Semi-Supervised Learning for Enhancing Ablation Outcomes in Persistent Atrial Fibrillation

Noor Qaqos[1], Fernando S Schlindwein[1,3], Ekenedirichukwu N Obianom[2], Shamsu I Abdullahi[1], Fan Feng[1], Abdulhamed M Jasim[1], G André Ng[2,3], Xin Li[1,3]

[1]University of Leicester, School of Engineering, Leicester, UK
[2]University of Leicester, Department of Cardiovascular Sciences, Leicester, UK
[3]National Institute for Health Research Leicester, UK

Abstract

Introduction: The outcome of ablations of AF targets in persistent atrial fibrillation (persAF) using dominant frequency (DF), rotors, and complex fractionated atrial electrograms (CFAEs) has been disappointing. Machine learning tools with novel electrogram (EGM) features from the three EGM domains (spectral, temporal, and statistical) by utilizing the labeled and unlabeled EGMs might enhance ablation outcomes. Methods: 20480 non-contact EGMs were collected from the left atrium of 10 persAF patients using the Ensite Array. 51 high dominant frequency locations (3206 nodes (EGMs)) were ablated. 1490 EGMs were labeled as positive 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, followed by feature selection and importance stages. Supervised learning (SL) was used to train and test the decision tree (DT) model. Subsequently, semi-supervised self-learning (SSSL) was used to take advantage of unlabelled data (17274 EGMs). Results: The accuracy and AUROC for SL were 72%, and 0.71; while for SSSL were 75.22%, and 0.73, respectively. Conclusions: SSSL played a significant role in increasing the model performance by 3.22% and 0.02 for accuracy and AUROC, respectively.

1. Introduction

Atrial fibrillation (AF) is the most common heart rhythm disorder, affecting about 1–2% of the general population. AF increases the risk of stroke up to 5-fold. Several strategies have been used to restore, control and maintain the sinus rhythm to terminate the AF such as antiarrhythmic medication, cardioversion, and catheter ablation. Catheter ablation effectively restores sinus rhythm in patients with paroxysmal AF, but it is less effective with persistent AF (persAF) and long-standing AF [1]. Ablation procedures have used individual features, including dominant frequency (DF), complex fractionated atrial electrograms (CFAEs), and rotors. Outcomes using these features have been suboptimal. Researchers have been seeking other features that might give better ablation outcomes. Features from the three electrogram domains (spectral, temporal, and statistical) were used to study and analyze the catheter ablation outcomes using supervised machine learning models. An acceptable performance was obtained from the labeled data [2, 3]. Most of the biomedical applications based on artificial intelligence tools suffer from not having sufficient label data which is time-consuming process, requires a big effort from clinicians to label a huge number of data, and also the most important point is cumbersome to obtain In contrast, unlabeled data are available cheaply [4]. Semi-supervised learning was superior to supervised learning techniques in many biomedical applications including AF [5, 6], by taking advantage of the label and unlabeled data in the classification process. In this work, we aim to investigate the performance of the DT classifier using SSSL in predicting the catheter ablation outcomes using new features from the three intracardiac signal domains by leveraging unlabeled EGMs data to train the ML model.

2. Materials and Methods

2.1. Dataset and Electrophysiology Study

A total of 20480 non-contact EGMs were collected from the left atrium (LA) of 10 persAF patients using an Ensite array catheter (Abbot, USA) (Figure 1). 51 locations were identified as high dominant frequency (HDF) regions to guide the catheter in the ablation procedure. The EGM signals for 20s duration were collected pre- and post-ablations. A total of 3206 out of 20480 EGMs from the 51 locations were labeled by cardiologists from the Leicester Glenfield Hospital into two classes: a positive response to catheter ablation: (AF termination or AF cycle length >=10ms; and negative

 ISSN: 2325-887X DOI: 10.22489/CinC.2024.297

responses: AFCL increased <10ms) [7]. Four out of ten patients had AF termination (3 flutter and one sinus rhythm) before the PVI procedure.

2.2. AF Signal Processing

The 20-second EGM signal was sampled at 2034.5Hz, and then resampled at 512 Hz using the cubic interpolation method to reduce the memory allocation and reduce the processing time. QRST subtraction was performed on each of the EGM signals to remove the far-field effect from ventricular that may affect the frequencies and other characteristics of the EGM signals. Figure 2 shows the positive and negative EGM responses to catheter ablation and the QRST subtraction process taking the ECG lead I as the reference.

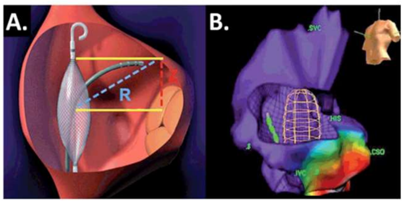


Figure 1. A. EnSite array catheter inserted in the atrium; B. EnSite array catheter shown on EnSite 3000 mapping system.

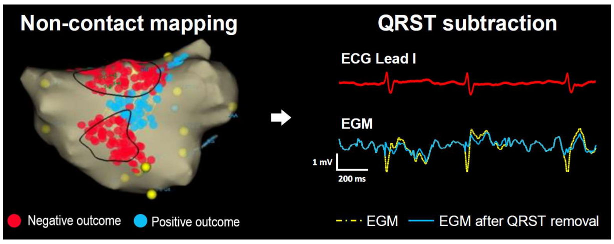


Figure 2. Left-hand side: LA regions where ablation had resulted in negative (red dots) and positive (blue dots) outcomes. Right-hand side: illustration of the QRST subtraction in noncontact EGMs (modified from [4]).

2.3. EGM Feature Extraction

A time series feature extraction library was used to extract 390 features from the three EGM signal domains (spectral, temporal, and statistical) as explained and identified previously in [3, 8]. These features were used to train and test the ML models.

2.4. EGM Feature Selection

The feature selection stage is extremely important to select the best features to train and test the machine learning models as features with high correlation and low variance might bias the model and will not give any benefit in the classification process. A threshold of 80% using the Pearson correlation method and a zero threshold were selected to remove highly correlated and low variance features, respectively.

2.5. EGM Feature Importance

Gini importance was used to order and sequence the features based on their importance in the prediction process using the decision tree (DT) model. Figure 3 reveals the top-ranking features for the classification process.

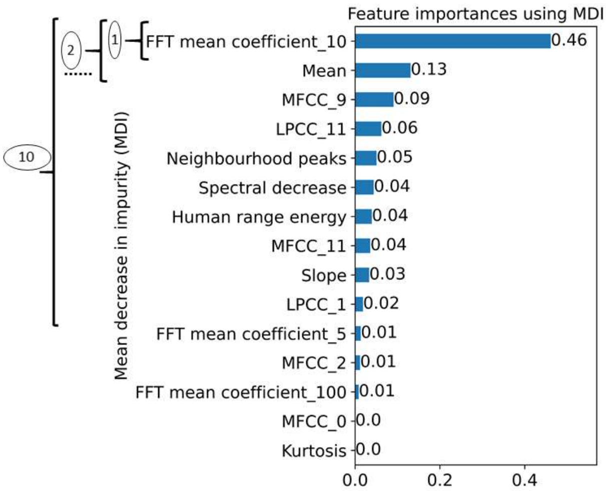


Figure 3. The top-ranking features, where (1,2, …., 10) are the feature sets1, 2, ……, 10; respectively.

2.6. SL and SSSL classification process

Ten sets of features were used to train and test the DT classifier (Figure 3). The highest performance (cross-validation accuracy of 72% and AUROC of 0.71) were obtained from our previous work using only the first important features (PSD10), which was based on the SL method in the training and testing of the model [8]. In this work, SSSL was used to take advantage of labeled data (3206 labeled EGMs) and the leverage of unlabeled data (2048*10 – 3206 = 17274 unlabeled EGMs). The DT model was built first with the label EGMs, and then used to label unlabeled EGMs. The new pseudo-label EGMs were added to the dataset to train the model again. This process was repeated on all unlabeled EGMs based on some criteria that we mentioned in section (3.1). Figure 4 shows the flowchart used in the SSSL process.

3. Results and Discussions

A total of 3206 non-contact EGMs were used to train and test the DT classifier using SL and SSSL techniques. The 10-fold cross-validation tool was applied taking into account the interpatient differences (EGMs from the 9 patients were used as a train set and one patient as a test set). This process was repeated 10 times and the average was taken to show the model performance. The highest performance (CV accuracy of 72% and the AUROC of 0.71) from SL was recorded from our previous work [8]. In the same way, the results were recorded for the SSSL to show the performance of this approach over the SL.

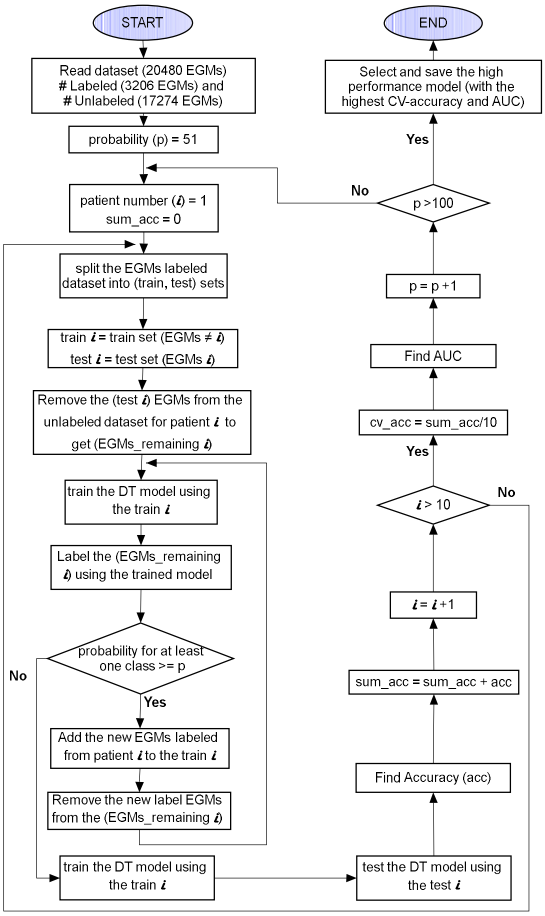


Figure 4. Flowchart for the proposed semi-supervised learning method

3.1. Probability Selection

The crucial aspect of the semi-supervised learning techniques is estimating the probability of classes in the classification process because it helps the ML models to make decisions about how to assign labels to unlabeled data. In this work, a probability of at least one class was applied for each of 10 sets of features to find the best probability that gives the highest model performance. The highest performance was obtained using the 5 most important features (feature set 5) (PSD10Hz, mean, MFCC9Hz, LPCC11Hz, and the neighborhood peaks). 82% probability was represented as the most effective to get a high model performance (CV accuracy of 75.22% and ROAUC of 0.73) (Figure 5).

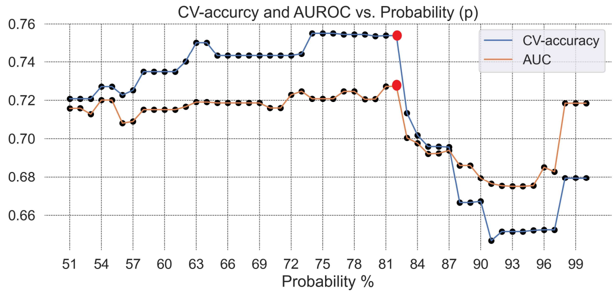


Figure 5. The CV-accuracy and the AUC of the SSSL model based on the probability for at least one class during the labeling. The probability of 82% (red dots) was recorded as the threshold for a high-performance model. The CV-accuracy and AUC were 75.22% and 0.73, respectively (Figures. 6, 7).

The highest performance for the SL and SSSL models was recorded using 10-fold cross-validation. CV accuracy and area under the curve (AUC) for SL were 72%, and 0.71, respectively; and for SSSL were 75.22%, and 0.73, respectively using the testing dataset. For SL, the sensitivity, specificity, precision, and F1_score were 66%, 78.9%, 73.1%, and 69.3%, respectively; while for SSSL were 76.5%, 74.5%, 72.3%, and 74.3%, respectively (Figures 6, 8).

3.2. CMs and ROCs

A bar chart and the receiver operating characteristic (ROC) curve (Figures 6, 7) show an improvement in the model performance using the DT-based on SSSL over the SL. The unlabeled data helps to refine the ML models and also to make them more robust and enhance the model generalization on unseen data by reducing overfitting [9]. Moreover, the SSSL made a good balance between the true positive and negative classes by improving the model's understanding of the class distributions[10]; the sensitivity and specificity were around 79% and 66% for SL, whereas using SSSL were 74% and 76%, respectively (Figure 8A,B).

3.3. Features Statistical Analysis

All five non-normally distributed features were analyzed using the Mann-Whitney unpaired t-test and were expressed as median and interquartile intervals (Figure 9). It can be seen that PSD10Hz and MFCC9Hz features were higher for the EGM positive responses than negative ones. The 10 and 9 Hz frequency power might be sufficiently large to reach the cell activation threshold, which makes the action potential trigger and produce the focal ectopic ring [11] as an AF driver in the AF frequency range of 3-15 Hz. Moreover, low voltage areas were more AF harbors (mean feature) [12]. Additionally, the neighborhood peaks were greater for positive responses, indicating the rapid focal ectopic in these EGM regions. In contrast, the LPCC11Hz feature was higher for the EGM negative responses.

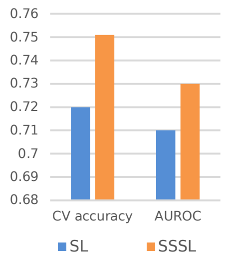


Figure 6. Model performance for DT-based on SL and SSST

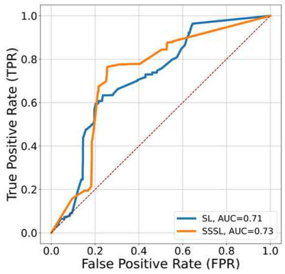
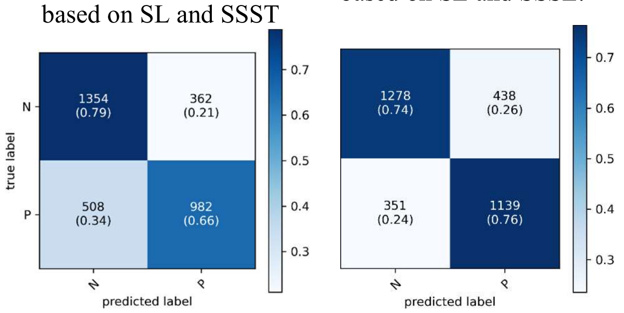


Figure 7. ROC for DT-based on SL and SSSL.



(A) (B)

Figure 8. CM using A) SL model and B) SSSL (P=positive response and N=negative response class).

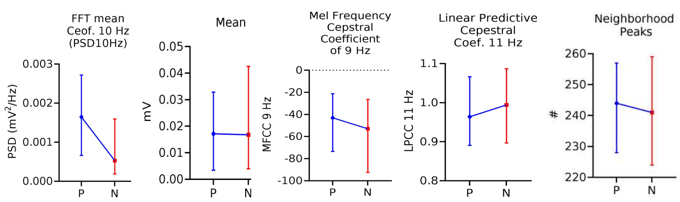


Figure 9. Statistical analysis of 5 most important features

4. Conclusions

Semi-supervised self-learning played a significant role in improving the DT model performance for predicting catheter ablation outcomes. This technique improved the model by about 3% and 0.02 for CV accuracy and AUC, respectively. Also, this technique produced a good balance in predicting the true positive and negative classes. Further studies focusing on unlabeled data from other patients should be considered.

Acknowledgments

This work was funded by the Higher Committee for Education Development (HCED) in Iraq, and the British Heart Foundation.

References

[1] D. w. t. S. C. o. t. E. H. R. Association, E. b. t. E. A. f. C.-T. Surgery, A. T. F. Members, A. J. Camm, P. Kirchhof, G. Y. Lip, U. Schotten, I. Savelieva, S. Ernst, and I. C. Van Gelder, "Guidelines for the Management of Atrial Fibrillation: the Task Force for the Management of Atrial Fibrillation of the European Society of Cardiology (ESC)," *European heart journal,* vol. 31, no. 19, pp. 2369-2429, 2010.

[2] N. Qaqos, F. S. Schlindwein, G. A. Ng, and X. Li, "Evaluating Electrograms Domain Knowledge for Enhancing Catheter Ablation Outcomes Based on Time Series Features." pp. 1-4.

[3] N. Qaqos, F. S. Schlindwein, G. A. Ng, and X. Li, "Choosing Electrogram Features for Predicting Catheter Ablation Outcomes in Persistent Atrial Fibrillation." pp. 1-4.

[4] Y. C A Padmanabha Reddy, P. Viswanath, and B. Eswara Reddy, "Semi-Supervised Learning: a Brief Review," *International Journal of Engineering & Technology,* vol. 7, no. 1.8, 2018.

[5] S. M. Rasmussen, M. E. Jensen, C. S. Meyhoff, E. K. Aasvang, and H. B. Słrensen, "Semi-Supervised Analysis of the Electrocardiogram Using Deep Generative Models." pp. 1124-1127.

[6] P. Zhang, Y. Chen, F. Lin, S. Wu, X. Yang, and Q. Li, "Semi-Supervised Learning for Automatic Atrial Fibrillation Detection in 24-Hour Holter Monitoring," *IEEE J Biomed Health Inform,* vol. 26, no. 8, pp. 3791-3801, Aug, 2022.

[7] L. X. Almeida TP, Sidhu B, Bezerra AS, Ehnesh M, Anton I, Nasser IA, Chu GS, Stafford PJ, Yoneyama T, Ng GA., "Dominant Frequency and Organization Index for Substrate Identification of Persistent Atrial Fibrillation," *Computing in Cardiology,* vol. 48, pp. 1-4, 2021.

[8] N. Qaqos, X. Li, and G. A. Ng, "Po-04-077 Electrogram Strength at 10 Hz Unveiled as Accurate Predictor of Regional Catheter Ablation Success in Persistent Atrial Fibrillation," *Heart Rhythm,* vol. 21, no. 5, pp. S486-S487, 2024.

[9] S. Zhang, M. Wang, S. Liu, P.-Y. Chen, and J. Xiong, "How Does Layer Theoretical Analysis," *arXiv preprint arXiv:2201.08514*, 2022.

[10] Y. Ouali, C. Hudelot, and M. Tami, "An Overview of Deep Semi-Supervised Learning," *arXiv preprint arXiv:2006.05278*, 2020.

[11] G. Leventopoulos, R. Koros, C. Travlos, A. Perperis, P. Chronopoulos, E. Tsoni, E. E. Koufou, A. Papageorgiou, A. Apostolos, P. Kaouris, P. Davlouros, and G. Tsigkas, "Mechanisms of Atrial Fibrillation: How Our Knowledge Affects Clinical Practice," *Life (Basel),* vol. 13, no. 6, May 25, 2023.

[12] Y. Huo, T. Gaspar, R. Schönbauer, M. Wójcik, L. Fiedler, F. X. Roithinger, M. Martinek, H. Pürerfellner, B. Kirstein, U. Richter, S. Ulbrich, J. Mayer, O. Krahnefeld, T. Agdirlioglu, A. Zedda, J. Piorkowski, and C. Piorkowski, "Low-Voltage Myocardium-Guided Ablation Trial of Persistent Atrial Fibrillation," *NEJM Evidence,* vol. 1, no. 11, 2022.

Address for correspondence:

Noor Qaqos
School of Engineering
University of Leicester, UK
nnq2@leicester.ac.uk