Weakly Supervised P Wave Segmentation in Pathological Electrocardiogram Signals Using Deep Multiple-instance Learning

Jakub Hejc[1,2,3], Richard Redina[1,3], David Pospisil[5], Ivana Rakova[1], Jana Kolarova[3], Zdenek Starek[1,4]

[1] International Clinical Research Center, St. Anne's University Hospital, Brno, Czech Republic
[2] Department of Pediatric, Children's Hospital, The University Hospital Brno, Brno, Czech Republic
[3] Department of Biomedical Engineering, Brno University of Technology, Brno, Czech Republic
[4] 1st Department of Internal Medicine, Cardio-Angiology, Faculty of Medicine, Masaryk University, Brno, Czech Republic
[5] Department of Internal Medicine and Cardiology, University Hospital Brno and Faculty of Medicine of Masaryk University, Brno, Czech Republic

Abstract

Detection of obscured P waves remains a largely unexplored topic. This study proposes a weakly supervised learning approach for P wave feature embedding by leveraging surrogate labels and 3265 eight-lead electrocardiographic (ECG) signals with diverse cardiac rhythms, including supraventricular tachycardias, atrial fibrillation, and paced rhythms.

The proposed method employs a temporal convolutional neural network and multiple instance learning to learn pyramidal feature embeddings that estimate both labeled and unlabeled instances of the P wave. The fine-tuned model achieved a temporally aggregated Dice score of 81.1%, outperforming the baseline model by 1.0%. On the subset with sinus rhythms and minor rhythm irregularities, the model consistently achieved recall and precision of around 84–85% for P wave onset and offset.

The framework can be used to learn embeddings correlated with the distribution of the atrial depolarization, using only a fraction of labeled samples. Surrogate labels allow us to embed more detailed context, which may enhance the performance and interpretability of deep neural networks in downstream tasks in the future.

1. Introduction

Alterations in the morphology and timing of P waves are important clinical markers for the diagnosis and risk stratification of a variety of cardiac conditions, including atrioventricular (AV) blocks, supraventricular tachycardias (SVTs), and atrial fibrillation (AF) [1, 2].

Several factors challenge P wave segmentation, namely, it has a relatively low amplitude, high inter-individual variability in morphology, and is often frequent obscured by the QRS complex and T wave, which exhibit similar frequency content [3].

A number of approaches have been proposed for P wave segmentation, such as correlation-based methods [4, 5], QRS-T suppression employing subtraction of a template [6] or source separation techniques [7], and deep learning approaches extracting P wave features by means of supervised learning [8]. To date, only two studies [3, 9] have investigated the detection of P waves in pathological cardiac rhythms, however, excluding obfuscated P waves due to the aforementioned challenges.

The aim of this work is to extract the microfeatures representing superimposed P waves by utilizing surrogate labels retrieved inexpensively from intracardiac electrograms (EGMs). This reformulates the problem from learning with only a subset of labeled P waves (instance-wise incomplete information) to a sample-wise incompleteness, where some part of each wave is labeled.

2. Methods

2.1. Data

3265 8-lead ECGs (708 patients indicated for electrophysiology procedure, 41.7% female, median age 36.6 years, median duration 11.0 seconds) recorded by the Abbott WorkMate 4.3 system at 2000 Hz with 78 nV/LSb resolution. Signals were undersampled to 250 Hz using anti-aliased decimation and processed with a bandpass filter (0.5–40Hz).

The ECGs were classified into four groups contained: (SR) 1771 sinus rhythms and less severe arrhythmias, such as premature beats, AV blocks, and bundle branch blocks; (SVT) 1036 SVTs, mostly caused by accessory pathways

and nodal reentry circuits; (AF) 458 cases of AF/flutter; (ST) 770 rhythms from atrial/ventricular artificial pacing.

A stratified randomized split into 70:30 training and validation sets (DB_V) was performed to ensure similar distribution of SR, SVT, AF and ST in both sets. Figure 1 shows the steps involved in orchestrating the dataset.

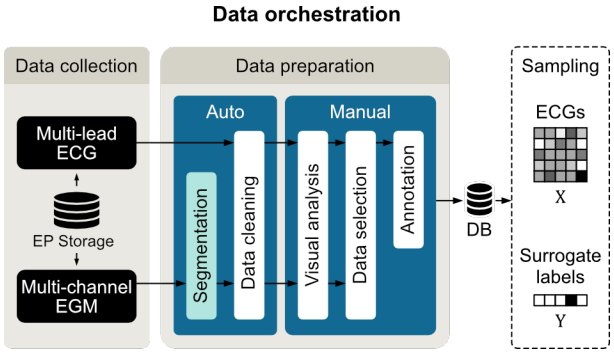

Figure 1. Schema of the dataset preparation.

2.2. Incomplete labeling

Surrogate labels of atrial beats were obtained by segmenting 5-channel intracardiac electrograms (EGM) recorded from a diagnostic catheter placed in the coronary sinus (CS). A custom-made algorithm [10] was used to detect early and late activation in the CS. Due to their predominantly random nature, f-waves were categorized into a negative class.

The measurements roughly indicate depolarization times of the left atrium, providing incomplete information about overall P wave duration. Formally, an ECG signal can be expressed as a set $X = \{x_1, \ldots, x_n\}$, where x_i is the *i*-th sample. The ground truth labels aren't available for all samples x_i, so the set of labeled data $Y = \{y_1, \ldots, y_k\}$, where $k \ll n$, represents known ground truths, and $n - k$ is the number of unidentified samples.

2.3. Feature extraction

We employed a temporal convolutional neural network (CNN) to learn a mapping $h : X \mapsto H$, where h is a feature extractor and H represents vector embeddings. The CNN incorporates a ResNet-50 backbone encompassing four residual stages with residual units combining preactivated 1×7 kernels and a 1×1 bottleneck with a 50% compression factor. The filter dimensions across encoder stages consist of 128, 256, 512, and 1024 filters. A feature pyramid network (FPN) [11] was integrated as a decoder allowing for the extraction of semantically rich features H^s from multiple scales. For more details please refer to the source code.

2.4. Multiple instance learning

The objective was to find $f : H \mapsto \hat{Y}$ given the ground truths Y, where $f(H)$ is an estimator used to model the posterior probability $\hat{y}_i \in$ of being a part of the P wave for both labeled and unlabeled instances.

We modeled $f(H)$ using a multiple instance learning (MIL) paradigm. MIL frames the original problem to learning $f(H)$ to predict the label $Y \in \{0, 1\}$ of a bag given its instances. A bag is defined as a subset $H_m \subseteq H$, and is considered positive if and only if at least one of its instances is labeled as positive.

Multi-scale pretraining (Figure 2, top part) employed embeddings H^s from the top three FPN layers. Bags were created by splitting each H^s into constant width regions H^s_m lasting approximately 96, 192, and 384 ms (33% overlap). A label was assigned to each H^s_m according to the MIL principle. A multilayer perceptron (MLP) projection network $f_{\text{proj}}(H^s_m)$ combined the features within the region m, followed by a sigmoid function σ to model the probability of the bag label denoted as Y^*_{MIL}.

Next, we performed the fine-tuning (F-T, Figure 2, bottom part) of the h using the MLP classifier f_{class}. The f_{class} reduces the rank of the top-most embeddings H^0 to $1 \times n$ creating a probability distribution map $\hat{Y}$. Similarly to a pretraining, a constant time window is used to splitting the $\hat{Y}$ into m regions 160 ms wide regions (50% overlap). Then, temporal max pooling is applied to aggregate the probabilities for each $\hat{Y}_m$, to model fine-grained Y^*_{MIL}.

Both h and f were optimized using the cross-entropy loss $\mathcal{L}_{\text{wBCE}}$, reweighted by the inverse of label temporal occurrences. Naive oversampling was applied to address the imbalance between the SR, SVT, AF, and ST groups.

2.5. Data augmentation

The function $g(X, Y)$ was utilized to reverse and shift the temporal axis to mitigate overfitting arising from the stochastic interrelationship within the P-QRS-T sequence, and from the constant P wave position relative to region splitting. Other data augmentations (DAs) involved random temporal and voltage scaling, inverse polarity, artificially generated Gaussian white noise, 50 Hz powerline interference, and stimulation artifacts.

2.6. Training setup

The AdamW algorithm with decoupled L_2 regularization and default β_1, β_2 was employed for the optimization. Convolutional and normalization layers were initialized using the Kaiming and constant initialization, respectively. The initial learning rate $\mu_0 = 0.001$ and a batch size of 64 were determined through the grid search. Learning rate was scheduled using warm-up phase for the

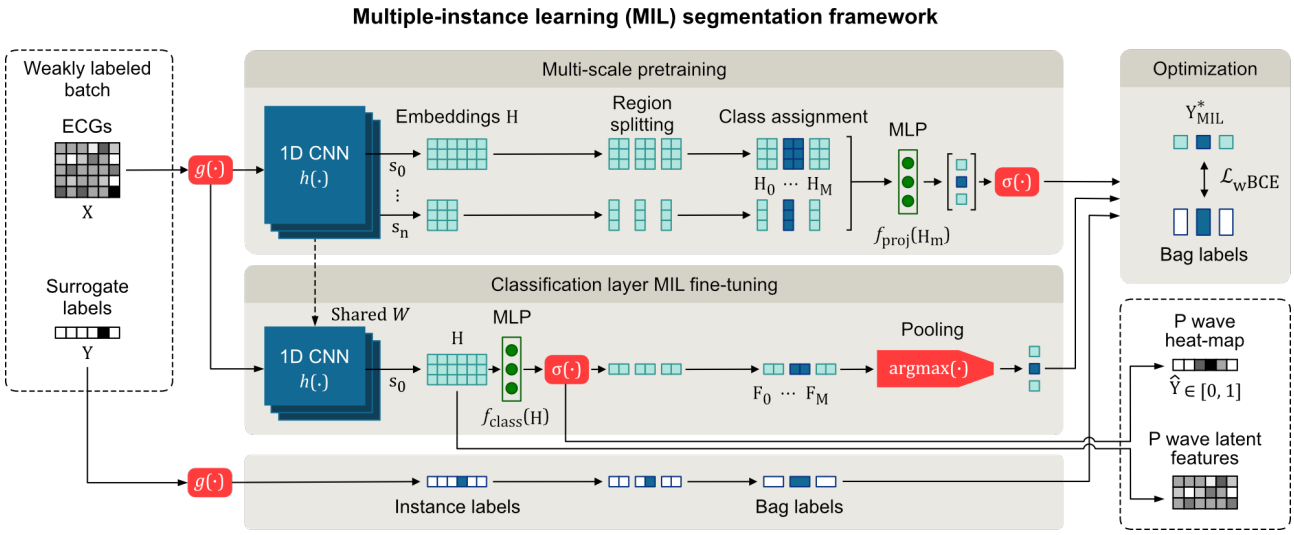

Figure 2. Multiple instance learning framework for the segmentation of hidden P waves using incomplete surrogate labels.

first 20 epochs, followed by a reduce-on-plateau strategy with decaying factor of 0.1. The code was implemented in Python 3.8.1 using the PyTorch 1.12.1, and is available at: www.github.com/jakubhejc/MILPNet.

3. Results and discussion

The Dice score DS_M [%] was computed for output scores Y^*_{MIL}. Thresholded sequence $\hat{Y}$ (default limit of 0.5) was employed to estimate recall (Rec.) and precision (P^+) for each fiducial point of the P wave regarding the surrogate references Y and tolerance of 60 ms. Overall results on the validation set DB_V are provided in Table 1.

Table 1. The performance of the multi-scale and fine-tuned (F-T) model on the validation set DB_V and SR subset DB_V:SR. Symbol + represents the presence of the P wave.

			DB_V			DB_V:SR		
	F-T	+	P_{on}	P_{off}	P_{peak}	P_{on}	P_{off}	P_{peak}
DS_M	Yes	81.1						
	No	80.1						
Rec.	Yes		0.63	0.64	0.63	0.85	0.85	0.84
	No		0.59	0.59	0.60	0.83	0.83	0.83
P^+	Yes		0.69	0.66	0.69	0.84	0.84	0.85
	No		0.62	0.60	0.62	0.83	0.83	0.83

P wave segmentation in the presence of arrhythmias is a non-trivial task, as demonstrated by our results. The F-T model outperformed the baseline model, achieving a Dice score of 0.811 compared to 0.801 without fine-tuning. For the P_{on}, P_{peak}, and P_{off}, the F-T model also performed slightly better, with a maximum recall and precision of 0.63 and 0.69, respectively. On the DB_V:SR subset, the F-T model consistently achieved recall and precision scores of around 0.84-0.85 across all fiducial points. This is consistent with Saclova et al.[9], who found that P wave detection performance varied widely, with recall and P^+ ranging from 78.1–93.1% and 72.0–88.6%, respectively. Higher performance was observed for datasets with easily distinguishable P waves.

Visual analysis showed that the model tended to approximate the temporal co-occurrence of the P-QRS-T waves, resulting in false P wave predictions in some cases. This is likely due to several factors: a) the complexity of the optimization goal and insufficient training data for the model to learn uncorrelated representations of superimposed P wave and QRS complex; b) suboptimal settings of temporal hyperparameters.

Despite these limitations, the study demonstrates that the MIL can be used to capture the temporal dynamics of the ECG signal and to learn embeddings correlated with the distribution of the atrial depolarization, even with only a fraction of labeled samples. Surrogate labels allows us to embed representations with more detailed context (see Figure 3), which may enhance the performance and interpretability of deep neural networks in downstream tasks.

4. Conclusion

Conventional weakly-supervised learning employing the MIL paradigm may not be sufficient for the effective extraction of microfeatures representing superimposed P waves. Novel regularization techniques may be required to address possible overfitting to the P-QRS-T temporal distribution, particularly in deep learning frameworks. ECG

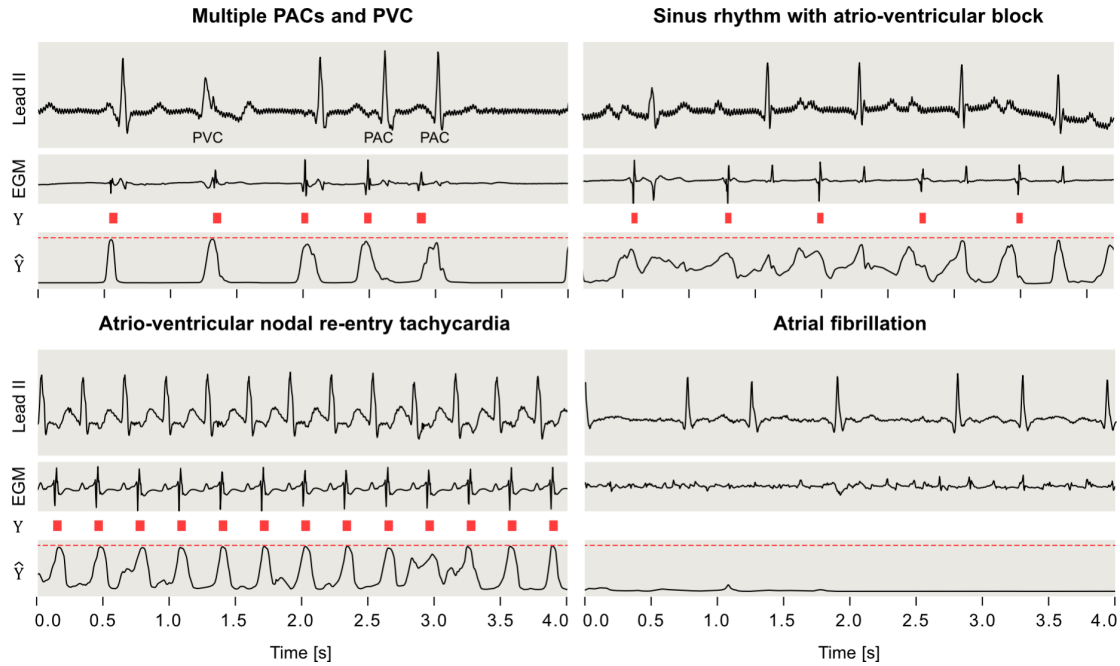

Figure 3. Examples of challenging pathological ECGs from DB_V with heat maps $\hat{Y}$ generated by CNN model. Red squares – surrogate labels Y; red dashed line – the upper bound of normalized output scores.

segmentation algorithms should be evaluated using comprehensive datasets incorporating diverse cardiac rhythms.

Acknowledgments

Brno Ph.D. Talent Scholarship Holder R.R. funded by the Brno City Municipality.

References

[1] Brady WJ, Mattu A, Tabas J, Ferguson JD. The differential diagnosis of wide QRS complex tachycardia. The American Journal of Emergency Medicine October 2017; 35(10):1525–1529.

[2] Stafford PJ, Turner I, Vincent R. Quantitative analysis of signal-averaged p waves in idiopathic paroxysmal atrial fibrillation. The American Journal of Cardiology September 1991;68(8):751–755.

[3] Portet F. P wave detector with PP rhythm tracking: evaluation in different arrhythmia contexts. Physiol Meas January 2008;29(1):141–155.

[4] Martinez J, Almeida R, Olmos S, Rocha A, Laguna P. A wavelet-based ecg delineator: evaluation on standard databases. IEEE Transactions on Biomedical Engineering 2004;51(4):570–581.

[5] Hesar HD, Mohebbi M. A multi rate marginalized particle extended kalman filter for p and t wave segmentation in ecg signals. IEEE Journal of Biomedical and Health Informatics 2019;23(1):112–122.

[6] Sippensgroenewegen A, Mlynash MD, Roithinger FX, Goseki Y, Lesh MD. Electrocardiographic analysis of ectopic atrial activity obscured by ventricular repolarization: P wave isolation using an automatic 62-lead QRST subtraction algorithm. Journal of Cardiovascular Electrophysiology July 2001;12(7):780–790.

[7] Vaya C, Rieta J, Sanchez C, Moratal D. Convolutive blind source separation algorithms applied to the electrocardiogram of atrial fibrillation: Study of performance. IEEE Transactions on Biomedical Engineering August 2007; 54(8):1530–1533.

[8] Malali A, Hiriyannaiah S, G.M. S, K.G. S, N.T. S. Supervised ECG wave segmentation using convolutional LSTM. ICT Express September 2020;6(3):166–169.

[9] Saclova L, Nemcova A, Smisek R, Smital L, Vitek M, Ronzhina M. Reliable P wave detection in pathological ECG signals. Sci Rep April 2022;12(1):6589.

[10] Hejc J, Pospisil D, Novotna P, Pesl M, Janousek O, Ronzhina M, Starek Z. Segmentation of atrial electrical activity in intracardiac electrograms (IECGs) using convolutional neural network (CNN) trained on small imbalanced dataset. In 2021 Computing in Cardiology (CinC). IEEE, September 2021; .

[11] Lin TY, Dollar P, Girshick R, He K, Hariharan B, Belongie S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). July 2017; .

Address for correspondence:

Jakub Hejc, ICRC – St. Anne's University Hospital, Pekarska 53, 602 00 Brno, Czech Republic, jakub.hejc@fnusa.cz