Improving the Predicting Catheter Ablation Outcomes in Persistent Atrial Fibrillation using Semi-Supervised Self-Training Technique.

Noor Qaqos, Fernando S Schlindwein, G André Ng, Xin Li
University of Leicester, Leicester, UK

Introduction:
The ablations of AF targets in persistent atrial fibrillation (persAF) using dominant frequency (DF), rotors, and complex fractionated atrial electrograms (CFAEs) have been disappointing. Machine learning tools with novel electrogram (EGM) features from the three EGM signal domains (spectral, temporal, and statistical) by utilizing the labeled and unlabeled EGMs data might enhance the prediction of catheter ablation outcomes.
Methods:
20480 non-contact EGMs were collected from the left atrium of 10 persAF patients using the Ensite Array (Abbot, USA). 51 high dominant frequency locations (3206 nodes (EGMs)) were ablated. 1490 EGMs were labeled as positive ablation responses (AF termination or AF cycle length (AFCL) increased (≥10msec), whereas 1716 EGMs had negative responses (AFCL increase (<10msec))) to catheter ablation. 390 features were extracted using a time series feature extraction library embedded in a Python environment, followed by feature selection and importance stages to sequence the features based on their importance in the prediction process. Decision Tree (DT) classifier was used for the training and testing processes. Subsequently, the semi-supervised self-training was used to take advantage of unlabeled (17274 EGMs) for improving the DT model performance by considering the differences among patients.
Results:
The 10-fold cross-validation accuracy and AUC for supervised learning were 72%, and 0.71, respectively; while for semi-supervised self-training were around 74%, and 0.73, respectively based on the testing dataset.
Conclusions:
Semi-supervised self-training played an important role in improving the DT model performance in prediction process. This technique improved the model by 2% and 0.02 for accuracy and AUC, respectively. Further study focusing on labeled and unlabeled data from other patients should be considered.