Uncertainty-Guided Boundary Localization Network for Weakly Supervised ECG Waveform Delineation

Jiaran Chen[1], Qianhui Bao[1], Jichao Zhao[1]

[1] Auckland Bioengineering Institute,the University of Auckland, Auckland, New Zealand

Abstract

Accurate localization of P-, QRS- and T-wave boundaries is essential for quantifying clinically relevant electrocardiogram (ECG) intervals. In conventional ECG analysis, waveform boundaries are commonly detected using rule-based algorithms, which often show limited accuracy under waveform variability and distortion. Developing more accurate supervised AI systems, however, requires extensive expert boundary annotations. We propose a weakly supervised framework in which multiple rule-based algorithms generate complementary candidate boundaries from heterogeneous data, knowledge and predictive subsets. Candidate uncertainty is incorporated into an iterative distributional clustering procedure to infer stable, high-confidence consensus boundaries for model training. On the Lobachevsky University Database, the framework correctly detected 96.42% of reference boundaries, with a mean localization error of 11.10 ms across P-, QRS- and T-wave onsets and offsets. Without dense expert annotations, the proposed method achieved performance comparable to conventional supervised networks, enabling accurate automated delineation of large-scale unlabeled ECG data and providing a reliable basis for quantifying clinically relevant waveform durations and cardiac intervals.

1. Introduction

Electrocardiography (ECG) is a non-invasive and cost-effective modality for recording cardiac electrical activity and is widely used in clinical practice [1]. Portable devices, such as Holter monitors, have extended ECG acquisition from short examinations to continuous monitoring, facilitating the detection of intermittent arrhythmias. In conventional ECG analysis, waveform boundaries are commonly identified using embedded rule-based algorithms. However, their accuracy can deteriorate in the presence of noise, morphological variability and arrhythmic waveform distortion, limiting their robustness in real-world applications. Artificial intelligence (AI) offers a promising alternative for automated ECG analysis, but supervised models typically require extensive expert annotations of waveform boundaries. To reduce this annotation burden, recent studies have investigated weakly supervised learning using limited expert annotations or large-scale noisy labels to guide model training [2, 3].

Weakly supervised learning nevertheless introduces two fundamental challenges: whether noisy weak labels can provide reliable supervision and how their errors can be corrected during training. Differences in data distributions, rule-based knowledge and model predictions may produce heterogeneous and systematically biased boundary candidates. If these errors are treated as reliable supervision, they can be propagated and amplified during model training. Existing uncertainty-aware methods generally estimate uncertainty at the sample level to reweight predictions or guide optimization [4, 5]. However, ECG delineation is a temporally structured and physiologically constrained task, in which waveform boundaries determine clinically relevant durations and intervals [6, 7]. Sample-level uncertainty alone cannot directly characterize the reliability or consistency of candidate boundary locations. It is therefore necessary to model uncertainty at the boundary level and use it to identify reliable consensus boundaries from heterogeneous weak labels.

To address these challenges, we propose a weakly supervised framework for structured ECG waveform delineation. The unlabeled data are partitioned into multiple subsets and annotated using different rule-based algorithms, generating complementary weak supervisory signals that capture variability in both data and prior knowledge. Separate models are trained on these weak-label subsets, and Monte Carlo dropout is applied during inference to generate stochastic predictions from different parameter configurations. Predictive entropy across repeated forward passes is then used to quantify uncertainty at each candidate boundary.

Candidate positions and their associated uncertainties from different data subsets, labeling strategies and stochastic predictions are jointly represented in a distributional space. An iterative clustering procedure progressively groups consistent candidates and identifies consensus regions characterized by low uncertainty. The resulting high-confidence boundaries are used as refined supervisory signals to train an independent delineation model. In this way, the framework transforms heterogeneous and potentially unreliable weak labels into stable boundary supervision without requiring manual expert annotations.

2. Materials and Methods

2.1. Dataset

The Lobachevsky University Electrocardiography Database (LUDB) contains 200 ten-second, twelve-lead ECG recordings acquired at 500 Hz from patients with diverse cardiovascular conditions in Nizhny Novgorod, Russia. The cohort comprises 115 male and 85 female patients aged 11–90 years. For each recording, expert cardiologists annotated the peaks, onsets and offsets of the P wave, QRS complex and T wave. These expert annotations were used exclusively as reference standards for evaluation and were not included in weak-label generation or model training.

2.2. Overview

The proposed waveform boundary localization framework is illustrated in Fig. 1 and comprises three stages: boundary uncertainty modeling, uncertainty-guided boundary clustering, and uncertainty-aware adaptive refinement. In the first stage, complementary sources of variability are introduced through rule-based annotation, data partitioning and stochastic parameter sampling to estimate predictive uncertainty. In the second stage, sample-level uncertainty is aggregated around candidate waveform boundaries. An iterative distance- and uncertainty-based clustering procedure then groups consistent candidates and identifies low-uncertainty consensus boundaries as reliable weak supervisory signals. In the final stage, confidence-dependent supervision strategies are applied to the refined weak labels to support stable training of an independent delineation network. A one-dimensional U-Net is used as the common backbone throughout the framework. The framework is designed to localize P-, QRS- and T-wave boundaries independently of beat-type labels; beat classification is outside the scope of this study. The implementation of each stage is described below.

Boundary Uncertainty Modeling: Three rule-based annotation strategies are used to generate weak labels for the training data. Models are trained on progressively larger subsets at 10% increments, and the variation between consecutive predictions is measured. The subset size producing the largest variation is used to partition the data, with each partition training an independent model to capture data-induced uncertainty. Distinct heuristic label sets provide complementary prior knowledge across annotation sources. Finally, Monte Carlo dropout samples multiple parameter configurations during inference to capture predictive uncertainty within each model.

Uncertainty-Guided Boundary Clustering: For each model prediction, the class with the highest probability at each sampling point is selected to determine waveform labels and derive the corresponding boundaries. Given N data subsets, M rule-based annotation strategies, and L Monte Carlo dropout passes, $N \times M \times L$ candidate boundary sets are obtained. Predictive entropy is computed from the probability maps to quantify uncertainty. As unreliable predictions are often accompanied by spatially extended high-uncertainty regions, boundary uncertainty is modeled from the local structure rather than individual sampling points.

Specifically, a region of length S is selected along the direction of significant entropy change relative to a baseline. Its uncertainty is aggregated using distance-weighted pooling:

$$u_t = \frac{\sum_{s=1}^{S} w_s u_{t_s}}{\sum_{s=1}^{S} w_s}, \tag{1}$$

where u_{t_s} denotes the predictive entropy at sampling point t_s, and w_s increases with its proximity to candidate boundary t. This formulation emphasizes uncertainty near the boundary while reducing the influence of isolated noise.

Candidate boundaries and their uncertainties are then mapped into a discrete space for uncertainty-driven iterative aggregation. Starting from an uncertainty-weighted initialization, each iteration updates candidate locations according to spatial proximity and boundary uncertainty until they converge to locally consistent, low-uncertainty solutions.

Uncertainty-Guided Adaptive Refinement: For each ECG recording, the uncertainties of all waveform boundaries are ranked. Significant relative jumps between consecutive values divide the boundaries into low-, medium-, and high-uncertainty groups, avoiding heuristic percentile thresholds. A single network is trained using uncertainty-dependent supervision. Low-uncertainty boundaries use binarized hard labels, medium-uncertainty boundaries use consistency supervision from an exponential moving average (EMA) network, and high-uncertainty boundaries use probability maps as soft labels.

3. Results

Waveform localization is commonly evaluated using accuracy, sensitivity, specificity, and F1-score. Boundary-level performance is further assessed using temporal error and onset/offset sensitivity, defined as the proportion of reference boundaries correctly detected within 50 ms. We compare the proposed method with a fully supervised

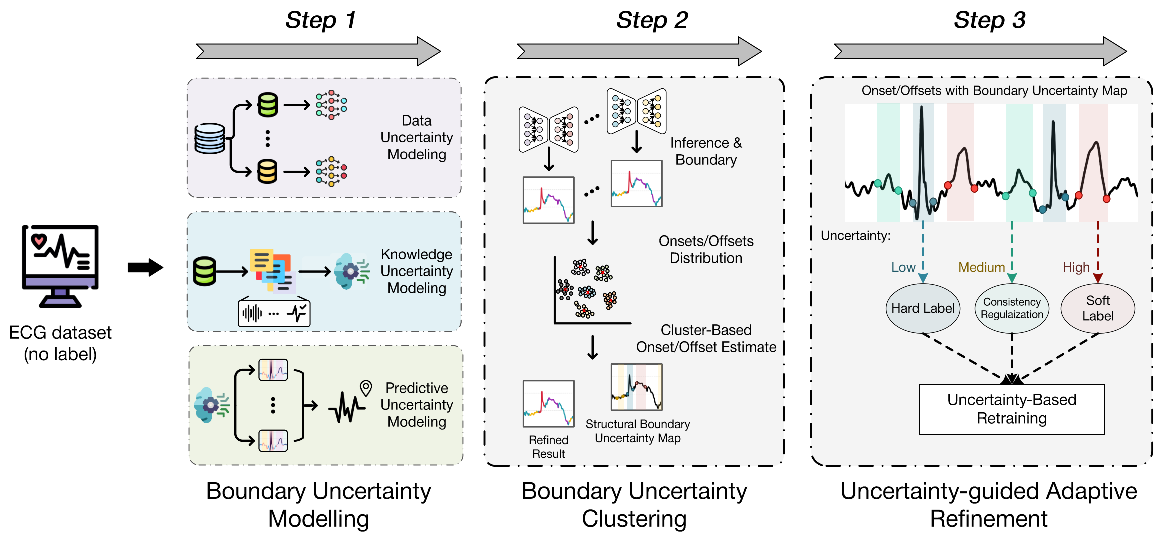

Figure 1: Overview of the proposed weakly supervised ECG waveform boundary localization framework.

UNet trained using expert annotations and two weakly supervised methods, EM-UNet and Att-EMD-UNet, trained without expert labels.

As shown in Table 1, boundary-level evaluation is particularly important because waveform onsets and offsets determine clinically relevant measurements, including PR interval, QRS duration, and QT interval. The proposed method detects over 94% of the boundaries for every waveform and over 98% of QRS boundaries, substantially outperforming the weakly supervised baselines while approaching the supervised UNet. Localization errors remain below 13 ms for P- and QRS-wave boundaries and approximately 15 ms for T-wave boundaries. In contrast, Att-EMD-UNet shows onset errors of 33.00 and 43.15 ms for P and T waves, while EM-UNet also produces larger errors for several offsets. These results indicate that heterogeneous weak supervision reduces source-specific bias and provides more reliable boundaries for electrophysiological measurement.

Table 2 presents the sample-level results. The proposed method performs consistently better than both weakly supervised baselines and remains close to the fully supervised model, achieving a mean F1-score of 0.9022 versus 0.8247 for EM-UNet, 0.7727 for Att-EMD-UNet, and 0.9242 for UNet. This suggests that refining weak labels is more influential than increasing backbone complexity and can provide reliable supervision for automated ECG analysis.

Table 1: Boundary-level delineation performance on the LU dataset. UNet is fully supervised, while the remaining methods use weak supervision.

Metric	Type	Method	P	QRS	T
Sensitivity	Onset	UNet [8]	0.9851	0.9825	0.9415
		EM-UNet [2]	0.8809	0.8227	0.8831
		Att-EMD-UNet [9]	0.9193	0.9300	0.7830
		Proposed	0.9683	0.9841	0.9403
	Offset	UNet [8]	0.9851	0.9825	0.9463
		EM-UNet [2]	0.8801	0.8198	0.8643
		Att-EMD-UNet [9]	0.9193	0.9348	0.7872
		Proposed	0.9683	0.9833	0.9410
Mean error (ms)	Onset	UNet [8]	7.10	5.76	13.19
		EM-UNet [2]	9.00	13.54	21.27
		Att-EMD-UNet [9]	33.00	15.34	43.15
		Proposed	10.16	6.73	15.17
	Offset	UNet [8]	6.41	7.38	11.57
		EM-UNet [2]	24.11	13.54	21.06
		Att-EMD-UNet [9]	15.21	9.86	20.04
		Proposed	8.68	12.49	13.39

4. Conclusion

In this study, we propose a weakly supervised framework for ECG waveform boundary localization without expert annotations. Uncertainty from data, annotation rules, and model predictions is aggregated at the boundary level to guide iterative clustering and generate reliable supervision. The framework correctly detects 96.42% of reference boundaries, with a mean localization error of 11.10 ms and errors below 20 ms across all waveform types. These results demonstrate that accurate boundary localization can be achieved without dense expert labeling, sup-

Table 2: Sample point-level delineation performance on the LU dataset. UNet is fully supervised, while the remaining methods use weak supervision.

Metric	Method	P	QRS	T
Accuracy	UNet [8]	0.9844	0.9867	0.9712
	EMUNet [2]	0.9659	0.9625	0.9522
	Att-EMD-UNet [9]	0.9569	0.9660	0.9278
	Proposed	0.9786	0.9812	0.9682
Sensitivity	UNet [8]	0.9078	0.9140	0.9103
	EMUNet [2]	0.7672	0.7660	0.8568
	Att-EMD-UNet [9]	0.6053	0.7424	0.6092
	Proposed	0.8781	0.9048	0.9008
Specificity	UNet [8]	0.9939	0.9939	0.9845
	EMUNet [2]	0.9913	0.9878	0.9884
	Att-EMD-UNet [9]	0.9989	0.9956	0.9946
	Proposed	0.9908	0.9892	0.9830
F1-score	UNet [8]	0.9276	0.9258	0.9191
	EMUNet [2]	0.8316	0.7801	0.8625
	Att-EMD-UNet [9]	0.7482	0.8292	0.7407
	Proposed	0.8987	0.8987	0.9091

porting automated analysis of large-scale unlabeled ECG data and quantification of clinically relevant waveform durations and intervals. Future work will incorporate beat-type information, improve P- and T-wave localization, and further optimize weak-label refinement and model training.

Acknowledgments

This work was supported by the Health Research Council of New Zealand, the National Heart Foundation of New Zealand and the China Scholarship Council program (Project ID: 202408500037)

References

[1] Salama A, Kachenoura A, Almuhammad Alali S, Carrault G, Senhadji L, Karfoul A. Test-time adaptation for a generalizable deep learning-based ecg segmentation. 12 2025; .

[2] Lin Y, Wang W. Att-emd-unet: A novel weakly supervised perspective for ecg segmentation. In 2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). 2024; 2170–2176.

[3] Hejc J, Redina R, Pospisil D, Rakova I, Kolarova J, Starek Z. Weakly supervised p wave segmentation in pathological electrocardiogram signals using deep multiple-instance learning. In 2023 Computing in Cardiology (CinC), volume 50. 2023; 1–4.

[4] Han K, Wang S, Chen J, Qian C, Lyu C, Ma S, Qiu C, Sheng VS, Huang Q, Liu Z. Region uncertainty estimation for medical image segmentation with noisy labels. IEEE Transactions on Medical Imaging 2025;44(12):5197–5207.

[5] Dong M, Yang A, Wang Z, Li D, Yang J, Zhao R. Uncertainty-aware consistency learning for semi-supervised medical image segmentation. Knowledge Based Systems 2025;309:112890. ISSN 0950-7051.

[6] Zhang Y, Lai J, Zhao C, Wang J, Yan Y, Chen M, Ji L, Guo J, Han B, Shi Y, Chen Y, Yang W, Feng Q. Abnormal recognition-assisted and onset-offset aware network for pathological wearable ecg delineation. Artificial Intelligence in Medicine 2024;157:102992. ISSN 0933-3657.

[7] Svehlikova J, Pribilova A, Zelinka J, Ondrusova B, Kromkova K, Hlivak P, Hatala R, Tysler M. The importance of ecg offset correction for premature ventricular contraction origin localization from clinical data. Measurement Science Review 2022;22(6):246–252.

[8] Chen Z, Wang M, Zhang M, Huang W, Gu H, Xu J. Post-processing refined ecg delineation based on 1d-unet. Biomedical Signal Processing and Control 2023;79:104106. ISSN 1746-8094.

[9] Xiong H, Liu S, Sharan RV, Coiera E, Berkovsky S. Weak label based bayesian u-net for optic disc segmentation in fundus images. Artificial Intelligence in Medicine 2022; 126:102261. ISSN 0933-3657.

Address for correspondence:

Jiaran Chen

jehc643@aucklanduni.ac.nz