster.ac.uk; h.qu5@lancaster.ac.uk; hongbinxi1013@gmail.com; wenchang.chai@connect.polyu.hk; zhiluo980@gmail.com; auyfzhang@mail.scut.edu.cn;

Abstract

Physiological signals are essential for clinical diagnosis and health monitoring. In recent years, many methods have been developed to automate their analysis. However, most methods are designed for specific tasks or modalities and rely on abundant labeled data. In practice, real-world clinical scenarios are more diverse. For instance, rare disease diagnose typically involves limited labeled data, requiring models to generalize from minimal supervision. In addition, integrated data repositories (IDRs) in modern hospitals contain abundant physiological signals across diverse modalities and diagnostic objectives. In this scenario, task and modality diversity makes it costly and inefficient to develop and maintain separate models for each, highlighting the need for a unified framework with high reusability and low deployment overhead. These two representative settings, (1) data-scarce scenario and (2) heterogeneous-task scenario, remain underexplored by current state-of-the-art methods. To investigate this gap, we evaluated several