

# A Topology Informed Random Forest Classifier for ECG Classification

Jay-Anne Bulauan, Paul Samuel Ignacio, John Rick Manzanares

University of the Philippines Baguio  
Baguio City, Philippines

We leverage mathematically computable topological signatures of 12-lead electrocardiograms (ECGs) as proxy for features informed by medical expertise to train a random forest model in a 9-class classification task. As has been shown for detecting Atrial Fibrillation using single-lead ECGs, this approach verifies the existence and viability of signal in the topology of ECGs for diagnosing cardiac conditions. Upscaling this to the use of all 12 leads of a standard ECG to diagnose multiple heart conditions improves accessibility to automated diagnostics by reducing expert-dependent input in feature extraction.

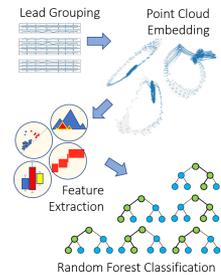
We view ECGs as multivariate time series data and convert different segments and groupings of leads to point cloud embeddings. This stores both local and global structures of ECGs, and encodes periodic information as attractor cycles in high-dimensional space. We then employ topological data analysis on these embeddings to extract topological features based on different summaries (persistence barcodes, landscapes, entropy) available in the literature. We supplement these features with demographic data and statistical moments of RR intervals based on the Pan-Tomkins algorithm for each lead to train the classifier.

We rank features by importance and narrow down a pool of 365 features to obtain an initial shortlist of 56 features of which 2 are demographic, 31 are RR-statistics, and 23 are topological. We use the shortlisted features to train a random forest classifier with 800 trees, and perform a stratified 10-fold cross validation using the Challenge metrics. A summary of the cross validated training performance appears in the table below.

Challenge Metrics on a Stratified 10-Fold Cross Validation

	AUROC	AUPRC	Acc.	$F_1$	$F_2$	$G_2$
Mean	0.876	0.585	0.833	0.473	0.590	0.287
Std. Dev.	0.007	0.031	0.006	0.016	0.017	0.014

Testing performed by the Challenge organizers on the trained classifier during the unofficial phase yields an  $F_2$  score of 0.577 and a  $G_2$  score of 0.274 for a geometric mean of 0.398.



Classification Pipeline.