

Cross-Modal Generative Framework for Signal Translation from Fetal–Maternal Electrocardiograms to Fetal Doppler Waveforms

Tongli Su¹, Alireza Rafiei¹, Marly van Assen¹, Reza Sameni^{1,2},
Gari D. Clifford^{1,2}, Faezeh Marzbanrad³, Nasim Katebi¹

¹Emory University, Atlanta, GA, USA, ²Georgia Institute of Technology, Atlanta, GA, USA

³Monash University, Clayton, VIC, Australia

Abstract—Fetal electrocardiogram (fECG) and Doppler ultrasound provide complementary views of fetal cardiovascular function: fECG captures electrical activity while Doppler reflects mechanical hemodynamics shaped by factors such as placental resistance and vascular compliance. Understanding the recoverable and unrecoverable Doppler components through reconstruction from fECG offers insight into the relative contributions of electrical versus mechanical factors in fetal circulation, thereby informing clinical decisions. In addition, clinical evidence of maternal–fetal cardiac coupling suggests that maternal cardiovascular dynamics may also inform fetal hemodynamics. To computationally model these relationships, we propose a cross-modal generative framework combining dilated convolutions with cross-modal attention to selectively incorporate maternal ECG and self-attention to capture long-range temporal dependencies. Trained on 885 synchronized fetal/maternal ECG and Doppler envelope segments from 39 pregnancies, our model synthesizes Doppler envelopes with power spectral density mean squared error (PSD MSE) of $49.9 \pm 15.8 \text{ dB}^2$ (51% lower than two-channel baseline) and heart-rate error of $4.71 \pm 0.77 \text{ bpm}$ (1.5% better than baseline; negligible relative to the 110–160 bpm physiological range). Cross-modal attention yields 39% PSD MSE reduction over naïve dual-channel concatenation, quantifying the contribution of maternal–fetal coupling. Our proposed framework advances computational modeling of the maternal–fetal cardiovascular system by enabling the synthesis of Doppler envelopes from dual-lead ECG. By analysis of both recoverable and residual Doppler components, this approach enables quantification of the purely mechanical contributions to Doppler waveforms—those not recoverable from electrical recordings—ultimately facilitating a more comprehensive fetal assessment.

Index Terms—Attention mechanism, cross-modal synthesis, deep generative learning, Doppler ultrasound, fetal health monitoring, maternal–fetal coupling

I. INTRODUCTION

Fetal electrocardiogram (fECG) and Doppler ultrasound provide complementary views of fetal cardiovascular function. The fECG captures the electrical activity of the heart, primarily heart rate and rhythm, while Doppler reflects mechanical hemodynamics including blood flow velocity patterns [1]. Clinical indices derived from Doppler waveforms characterize diastolic function and overall cardiac performance [2], [3].

However, Doppler acquisition requires expensive equipment and skilled sonographers, limiting accessibility in resource-constrained settings [1]. In contrast, fECG can be collected widely and inexpensively but lacks direct measurements of mechanical function. Rather than treating these modalities as interchangeable, we recognize that fECG and Doppler capture fundamentally different aspects of fetal cardiac physiology. Excitation-contraction coupling links electrical depolarization to mechanical contraction: each R-peak marks systolic onset after an electromechanical delay of roughly 60–80 ms [4], [5]. Correlations between ECG features and contraction vigor suggest that fECG encodes partial information about mechanical events [6]. However, Doppler waveforms are also shaped by purely mechanical factors, including ventricular compliance, valve function, and preload conditions, which have no direct electrical signature [2]. Understanding which Doppler-based signatures can be recovered from fECG, and which cannot, offers insight into the relative contributions of electrical versus mechanical factors in fetal cardiac function. Clinical evidence further suggests that maternal cardiovascular dynamics may inform fetal hemodynamics through maternal–fetal cardiac coupling (MFCC). Van Leeuwen *et al.* demonstrated phase synchronization between maternal and fetal heartbeats [33], while transfer entropy analysis has quantified significant bidirectional information transfer that strengthens with gestational age [8]. A scoping review reported that 91.3% of studies investigating MFCC detected at least occasional coupling, where the failure of traditional linear methods to identify this interaction suggests a complex and non-linear relationship [9]. This coupling likely occurs through hemodynamic interactions within the circulatory system, vibroacoustic transmission or autonomic links, though the underlying physiological pathways have not been definitively verified [9]. These findings motivate including maternal ECG as an auxiliary input for modeling fetal hemodynamics.

A. Computational Modeling of Maternal–Fetal Cardiac Relationships

To computationally model relationships between fECG, maternal ECG, and fetal Doppler, we require an architecture

This work was supported in part by the NIH (R01HD110480) and Google.org AI for the Global Goals Impact Challenge. Email: {ken.su}@emory.edu

capable of selectively integrating information from multiple sources. We hypothesize that maternal influences on fetal hemodynamics are conditional and time-varying, occurring intermittently rather than uniformly. Simply concatenating maternal and fetal signals treats all maternal information as equally relevant, failing to exploit this selective coupling.

Cross-modal attention addresses this limitation by treating fetal features as queries and maternal features as keys/values, enabling the model to emphasize maternal ECG segments temporally correlated with fetal blood flow while suppressing unrelated noise [10]. Self-attention then captures long-range dependencies within the fetal signal, modeling how early cardiac events influence later phases. Together, these mechanisms enable investigation of both the recoverable Doppler components, reflecting excitation-contraction coupling and maternal-fetal interactions, and residual components that may reveal purely mechanical influences.

B. Contributions

Our main contributions are:

- 1) We propose a novel cross-modal generative framework combining dilated convolutions with cross-modal and self-attention to model the relationship between fetal/maternal ECG and fetal Doppler waveforms. This approach enables the computational investigation of maternal-fetal cardiac coupling and the decomposition of Doppler components into electrically recoverable and residual mechanical components.
- 2) We demonstrate that naïve inclusion of maternal ECG provides no benefit and can degrade performance; only when the model learns to selectively attend to maternal signals does the coupling information improve Doppler reconstruction, quantifying the contribution of maternal-fetal cardiac interactions.
- 3) We introduce a composite loss function for physiological waveform reconstruction by balancing pointwise error, derivative error, and correlation to enforce both amplitude accuracy and morphological fidelity.

II. BACKGROUND AND RELATED WORK

A. Physiology and Doppler Waveforms

Pulsed-wave Doppler ultrasound estimates blood-cell velocity from the frequency shift between transmitted and received waves. For target velocity v , $\Delta f = (2vf_0 \cos \theta)/c_s$, where f_0 is transmit frequency, θ the insonation angle, and c_s the speed of sound in tissue [13]. In fetal applications, range gating samples specific vessels to produce a one-dimensional velocity envelope with sawtooth morphology: a rapid rise to peak systolic velocity (PSV), followed by decay toward end-diastolic velocity (EDV) during relaxation [14]. EDV reflects placental vascular resistance; placental dysfunction reduces EDV, with absent end-diastolic flow indicating approximately 40% loss of the placental vascular tree and reversed flow associated with a tenfold increase in perinatal mortality [14].

B. ECG-Doppler Temporal Relationships

Electromechanical coupling creates reproducible timing between ECG and Doppler features. QRS onset precedes PSV by the pre-ejection period (PEP) plus rapid ejection time; in near-term fetuses, PEP is approximately 70.0 ± 2.4 ms and comprises electromechanical delay plus isovolumetric contraction [6]. Ventricular systole ends with aortic valve closure about 225 ± 13 ms after QRS onset, coinciding with the terminal T-wave and transition to diastole [7].

C. Generative Models for Physiological Signals

Early attempts to synthesise physiological waveforms used mathematical and dynamical models. McSharry *et al.* described a two-stage dynamical model that generates synthetic ECG by summing Gaussian functions representing the P, QRS, and T waves and controlling RR interval spectra [16]. Sameni *et al.* extended this approach to multichannel ECG using a nonlinear Bayesian framework [17]. While these models provide control over beat morphology, they cannot capture the full complexity of real signals.

The advent of deep learning enabled data-driven generative models. Dilated convolution networks such as those introduced in the WaveNet architecture were initially developed for raw-audio synthesis and have been adapted to produce realistic ECG segments without manual feature engineering [18], [19]. These networks achieved large receptive fields through dilation, capturing long-range dependencies without excessive parameter counts. Generative adversarial networks (GANs) have been used to synthesise fetal heart rate traces [20] and multi-lead ECGs, while diffusion models extend to multimodal conditional generation [21].

D. Cross-Modal Synthesis of Physiological Signals

Cross-modal synthesis seeks to infer an inaccessible modality from an easily acquired one. Photoplethysmogram (PPG) signals have been used to reconstruct ECG waveforms via deep learning, leveraging cycle-level correlation between the modalities [22]. Motion sensors such as radar and wrist accelerometers have also been used for ECG and pulse waveform reconstruction [24]. For fetal monitoring, Verma *et al.* recently addressed Doppler envelope synthesis from fECG using a U-Net architecture [25], and Rafiei *et al.* introduced Auto-FEDUS, an autoregressive dilated convolution model producing high-quality Doppler signals from fetal ECG alone [26]. While this approach demonstrated the feasibility of fECG-to-Doppler translation, it did not incorporate maternal ECG despite physiological evidence of coupling between the maternal and fetal cardiovascular systems.

E. Attention Mechanisms and Multi-Channel Fusion

Attention mechanisms enable models to assign varying weights to different inputs based on learned relevance [10]. In maternal-fetal signal processing, attention has been used to extract fetal ECG from mixed maternal-fetal abdominal recordings. Wang *et al.* proposed a correlation-aware attention CycleGAN to emphasise fetal segments when maternal ECG

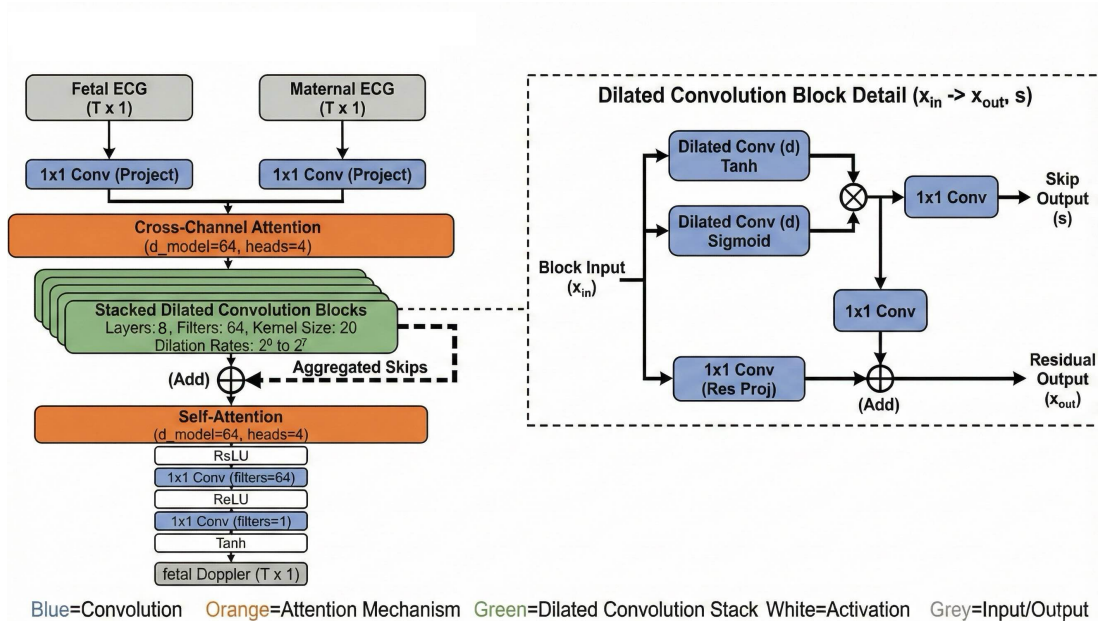


Fig. 1. Overview of the proposed cross-modal generative model. Dilated convolution layers extract features from fetal and maternal ECG, cross-modal attention selectively fuses maternal information based on fetal queries, and self-attention captures long-range dependencies to generate Doppler velocity envelopes.

components overlap [27]. Qu *et al.* integrated multi-head self-attention and channel-wise attention with contrastive learning for fetal ECG extraction [28], and Nguyen *et al.* applied conditional GANs on spectrograms for improved fECG separation [29]. Beyond ECG, attention mechanism has also been applied to fetal Doppler signals in different tasks such as gestational age estimation, quality assessment and heart rate estimation [35]–[37]. These works highlight the need for selective processing of maternal and fetal signals. To our knowledge, attention has not previously been applied to cross-modal synthesis of fetal Doppler envelopes from joint maternal–fetal ECG.

III. METHODS

A. Dataset and Preprocessing

We used the NInFEA dataset [30], which provides synchronized abdominal ECG and pulsed-wave Doppler recordings from 39 pregnant women between 21 and 27 weeks of gestation. The dataset was collected under approval from the Independent Ethics Committee of the Cagliari University Hospital (AOU Cagliari) following the principles outlined in the Helsinki Declaration; all participants provided signed informed consent [30]. From the available recordings, we extracted 885 paired segments, each 3.75 s long. At typical fetal heart rates (110–160 bpm), this duration spans six to ten cardiac cycles, capturing sufficient temporal context for estimating indices such as PSV and EDV and allowing our model to learn interbeat patterns; prior work on cross-modal ECG reconstruction found that multiple consecutive beats yield more accurate generative models than single-beat inputs [23]. Each pair comprised one fetal ECG channel and the corresponding maternal ECG channel. We applied envelope

extraction as described in the dataset paper [30] to extract Doppler velocity envelopes, which were then low-pass filtered (4th-order Butterworth, 10 Hz cutoff) to attenuate high-frequency artifacts. All signals were resampled to 284 Hz and normalised per segment to zero mean and unit variance.

B. Architecture Design

Figure 1 illustrates the proposed cross-modal generative architecture, which comprises three main components: cross-modal attention for selective maternal-fetal fusion, dilated residual blocks for multi-scale temporal modeling, and self-attention for capturing long-range dependencies.

Cross-Modal Attention: Convolutional encoders extract features from fetal (F_f) and maternal (F_m) ECG. Multi-head cross-attention [10] with four heads computes:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right) \mathbf{V} \quad (1)$$

where $\mathbf{Q} = F_f \mathbf{W}^Q$, $\mathbf{K} = F_m \mathbf{W}^K$, $\mathbf{V} = F_m \mathbf{W}^V$, and $d_k = 64$. This selective fusion conditions maternal information on fetal features, reflecting physiological maternal-fetal coupling.

Dilated Residual Blocks: Eight residual blocks with dilated convolutions [11] at rates $d \in \{2^0, \dots, 2^7\}$ expand the receptive field to spanning multiple cardiac cycles (0.88 seconds) without significant parameter growth. Each block applies gated activation, which empirically outperforms standard ReLU activations for sequential signal modeling [12]:

$$\mathbf{z}_k = \tanh(\mathbf{W}_{f,k} *_{d_k} \mathbf{x}_k) \odot \sigma(\mathbf{W}_{g,k} *_{d_k} \mathbf{x}_k) \quad (2)$$

where $*_{d_k}$ denotes dilated convolution. Outputs split into residual ($\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{W}_{r,k}^{1 \times 1} \mathbf{z}_k$) and skip ($\mathbf{s}_k = \mathbf{W}_{s,k}^{1 \times 1} \mathbf{z}_k$) paths. Aggregated skips $\mathbf{S} = \sum_{k=1}^8 \mathbf{s}_k$ combine multi-scale

TABLE I
SIGNAL EXTRAPOLATION PERFORMANCE METRICS COMPARISON ACROSS ARCHITECTURES. VALUES: MEAN $\pm$ SD ACROSS 5-FOLD CROSS-VALIDATION.

Metric	Single-Channel (1-ch, no attn)	Two-Channel (2-ch, no attn)	Cross-Attention (2-ch, cross-attn)	Combined Attention (2-ch, cross+self)
<i>Pointwise error</i>				
MAE $\downarrow$	0.407 $\pm$ 0.100	0.403 $\pm$ 0.082	0.404 $\pm$ 0.096	0.403 $\pm$ 0.103
MSE $\downarrow$	0.243 $\pm$ 0.108	0.235 $\pm$ 0.085	0.244 $\pm$ 0.104	0.245 $\pm$ 0.111
RMSE $\downarrow$	0.481 $\pm$ 0.102	0.476 $\pm$ 0.082	0.483 $\pm$ 0.098	0.482 $\pm$ 0.107
Gradient MAE $\downarrow$	0.034 $\pm$ 0.007	0.034 $\pm$ 0.007	0.035 $\pm$ 0.007	0.035 $\pm$ 0.007
<i>Correlation</i>				
Cross-correlation (zero lag) $\uparrow$	0.287 $\pm$ 0.378	0.268 $\pm$ 0.365	0.279 $\pm$ 0.364	0.270 $\pm$ 0.384
<i>Sequence alignment</i>				
DTW (a.u.) $\downarrow$	287.2 $\pm$ 66.9	340.0 $\pm$ 69.4 $\dagger$	261.8 $\pm$ 60.8	253.7 $\pm$ 59.7
<i>Distribution similarity</i>				
KLD $\downarrow$	10.64 $\pm$ 2.38	13.17 $\pm$ 2.08	9.12 $\pm$ 2.43	9.19 $\pm$ 2.39
<i>Frequency domain</i>				
PSD MSE (dB ²) $\downarrow$	84.1 $\pm$ 17.1	102.0 $\pm$ 20.0	61.8 $\pm$ 16.6	49.9 $\pm$ 15.8
PSD correlation $\uparrow$	0.957 $\pm$ 0.012	0.948 $\pm$ 0.012	0.962 $\pm$ 0.011	0.970 $\pm$ 0.010

Bold: Best performance. $\uparrow$: higher is better; $\downarrow$: lower is better.

$\dagger$ Two-channel DTW is highest despite best MSE/RMSE, indicating that minimising pointwise error yields amplitude-correct but temporally misaligned envelopes.

features from dilated convolution blocks to model both local pattern and global temporal structure in Doppler waveforms.

Self-Attention: Multi-head self-attention processes $\mathbf{S}$ with $\mathbf{Q} = \mathbf{K} = \mathbf{V} = \mathbf{S}\mathbf{W}^{QKV}$ to capture long-range temporal dependencies across cardiac cycles before dense layers produce the Doppler waveform.

Model Variants: To quantify the contributions of each component, we implemented four variants: (1) *single-channel*, using only fetal ECG; (2) *two-channel*, concatenating fetal and maternal ECG without attention; (3) *cross-attention*, employing cross-modal attention but no self-attention; and (4) *combined attention*, incorporating both cross-modal and self-attention.

C. Loss Function Design

We trained the model using a composite loss that encourages pointwise fidelity, smooth temporal gradients and morphological similarity:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{MAE}} + \alpha \mathcal{L}_{\text{deriv}} + \alpha \mathcal{L}_{\text{corr}}, \quad (3)$$

where $\mathcal{L}_{\text{MAE}} = \frac{1}{T} \sum_t |\hat{y}_t - y_t|$ is the mean absolute error, $\mathcal{L}_{\text{deriv}} = \frac{1}{T} \sum_t |(\hat{y}_t - \hat{y}_{t-1}) - (y_t - y_{t-1})|$ penalises differences in first derivatives to promote physiologically realistic systolic upstrokes and diastolic decay, and $\mathcal{L}_{\text{corr}} = 1 - \rho(\hat{y}, y)$ encourages high Pearson correlation between generated and real envelopes. We set $\alpha = 0.5$ to assign equal weight to each auxiliary objective, ensuring approximately balanced gradient contributions from pointwise error, derivative matching, and correlation alignment. This choice was informed by monitoring individual loss component magnitudes during preliminary training runs, which confirmed that equal weighting prevented any single term from dominating the total gradient and yielded stable convergence.

D. Implementation Details

Models were trained using the Adam optimiser with an initial learning rate of 10^{-3} . The learning rate was reduced by a factor of 0.5 if the validation loss plateaued for five epochs, with a minimum of 10^{-5} . Early stopping with a patience of 20 epochs prevented overfitting. We used a batch size of 8 and trained for up to 100 epochs per fold. Fivefold cross-validation with patient-level splitting assessed generalisation; no patient overlapped between training and validation sets.

E. Evaluation Metrics

Reconstruction performance was assessed using metrics that capture time-domain, frequency-domain, and distributional properties. Pointwise errors included the mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), and gradient MAE (error in first derivatives). Correlation measures comprised the Pearson correlation coefficient and zero-lag cross-correlation. Temporal alignment was quantified using dynamic time warping (DTW) distance; since all signals are normalised to zero mean and unit variance prior to evaluation, DTW values are in normalised amplitude units (a.u.). Distributional similarity was measured by the Kullback–Leibler divergence (KLD) between histograms of generated and real envelopes. Frequency-domain fidelity was evaluated by the MSE and correlation of the power spectral density (PSD). PSD was estimated using Welch’s method [38] with a sampling rate of $f_s = 284$ Hz and a segment length of $n_{\text{seg}} = 256$ samples, yielding a frequency resolution of approximately 1.1 Hz. PSD values were converted to decibels ($10 \log_{10}(\text{PSD} + \epsilon)$, where $\epsilon = 10^{-10}$) before computing the MSE, so PSD MSE is reported in units of dB². Clinical relevance was assessed by computing MAE and correlation

for PI, RI, S/D ratio, PSV, EDV, time-averaged maximum (TAMX), and heart rate.

IV. RESULTS

A. Overall Performance

Table I summarizes reconstruction performance. The combined attention model achieves the lowest DTW (253.7 ± 59.7 a.u.) and best frequency-domain fidelity: PSD MSE of 49.9 ± 15.8 dB², which is 51% lower than the two-channel baseline, and PSD correlation of 0.970 ± 0.010 . Naïvely concatenating maternal and fetal ECG yields the lowest MSE and RMSE but the worst DTW, indicating that pointwise accuracy alone does not guarantee morphological alignment. By contrast, cross-modal and self-attention improve both temporal alignment and frequency-domain fidelity.

B. Ablation Analysis

Maternal ECG paradox: Single-channel and two-channel baselines yield nearly identical pointwise error (MAE 0.407 vs. 0.403) and correlation, while the two-channel variant has a substantially worse DTW. This shows that naïve inclusion of maternal ECG provides no benefit and can degrade temporal structure.

Cross-attention impact: Introducing cross-channel attention dramatically improves performance relative to the naïve two-channel model. As shown in Table I, cross-attention reduces DTW by 23%, KLD by 31% and PSD MSE by 39% compared with simple concatenation, confirming that selective incorporation of maternal information is critical for capturing maternal–fetal coupling.

Self-attention contribution: Adding self-attention on top of cross-attention yields further gains. Self-attention provides an additional 19% improvement in PSD MSE (from 61.8 ± 16.6 to 49.9 ± 15.8 dB²) and a modest 3% improvement in DTW, reflecting its ability to capture long-range temporal dependencies and beat-to-beat consistency.

Example Doppler velocity envelopes generated by each model variant are shown in Figure 2. Each panel compares the generated envelope (blue) with the ground truth (red) for a representative segment, illustrating how attention mechanisms progressively improve temporal alignment and morphological fidelity.

C. Clinical Metrics Performance

As summarized in Table II, all models achieved clinically acceptable heart rate accuracy (MAE 4.69–4.80 bpm; correlation 0.55–0.58). Given that fetal bradycardia and tachycardia are defined as heart rates below 110 bpm and above 160 bpm, respectively [15], an absolute error of less than 5 bpm is negligible relative to the physiological range. Combined attention achieved the best performance in restoring S/D ratio (MAE 0.66), which is crucial for detecting elevated placental resistance.

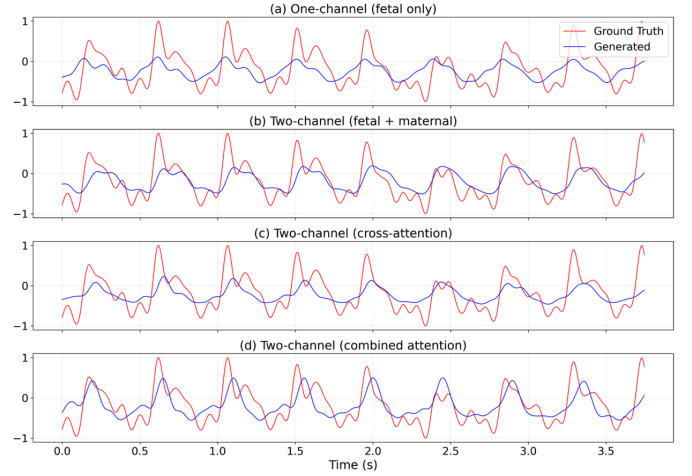


Fig. 2. Example Doppler velocity envelopes generated by each model variant. Each panel shows the generated envelope (blue) superimposed on the ground-truth envelope (red): (a) single-channel model (fetal ECG only), (b) two-channel model (fetal and maternal ECG) without attention, (c) two-channel cross-attention model, and (d) two-channel combined attention model. The x-axis is time in seconds (at 284 Hz sampling rate; 1065 samples $\approx$ 3.75 s); the y-axis shows normalised Doppler amplitude (zero mean, unit variance). Progressive improvements in temporal alignment and morphological fidelity are visible from (a) to (d); note how the systolic peaks in (d) align more closely with the ground truth than in (a) or (b).

V. DISCUSSION

A. Recoverable and Residual Components of Doppler Waveforms

Our framework decomposes fetal Doppler into components recoverable from ECG and residual components that cannot be reconstructed. The strong single-channel performance demonstrates that fetal ECG encodes substantial Doppler information through excitation-contraction coupling [5]: R–R intervals determine pulsation frequency, and QRS timing predicts systolic onset via the pre-ejection period [6]. These electrically recoverable components confirm that electromechanical coupling provides a robust basis for cross-modal synthesis.

The pattern of metric improvements reveals which Doppler features depend on electrical versus mechanical factors. Attention mechanisms yielded substantial gains in frequency-domain fidelity (PSD MSE: 49.9 vs. 84.1 dB² for single-channel, 41% improvement) and temporal alignment (DTW: 253.7 vs. 287.2 a.u., 12% improvement), indicating that overall waveform morphology and timing are electrically recoverable when maternal coupling is appropriately modeled. Heart rate estimation also benefited modestly (4.71 vs. 4.78 bpm error), consistent with its direct dependence on R–R intervals.

However, clinical indices derived from specific velocity components showed limited or no improvement. PI and RI correlations remained near zero across all architectures (PI: $r = -0.04$; RI: $r = 0.06$), and velocity metrics (PSV, EDV, TAMX) showed weak correlations despite reasonable MAE values. This dissociation between good waveform reconstruction and poor index correlation suggests that these indices depend on beat-to-beat mechanical variations, including ven-

TABLE II
CLINICAL/PHYSIOLOGICAL PERFORMANCE METRICS COMPARISON ACROSS ARCHITECTURES. FOR EACH METRIC, WE REPORT THE MAE BETWEEN REAL AND GENERATED VALUES AND THE PEARSON CORRELATION. VALUES REPORTED AS MEAN $\pm$ SD ACROSS 5-FOLD CROSS-VALIDATION.

Metric	Single-Channel (1-ch, no attn)		Two-Channel (2-ch, no attn)		Cross-Attention (2-ch, cross-attn)		Combined Attention (2-ch, cross+self)	
	MAE $\downarrow$	Corr. $\uparrow$	MAE $\downarrow$	Corr. $\uparrow$	MAE $\downarrow$	Corr. $\uparrow$	MAE $\downarrow$	Corr. $\uparrow$
<i>Hemodynamic indices</i>								
PI	48.91 $\pm$ 40.09	-0.02 $\pm$ 0.09	49.40 $\pm$ 39.13	-0.02 $\pm$ 0.08	49.11 $\pm$ 39.65	-0.08 $\pm$ 0.13	48.78 $\pm$ 39.36	-0.04 $\pm$ 0.08
RI	19.11 $\pm$ 28.28	0.07 $\pm$ 0.05	26.39 $\pm$ 31.06	0.00 $\pm$ 0.03	91.99 $\pm$ 106.68	-0.04 $\pm$ 0.09	30.41 $\pm$ 40.56	0.06 $\pm$ 0.06
S/D ratio	0.74 $\pm$ 0.33	-0.02 $\pm$ 0.06	0.92 $\pm$ 0.27	-0.01 $\pm$ 0.10	0.72 $\pm$ 0.15	-0.02 $\pm$ 0.05	0.66 $\pm$ 0.19	0.04 $\pm$ 0.09
<i>Cardiac function</i>								
Heart rate (bpm)	4.78 $\pm$ 0.69	0.58 $\pm$ 0.15	4.80 $\pm$ 1.13	0.55 $\pm$ 0.20	4.69 $\pm$ 0.71	0.58 $\pm$ 0.14	4.71 $\pm$ 0.77	0.57 $\pm$ 0.12
<i>Velocity components</i>								
PSV	0.61 $\pm$ 0.15	0.09 $\pm$ 0.17	0.71 $\pm$ 0.11	0.13 $\pm$ 0.17	0.64 $\pm$ 0.11	-0.06 $\pm$ 0.13	0.60 $\pm$ 0.12	0.03 $\pm$ 0.14
EDV	0.41 $\pm$ 0.18	0.15 $\pm$ 0.10	0.52 $\pm$ 0.19 [‡]	0.11 $\pm$ 0.10	0.33 $\pm$ 0.08	0.03 $\pm$ 0.08	0.34 $\pm$ 0.09	-0.04 $\pm$ 0.10
TAMX	0.12 $\pm$ 0.02	0.00 $\pm$ 0.05	0.12 $\pm$ 0.01	-0.02 $\pm$ 0.14	0.12 $\pm$ 0.01	-0.07 $\pm$ 0.15	0.12 $\pm$ 0.02	-0.04 $\pm$ 0.17

Bold: Best performance (lowest MAE or highest correlation). $\uparrow$: higher is better; $\downarrow$: lower is better.

[‡] The two-channel model yields higher EDV error due to misaligned diastolic morphology, consistent with its poor DTW alignment.

Legend: PI: pulsatility index; RI: resistance index; S/D: systolic/diastolic ratio; PSV: peak systolic velocity; EDV: end diastolic velocity; TAMX: time-averaged maximum velocity.

tricular compliance, placental resistance, and preload, that are invisible to electrical recordings [31], [32]. The model learns average morphology well but cannot capture the mechanical factors driving inter-beat variability in peak and end-diastolic velocities. These residual components may be clinically significant precisely because they reflect mechanical properties inaccessible to fECG monitoring.

B. Selective Attention for Maternal–Fetal Cardiac Coupling

The finding that naïve maternal ECG concatenation provides zero benefit, and degrades DTW alignment by 18%, parallels physiological evidence that MFCC occurs intermittently [33]. When coupling is weak, maternal signals contribute noise. Cross-modal attention addresses this by learning to emphasize temporally correlated maternal segments while suppressing irrelevant portions.

The 39% PSD MSE improvement from cross-modal attention over naïve concatenation (from 101.9 to 61.8 dB²) quantifies the value of learned selectivity. This aligns with transfer entropy analyses showing that maternal-to-fetal information transfer varies with gestational age and fetal behavioral state [8]. By treating fetal features as queries and maternal features as keys/values, cross-modal attention provides an architectural analogue to this conditional coupling. Self-attention contributes an additional 19% frequency-domain improvement (from 61.8 to 49.9 dB²) by capturing long-range temporal dependencies within the fetal signal.

Notably, the benefits of attention were domain-specific: frequency-domain and alignment metrics improved substantially, while pointwise error metrics (MAE, MSE) showed minimal differences across architectures. This dissociation arises because all architectures learn similar average waveform morphologies, yielding comparable pointwise error, whereas frequency-domain metrics and DTW capture spectral fidelity and temporal alignment of cardiac events, properties that depend on modeling the *timing* of systolic and diastolic

phases across beats. Cross-modal attention improves these timing-sensitive metrics by selectively incorporating maternal heart rate information that modulates beat timing, while self-attention enforces beat-to-beat temporal consistency. This pattern is consistent with the physiological expectation that electromechanical coupling primarily determines *when* cardiac events occur, while local amplitude depends on mechanical properties such as contractility and vascular impedance.

C. Comparison to Prior Work

Existing fetal Doppler synthesis methods use fetal ECG exclusively [25], [26]. Our work is the first to incorporate maternal ECG through attention mechanisms, explicitly modeling MFCC. Beyond improved reconstruction, this framework enables new scientific questions: What proportion of Doppler variability reflects maternal–fetal interactions? Which features require direct hemodynamic measurement? The attention-based decomposition provides quantitative answers: cross-modal attention contributes 39% of frequency-domain improvement, while residual errors identify features inaccessible to electrical recordings.

D. Limitations and Future Directions

Validation on pathological cohorts remains essential, as maternal–fetal cardiac coupling may differ under pathological conditions such as fetal growth restriction or pre-eclampsia, where autonomic regulation and placental resistance alter ECG–Doppler relationships [34]. The difficulty in reconstructing clinical indices suggests limits of ECG-based Doppler synthesis for mechanically dependent metrics, and future work should include formal per-segment statistical testing beyond descriptive mean $\pm$ SD comparisons across folds to strengthen the evidence for architectural differences.

From a clinical perspective, this framework has potential applications in continuous fetal monitoring where Doppler ultrasound is impractical. Since fECG can be acquired continuously via wearable electrodes while Doppler requires episodic

assessment by trained personnel, fECG-derived envelopes could enable real-time heart rate and waveform morphology estimation during ambulatory monitoring. However, the limited correlation in clinical indices indicates that synthesized waveforms should not replace direct Doppler for detecting placental dysfunction. Future work should incorporate physiological constraints, uncertainty quantification, residual analysis for Doppler-only mechanical abnormalities, and clinician-based qualitative assessment of waveform morphology.

VI. CONCLUSION

We presented a cross-modal generative framework for investigating relationships between fetal Doppler and fetal-maternal ECG. Critical ablation studies established that naïve maternal ECG concatenation provides no benefit, while selective attention yields 39% frequency-domain improvement, validating that MFCC requires learned selectivity for computational exploitation. Meanwhile, the differential response to attention mechanisms, with substantial gains in waveform morphology but limited improvement in clinical indices (PI, RI), reveals the boundary between electrically recoverable and mechanically determined Doppler components.

REFERENCES

- [1] J. Behar, A. Johnson, G. D. Clifford, and J. Oster, "A practical guide to non-invasive foetal electrocardiogram extraction and analysis," *Physiol. Meas.*, vol. 35, no. 8, pp. R1–R35, 2014.
- [2] E. Hernandez-Andrade *et al.*, "Evaluation of conventional Doppler fetal cardiac function parameters: E/A ratios, outflow tracts, and myocardial performance index," *Fetal Diagn. Ther.*, vol. 32, no. 1–2, pp. 22–29, 2012.
- [3] C. Tei *et al.*, "New index of combined systolic and diastolic myocardial performance: A simple and reproducible measure of cardiac function," *J. Cardiol.*, vol. 26, no. 6, pp. 357–366, 1995.
- [4] C. Kähler *et al.*, "Fetal magnetocardiography: Development of the fetal cardiac time intervals," *Prenat. Diagn.*, vol. 22, no. 5, pp. 408–414, 2002.
- [5] D. M. Bers, "Cardiac excitation-contraction coupling," *Nature*, vol. 415, no. 6868, pp. 198–205, 2002.
- [6] N. A. Mensah-Brown *et al.*, "Assessment of left ventricular pre-ejection period in the fetus using simultaneous magnetocardiography and echocardiography," *Fetal Diagn. Ther.*, vol. 28, no. 3, pp. 167–174, 2010.
- [7] A. H. Khandoker, Y. Kimura, T. Ito, N. Sato, K. Okamura, and M. Palaniswami, "Antepartum non-invasive evaluation of opening and closing timings of the cardiac valves in fetal cardiac cycle," *Med. Biol. Eng. Comput.*, vol. 47, no. 10, pp. 1075–1082, 2009.
- [8] F. Marzbanrad, Y. Kimura, M. Palaniswami, and A. H. Khandoker, "Quantifying the interactions between maternal and fetal heart rates by transfer entropy," *PLoS ONE*, vol. 10, no. 12, e0145672, 2015.
- [9] T. J. Nitching *et al.*, "Evidence and clinical relevance of maternal-fetal cardiac coupling: A scoping review," *PLoS ONE*, vol. 18, no. 7, e0287245, 2023.
- [10] A. Vaswani *et al.*, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.
- [11] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2016.
- [12] A. van den Oord, N. Kalchbrenner, and K. Kavukcuoglu, "Pixel recurrent neural networks," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2016, pp. 1747–1756.
- [13] D. H. Evans and W. N. McDicken, *Doppler Ultrasound: Physics, Instrumentation and Signal Processing*, 2nd ed. Wiley, 2000.
- [14] W. B. Giles *et al.*, "Fetal umbilical artery flow velocity waveforms and placental resistance: Pathological correlation," *Br. J. Obstet. Gynaecol.*, vol. 92, no. 1, pp. 31–38, 1985.
- [15] A. Bhide *et al.*, "ISUOG practice guidelines: Use of Doppler ultrasonography in obstetrics," *Ultrasound Obstet. Gynecol.*, vol. 41, no. 2, pp. 233–239, 2013.
- [16] P. E. McSharry, G. D. Clifford, L. Tarassenko, and L. A. Smith, "A dynamical model for generating synthetic electrocardiogram signals," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 3, pp. 289–294, 2003.
- [17] R. Sameni, G. D. Clifford, C. Jutten, and M. B. Shamsollahi, "Multichannel ECG and noise modeling: Application to maternal and fetal ECG signals," *EURASIP J. Adv. Signal Process.*, vol. 2007, art. 43407, 2007.
- [18] A. van den Oord *et al.*, "WaveNet: A generative model for raw audio," *arXiv preprint arXiv:1609.03499*, 2016.
- [19] N. Wulan, W. Wang, P. Sun, W. Wang, S. Xia, and W. Jiang, "Generating electrocardiogram signals by deep learning," *Neurocomputing*, vol. 404, pp. 122–136, 2020.
- [20] Y. Zhang, Z. Zhao, Y. Deng, and X. Zhang, "FHRGAN: Generative adversarial networks for synthetic fetal heart rate signal generation in low-resource settings," *Inf. Sci.*, vol. 594, pp. 136–150, 2022.
- [21] Y. Lai *et al.*, "DiffuSETS: 12-lead ECG generation conditioned on clinical text reports and patient-specific information," *Patterns*, vol. 6, no. 10, 101291, 2025.
- [22] Q. Zhu, X. Tian, C. Wong, and M. Wu, "Learning your heart actions from pulse: ECG waveform reconstruction from PPG," *IEEE Internet Things J.*, vol. 8, no. 23, pp. 16734–16748, 2021.
- [23] O. Omer, M. Alammar, H. Esmaili, and M. Salah, "Beat-by-beat ECG monitoring from PPG using spatio-temporal information with WaveNet," *Preprint (ResearchSquare)*, Nov. 2025.
- [24] J. Zschocke *et al.*, "Reconstruction of pulse wave and respiration from wrist accelerometer during sleep," *IEEE Trans. Biomed. Eng.*, vol. 69, no. 2, pp. 830–839, 2022.
- [25] A. Verma, D. S. S. Chintapenta, and S. Mujumder, "Towards reconstruction of pulsed-wave Doppler signals from non-invasive fetal ECG," *arXiv preprint arXiv:2501.16651*, Jan. 2025.
- [26] A. Rafiei, G. D. Clifford, and N. Katebi, "Auto-FEDUS: Autoregressive generative modeling of Doppler ultrasound signals from fetal electrocardiograms," *arXiv preprint arXiv:2504.13233*, 2025.
- [27] X. Wang *et al.*, "Correlation-aware attention CycleGAN for accurate fetal ECG extraction," *IEEE Trans. Instrum. Meas.*, vol. 72, 2023.
- [28] R. Qu *et al.*, "Integrating contrastive learning and CycleGAN for non-invasive fetal ECG extraction," *Pediatr. Cardiol.*, vol. 46, no. 7, pp. 2078–2088, 2025.
- [29] V. D. Nguyen, "Fetal ECG extraction on time-frequency domain using conditional GAN," in *Proc. CVPR Workshops*, 2024.
- [30] E. Sulas *et al.*, "NInFEA: Non-invasive multimodal foetal ECG-Doppler dataset for antenatal cardiology research," *Sci. Data*, vol. 8, no. 30, 2021.
- [31] K. Hecher, R. Campbell, P. Doyle, K. Harrington, and K. Nicolaides, "Assessment of fetal compromise by Doppler ultrasound investigation of the fetal circulation," *Circulation*, vol. 91, no. 1, pp. 129–138, 1995.
- [32] A. A. Baschat, "Doppler application in the delivery timing of the preterm growth-restricted fetus: Another step in the right direction," *Ultrasound Obstet. Gynecol.*, vol. 23, no. 2, pp. 111–118, 2004.
- [33] P. Van Leeuwen, D. Geue, S. Lange, D. Cysarz, H. Bettermann, and D. Grönemeyer, "Influence of paced maternal breathing on fetal-maternal heart rate coordination," *Proc. Natl. Acad. Sci. USA*, vol. 106, no. 33, pp. 13661–13666, 2009.
- [34] A. H. Khandoker *et al.*, "Alterations in maternal-fetal heart rate coupling strength and directions in abnormal fetuses," *Front. Physiol.*, vol. 10, p. 482, 2019.
- [35] N. Katebi, R. Sameni, P. Rohloff, and G. D. Clifford, "Hierarchical attentive network for gestational age estimation in low-resource settings," *IEEE J. Biomed. Health Inform.*, vol. 27, no. 5, pp. 2501–2511, 2023.
- [36] M. Motie-Shirazi, E. Ramos, S. Chandrasekaran, R. Sameni, P. Rohloff, G. Clifford, and N. Katebi, "Self-supervised learning for gestational age estimation from low-cost Doppler ultrasound in low-resource settings," in *Proc. NeurIPS Workshop Learn. Time Ser. Health*, 2025.
- [37] A. Rafiei, M. Motie-Shirazi, R. Sameni, G. D. Clifford, and N. Katebi, "Next-generation fetal heart monitoring: Leveraging neural sequential modeling for ultrasound analysis," *IEEE Trans. Biomed. Eng.*, 2025.
- [38] P. D. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Trans. Audio Electroacoust.*, vol. 15, no. 2, pp. 70–73, 1967.