A Peak-Aware Envelope-Guided Loss for Fetal ECG Detection and R-Peak Localization

Ngoc-Son Nguyen[1], Trung-Kien Le[1], Tai Le[2],
Huy-Dung Han[1], Loan Pham-Nguyen[1], Hung Cao[2,3]
[1]*School of Electrical and Electronic Engineering (Hanoi University of Science and Technology), Hanoi, Vietnam*
[2]*Department of Biomedical Engineering, University of California Irvine, CA, USA*
[3]*Department of Electrical Engineering and Computer Science, University of California Irvine, CA, USA*
Email: son.nn6303@gmail.com, kien.lt2414202@sis.hust.edu.vn, tail3@uci.edu
dung.hanhuy@hust.edu.vn, loan.phamnguyenthanh@hust.edu.vn, hungcao@uci.edu

***Abstract*—Fetal ECG (fECG) analysis is important for monitoring fetal heart activity and detecting cardiac anomalies, but its extraction from abdominal ECG (aECG) is challenging due to the low amplitude of fECG and the dominant maternal ECG interference. This paper introduces Envelope-based Fetal ECG Network (EnFE-Net) for fetal ECG detection, combining wavelet-based time-frequency analysis with residual learning to capture complementary signal features while suppressing maternal interference. A Gaussian envelope-guided supervision strategy further refines fetal R-peak localization by optimizing against smooth target masks derived from fetal R-peak annotations. Experiments on benchmark datasets demonstrate robust and accurate detection with an average F1 score of $97.27 \pm 0.81\%$ and a SNR of 9.78 ± 0.36 (dB) on the ADFECGDB dataset, highlighting the effectiveness of this hybrid approach for non-invasive fetal monitoring.**

***Index Terms*—Non-invasive fetal electrocardiogram; Fetal heart rate monitoring.**

I. Introduction

Non-invasive fetal electrocardiography (NI-fECG) is a valuable tool for monitoring fetal well-being during pregnancy and labor, as it provides cardiac information at the beat-level without invasive instrumentation. However, reliable fetal ECG detection from abdominal recordings remains a challenging problem, since the measured signal is a composite mixture of weak fetal cardiac activity, dominant maternal ECG, and various noise sources that overlap substantially in both the temporal and spectral domains [1]. This overlap makes fetal QRS localization difficult, especially when the fetal component is easily masked by stronger maternal interference or background noise [1].

To address this challenge, previous studies have explored a wide range of approaches, including conventional signal processing, wavelet-based analysis, blind source separation, hybrid pipelines, and deep learning methods [2], [3]. However, many recent learning-based methods still focus mainly on signal extraction or waveform reconstruction. Their training objectives are more closely related to recovering the general fetal waveform than to explicitly identifying the fetal cardiac peaks [3]. In practice, the fetal QRS annotations available in benchmark datasets provide direct information about the target events of interest, yet they are not always fully exploited as the primary supervisory signal for detection. As a result, there remains a gap between the optimization objective used during training and the final goal of accurately locating fetal QRS complex.

Motivated by this observation, we formulate fetal ECG detection as a peak-aware learning problem and propose EnFE-Net (Envelope-based Fetal ECG Network). The proposed framework employs the Discrete Wavelet Transform to provide multiscale representations of abdominal ECG signals, allowing the model to capture transient fetal cardiac structures at different resolutions while reducing the impact of irrelevant components. Rather than relying only on waveform-level supervision, we propose an envelope-guided strategy that directs the model toward fetal peak localization. In this framework, the annotated fetal R-peaks are first converted into a smooth Gaussian target mask, creating a probabilistic representation of peak locations. During training, the envelope of the model's output, derived via the Hilbert transform, is normalized and optimized against this Gaussian target using MSE loss. This formulation directly aligns the learning objective with the clinical goal of identifying fetal heartbeats, thereby improving detection sensitivity and reliability in noisy abdominal recordings.

The main contributions of this work are as follows:

- A hybrid architecture, EnFE-Net, is proposed, integrating multi-resolution wavelet decomposition with residual learning for robust fetal ECG feature extraction.
- An envelope-guided supervision strategy is utilized, whereby fetal QRS annotations are used to construct a Gaussian target mask for optimizing peak localization alongside waveform reconstruction.
- An end-to-end detection framework is developed, enabling abdominal ECG recordings to be directly mapped to fetal peak predictions without relying on a separate handcrafted multi-stage detection pipeline.

II. Methodology

A. Preprocessing of Abdominal ECG signal

To ensure consistency across recordings and facilitate downstream learning, all signals were processed using a unified

This work was funded by Vingroup Joint Stock Company and supported by Vingroup Innovation Foundation (VINIF) under project code VINIF.2021.DA00138.

preprocessing pipeline consisting of normalization, bandpass filtering, resampling, and fixed-length segmentation. Clinical ECG signals typically require a minimum bandwidth of 100 Hz [4]. In addition, Sameni and Clifford [5] reported that most ECG spectral energy lies below 35 Hz, while the fetal QRS complex is mainly concentrated in the 10-15 Hz range. Guided by these observations, a band-pass filter with cutoff frequencies of 10 Hz and 100 Hz was applied to suppress baseline wander and high-frequency noise while preserving the relevant fetal ECG components. The filtered signals were then resampled to 250 Hz using interpolation-based resampling so that all modalities shared the same temporal resolution.

Each dataset sample consists of three temporally corresponding components: an abdominal ECG (aECG) signal, a reference fetal ECG (fECG) target, and annotated fetal R-peak locations. After preprocessing, all signals were divided into fixed 4-second segments, aligned on a common time axis, and normalized to the range of [-1, 1]. For a given segment, the preprocessed aECG input is denoted by $x[n]$, where $n = 1, \ldots, N$ indexes the samples within a segment of length N. The corresponding reference fECG target at the same time instants is denoted by $\hat{y}[n]$. The fetal R-peak annotations are further transformed into a Gaussian-smoothed supervision sequence following [6], and this sequence is denoted by $\hat{p}[n]$.

B. Proposed Network Architecture

1) Notation and layer definitions: This section introduces the notation and the main building blocks used in the mode

1D convolution and convolution block: A one-dimensional convolution applied to an input feature vector denoted by $X \in \mathbb{R}^{C\times N}$ is expressed by the formula:

$$Y = \mathrm{Conv}_{k,s}^{C\to C'}(X), \tag{1}$$

where C and C' are the numbers of input and output channels, respectively, k is the kernel size, and s is the stride. The output is the vector $Y \in \mathbb{R}^{C'\times\lfloor N/s\rfloor}$. In the proposed model, all convolutions use kernel size $k = 3$, stride $s = 1$, and padding chosen to preserve the temporal length.

For notational simplicity, we define a convolution block as a convolution followed by batch normalization and ReLU activation:

$$\mathrm{ConvB}_{3,1}^{C\to C'}(X) = \mathrm{ReLU}\Big(\mathrm{BN}\Big(\mathrm{Conv}_{3,1}^{C\to C'}(X)\Big)\Big), \tag{2}$$

which maps $X \in \mathbb{R}^{C\times N}$ to an output in $\mathbb{R}^{C'\times N}$.

Residual block: For an intermediate feature vector $H \in \mathbb{R}^{C\times N}$, the residual block is defined as

$$U = \mathrm{ReLU}\big(\mathrm{BN}\big(\mathrm{Conv}_{3,1}^{C\to C}(H)\big)\big), \tag{3}$$

$$V = \mathrm{BN}\big(\mathrm{Conv}_{3,1}^{C\to C}(U)\big), \tag{4}$$

$$\mathrm{Res}(H) = \mathrm{ReLU}(V + H). \tag{5}$$

Squeeze-and-excitation block: Given an input feature vector $X \in \mathbb{R}^{C\times N}$, the squeeze-and-excitation block first compresses the temporal dimension by global average pooling. For the c-th channel, let $x_c \in \mathbb{R}^{1\times N}$ denote the temporal feature vector of that channel. The pooled descriptor is computed as

$$z_c = \frac{1}{N}\sum_{n=1}^{N} x_c(n), \qquad c = 1, 2, \ldots, C. \tag{6}$$

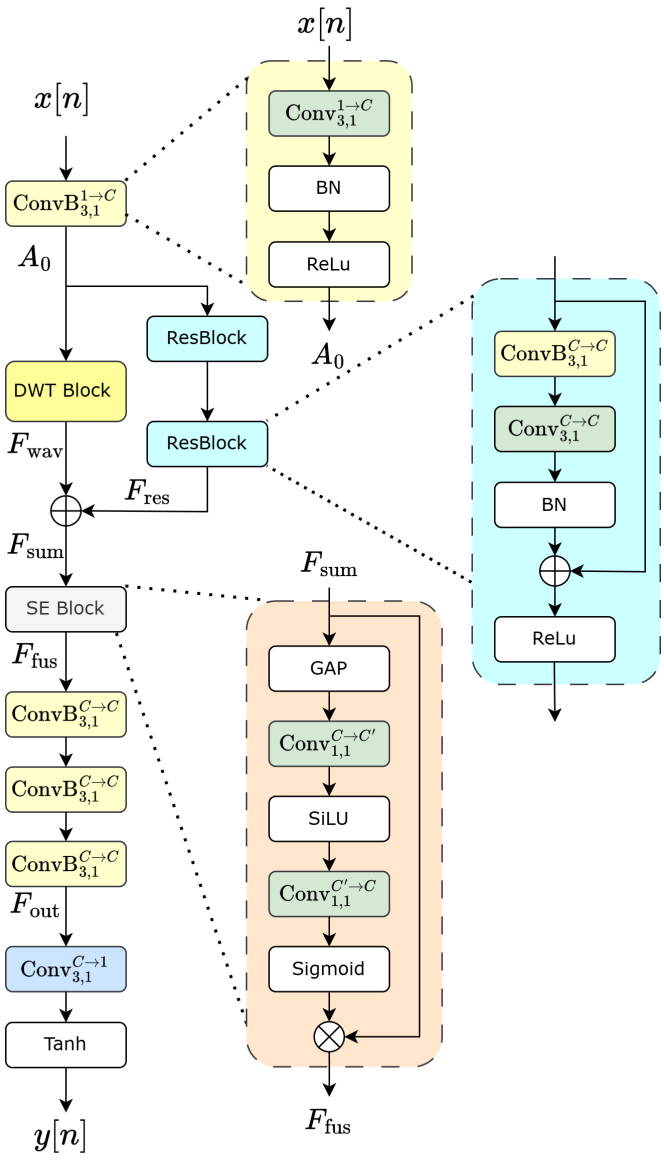

Fig. 1: The proposed EnFe framework for fetal ECG detection, incorporating DWT, residual, and SE blocks.

By stacking all channel descriptors, we obtain

$$Z = [z_1, z_2, \ldots, z_C]^\top \in \mathbb{R}^{C\times 1}. \tag{7}$$

The descriptor is then passed through a pointwise convolution to reduce the channel dimension, followed by a SiLU activation:

$$U = \mathrm{SiLU}\Big(\mathrm{Conv}_{1,\,1}^{C\to C'}(Z)\Big), \qquad U \in \mathbb{R}^{C'\times 1}, \tag{8}$$

where C' denotes the reduced channel dimension. The compressed representation is then mapped back to the original channel dimension:

$$Y = X \odot \sigma\Big(\mathrm{Conv}_{1,\,1}^{C'\to C}(U)\Big), \qquad Y \in \mathbb{R}^{C\times N}. \tag{9}$$

This mechanism explicitly models inter-channel dependencies and adaptively emphasizes informative channels while suppressing less useful ones [7].

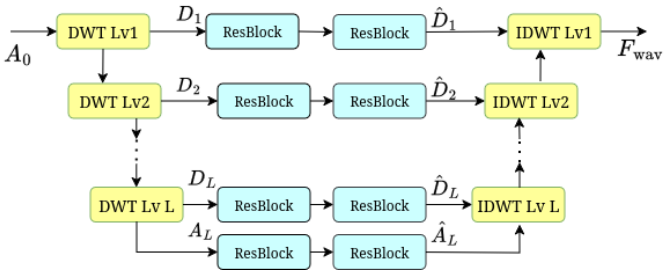

Fig. 2: Internal structure of the Discrete Wavelet Transform (DWT) Block. The input feature A_0 undergoes an L-level discrete wavelet decomposition into approximation and detail coefficients, which are independently refined by Residual Blocks before inverse reconstruction.

2) Multi-level wavelet decomposition and reconstruction: Let $A_0 \in \mathbb{R}^{C\times N}$ denote the input feature vector of the wavelet branch, where C is the number of channels and T is the temporal length. A L-level discrete wavelet transform recursively decomposes the approximation component as

$$A_l[n] = \sum_k h[k]\, A_{l-1}[2n-k], \qquad l = 1,\dots,L. \tag{10}$$

$$D_l[n] = \sum_k g[k]\, A_{l-1}[2n-k], \qquad l = 1,\dots,L. \tag{11}$$

where $h[k]$ and $g[k]$ are the low-pass and high-pass analysis filters, respectively.

After the l-th decomposition level, both A_l and D_l have temporal length $N/2^l$, i.e., $A_l,\ D_l \in \mathbb{R}^{C\times N/2^l}$. Here, A_l denotes the low-frequency approximation at level l, while D_l denotes the corresponding detail component. Hence, D_1 contains the highest-frequency details, and larger l corresponds to lower-frequency subbands, with A_L being the coarsest approximation. Thus, the complete output of the L-level decomposition is written as

$$\{A_L, D_1, \dots, D_L\} = \mathrm{DWT}^{(L)}(A_0). \tag{12}$$

The inverse transform then recursively reconstructs the full-resolution feature:

$$A_{l-1} = \mathrm{IDWT}(A_l, D_l), \qquad l = L,\dots,1, \tag{13}$$

where A_l and D_l denote the approximation and detail coefficients at level l. respectively.

3) Network formulation: Let $\mathbf{x} \in \mathbb{R}^{1\times N}$ represent a single preprocessed abdominal ECG (aECG) segment, where N denotes the number of temporal samples. This segment is then mapped into the feature space through the initial convolution block:

$$A_0 = \mathrm{ConvB}_{3,1}^{1\to \mathrm{C}}(\mathbf{x}), \qquad A_0 \in \mathbb{R}^{C\times N}, \tag{14}$$

where C denotes the base number of channels, the network then splits into two parallel branches.

a) Wavelet branch.: The first branch applies a L-level wavelet transform to A_0:

$$\{A_L, D_1, \dots, D_L\} = \mathrm{DWT}^{(L)}(A_0). \tag{15}$$

The deepest approximation coefficients and each detail subband are refined independently by two consecutive residual blocks:

$$\hat{A}_L = \mathrm{Res}(\mathrm{Res}(\mathrm{A_L})) \qquad l = 1,\dots,L. \tag{16}$$

$$\hat{D}_l = \mathrm{Res}(\mathrm{Res}(\mathrm{D_l})), \qquad l = 1,\dots,L. \tag{17}$$

The refined subbands are then recombined through inverse wavelet reconstruction:

$$F_{\text{wav}} = \mathrm{IDWT}^{(L)}(\hat{A}_L, \hat{D}_1, \dots, \hat{D}_L), \qquad F_{\text{wav}} \in \mathbb{R}^{C\times N}. \tag{18}$$

b) Direct residual branch.: In parallel, the same shared feature vector is processed directly in the time domain:

$$F_{\text{res}} = \mathrm{Res}(\mathrm{Res}(\mathrm{A_0})), \qquad F_{\text{res}} \in \mathbb{R}^{C\times N}. \tag{19}$$

c) Feature fusion: The outputs of the two branches are fused by element-wise addition:

$$F_{\text{sum}} = F_{\text{wav}} + F_{\text{res}}. \tag{20}$$

A squeeze-and-excitation block is then used for channel-wise recalibration:

$$F_{\text{fus}} = \mathrm{SE}(F_{\text{sum}}), \qquad F_{\text{fus}} \in \mathbb{R}^{C\times N}. \tag{21}$$

d) Output mapping.: The fused representation is refined as

$$F_{\text{out}} = \mathrm{ConvB}_{3,1}^{C\to C}\big(\mathrm{ConvB}_{3,1}^{C\to C}\big(\mathrm{ConvB}_{3,1}^{C\to C}(F_{\text{fus}})\big)\big). \tag{22}$$

The final output is then obtained by

$$y[n] = \tanh\big(\mathrm{Conv}_{3,1}^{C\to 1}(F_{\text{out}})\big), \qquad y[n] \in \mathbb{R}^{1\times N}. \tag{23}$$

C. Loss Function

Let $y[n]$ denote the predicted fECG signal and $\hat{y}[n]$ the corresponding reference target, for $n = 1,\dots,N$. Following the Hilbert-transform-based envelope analysis in [8], the amplitude envelope of $y[n]$ is extracted and normalized to $[0,1]$, yielding $\tilde{p}[n]$. In parallel, the binary fetal QRS annotations, $p[n]$, are converted into a Gaussian-smoothed target sequence $\hat{p}[n]$ following [6]. The overall loss is defined as

$$\mathcal{L} = \alpha\mathcal{L}_{\text{MSE}} + \beta\mathcal{L}_{\text{MAE}} + \gamma\mathcal{L}_{\text{env}}, \tag{24}$$

where

$$\mathcal{L}_{\text{MSE}} = \frac{1}{N}\sum_{n=1}^{N}\left(y[n]-\hat{y}[n]\right)^2, \tag{25}$$

$$\mathcal{L}_{\text{MAE}} = \frac{1}{N}\sum_{n=1}^{N}\left|y[n]-\hat{y}[n]\right|, \tag{26}$$

and

$$\mathcal{L}_{\text{env}} = \frac{1}{N}\sum_{n=1}^{N}\left(\tilde{p}[n]-\hat{p}[n]\right)^2. \tag{27}$$

The first two terms jointly enforce waveform fidelity, enabling the model to learn both the overall morphology and local variations of the target fECG signal. The envelope loss $\mathcal{L}_{\text{env}}$ is computed as the mean squared error (MSE) between the normalized predicted envelope $\tilde{p}[n]$ and the Gaussian-smoothed target sequence $\hat{p}[n]$. This term provides peak-aware supervision by encouraging stronger responses around fetal R-peak locations, thereby improving temporal localization accuracy.

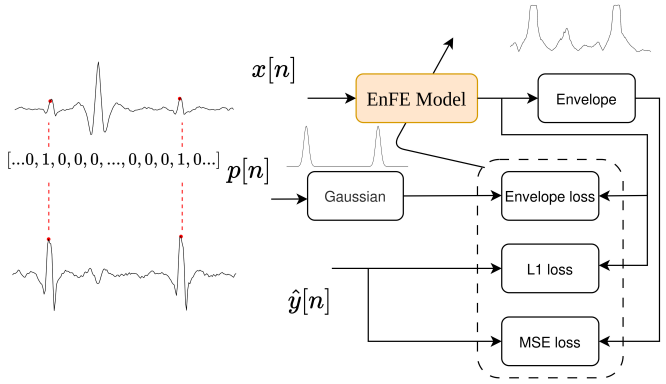

Fig. 3: Training framework combining waveform reconstruction and Gaussian envelope supervision.

III. Experiment

A. Dataset description

The proposed fQRS detection model was trained and evaluated using two independent datasets. The first dataset, the Abdominal and Direct Fetal Electrocardiogram Database (ADFECGDB), is available from PhysioNet [9] and contains five recordings collected from five pregnant women in the 38th to 41st week of pregnancy, denoted as R01, R04, R07, R08, and R10. Each recording includes four abdominal ECG (aECG) channels and one direct fetal ECG (fECG) channel measured from the fetal head. All signals were sampled at 1000 Hz, with a duration of five minutes per recording. Since ADFECGDB provides a reference fetal ECG signal, it was used for supervised training. Following prior work [10], the recordings were partitioned into training, validation, and testing sets with a ratio of 70%, 15%, and 15%, respectively. To avoid temporal leakage, the split was performed at the temporal-block level, such that segments originating from the same contiguous time interval were assigned exclusively to a single subset, i.e., train, validation, or test.

The second dataset was the clinical ECG dataset from the PhysioNet/Computing in Cardiology Challenge 2013 [11]. The PCDB / Challenge 2013 Training Set A (Challenge/2013/set-a) consists of 25 recordings, each containing four abdominal mixed ECG signals acquired from the maternal abdomen, with a sampling frequency of 1000 Hz and a duration of 60s. Unlike ADFECGDB, this dataset does not provide a direct reference fetal ECG waveform for supervised reconstruction. Therefore, it was used only for evaluation in order to assess the generalization capability of the proposed model on unseen recordings.

B. Training and Evaluation

1) Training setup: Our proposed EnFE-Net framework was implemented in PyTorch to extract fetal ECG R-peaks from 1D abdominal recordings. To analyze the impact of time–frequency representations, multiple wavelet families were evaluated with decomposition levels of 4 and 5. For network optimization, we selected the Adam optimizer with a learning rate of 1×10^{-3} and zero weight decay, using a batch size of 64 and training for up to 100 epochs. An early stopping mechanism was applied to prevent overfitting and halt training after 15 epochs of non-improving validation loss. All computational experiments were accelerated using NVIDIA GPUs.

2) Evaluation setup: We evaluated the model performance through two aspects: signal reconstruction quality and fetal QRS detection accuracy. Firstly, Signal-to-Noise Ratio (SNR) was used to quantify the fidelity of waveform reconstruction, according to Yang's formulation [12], in which a higher SNR value indicates better model performance. Secondly, for beat-level fetal QRS detection, we classified detected R-peaks as true positives (TP), false positives (FP), or false negatives (FN) via their correspondence with the reference position. If an R-peak occurred within a tolerance window of ±50 ms around the reference R-peak annotations, the detected fQRS complex would be considered correct [13]. The evaluation metrics, including Sensitivity (Se), Positive Predictive Value (PPV) and the F1-score, were computed based on these definitions [12] to assess the accuracy and reliability of our model.

TABLE I: Comparison with published methods on the PCDB and ADFECGDB test sets. The proposed method uses the Coiflet1 wavelet with five decomposition levels and loss weights of $\alpha = 0.2$, $\beta = 0.5$, and $\gamma = 0.3$, as defined in (24).

Study	Se (%)	PPV (%)	F1 (%)
PCDB			
Zhong [14]	67.05 ± 30.23	68.18 ± 30.02	67.61 ± 30.13
Barnova [15]	81.79 ± 25.29	87.16 ± 20.98	84.08 ± 23.27
Proposed method	$\mathbf{83.18 \pm 22.27}$	$\mathbf{85.25 \pm 19.42}$	$\mathbf{84.12 \pm 20.91}$
ADFECGDB			
Su [16]	90.37 ± 10.26	89.08 ± 11.00	89.71 ± 10.61
Zhong [14]	92.45 ± 10.59	88.56 ± 9.87	90.45 ± 10.19
Zhang [17]	96.55 ± 1.33	96.56 ± 2.67	96.56 ± 1.76
Proposed method	$\mathbf{98.67 \pm 0.61}$	$\mathbf{95.91 \pm 1.12}$	$\mathbf{97.27 \pm 0.81}$

IV. Results

Table I presents the performance comparison of the proposed method against several published approaches on the PCDB and ADFECGDB testing sets. The evaluation metrics include sensitivity (Se), positive predictive value (PPV), and F1-score, reported as mean ± standard deviation.

In the PCDB dataset, the proposed method achieved a sensitivity of $83.18 \pm 22.27\%$, a PPV of $85.25 \pm 19.42\%$, and an F1-score of $84.12 \pm 20.91\%$. The large standard deviations reflect substantial signal quality variability in this dataset, where fetal ECG components are often heavily contaminated by maternal ECG and noise. Despite these challenging conditions, the method maintains a competitive average performance compared to existing approaches.

In the ADFECGDB dataset, the proposed method demonstrated substantial improvements, achieving a sensitivity of $98.67 \pm 0.61\%$, a PPV of $95.91 \pm 1.12\%$, and an F1-score of $97.27 \pm 0.81\%$. These results compare favorably with prior works, including Zhang [17], which reported an F1-score of $96.56 \pm 1.76\%$. The high and consistent performance across all three metrics highlights the robustness and generalizability of the proposed method for fetal ECG extraction.

As illustrated in Fig. 4, different wavelet families exhibit varying trade-offs between signal reconstruction quality (SNR) and fetal R-peak detection accuracy (F1-score). While some wavelets achieve higher SNR values, this does not necessarily translate into improved detection performance. Noticeably, Reverse Biorthogonal 4.4 (rbio4.4) produces a slightly higher SNR than Coiflet-1 (coif1), indicating stronger waveform reconstruction, yet its F1-score is lower. This suggests that better morphological reconstruction does not always guarantee more accurate peak localization. Among the evaluated wavelets, Coiflet-1 achieves the most balanced performance by maintaining competitive SNR while providing the highest detection accuracy, likely due to its near-symmetric filters and good time–frequency localization that preserve transient fetal QRS structures.

In addition, the weight loss analysis in Fig. 5 reveals a similar trade-off. Increasing the envelope loss weight tends to improve fetal beat detection performance, as reflected by higher F1-scores, because the supervision emphasizes peak-related features. However, this often leads to a slight degradation in waveform morphology, resulting in reduced SNR. Notably, the envelope loss weight appears to be the primary factor governing this trade-off, as configurations with different envelope weights form distinct operating regions in the F1–SNR space. In contrast, varying the relative contributions of MSE and MAE mainly fine-tunes performance within each region. Therefore, selecting appropriate loss weights is important to balance accurate peak detection and faithful signal reconstruction.

Figure 6 illustrates representative examples of both successful and difficult detection scenarios. Strong maternal interference and noise may lead to spurious peaks or missed detections, as highlighted by the red and purple regions in the figure. Nevertheless, the proposed method demonstrates robust and efficient performance overall, confirming that the combination of wavelet-based feature extraction and peak-aware supervision enables reliable and precise fetal QRS complex detection in non-invasive fetal ECG monitoring.

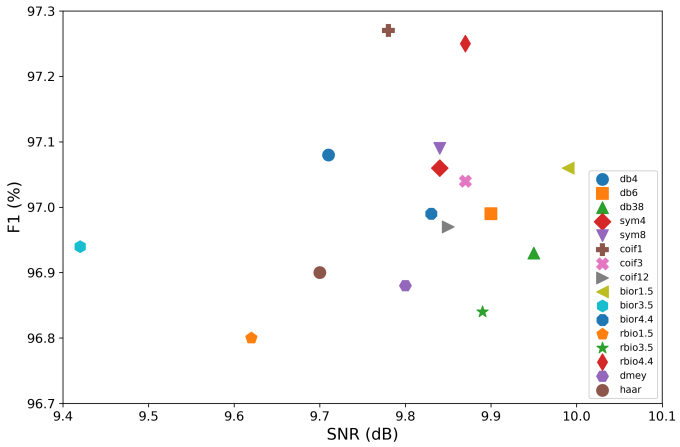

Fig. 4: **C**omparison of different wavelet families in terms of SNR and F1-score. Each marker represents the average reconstruction performance of a model trained with a specific wavelet basis, evaluated in terms of output SNR (horizontal axis) and F1 (vertical axis). The results were obtained using the ADFECGDB dataset.

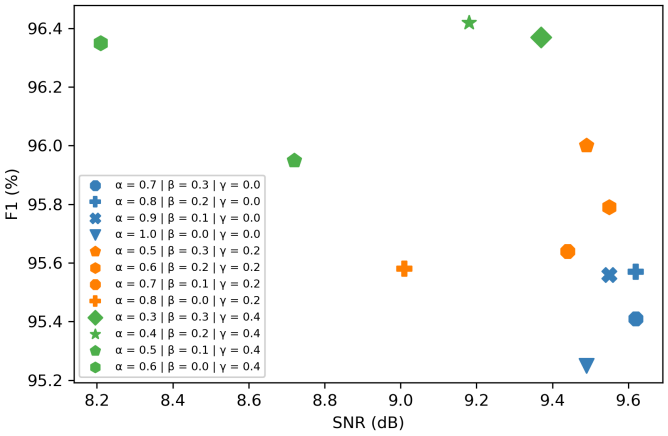

Fig. 5: Relationship between signal reconstruction quality (SNR) and fetal R-peak detection accuracy (F1-score) for different loss weight configurations. The parameters α, β, and γ correspond to the weights of the MSE, MAE, and envelope loss terms. Experiments were conducted on the ADFECGDB dataset, with results averaged across the wavelets evaluated in Fig. 4.

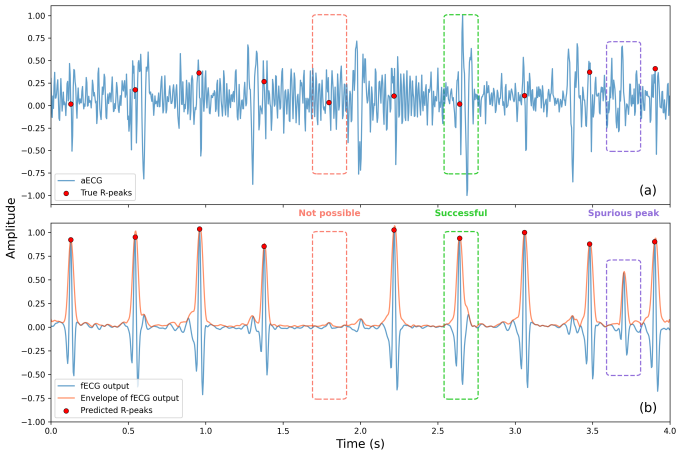

Fig. 6: **Performance overview.** Panel A represents the raw input of abdECG signals, and panel B illustrates the reconstructed fECG signal with detected R-peaks. The green rectangle indicates successful fECG separation, while the red rectangle highlights areas where fECG separation is not possible. The purple box marks a spurious peak caused by strong noise in the input signal.

V. Discussion

This study evaluates the performance of the proposed EnFE-Net framework for fetal ECG extraction and R-peak detection from abdominal recordings. The results demonstrate that combining wavelet-based representations with envelope-guided supervision enables reliable fetal heartbeat localization even in noisy abdominal recordings.

The key aspect of our approach is an envelope-guided supervision strategy that directs the model to focus on fetal QRS complexes rather than only reconstructing the waveform. Annotated R-peaks are transformed into smooth Gaussian masks, and the amplitude output of the envelope is optimized against these targets. This enables the network to learn both peak locations and general signal structure. However, increasing the envelope loss introduces a trade-off: it improves peak detection but may slightly degrade waveform morphology. This occurs because the normal baseline reconstruction loss (MAE, MSE) preserves the entire fECG morphology, including low-amplitude structures like P-waves and T-waves. In contrast, Gaussian envelope supervision acts as a temporal attention weighting function around R-peaks, causing gradients to be concentrated within QRS regions. Consequently, the model favors peak-centric representations that enhance the temporal localization and prominence of fetal QRS complexes, while non-QRS waveform components receive comparatively weaker supervision and may become attenuated or distorted. Accordingly, the loss weights can be adjusted based on the clinical objective—prioritizing waveform fidelity for morphological analysis or emphasizing peak detection for accurate heart rate estimation.

Another important aspect of the architecture is its ability to jointly capture temporal and frequency features for fetal ECG extraction. The wavelet branch performs multi-level decomposition using the Discrete Wavelet Transform, allowing the network to learn frequency-specific representations through band-wise residual processing of the approximation and detail coefficients. This facilitates the isolation of transient fetal QRS activity, which typically appears in higher-frequency detail bands. In parallel, a residual branch operates directly in the time domain to preserve temporal structures associated with fetal cardiac activity. The outputs of these two branches are then combined and passed through a squeeze-and-excitation (SE) block, which models inter-channel dependencies and adaptively recalibrates feature responses. Thus, this mechanism helps preserve the overall waveform structure and partially compensates for the morphology degradation introduced by the envelope-guided supervision.

Despite these promising results, the envelope-guided supervision introduces a trade-off between waveform morphology preservation and peak detection accuracy, and future work will explore strategies to better balance these objectives in larger datasets and real-time settings.

VI. Conclusion

This paper presented EnFE-Net, a lightweight hybrid framework for fetal ECG detection that combines Discrete Wavelet Transform-based time-frequency analysis with residual learning and an envelope-guided supervision strategy using Gaussian target masks. Systematic evaluation identified Coiflet-1 with five decomposition levels as the most effective configuration. The proposed method achieves state-of-the-art performance with F1-scores of 84.12% on PCDB and 97.27% on ADFECGDB, achieving competitive performance compared with existing approaches. These results demonstrate that integrating wavelet-based feature extraction with peak-aware supervision provides a robust solution for non-invasive fetal ECG monitoring. However, the trade-off between waveform morphology preservation and fetal beat detection accuracy remains a potential limitation, which future studies may address by exploring approaches that better balance both aspects.

References

[1] G. D. Clifford, I. Silva, J. Behar, and G. B. Moody, "Non-invasive fetal ecg analysis," *Physiological Measurement*, vol. 35, no. 8, pp. 1521–1536, 2014.

[2] R. Kahankova, R. Martinek, R. Jaros, K. Behbehani, A. Matonia, M. Jezewski, and J. A. Behar, "A review of Signal Processing Techniques for Non-Invasive Fetal Electrocardiography," *IEEE Reviews in Biomedical Engineering*, vol. 13, pp. 51–73, 8 2019.

[3] X. Li, J. Wan, and X. Peng, "Review of non-invasive fetal electrocardiography monitoring techniques," *Sensors*, vol. 25, no. 5, p. 1412, 2025.

[4] J. J. Bailey, A. S. Berson, A. Garson, L. G. Horan, P. W. Macfarlane, D. W. Mortara, and C. Zywietz, "Recommendations for standardization and specifications in automated electrocardiography: bandwidth and digital signal processing. A report for health professionals by an ad hoc writing group of the Committee on Electrocardiography and Cardiac Electrophysiology of the Council on Clinical Cardiology, American Heart Association.," *Circulation*, vol. 81, pp. 730–739, 2 1990.

[5] R. Sameni and G. D. Clifford, "A review of fetal ecg signal processing; issues and promising directions," *The open pacing, electrophysiology & therapy journal*, vol. 3, p. 4, 2010.

[6] Q. Huang, H. Yang, E. Zeng, and Y. Chen, "A deep-learning-based multimodal ecg and pcg processing framework for label efficient heart sound segmentation," in *2024 IEEE/ACM Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*, pp. 109–119, IEEE, 2024.

[7] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation Networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7132–7141, 6 2018.

[8] N. M. Arzeno, Z.-D. Deng, and C.-S. Poon, "Analysis of first-derivative based qrs detection algorithms," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 2, pp. 478–484, 2008.

[9] J. Jeżewski, A. Matonia, T. Kupka, D. Roj, and R. Czabański, "Abdominal and direct fetal ecg database (version 1.0.0)," 2012.

[10] H. Ghonchi and V. Abolghasemi, "A dual attention-based autoencoder model for fetal ecg extraction from abdominal signals," *IEEE Sensors Journal*, vol. 22, no. 23, pp. 22908–22918, 2022.

[11] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[12] Y. Yang, L. Chen, and S. Wu, "Enhancing fetal electrocardiogram signal extraction accuracy through a cyclegan utilizing combined cnn–bilstm architecture," *Sensors*, vol. 24, no. 9, p. 2948, 2024.

[13] L. Billeci and M. Varanini, "A Combined Independent Source Separation and Quality Index Optimization Method for Fetal ECG Extraction from Abdominal Maternal Leads," *Sensors*, vol. 17, p. 1135, 5 2017.

[14] W. Zhong, Z. Cao, X. Guo, and G. Wang, "A Novel Framework for Maternal ECG Removal from Single-Channel Abdominal Recording," in *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine*, pp. 616–621, 11 2019.

[15] K. Barnova, R. Martinek, R. Jaros, R. Kahankova, A. Matonia, M. Jezewski, R. Czabanski, K. Horoba, and J. Jezewski, "A novel algorithm based on ensemble empirical mode decomposition for non-invasive fetal ECG extraction," *PLoS ONE*, vol. 16, p. e0256154, 8 2021.

[16] L. Su and H.-T. Wu, "Extract Fetal ECG from Single-Lead Abdominal ECG by De-Shape Short Time Fourier Transform and Nonlocal Median," *Frontiers in Applied Mathematics and Statistics*, vol. 3, 2 2017.

[17] W.-T. Zhang, Z.-Z. Huang, Y.-Y. Ma, and D.-J. Zhang, "A fast adaptive LPCA method for fetal ECG extraction based on multichannel signals," *IEEE Transactions on Instrumentation and Measurement*, vol. 73, pp. 1–11, 12 2023.