Topology-Informed Machine Learning Models for ECG Classification

John Rick Manzanares[1, 2],
joint work with Paul Samuel Ignacio[2], Jay-Anne Bulauan[2]

[1] *Dioscuri Centre of Topological Data Analysis, Institute of Mathematics of the Polish Academy of Sciences, 00-656 Warsaw, Poland*
[2] *Department of Mathematics and Computer Science, University of the Philippines - Baguio, Baguio City, Philippines 2600*

In this talk, we present a method for multi-class classification of twelve-lead electrocardiograms (ECGs) developed as part of the 2020 Computing in Cardiology Challenge, focusing on classifying ECG abnormalities. This approach was among the first to incorporate topological features, derived through persistent homology, into a machine learning model for ECG classification. ECGs are modeled as multivariate time series, with different segments and lead groupings transformed into point cloud embeddings that retain both local and global structures, encoding periodic information as attractor cycles in high-dimensional space. By applying persistent homology to these embeddings, we generate topological features, such as persistence diagrams and landscapes, which capture the complex temporal dynamics of ECG signals. To improve predictive power, we combine these topological features with demographic data and RR interval statistics from the Pan-Tompkins algorithm. Our two-level random forest model first selects significant features from this combined set, and the second stage uses them for final classification. This integration of topological, demographic, and statistical information demonstrates the potential of topological data analysis in medical signal processing. While the model enhances classification for data similar to the training set, it faces challenges in generalizing effectively to a hidden test set. Lastly, we discuss recent extensions and refinements inspired by our approach that aim to further improve model performance in this challenging task.