Beyond Deep Learning: An Efficient and Explainable Model for ECG Wave Segmentation

Alaa Salama[1], Amar Kachenoura[1], Lotfi Senhadji[1], Guy Carrault[1], Alfredo Hernandez[1], Ahmad Karfoul[1]

[1] Univ Rennes, Inserm, LTSI-UMR 1099, Rennes, France

Abstract

Detecting the onsets and offsets of the P, QRS, and T waves in electrocardiogram (ECG) signals is essential for clinical applications. However, ECG segmentation remains challenging due to noise and artefacts. Our previous ConvLSTM$_{MA}$ pipeline outperformed state-of-the-art methods but lacked explainability and had high computational complexity. This study presents a lightweight and explainable ECG segmentation method based on time-frequency representations and ECG physiology. The method combines: i) a time-frequency representation of a band-filtered ECG signal, ii) sample-wise onset/offset detection using a lightweight linear MLP, and iii) adaptive soft-thresholding to suppress low-confidence responses. Forward causal and backward anti-causal estimations are used to improve robustness. Using two PhysioNet databases (QT for training and LU for testing), a database-independent evaluation strategy was employed to compare our method our method with three supervised methods (UNet, ConvLSTM$_{SA}$, and ConvLSTM$_{MA}$) and two unsupervised methods (ECGDeli and ECGKit). The proposed method achieves linear computational complexity compared with the quadratic complexity of ConvLSTM$_{MA}$. These results demonstrate its potential for accurate segmentation and generalization to unseen data.

1. Introduction

Early detection of abnormalities and automated interpretation of electrocardiogram (ECG) signals are essential for clinical applications. Accurate segmentation of the P, QRS, and T waves is a key step in reliable ECG analysis. However, this task remains challenging due to noise and artefacts. ECG segmentation methods fall into two main categories: unsupervised and supervised. Unsupervised methods mainly based on signal processing techniques [1, 2]. Among supervised approaches, Deep Learning (DL) methods have achieved remarkable performance in ECG segmentation by learning complex representations from large datasets. Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, ConvLSTM architectures, and UNet-based frameworks have achieved state-of-the-art performance across several public databases [3, 4]. In previous work [5], we proposed a three-stage ConvLSTM$_{\text{MA}}$ ECG segmentation pipeline combining preprocessing, a CNN–LSTM model with multi-head attention, and physiology-informed post-processing.

Although outperforming state-of-the-art supervised and unsupervised methods [5], this approach has high computational cost and operates as a black-box system. This lack of explainability limits the understanding of model predictions and reduces clinical trust. Additionally, these highly expressive models have limited interpretability, as their predictions are difficult to relate to input features compared with simpler models such as linear regression. Their large number of parameters also increases computational complexity, limiting deployment on resource-constrained devices. Moreover, most DL models are evaluated on a single ECG database, limiting their generalizability. These limitations highlight the need for ECG segmentation methods that combine transparent decision-making, interpretable feature representations, low model complexity, and strong generalization performance while maintaining high segmentation accuracy.

To address these limitations, we propose **SPWVD$_{\text{BiLin}}$**, a lightweight and explainable ECG segmentation method that uses ECG physiology and formulates segmentation as a linear regression problem. The method consists of three main steps: **(i) SPWVD-based Time–Frequency Representation**, which captures the spectral and local temporal characteristics of the ECG signal; **(ii) Bidirectional Linear Delineation**,which estimates continuous delineation scores from frequency features within a sliding window using two linear branches: a forward (causal) branch based on preceding samples and a backward (anti-causal) branch based on future samples; and **(iii) Adaptive Soft-Thresholding**, which identifies wave onsets and offsets. This physiologically grounded method provides in-

terpretable decisions with low computational complexity and good generalization.

Finally, a database-independent evaluation strategy, using two databases from PhysioNet (QT [6] for training and LU [7] for testing) was exploited to compare our pipeline with with three supervised methods ($\text{ConvLSTM}_{\text{MA}}$ [5],$\text{ConvLSTM}_{\text{SA}}$ [3], and UNet [4]) and two commonly used unsupervised ones (ECGDeli [8] and ECGKit [9]).

2. Methodology

2.1. The proposed $\text{SPWVD}_{\text{BiLin}}$ method

SPWVD-based Time–Frequency Representation: In this step, the raw ECG signals were normalized using absolute-value normalization and filtered with 50-Hz notch and 0.5-Hz high-pass filters to reduce power-line interference and baseline drift, as shown in Figure 1. Butterworth bandpass filters were then applied to emphasise the dominant frequency ranges of the P, QRS and T waves[10] (3–11 Hz, 8–50 Hz and 1–10 Hz, respectively). The filtered signals were then transformed using the Smoothed Pseudo Wigner–Ville Distribution (SPWVD) [11], which provides a balanced time–frequency representation while reducing interference terms through time and frequency smoothing. The SPWVD is defined as $SPWVD_x(t,f) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(u)h(\tau)x\left(t+u+\frac{\tau}{2}\right)x^*\left(t+u-\frac{\tau}{2}\right)e^{-j2\pi f\tau}\,d\tau\,du$ where $g(u)$ and $h(\tau)$ denote the time- and frequency-smoothing windows, respectively. Hanning and Hamming windows were used for temporal and frequency smoothing. The window lengths were adapted to each ECG component: $(L_t, L_f) = (161, 7)$ and $(111, 7)$ for P-wave onset and offset detection, $(99, 11)$ for QRS detection, and $(61, 9)$ for T-wave detection. The resulting SPWVD provides local time–frequency features used in the next step.

Bidirectional Linear Delineation: At this stage, a separate model is trained for each of the six ECG boundaries: P-onset, P-offset, QRS-onset, QRS-offset, T-onset, and T-offset. For each candidate sample (see Figure 1), two complementary linear regressors process the local SPWVD representation: a causal branch using preceding samples and an anti-causal branch using following samples. For an input window $\mathbf{X} \in \mathbb{R}^{\mathbb{N}_f \times \mathbb{L}}$, both branches are expressed using a unified formulation. Let $r \in c, a$ denote the branch, where c and a represent the causal and anti-causal branches, respectively. The corresponding linear score is given by: $a^r = \sum_{f=1}^{N_f} w_f^r(\sum_{t=1}^{L} v_t^r x_{f,t}^r + b_1^r) + b_2^r \qquad r \in c, a$ where $x^r f, t$ denotes the SPWVD value at frequency bin f and temporal position t for branch r. Here, $N_f = 256$ is the number of frequency bins and $L = 31$ is the temporal window length. The parameters w_f^r and v_t^r are the learned frequency and temporal weights, while b_1^r and b_2^r are the inner and outer biases, respectively. Thus, a^c and a^a denote the linear scores from the causal and anti-causal branches. Each branch performs linear regression on the SPWVD features. Their outputs are then combined linearly and passed through a sigmoid activation: $\hat{y} = \sigma\left(\alpha_c a^c + \alpha_a a^a + b_o\right)$ where α_c and α_a weight the causal and anti-causal contributions, respectively. The model has only 581 trainable parameters, with no nonlinear hidden layers. Since the original delineation target is binary (i.e., one for the presence of the corresponding landmark and zero otherwise), it is converted into a continuous target by convolving each annotation with a Gaussian-shaped window $g(n) = \exp\left(\ln(k)\left(\frac{2n}{N-1}\right)^2\right)$. Here, N denotes the window length and $k \in (0, 1)$ controls the boundary attenuation. The window peaks at 1 for $n = 0$ and is equivalent to a Gaussian function with $\sigma = 1/\sqrt{-2\ln(k)}$. In this study, $k = 0.005$ and $N = 21$ were empirically selected for optimal delineation performance.

Adaptive Soft-Thresholding: This stage applies adaptive soft-thresholding to the continuous delineation scores to suppress low-confidence responses while preserving high-confidence onset and offset landmarks. For ECG wave $\kappa \in \{P, QRS, T\}$ and landmark $j \in \{\text{onset}, \text{offset}\}$, the refined score $\hat{y}_{\text{soft}}^{(\kappa,j)}$ is computed as $\hat{y}_{\text{soft}}^{(\kappa,j)} = \hat{y}^{(\kappa,j)}\,\sigma\big(\alpha\left(|\hat{y}^{(\kappa,j)}| - \beta\right)\big)$ where $\sigma(\cdot)$ is the sigmoid function, β is a learnable threshold, and α controls the transition sharpness. The landmark is selected as the highest-scoring sample above the learned threshold. This suppresses spurious responses and improves delineation robustness, particularly for weak or absent waves. Finally, the detected onset and offset landmarks are paired to reconstruct the P, QRS, and T waves.

2.2. Dataset and training strategy

In this study, a database-independent evaluation strategy is adopted. The proposed $\text{SPWVD}_{\text{BiLin}}$ method is trained on the QT Database (QTDB) [6] and evaluated on the LU Database (LUDB) [7]. QTDB contains 105 15-min recordings sampled at 250 Hz, while LUDB comprises 200 10-s, 12-lead recordings sampled at 500 Hz. The method is trained using 80% of QTDB, with the remaining 20% used for validation. It consists of two stages: first, six bidirectional linear regressors are trained to estimate the onset and offset scores of the P, QRS, and T waves; second, six adaptive soft-thresholding functions are independently optimized on the validation set using MSE loss with early stopping. The learned thresholds are then fixed during inference. Finally, $\text{SPWVD}_{\text{BiLin}}$ is evaluated on the entire LUDB without retraining, fine-tuning, or parameter adaptation. All supervised and unsupervised methods are eval-

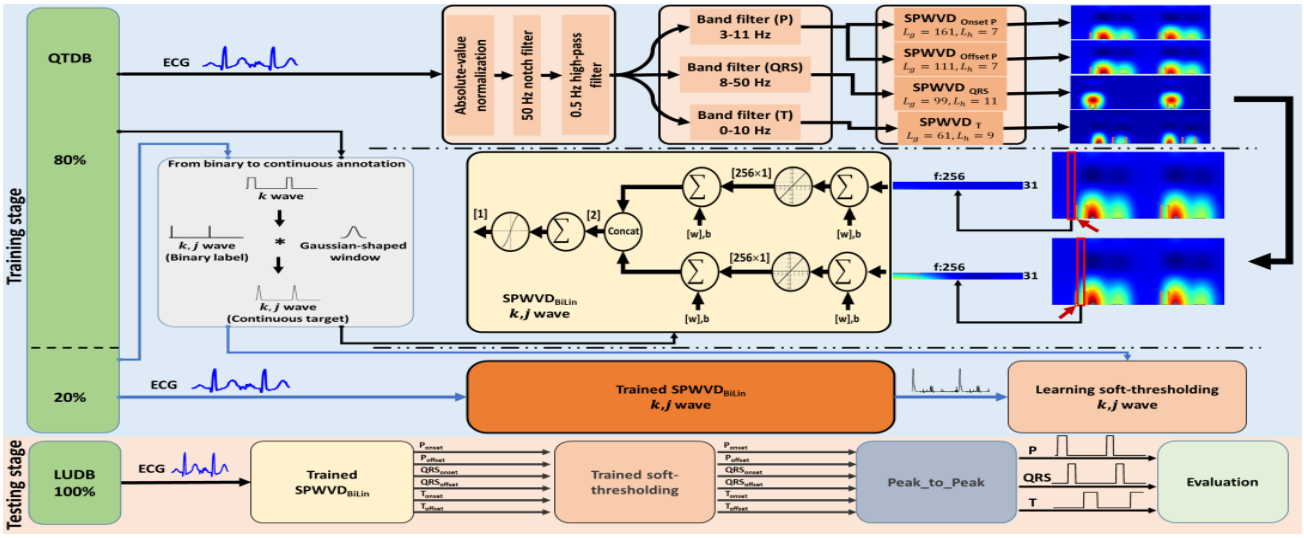

Figure 1. Proposed SPWVD$_{\text{BiLin}}$ method for ECG waves segmentation.

uated on LUDB for fair comparison.

3. Experiments and results

ECG segmentation performance is evaluated using Sensitivity, Precision, and F1-score, collectively referred to as **Criterion 1**. To assess incorrect wave detections, **Criterion 2** measures the False Detection Rate (FDR) as: $FDR = \frac{(n_p - n_t)*100}{n_t}$, where n_p and n_t denote the numbers of detected and expert-annotated waves, respectively. Computational complexity is assessed using two metrics: the number of trainable parameters (**Criterion 3**) and Floating-Point Operations (FLOPs, **Criterion 4**), with one FLOP defined as a multiplication followed by an addition.

The results for Criteria 1 and 2 are presented in Table 1. According to Criterion 1, supervised methods outperform unsupervised methods. The best F1-scores achieved by unsupervised methods are 79.2%, 80.6%, and 77.6% for the P, QRS, and T waves, respectively, while supervised methods achieve 84–89%. The proposed SPWVD$_{\text{BiLin}}$ achieves the best performance across all waves, with F1-scores of 88.5%, 88.9%, and 88.9% for the P, QRS, and T waves, respectively. Regarding Criterion 2, the DL-based methods ConvLSTM$_{\text{SA}}$ and UNet tend to overdetect the P, QRS, and T waves, while the unsupervised ECGKit underdetects the P and T waves. Note that, SPWVD$_{\text{BiLin}}$ consistently achieves lower over/underdetection than other DL-based methods without requiring a postprocessing stage, such as ConvLSTM$_{\text{MA}}$. Table 2 compares the computational complexity of the DL-based methods. The proposed SPWVD$_{\text{BiLin}}$ is substantially more lightweight, with only 3,486 trainable parameters, representing over 97% fewer parameters than the other methods. It also has linear computational complexity, requiring $104490N$ FLOPs, where N is the number of ECG samples, compared with the quadratic complexity of ConvLSTM$_{\text{MA}}$ and ConvLSTM$_{\text{SA}}$. These results demonstrate the computational efficiency of SPWVD$_{\text{BiLin}}$ while maintaining strong segmentation performance.

The proposed method provides an interpretable linear representation of the causal and anti-causal windows. Each branch independently estimates the contribution of temporal samples, which are combined with frequency information to produce a scalar score, a^c and a^a. The final prediction combines both scores using learned weights, allowing the contribution of each temporal direction to be directly interpreted. To improve explainability, saliency maps are used to identify the temporal regions contributing most to the predictions. As shown in Fig. 2. For P-wave onset and offset detection, the method mainly relies on the region surrounding the target event. QRS-onset detection considers the region around the QRS onset, P-wave peak, and QRS end, while QRS-offset detection relies mainly on the R peak and ST segment. T-wave onset and offset detection, the methodmainly relies on the regions around the S point of the QRS complex and the T-wave peak.

4. Conclusion

In this study, we propose a lightweight, efficient, and fully explainable method for ECG wave segmentation. The method combines SPWVD-based Time–Frequency Representation, Bidirectional Linear Delineation, and Adaptive Soft-Thresholding to suppress low-confidence responses. Experiments under a database-independent evaluation protocol demonstrated that the proposed method outperforms existing methods. In addition, SPWVD$_{\text{BiLin}}$ is more computationally efficient, requiring fewer FLOPs while main-

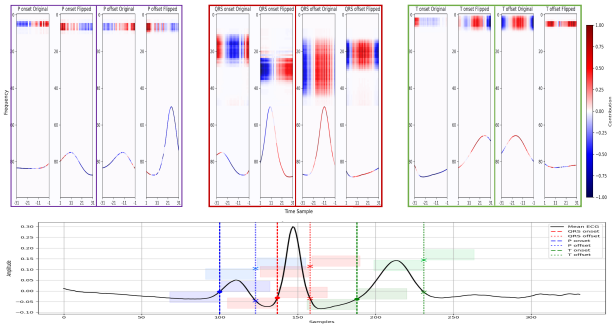

Figure 2. Saliency maps highlighting the temporal regions contributing most to the method predictions.

Table 1. Criteria 1 and 2 of the six tested methods.

Criteria	Metric	Model	P	QRS	T
Criterion 1	Sensitivity	ECGDeli	96.30	95.42	84.29
		ECGkit	78.42	82.59	73.72
		UNet	83.53	81.39	86.40
		$\text{ConvLSTM}_{\text{SA}}$	85.39	95.37	90.92
		$\text{ConvLSTM}_{\text{MA}}$	82.45	88.21	87.04
		$\text{SPWVD}_{\text{BiLin}}$	83.17	88.12	87.85
	Precision	ECGDeli	67.31	69.76	71.15
		ECGkit	72.36	59.09	81.83
		UNet	84.69	92.89	82.70
		$\text{ConvLSTM}_{\text{SA}}$	90.16	78.79	83.68
		$\text{ConvLSTM}_{\text{MA}}$	94.71	88.21	88.50
		$\text{SPWVD}_{\text{BiLin}}$	94.46	89.74	90.06
	F1-score	ECGDeli	79.24	80.60	77.16
		ECGkit	75.27	68.89	77.56
		UNet	84.11	86.76	84.51
		$\text{ConvLSTM}_{\text{SA}}$	87.71	86.29	87.15
		$\text{ConvLSTM}_{\text{MA}}$	88.15	88.34	87.76
		$\text{SPWVD}_{\text{BiLin}}$	88.46	88.92	88.94
Criterion 2	FDR [%]	ECGDeli	2.477	2.304	0.987
		ECGKit	-3.633	12.757	-18.106
		UNet	24.772	19.423	24.444
		$\text{ConvLSTM}_{\text{SA}}$	16.597	11.604	18.930
		$\text{ConvLSTM}_{\text{MA}}$	-1.816	0.164	0.246
		$\text{SPWVD}_{\text{BiLin}}$	-0.740	-0.247	-1.316

Table 2. Performances of the DL-based approaches in terms of complexity.

Model	Number of Parameters	FLOPs (N time samples)	FLOPs ($N = 1$ time sample)
Unet	167,748	$212162 \times N$	212162
$\text{ConvLSTM}_{\text{SA}}$	364,100	$256 \times N^2 + 213504 \times N$	213760
$\text{ConvLSTM}_{\text{MA}}$	461,188	$544 \times N^2 + 361728 \times N$	362272
$\text{SPWVD}_{\text{BiLin}}$	3,486	$104490 \times N$	104490

taining robust segmentation performance. The proposed method provides an interpretable linear representation of the causal and anti-causal windows, allowing the contribution of each temporal direction to the delineation decision to be directly analysed. This study also uses saliency maps to enhance explainability by identifying the temporal regions that most influence the model's predictions. The study's findings show that the proposed lightweight and fully explainable method improves the interpretability, computational efficiency, and cross-database generalization of ECG segmentation.

5. Acknowledgments

This work was supported in part by the PEPR Santé Numérique; Projet DIIP-HEART: ANR-22-PESN-0018 and in part by the PrepRisC project of the ARED program of Région Bretagne, France.

References

[1] Lenis G, Pilia N, Oesterlein T, Luik A, Schmitt C, Dössel O. P wave detection and delineation in the ecg based on the phase free stationary wavelet transform and using intracardiac atrial electrograms as reference. Biomedical Engineering Biomedizinische Technik 2016;61(1):37–56.

[2] Murthy I, Niranjan U. Component wave delineation of ecg by filtering in the fourier domain. Medical and Biological Engineering and Computing 1992;30:169–176.

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

[4] Moskalenko V, Zolotykh N, Osipov G. Deep learning for ecg segmentation. In Advances in Neural Computation, Machine Learning, and Cognitive Research III: Selected Papers from the XXI International Conference on Neuroinformatics, October 7-11, 2019, Dolgoprudny, Moscow Region, Russia. Springer, 2020; 246–254.

[5] Salama A, Kachenoura A, Alali SA, Carrault G, Senhadji L, Karfoul A. A physiology-driven deep learning-based pipeline for ecg segmentation. Biomedical Signal Processing and Control 2026;112:108608.

[6] Laguna P, Mark RG, Goldberger AL, Moody GB. A database for evaluation of algorithms for measurement of qt and other waveform intervals in the ECG. Computers in Cardiology 1997;24:673–676.

[7] Kalyakulina AI, Yusipov II, Moskalenko VA, Nikolskiy AV, Kosonogov KA, Osipov GV, Zolotykh NY, Ivanchenko MV. Ludb: a new open-access validation tool for electrocardiogram delineation algorithms. IEEE Access 2020;8:186181–186190.

[8] Pilia N, Nagel C, Lenis G, Becker S, Dössel O, Loewe A. Ecgdeli-an open source ECG delineation toolbox for matlab. SoftwareX 2021;13:100639.

[9] Demski A, Llamedo MS. ECG-kit a matlab toolbox for cardiovascular signal processing. Journal of Open Research Software 2016;4:e215–e220.

[10] Tereshchenko LG, Josephson ME. Frequency content and characteristics of ventricular conduction. Journal of electrocardiology 2015;48(6):933–937.

[11] Wu X, Liu T. Spectral decomposition of seismic data with reassigned smoothed pseudo wigner–ville distribution. Journal of Applied Geophysics 2009;68(3):386–393.

Address for correspondence:

Ahmad Karfoul
LTSI-Inserm UMR 1099, Batiment 22, Campus de Beaulieu, Universite de Rennes, 35042 Rennes, France
ahmad.karfoul@univ-rennes.fr