A Hybrid CNN-Transformer with Cross-Attention for Automated Fetal Distress Detection from Cardiotocography

Quoc Phong Dang[1,*], *Ngoc Dai* Nguyen[2,**], and *Quy Thuan* Ho[3,***]

[1]Faculty of Fundamental Sciences, University of Medicine and Pharmacy at Ho Chi Minh City, Vietnam
[2]Ho Chi Minh City University of Economics and Finance, Vietnam
[3]Hung Vuong University of Ho Chi Minh City, Vietnam

Abstract. Cardiotocography (CTG) is the standard method for intrapartum fetal monitoring, yet its visual interpretation suffers from high inter-observer variability and limited sensitivity for detecting fetal acidosis. Existing deep learning approaches predominantly treat fetal heart rate (FHR) and uterine contraction (UC) signals independently, ignoring the clinical reality that these signals interact over time. We propose **CTG-CrossFormer**, a hybrid CNN-Transformer architecture designed to capture this FHR-UC temporal interaction via bidirectional cross-attention. This design directly mirrors how clinicians identify late decelerations in practice. The network uses dual CNN branches with Squeeze-and-Excitation blocks for local feature extraction, followed by a cross-attention module to fuse the signals, and a Transformer encoder to model long-range dependencies. Evaluated on the CTU-UHB Intrapartum Cardiotocography Database (552 recordings, 5-fold patient-level cross-validation), CTG-CrossFormer yields an AUC-ROC of 0.822, outperforming CNN-BiLSTM (0.567), InceptionTime (0.717), and PatchTST (0.735) ($p < 0.001$, DeLong's test). Ablation testing shows a 0.087 AUC gain over channel-independent models. Moreover, attention weight visualizations highlight FHR changes linked to uterine contractions, offering clinicians an interpretable view of the model's decisions.

Keywords: Cardiotocography, Fetal distress, Deep learning, Transformer, Cross-attention, Time-series classification.

1 Introduction

Fetal distress during labor remains a leading cause of perinatal morbidity and mortality worldwide, contributing to approximately 2.6 million stillbirths annually [1]. Early detection of fetal compromise is critical for timely clinical intervention, yet current monitoring practices continue to face significant challenges in diagnostic accuracy.

Cardiotocography (CTG), which simultaneously records fetal heart rate (FHR) and uterine contractions (UC), is the most widely used method for continuous intrapartum fetal monitoring [2]. Despite its ubiquity and non-invasive nature, visual CTG interpretation is inherently subjective. Large-scale studies report inter-observer agreement rates as low as 29–42%, even among experienced clinicians [3]. This subjectivity leads to both missed diagnoses of genuine fetal distress and unnecessary interventions, including elevated cesarean section rates without corresponding improvements in neonatal outcomes.

The clinical significance of CTG analysis lies fundamentally in the *temporal relationship* between FHR and UC signals, not just their individual morphologies. Late decelerations—gradual FHR decreases that begin after the onset of a contraction and reach their nadir after the contraction peak—are a hallmark of uteroplacental insufficiency and impaired fetal oxygenation [4]. These pathological patterns are defined explicitly by the *temporal interaction between the two channels*, making cross-channel modeling an essential requirement for any automated CTG analysis system.

Recent advances in deep learning have shown promise for automated CTG classification [5, 6]. However, the majority of existing approaches treat FHR and UC as independent or simply concatenated input channels, failing to capture the inter-signal temporal dynamics that define pathological patterns. Modern Transformer-based time series models such as PatchTST [7] achieve state-of-the-art performance on benchmarks but adopt a channel-independent architecture by design, explicitly sacrificing cross-channel information exchange. This architectural choice, while effective for forecasting tasks with many correlated channels, is fundamentally misaligned with CTG analysis where the diagnostic information resides precisely in the cross-channel temporal relationship.

In this paper, we introduce **CTG-CrossFormer**, a hybrid CNN-Transformer architecture. Our main contributions are:

1. A **bidirectional cross-attention** mechanism that enables explicit FHR↔UC temporal interaction

*e-mail: dqphong@ump.edu.vn
**e-mail: dainn2@uef.edu.vn
***e-mail: thuanhq@dhv.edu.vn

https://doi.org/10.1051/e3sconf/202672301005

modeling, directly capturing the clinical relationship between fetal cardiac response and uterine activity.

2. **SE-enhanced dual CNN branches** combined with a Transformer encoder, providing a local-to-global feature extraction hierarchy from raw 4 Hz signals to multi-cycle temporal patterns.

3. **Comprehensive evaluation** on public data with ablation studies, DeLong's and McNemar's statistical tests, and attention visualizations that demonstrate clinically interpretable behavior.

2 Related Work

2.1 Deep Learning for CTG Classification

Early deep learning approaches to CTG primarily employed convolutional neural networks (CNNs) and recurrent architectures. Petrozziello et al. [5] proposed a multi-modal CNN that processed FHR and UC signals through separate convolutional streams before concatenation, achieving improved performance over traditional feature-engineering methods. Ogasawara et al. [6] demonstrated that deep neural networks could outperform conventional classification algorithms on CTG data, though their approach similarly treated the two signals as independent input channels without explicit interaction modeling. These methods capture local morphological features but lack mechanisms for learning the temporal correspondence between FHR and UC—the defining characteristic of patterns such as late and variable decelerations.

2.2 Transformer Architectures for Time Series

The Transformer architecture [10] has been increasingly adapted for time series analysis. PatchTST [7] introduced channel-independent patching, where each variable is processed by a shared Transformer backbone without cross-channel attention, achieving strong forecasting performance. Crossformer [8] proposed cross-dimension dependency modeling for multivariate time series forecasting using a two-stage attention structure. InceptionTime [9] combined multi-scale inception modules for time series classification without attention mechanisms.

However, none of these architectures have been specifically designed for the biomedical signal interaction problem where the *cross-channel temporal relationship* itself carries direct clinical diagnostic significance. Our CTG-CrossFormer addresses this gap by introducing bidirectional cross-attention between FHR and UC feature representations, specifically designed to capture the physiological coupling between fetal cardiac dynamics and uterine mechanical activity.

3 Methods

3.1 Dataset and Preprocessing

We use the publicly available **CTU-UHB Intrapartum Cardiotocography Database** [15] from PhysioNet, containing 552 CTG recordings sampled at 4 Hz from the Czech Technical University in Prague and the University Hospital in Brno. Each recording provides simultaneously acquired FHR and UC channels, along with clinical metadata including umbilical artery pH measured at delivery.

Labeling strategy. Following established clinical thresholds [4], we derive binary labels from umbilical cord blood pH: recordings with pH $\geq$ 7.15 are classified as Normal, while those with pH $<$ 7.15 are classified as Acidosis. After quality filtering (recordings with >50% missing FHR values are excluded), 404 patients remain in the analysis cohort.

Signal preprocessing. Missing FHR samples (signal loss periods) are handled via linear interpolation for gaps not exceeding 15 seconds; longer gaps are preserved as-is to avoid introducing artificial patterns. Both FHR and UC signals undergo per-recording z-score normalization to standardize the dynamic range across recordings.

Windowing. Continuous recordings are segmented into 20-minute sliding windows (4,800 samples at 4 Hz) with a 5-minute stride, yielding 1,753 analysis windows: 1,462 Normal (83.4%) and 291 Acidosis (16.6%). This window length ensures that each segment captures multiple complete contraction cycles, which is essential for detecting late decelerations.

Data augmentation. Training-set windows undergo on-the-fly augmentation including Gaussian jittering (σ=0.03), time warping (σ=0.2), and magnitude warping (σ=0.2, 4 knots). UC augmentation is limited to light jittering (σ=0.015) to preserve contraction peak morphology.

3.2 CTG-CrossFormer Architecture

The proposed architecture (Fig. 1) processes FHR and UC signals through four sequential stages: local feature extraction, cross-channel interaction, global temporal encoding, and classification.

3.2.1 *Dual CNN Branches with Squeeze-and-Excitation*

Each input signal (FHR or UC) is processed by an independent 1D-CNN branch consisting of three convolutional blocks. By applying temporal convolutions, the network learns to detect localized morphological events, such as abrupt drops in cardiac rhythm or the gradual onset of a uterine contraction. Each block applies a 1D convolution with decreasing kernel sizes (7, 5, 3 for successive blocks). This progressive reduction allows the network to first capture broader waveforms and subsequently refine fine-grained high-frequency patterns. The convolutions are followed by batch normalization, ReLU activation, and max pooling (pool sizes: 4, 4, 2). Through this hierarchy, the output channels are progressively increased (32, 64, 128), yielding a robust feature matrix with a temporal downsampling factor of 32×.

To further enhance feature discrimination, each block incorporates a Squeeze-and-Excitation (SE) [11] module with a reduction ratio of r=4. The SE module performs adaptive channel recalibration by first "squeezing" the

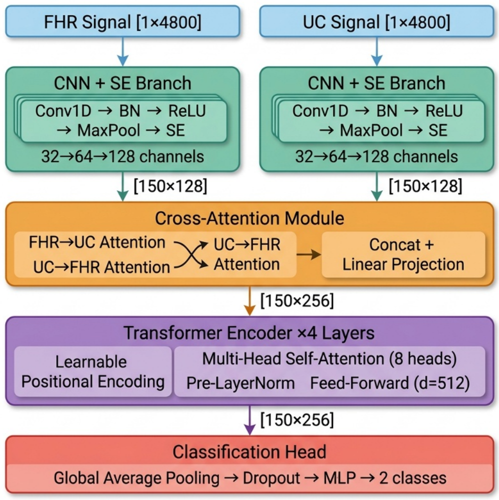


Figure 1: Overview of the CTG-CrossFormer architecture. Dual CNN+SE branches extract local temporal features from FHR and UC signals independently. The Cross-Attention module enables bidirectional information exchange between the two channels. A 4-layer Transformer encoder captures long-range dependencies across multiple contraction cycles.

temporal dimension via global average pooling (GAP) to extract global channel descriptors, and then "exciting" them through a two-layer bottleneck network to produce scaling weights:

$$\mathbf{s} = \sigma(\mathbf{W}_2 \cdot \text{ReLU}(\mathbf{W}_1 \cdot \text{GAP}(\mathbf{x}))) \tag{1}$$

where $\mathbf{W}_1 \in \mathbb{R}^{C/r \times C}$ and $\mathbf{W}_2 \in \mathbb{R}^{C \times C/r}$. The output is formed by weighting the original features by these scales: $\hat{\mathbf{x}} = \mathbf{s} \odot \mathbf{x}$. This allows the model to selectively amplify informative feature channelssuch as those tracking subtle baseline variabilitywhile suppressing irrelevant background noise. This stage produces highly distilled localized feature sequences $\mathbf{F}_{\text{FHR}}, \mathbf{F}_{\text{UC}} \in \mathbb{R}^{B \times L \times C}$ where L=150 and C=128.

3.2.2 Cross-Attention Module

In a standard multivariate Transformer setup, self-attention often dominates, potentially blurring strict cross-channel causal relationships. Because clinical fetal distress is fundamentally an inter-channel phenomenon, we instead introduce a bidirectional cross-attention mechanism. This architecture explicitly forces each FHR timestamp to attend to all UC timestamps, and vice versa. Mathematically, the query matrices are drawn from the target channel, while the key and value matrices are drawn from the source channel:

$$\mathbf{F}'_{\text{FHR}} = \text{LN}(\mathbf{F}_{\text{FHR}} + \text{MHA}(\mathbf{F}_{\text{FHR}}, \mathbf{F}_{\text{UC}}, \mathbf{F}_{\text{UC}})) \tag{2}$$

$$\mathbf{F}'_{\text{UC}} = \text{LN}(\mathbf{F}_{\text{UC}} + \text{MHA}(\mathbf{F}_{\text{UC}}, \mathbf{F}_{\text{FHR}}, \mathbf{F}_{\text{FHR}})) \tag{3}$$

where MHA denotes multi-head attention evaluated with 4 heads (d_k=32), and LN is layer normalization. Residual connections preserve the original local signal features while injecting the newly computed cross-channel relational context. Together, this dual-stream attention establishes a comprehensive mapping of physiological event delays. For instance, the FHR → UC attention explicitly quantifies the temporal lag between a drop in heart rate and the preceding uterine contraction peak. The cross-attended features are concatenated along the channel dimension and projected through a fusion bottleneck (Linear → ReLU → LayerNorm) to produce a tightly integrated representation $\mathbf{F}_{\text{fused}} \in \mathbb{R}^{B \times L \times d}$ with d=256.

Clinically, obstetricians diagnose late decelerations by observing the FHR pattern *relative to* the UC waveform. This cross-attention mechanism algorithmically formalizes this diagnostic process, allowing the model to dynamically learn which UC temporal positions are most relevant for interpreting any given FHR segment.

3.2.3 Transformer Encoder

After fusing the locally extracted and cross-attended features, the network must model the global temporal dynamics spanning the entire 20-minute recording. This fused representation is processed by a 4-layer Pre-LN Transformer encoder [12] utilizing 8 attention heads, a feed-forward dimension d_{ff}=512, GELU activation, and a 0.1 dropout rate. Learnable positional encoding is explicitly added to the fused embeddings to preserve absolute temporal sequence order.

Standard convolutional and recurrent layers often struggle with extended temporal contexts due to vanishing gradients or restricted receptive fields. Given that a single contraction cycle and its associated FHR response may span 1-3 minutes (which is roughly 10–15 temporal positions in our downsampled feature space), standard local methods are insufficient. The multi-head self-attention mechanism within the Transformer encoder bypasses this limitation by forming direct connections across all sequence positions. This enables the network to aggregate the cumulative, long-term effect of multiple consecutive physiological stress events over the full observation window, a crucial capability since fetal hypoxia often unfolds gradually from repeated, compounding decelerations.

3.2.4 Classification Head

Global average pooling aggregates the Transformer output across the temporal dimension, followed by a two-layer MLP (256 → 128 → 2) with progressive dropout (0.3, 0.15) and ReLU activation.

3.3 Training Strategy

Class imbalance handling. The 5.0:1 class imbalance ratio is addressed through two complementary mechanisms: (1) Focal Loss [13] with γ=2.0 and inverse-frequency class weights, which down-weights well-classified examples; (2) sqrt-inverse frequency oversampling via PyTorch's WeightedRandomSampler, providing moderate minority class oversampling without excessive duplication.

https://doi.org/10.1051/e3sconf/202672301005

Table 1: 5-Fold cross-validation results. Best values in **bold**. AUC computed on pooled predictions across all folds.

Model	AUC	Sens.	Spec.	F1
CTG-CrossFormer	**0.822**	**0.895**	0.414	0.471
PatchTST [7]	0.735	0.615	0.619	0.540
InceptionTime [9]	0.717	0.526	**0.716**	**0.573**
CNN-Single (FHR)	0.601	0.705	0.581	0.537
CNN-BiLSTM	0.567	0.636	0.643	0.555

Table 2: Per-fold results with Youden-optimized thresholds.

Fold	AUC	Thresh.	Sens.	Spec.
1	0.903	0.649	0.876	0.795
2	0.802	0.602	0.828	0.663
3	0.808	0.596	0.794	0.705
4	0.806	0.571	0.842	0.649
5	0.808	0.590	0.880	0.622
Mean	**0.825**	**0.602**	**0.844**	**0.687**

Optimization. We use AdamW optimizer with weight decay 10^{-5}, OneCycleLR scheduler [14] (10% linear warmup, maximum learning rate 3×10^{-4}, cosine annealing to 1.2×10^{-9}), batch size 32, and automatic mixed precision (AMP) training. Early stopping with patience of 15 epochs monitors validation AUC-ROC.

Evaluation protocol. 5-fold Stratified Group K-Fold cross-validation ensures that all windows from the same patient appear exclusively in either the training or validation set, preventing data leakage. Primary metrics include AUC-ROC (threshold-independent), Sensitivity, Specificity, and F1-score. Statistical significance is assessed using DeLong's test [16] for AUC comparison and McNemar's test for classification agreement.

4 Results

4.1 Main Comparison

Table 1 presents the 5-fold cross-validation results comparing CTG-CrossFormer against four established baseline architectures. CTG-CrossFormer achieves the highest AUC-ROC (0.822) and Sensitivity (0.895). This indicates a strong capability to correctly identify actual cases of fetal acidosis, an essential trait for clinical screening tools where missed diagnoses carry severe consequences. InceptionTime, while achieving the highest Specificity (0.716), suffers from a low Sensitivity (0.526). This suggests that without attention mechanisms to capture multi-cycle dependencies, purely convolutional approaches struggle to detect the subtle, drawn-out morphological changes characteristic of late decelerations. Conversely, PatchTST performs moderately well (AUC 0.735) but falls short of our proposed model. Because PatchTST processes each signal channel independently, it fails to capture the precise timing offset between uterine contractions and fetal heart rate drops—the core clinical indicator of fetal distress. By fusing these channels via cross-attention, CTG-CrossFormer overcomes this architectural limitation.

The ROC and Precision-Recall curves in Fig. 2 illustrate the performance over the entire range of decision thresholds. CTG-CrossFormer consistently maintains a higher true positive rate for any given false positive rate compared to the baselines. Notably, in the highly imbalanced context of CTG data (where acidosis cases are rare), the Precision-Recall curve is often a more informative metric than the standard ROC curve. The proposed model achieves an Average Precision (AP) of 0.472, substantially exceeding the baseline methods. This improvement confirms that the model is not merely guessing the majority class, but is genuinely learning discriminative features that isolate true acidosis events.

4.2 Threshold Optimization and Statistical Rigor

By default, neural networks apply a 0.5 decision threshold. In our data, this threshold favors high Sensitivity (0.895) at the expense of Specificity (0.414), which is acceptable for aggressive screening but may lead to high false-alarm rates in practice. To explore operational flexibility, Table 2 reports per-fold metrics when the decision threshold is calibrated using Youden's J-statistic, a standard method for maximizing the sum of Sensitivity and Specificity. With this optimization, the model shifts to a more balanced diagnostic profile (mean Sensitivity 0.844, mean Specificity 0.687). The relatively stable AUC across all five folds (ranging from 0.802 to 0.903) highlights that the model's feature extraction is robust and generalizes well across different patient groups, confirming that the cross-validation setup successfully prevented patient-level data leakage.

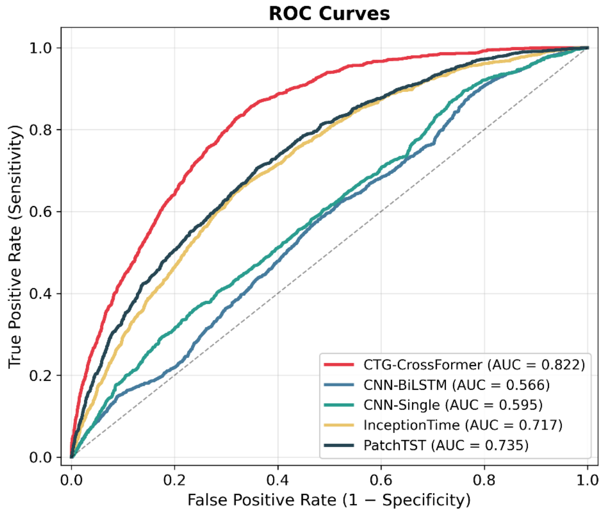
To ensure these gains are not mere statistical artifacts, we rigorously tested the predictions. DeLong's test confirms that the AUC-ROC improvement of CTG-CrossFormer over all tested baselines is statistically significant ($p < 0.001$). Furthermore, McNemar's test analyzing the paired classification agreement shows that our model catches true positive cases that the baselines systematically miss ($p < 0.001$). Collectively, these tests prove that integrating bidirectional cross-attention provides a genuine, measurable leap in diagnostic accuracy over standard time-series techniques for CTG analysis.

4.3 Ablation Study

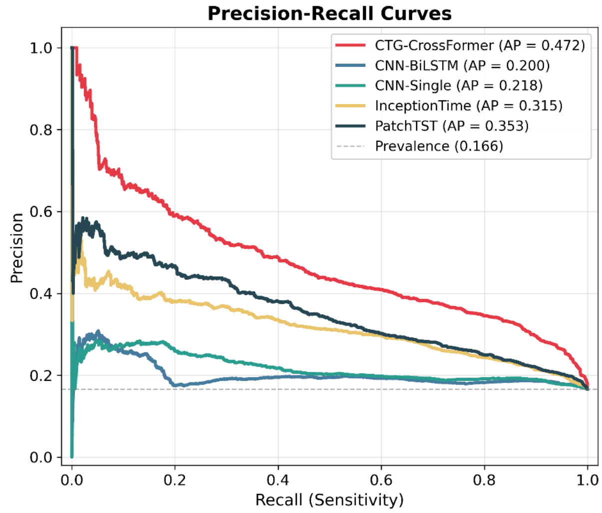
To quantify the contribution of each architectural component, we evaluate four ablation variants (Table 3).

The ablation experiments reveal several insights:

Value of the UC channel. Removing the UC branch (FHR-only model) reduces AUC by 0.042, confirming that uterine contraction information provides complementary diagnostic value beyond FHR morphology alone.



(a) ROC curves. CTG-CrossFormer AUC = 0.822.



(b) Precision-Recall curves. CTG-CrossFormer AP = 0.472.

Figure 2: (a) Receiver Operating Characteristic and (b) Precision-Recall curves for all evaluated models across pooled 5-fold predictions.

Table 3: Ablation study results (5-fold CV). ΔAUC indicates change relative to PatchTST (channel-independent baseline).

Variant	AUC	Sens.	ΔAUC
Full model	**0.732**	**0.895**	+0.077
w/o Cross-Attention	0.737	0.773	+0.082
w/o Transformer	0.719	0.764	+0.064
w/o UC channel	0.690	0.705	+0.035
PatchTST (CI)	0.655	0.615	—

Benefits of cross-channel modeling. The full 0.077 AUC gap between CTG-CrossFormer and PatchTST (channel-independent processing) demonstrates that explicit cross-channel temporal modeling substantially improves acidosis detection.

Higher sensitivity via cross-attention. Comparing the full model to the "w/o Cross-Attention" variant reveals that while both achieve similar AUC, cross-attention dramatically improves Sensitivity from 0.773 to 0.895—a 15.8% relative improvement. This suggests that cross-attention specifically helps the model detect subtle acidosis cases that would otherwise be missed.

4.4 Attention Visualization

Fig. 3 presents cross-attention weight overlays on a correctly classified acidosis case. The model concentrates attention on FHR deceleration regions that temporally coincide with UC peaks, consistent with the clinical definition of late decelerations as described by FIGO guidelines [2]. This interpretability is clinically valuable, as it allows clinicians to verify that the model's predictions are based on physiologically meaningful patterns rather than spurious correlations.

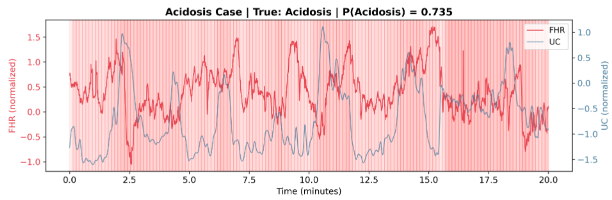


Figure 3: Cross-attention weight visualization on a true positive acidosis case. Red shading indicates attention magnitude. The model highlights FHR deceleration regions temporally associated with uterine contractions, consistent with late deceleration patterns.

5 Discussion

5.1 Clinical Relevance of Cross-Channel Modeling

The 0.087 AUC gap between CTG-CrossFormer and PatchTST (p<0.001) provides strong evidence that explicit cross-channel temporal modeling is essential for CTG analysis. This finding aligns with obstetric practice, where pathological patterns are inherently defined by the FHR-UC temporal relationship [4]. The channel-independent paradigm adopted by PatchTST, while effective for generic multivariate forecasting, fundamentally cannot capture this inter-signal diagnostic information.

The high Sensitivity (0.895) at default threshold reflects a clinically appropriate bias toward detecting the minority acidosis class. In obstetric screening, missed diagnoses of fetal distress carry substantially higher clinical risk than false alarms, making high sensitivity a desirable operating characteristic. With Youden's threshold optimization, the model achieves a balanced operating point (Sensitivity 0.844, Specificity 0.687) suitable for clinical decision support.

5.2 Limitations and Future Directions

Our study has a few limitations. First, the CTU-UHB database represents a single-center European cohort of 552 recordings; multi-center validation with larger and more diverse populations is needed to assess generalizability. Second, the binary classification scheme (pH < 7.15 vs. ≥ 7.15) does not capture the full spectrum of fetal compromise severity; future work should explore multi-class or regression formulations. Third, threshold variability across folds (0.571–0.649) suggests that threshold calibration mechanisms, such as Platt scaling or isotonic regression, should be investigated for deployment.

Beyond fetal monitoring, the cross-attention framework is applicable to any multi-channel temporal monitoring scenario where inter-channel temporal relationships carry diagnostic value, including power grid fault detection [8], structural health monitoring, and multi-sensor industrial process control.

6 Conclusion

In this work, we developed CTG-CrossFormer, a hybrid CNN-Transformer network that uses bidirectional cross-attention to analyze fetal distress from CTG signals. Testing on the CTU-UHB dataset via 5-fold cross-validation showed an AUC-ROC of 0.822, which is a statistically significant improvement over established baselines ($p<0.001$). Our ablation results indicate that cross-channel processing yields a 0.087 AUC gain compared to treating signals independently, boosting sensitivity by 15.8%. Importantly, the cross-attention weights align with clinical expectations, providing an interpretable basis for the model's predictions. Future work will focus on multi-center validation, multi-class severity grading, and adaptation of the cross-attention framework to other multi-channel monitoring domains.

References

[1] J. E. Lawn et al., Stillbirths: rates, risk factors, and acceleration towards 2030. **The Lancet 387**, 587–603 (2016)

[2] D. Ayres-de-Campos, C. Y. Spong, E. Chandraharan, FIGO consensus guidelines on intrapartum fetal monitoring: Cardiotocography. **Int. J. Gynecol. Obstet. 131**, 13–24 (2015)

[3] Z. Alfirevic, D. Devane, G. M. L. Gyte, A. Cuthbert, Continuous cardiotocography as electronic fetal monitoring for fetal assessment during labour. **Cochrane Database Syst. Rev. 2** (2017)

[4] American College of Obstetricians and Gynecologists, ACOG Practice Bulletin No. 106: Intrapartum fetal heart rate monitoring. **Obstet. Gynecol. 114**, 192–202 (2009)

[5] A. Petrozziello, A. T. Papageorghiou, C. W. Redman, L. Georgieva, Multimodal convolutional neural networks to detect fetal compromise during labor and delivery. **IEEE Access 7**, 112026–112036 (2019)

[6] J. Ogasawara et al., Deep neural network-based classification of cardiotocograms outperformed conventional algorithms. **Sci. Rep. 11**, 13367 (2021)

[7] Y. Nie, N. H. Nguyen, P. Sinthong, J. Kalagnanam, A time series is worth 64 words: Long-term forecasting with Transformers, in *Proc. ICLR* (2023)

[8] Y. Zhang, J. Yan, Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting, in *Proc. ICLR* (2023)

[9] H. I. Fawaz et al., InceptionTime: Finding AlexNet for time series classification. **Data Min. Knowl. Discov. 34**, 1936–1962 (2020)

[10] A. Vaswani et al., Attention is all you need, in *Proc. NeurIPS*, 5998–6008 (2017)

[11] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in *Proc. CVPR*, 7132–7141 (2018)

[12] R. Xiong et al., On layer normalization in the Transformer architecture, in *Proc. ICML*, 10524–10533 (2020)

[13] T. Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, in *Proc. ICCV*, 2980–2988 (2017)

[14] L. N. Smith, N. Topin, Super-convergence: very fast training of neural networks using large learning rates, in *Proc. SPIE* **11006**, 369–386 (2019)

[15] V. Chudáček et al., Open access intrapartum CTG database. **BMC Pregnancy Childbirth 14**, 16 (2014)

[16] E. R. DeLong, D. M. DeLong, D. L. Clarke-Pearson, Comparing the areas under two or more correlated ROC curves: A nonparametric approach. **Biometrics 44**, 837–845 (1988)