Fetal Well-Being Prediction with One-Class Gaussian Process Anomaly Detection

Taraneh G. Azarnir,* Cassandra Heiselman†, and Petar M. Djurić*
* Department of Electrical and Computer Engineering, Stony Brook University, New York, USA
† Department of Obstetrics, Gynecology, and Reproductive Medicine, Stony Brook Medicine, New York, USA

***Abstract*—Fetal heart rate (FHR) monitoring is vital to assess fetal well-being during labor. However, clinical decisions based on subjective visual interpretations can lead to inconsistencies, unnecessary cesarean sections, and legal disputes. The key challenges in the computerized analysis of FHR include class imbalance, where healthy cases vastly outnumber distress cases, lack of confidence score, as most approaches focus on classification rather than continuous fetal health assessment, and limited feature interpretability, which hinders clinical adoption. To address these challenges, we propose a One-Class Gaussian Process model trained on interpretable features from healthy FHR segments. This model learns the healthy FHR distribution and identifies potential anomalies. We further introduce the health confidence score (HCS), a continuous metric quantifying fetal well-being. This score offers clinicians an intuitive and interpretable measure of the fetus's condition, thereby supporting timely and informed clinical decision-making. The results demonstrate the model's robust 96% accuracy in classifying FHR segments.**

***Index Terms*—Cardiotocography (CTG), Fetal Heart Rate (FHR), Gaussian Processes, Anomaly Detection**

I. Introduction

Since the 1960s, monitoring fetal heart rate (FHR) and uterine contractions (UC) during labor has been a routine practice [1]. Doctors screen Cardiotocography (CTG) tracings continuously before birth to detect any unusual activity indicating a risk to neonatal and maternal health [2]. These continuous signals are a valuable asset in obstetrics, providing critical insights in the delivery room to determine the optimal timing for interventions in high-risk situations that could lead to temporary or permanent harm to the infant or the mother. The complexity of visual examination of these signals prompted the National Institute of Child Health and Human Development and the International Federation of Gynecology and Obstetrics (FIGO) to publish guidelines for interpreting FHR and UC [3]. However, the proposed CTG evaluation guidelines have faced criticism for their simplistic approach to interpretation [4].

Computerized analysis of FHR, UC, and their interrelation has been employed since the 1980s to support obstetricians in decision-making during labor [5]–[8]. The studies explore diverse methodologies, including morphological patterns, which involve extracting features based on the available guidelines [9] and nonlinear feature extraction methods that capture subtle patterns not easily discernible visually. Additionally, artificial neural networks have been utilized to analyze these time series data [10].

One key challenge in analyzing CTG signals is the extreme class imbalance in available datasets. Most recorded cases correspond to healthy outcomes, leading to a significant underrepresentation of the unhealthy category. This underrepresentation hinders robust model training and accurate anomaly detection, which results in poor classification performance, as many algorithms mistakenly assume the balanced distribution between classes [11]. There have been different attempts to handle imbalanced data. Most are based on eliminating bias by data resampling (undersampling or oversampling) [12]–[15]. Additionally, boost ensemble learning has been proposed in [16], which uses the distribution of classification errors to guide training. In each iteration, the most informative examples from the majority class are selectively sampled to improve classification performance.

It has been argued that deep neural networks achieve the highest accuracy in classifying CTG recordings [17], [18]. However, the features learned by these models are often complex and lack interpretability, which makes it challenging to understand the decision-making process based on them [19]. This underscores the need for explainable features that provide meaningful insights into fetal health assessment [20]. According to FIGO guidelines, FHR tracings are classified into three categories: normal, abnormal, and indeterminate [3]. However, since FHR signals are continuous time series, they can transition between these categories at different stages of labor. This raises a critical question: how reliably and accurately do these classifications reflect fetal health status at any given time?

This paper tackles the issue of class imbalance by adopting a one-class training approach. Our model is trained exclusively on the majority class that represents healthy cases and which allows effective detection of anomalies without being skewed by the limited availability of unhealthy recordings. Moreover, the features we extract from FHR signals are inherently interpretable, which in turn improves clinical understanding and decision-making. Additionally, we introduce a continuous confidence score from 0 to 1 that represents the confidence of the classification. This assessment provides clinicians with a valuable tool to monitor CTG recordings and make timely interventions.

The remainder of this article is organized as follows. Section II provides an overview of existing methods for automatic

This work has been supported by NIH under Award 1RO1HD097188-01.

FHR analysis, and Section III describes the workflow for designing the model. In Section IV we provide details of the experimental setup and evaluation metrics and present results along with a discussion of the model's performance. Finally, Section V summarizes the key findings, discusses the clinical implications, and outlines potential future research directions.

II. Background

The primary goal of automated FHR analysis is to detect fetal distress that requires immediate intervention during labor while minimizing overdiagnosis, which can lead to unnecessary cesarean sections. In clinical practice, fetal well-being is typically assessed based on the postnatal outcomes of newborns and laboratory tests. The key challenge lies in reliably estimating these indicators during labor using intrapartum-acquired signals. Most studies classify FHR outcomes based on the umbilical cord blood pH value of the fetus at birth [21]. However, there is an ongoing debate regarding the validity of pH labels and the precise threshold required for accurate classification [22]. This section briefly reviews common FHR features and introduces Gaussian Processes (GPs) and their extension for anomaly detection.

A. Fetal Heart Rate Feature Extraction

A wide range of features has been explored in the computerized analysis of FHR signals, including time-domain, frequency-domain, and non-linear features. Time-domain features, such as short- and long-term variability, provide valuable insights into overall heart rate trends and beat-to-beat fluctuations and can be critical for assessing fetal well-being [23]. Frequency-domain features, including energy in different frequency bands, offer a deeper understanding of autonomic nervous system regulation by analyzing the power spectral density (PSD) of FHR signals [24]. Furthermore, non-linear features, such as those derived from Poincaré plot analysis, capture the complexity and irregularity of FHR patterns and contribute to a more robust assessment of fetal status [25]. In particular, features that reflect FHR variability are of special importance in clinical evaluation and computerized analysis, as they provide critical insights into fetal autonomic regulation and adaptive responses to stressors [26]. These features improve the interpretability of automated analysis systems and help clinicians make informed, data-driven decisions during labor.

B. Gaussian Process

A Gaussian Process (GP) is a powerful non-parametric Bayesian model used for regression, classification, and anomaly detection. It provides a probabilistic framework for modeling complex functions while incorporating uncertainty estimates. Unlike traditional parametric models that assume a fixed functional form, a GP defines a distribution over functions, which allows for flexible adaptation to data. GPs have been successfully applied in supervised and unsupervised learning tasks [27], including FHR analysis [28]. A Gaussian Process (GP) is a stochastic process in which any finite set of random variables follows a joint Gaussian distribution. Given an input $\boldsymbol{x} \in \mathbb{R}^{d_x}$ and an output y such that $y = f(\boldsymbol{x})$, a GP defines a prior over the function $f(\boldsymbol{x})$ as $f(\boldsymbol{x}) \sim \mathcal{GP}(m(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}'))$, where $m(\boldsymbol{x})$ is the mean function and $k(\boldsymbol{x}, \boldsymbol{x}')$ is the covariance function (or kernel).

The choice of kernel $k(\boldsymbol{x}, \boldsymbol{x}')$ is crucial as it defines the properties of the function space. A commonly used covariance function is the Radial Basis Function (RBF) kernel, defined as $k(\boldsymbol{x}, \boldsymbol{x}') = \sigma_f^2 \exp\left(-\frac{\|\boldsymbol{x}-\boldsymbol{x}'\|^2}{2l^2}\right)$, where σ_f^2 represents the signal variance and l is the l is the length scale parameter that controls how quickly the function varies with $\boldsymbol{x}$.

Given noisy observations, we model the outputs as $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{X}) + \boldsymbol{\epsilon}$, where the noise term is distributed as $\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \sigma_\epsilon^2 \boldsymbol{I})$, $\boldsymbol{y} \in \mathbb{R}^{N\times 1}$, with N being the number of observed data, $\boldsymbol{X} \in \mathbb{R}^{N\times d_x}$, σ_ϵ^2 is the noise variance, and the prior distribution over function values follows: $\boldsymbol{f}(\boldsymbol{X}) \sim \mathcal{N}\big(\boldsymbol{m}(\boldsymbol{X}), \boldsymbol{K}(\boldsymbol{X}, \boldsymbol{X})\big)$. By conditioning the joint Gaussian distribution, we obtain the predictive posterior distribution for new inputs $\boldsymbol{X}^*$, given the training data $(\boldsymbol{X}, \boldsymbol{y})$, as $\boldsymbol{f}^* \mid \boldsymbol{X}, \boldsymbol{y}, \boldsymbol{X}^* \sim \mathcal{N}(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$, where $\boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}^*$ are the predictive mean and covariance, respectively.

Thus, GPs provide predictions and estimate the uncertainty of the made predictions, which is particularly useful for applications that require confidence quantification. Although GPs are computationally expensive due to their $\mathcal{O}(n^3)$ time and $\mathcal{O}(n^2)$ memory complexity [29], various approximation techniques have been developed to make them suitable for real-time applications. Methods such as sparse GPs [30] and online or streaming variants [31], [32] significantly reduce computational demands while retaining the core advantages of GPs, including probabilistic interpretability and uncertainty quantification [27].

C. Automatic Relevance Determination (ARD) Kernel

The ARD kernel extends the standard RBF kernel by assigning a separate length-scale parameter l_d to each input dimension d, which allows automatic feature relevance selection. The ARD kernel is given by [33]:

$$k_{\text{ARD}}(\boldsymbol{x}, \boldsymbol{x}') = \sigma_f^2 \exp\left(-\frac{1}{2}\sum_{d=1}^{d_x} \frac{(x_d - x'_d)^2}{l_d^2}\right) \tag{1}$$

where σ_f^2 is the signal variance that controls the output scale, and l_d is the characteristic length-scale for each input dimension d. The characteristic length-scale determines how quickly the function values change along that dimension.

The ARD kernel allows the model to identify the most relevant input features, and it assigns larger length scales to less relevant dimensions and smaller length scales to more important ones. The GP model involves key hyperparameters that require tuning, and they are the length-scales (l_d) in the ARD kernel, the signal variance (σ_f^2), and the noise variance (σ_n^2). These hyperparameters are optimized by maximizing the log marginal likelihood.

D. Gaussian Processes for Anomaly Detection

The one-class GP (OCGP) is an extension of GPs for anomaly detection, where the model is trained exclusively on data from a single class and identifies deviations as anomalies [27], [28]. The key idea in OCGP is to model a probability distribution over function values that represent the training class. The function $f(\boldsymbol{x})$ is a latent variable that defines the decision boundary [26]. Anomaly detection can be achieved by assessing the predictive variance of the model. Let $\sigma^2(\boldsymbol{x}^*)$ denote the predictive variance for a test input $\boldsymbol{x}$, and let T be a threshold. Then, we can define an anomaly detection function $\delta(\boldsymbol{x})$ as

$$\delta(\boldsymbol{x}) = \begin{cases} 1, & \text{if } \sigma^2(\boldsymbol{x}) < T, \\ 0, & \text{otherwise,} \end{cases} \tag{2}$$

where $\delta(\boldsymbol{x}) = 0$ indicates that the test sample is flagged as an anomaly [27]. The advantages of using the OCGP model are numerous: first, its probabilistic interpretation sets it apart from other one-class classifiers, such as the One-Class SVM, because OCGP provides uncertainty estimates that are particularly valuable in medical applications [19]. In addition, as a non-parametric model, OCGP does not assume a fixed functional form, which allows for the flexible modeling of complex normal data distributions. OCGP can detect deviations from normal FHR patterns as potential fetal distress indicators and provides continuous HCS for clinicians to assess fetal well-being.

III. The Proposed Method

This section describes the dataset, features, OCGP model, detection threshold, and HCS.

A. Data Description and Preprocessing

An open-access intrapartum CTG database, compiled from recordings collected between January 2018 and December 2020 at Stony Brook University Hospital in New York, contains CTG recordings from 10,314 pregnant women and serves as the basis for all experiments conducted in this study [34]. A subset of FHR signals was selected from the database based on the criterion that no more than 50% of values were missing in the last 30 minutes before birth. The missing values in the selected signals were linearly interpolated to ensure data continuity. Each signal was then segmented into three nonoverlapping 10-minute segments. These segments were randomly presented to an expert to label as healthy or unhealthy without providing additional information about the newborn or maternal status. The expert was blinded to the source of each segment and unaware of any relationships between segments of the same signal, which ensured an unbiased labeling process. The labels provided by the expert were used as the ground truth. We extracted 23 interpretable features from the FHR segments that capture various characteristics. Time-domain features include the estimated baseline, counts and durations of accelerations and decelerations, as well as central tendency and variability metrics (mean, median, standard deviation, minimum, maximum, range and root mean square of successive differences). We computed the dominant frequency and power in the low- and high-frequency bands and their ratio in the frequency domain. Nonlinear measures such as approximate entropy, sample entropy, and detrended fluctuation analysis (DFA) were calculated to capture signal complexity. In addition, statistical dispersion metrics like variance, interquartile range, and the 25th and 75th percentiles were also derived. Together, these features provide a robust, multi-dimensional representation of the FHR signal.

B. One Class Gaussian Process (OCGP) Model

Extracted features from healthy segments were used as model inputs, with all training output labels uniformly set to 1 to represent the healthy class. The ARD kernel was selected to allow feature-wise weighting to allow for improved interpretability after assessing the relative importance of each feature. The Python GPyTorch library was used to implement the model. The OCGP model learns a probabilistic representation of healthy FHR dynamics by training exclusively on healthy segments. Gamma distribution priors were also applied to the length scale and output scale parameters of the ARD kernel to enforce regularization, while a Gaussian prior was set on the mean function. The model was optimized by maximizing the Exact Marginal Log-Likelihood using the Adam optimizer, with a Step Learning Rate Scheduler that ensures stable convergence [35]. This probabilistic framework supports confidence-based anomaly detection, where deviations from the healthy FHR distribution indicate potential fetal distress.

C. Outlier Threshold and Confidence Score

The choice of threshold T significantly affects the sensitivity and specificity of the detector. To derive a robust anomaly detection threshold, we applied 5-fold cross-validation on the training set, evaluating predictive variance distributions across folds. The threshold that best balanced sensitivity and specificity corresponded to the 95th percentile of training variances, effectively minimizing false positives while reliably flagging segments with elevated uncertainty indicative of potential fetal distress. The HCS quantifies how similar a segment's features are to those in the healthy training set, as learned by the GP model. It provides a continuous score between 0 and 1, where values near 1 correspond to low predictive variance (indicating the model's high confidence that the segment resembles healthy patterns), and lower values indicate increased uncertainty and possible deviation from the healthy distribution, which suggests potential anomalies.

We define the health confidence score (HCS) as follows:

$$\text{HCS} = 1 - \frac{\sigma^2(\mathbf{x}^*) - \sigma^2_{\min}}{\max\{\sigma^2_{\max}, \sigma^2(\mathbf{x}^*)\} - \sigma^2_{\min}}, \tag{3}$$

where

$$\sigma^2_{\max} = \max_{\boldsymbol{x}\in\mathcal{D}} \sigma^2(\boldsymbol{x}), \quad \sigma^2_{\min} = \min_{\boldsymbol{x}\in\mathcal{D}} \sigma^2(\boldsymbol{x}),$$

$\sigma^2(\boldsymbol{x}^*)$ is the predictive variance at the new input $\boldsymbol{x}^*$, $\mathcal{D}$ denotes the training set, and $\max_{\boldsymbol{x}\in\mathcal{D}} \sigma^2(\boldsymbol{x})$, $\min_{\boldsymbol{x}\in\mathcal{D}} \sigma^2(\boldsymbol{x})$ are the maximum and minimum predictive variance observed

over the training data, respectively. Thus, a lower predictive variance (i.e., higher model confidence) corresponds to a higher score. Scaling the confidence score between 0 and 1 improves the interpretability of fetal health status. This allows clinicians to distinguish more clearly between normal and potentially concerning patterns.

IV. Experiments and Model Evaluation

In this section, we evaluate the performance of the OCGP model on a test set and demonstrate the utility of the continuous HCS in fetal health assessment.

A. Outlier Detection on a Test Set

A test set containing a mix of previously unseen healthy and unhealthy 10-minute nonoverlapping segments was used to evaluate the model's performance. The same feature extraction process applied during training transformed these test segments into model inputs for the OCGP model. The expert-labeled test set included 14 segments for each group (healthy, unhealthy), which provided a balanced test set. The classification results of the OCGP and one-class SVM models are presented in Table I. One-class SVM, a widely used method for anomaly detection, was originally introduced by Schölkopf et al. [36]. The OCGP model demonstrates superior performance in distinguishing between healthy and unhealthy FHR segments. It achieved an overall accuracy of 96%, with precision, recall, and F1-scores indicating balanced classification across both classes. Figure 1 presents a scatter plot of predictive variance for the test set. Each point is a 10-minute nonoverlapping segment of FHR. The dashed line denotes the anomaly detection threshold, set at the 95th percentile of the training variance. Test points with a variance above this threshold are classified as unhealthy (anomalies), whereas segments below the threshold are considered healthy (normal).

TABLE I: Classification Results of the OCGP and One-Class SVM Models

	Precision	Recall	F1-Score	Support
OCGP				
Healthy	0.93	1.00	0.97	14
Unhealthy	1.00	0.93	0.96	14
Average	0.97	0.96	0.96	28
Accuracy		0.96 (28 samples)		
One-Class SVM				
Healthy	0.88	1.00	0.93	14
Unhealthy	1.00	0.86	0.92	14
Average	0.94	0.93	0.93	28
Accuracy		0.93 (28 samples)		

B. Monitoring FHR with Health Confidence Score

A 30-minute FHR signal that satisfied the criterion was used to monitor the HCS. The signal was segmented into 10-minute segments, with each consecutive segment shifting by only one second (four samples at a 4 Hz sampling rate). This allows for smooth temporal tracking. For each segment, relevant features were extracted and fed into the OCGP model. Monitoring

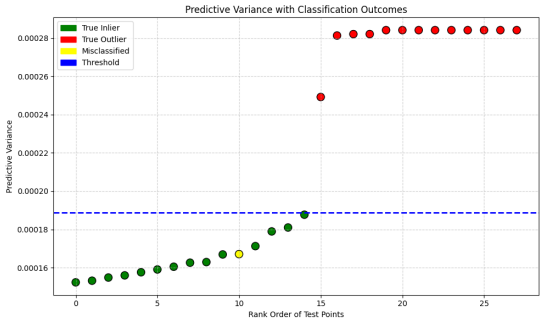

Fig. 1: Predictive variance for test points

how these features evolve over time closely mirrors the visual assessment of clinicians when evaluating FHR patterns. The predictive variance was transformed into the health confidence measure using equation 3. Figure 2 illustrates a 30-minute FHR signal alongside its corresponding HCS during labor. This visualization demonstrates the model's ability to capture rapid, dynamic changes in fetal health due to its high temporal resolution. The HCS drops sharply when the segment transitions from a healthy state to a more concerning one. This reflects the model's sensitivity to shifts in the underlying physiological state.

When two obstetricians annotated segments of the FHR signal they found concerning, the HCS decreased accordingly, suggesting that the model reliably captures clinically relevant patterns. Figure 3 shows an example FHR signal with the expert-identified regions highlighted in yellow and green.

Fig. 2: Fetal Heart Rate and Health Confidence Score

V. Conclusion

This study presents an OCGP model for FHR monitoring that leverages machine learning for confidence-based anomaly detection. Trained exclusively on healthy data, the model detects deviations as potential signs of fetal distress and introduces the Health Confidence Score (HCS) as a proxy for fetal well-being. While initial results are promising, further validation with a larger test set and broader expert input is underway. Future work will incorporate uterine contraction signals and additional clinical insights to enhance model robustness.

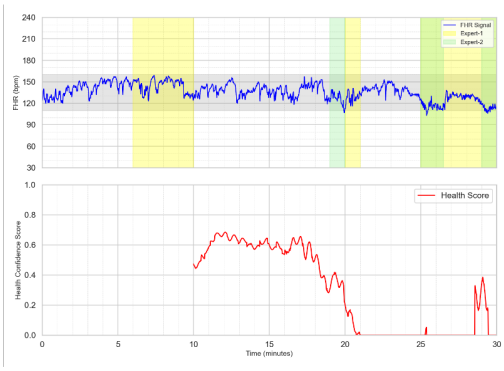

Fig. 3: Health Confidence Score and Expert's Concern

REFERENCES

[1] M. P. Nageotte, "Fetal heart rate monitoring," *Seminars in Fetal and Neonatal Medicine*, vol. 20, no. 3, pp. 144–148, 2015.

[2] M. J. Stout and A. G. Cahill, "Electronic fetal monitoring: past, present, and future," *Clinics in Perinatology*, vol. 38, no. 1, pp. 127–142, 2011.

[3] D. Ayres-de Campos, C. Y. Spong, E. Chandraharan, and F. I. F. M. E. C. Panel, "FIGO consensus guidelines on intrapartum fetal monitoring: Cardiotocography," *International Journal of Gynecology & Obstetrics*, vol. 131, no. 1, pp. 13–24, 2015. [Online]. Available: http://dx.doi.org/10.1016/j.ijgo.2015.06.020

[4] A. Ugwumadu, "Are we (mis)guided by current guidelines on intrapartum fetal heart rate monitoring? Case for a more physiological approach to interpretation," *BJOG: An International Journal of Obstetrics & Gynaecology*, vol. 121, no. 9, pp. 1063–1070, 2014.

[5] J. Bernardes, C. Moura, J. P. M. de Sa, and L. P. Leite, "The porto system for automated cardiotocographic signal analysis," *Journal of Perinatal Medicine-Official Journal of the WAPM*, vol. 19, no. 1-2, pp. 61–65, 1991.

[6] G. Dawes, G. Visser, J. Goodman, and C. Redman, "Numerical analysis of the human fetal heart rate: The quality of ultrasound records," *American Journal of Obstetrics and Gynecology*, vol. 141, no. 1, pp. 43–52, 1981.

[7] G. Dawes, C. Redman, and J. Smith, "Improvements in the registration and analysis of fetal heart rate records at the bedside," *BJOG: An International Journal of Obstetrics & Gynaecology*, vol. 92, no. 4, pp. 317–325, 1985.

[8] E. Hamilton and E. K. Kimanani, "Intrapartum prediction of fetal status and assessment of labour progress," *Baillière's Clinical Obstetrics and Gynaecology*, vol. 8, no. 3, pp. 567–581, 1994.

[9] S. Dash, J. Muscat, J. G. Quirk, and P. M. Djurić, "Implementation of nichd diagnostic criteria for feature extraction and classification of fetal heart rate signals," in *Signals, Systems and Computers (ASILOMAR)*, 2011, pp. 1684–1688.

[10] A. Georgieva, S. J. Payne, M. Moulden, and C. W. Redman, "Artificial neural networks applied to fetal monitoring in labour," *Neural Computing and Applications*, vol. 22, no. 1, pp. 85–93, 2013.

[11] Z. Liu, W. Cao, Z. Gao, J. Bian, H. Chen, Y. Chang, and T.-Y. Liu, "Self-paced ensemble for highly imbalanced massive data classification," in *2020 IEEE 36th International Conference on Data Engineering (ICDE)*. IEEE, 2020, pp. 841–852.

[12] G. Haixiang, L. Yijing, J. Shang, G. Mingyun, H. Yuanyue, and G. Bing, "Learning from class-imbalanced data: Review of methods and applications," *Expert Systems with Applications*, vol. 73, pp. 220–239, 2017.

[13] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2980–2988.

[14] X.-Y. Liu and Z.-H. Zhou, "The influence of class imbalance on cost-sensitive learning: An empirical study," in *Sixth International Conference on Data Mining (ICDM'06)*. IEEE, 2006, pp. 970–974.

[15] J. Shu, Q. Xie, L. Yi, Q. Zhao, S. Zhou, Z. Xu, and D. Meng, "Meta-weight-net: Learning an explicit mapping for sample weighting," *arXiv preprint arXiv:1902.07379*, 2019.

[16] M. Ajirak, C. Heiselman, J. G. Quirk, and P. M. Djurić, "Boost ensemble learning for classification of ctg signals," in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 1316–1320.

[17] Z. Zhao, Y. Deng, Y. Zhang, Y. Zhang, X. Zhang, and L. Shao, "DeepFHR: Intelligent prediction of fetal acidemia using fetal heart rate signals based on convolutional neural network," *BMC Medical Informatics and Decision Making*, vol. 19, no. 1, pp. 1–15, 2019.

[18] A. Petrozziello, I. Jordanov, T. A. Papageorghiou, W. C. Redman, and A. Georgieva, "Deep learning for continuous electronic fetal monitoring in labor," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2018, pp. 5866–5869.

[19] R. Miotto, L. Li, B. A. Kidd, and J. T. Dudley, "Deep patient: An unsupervised representation to predict the future of patients from electronic health records," *Scientific Reports*, vol. 6, no. 1, pp. 1–10, 2016.

[20] B. D. Fulcher and N. S. Jones, "Highly comparative feature-based time-series classification," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 12, pp. 3026–3037, 2014.

[21] A. Georgieva, P. Abry, V. Chudáček, P. M. Djurić, M. G. Frasch, R. Kok, C. A. Lear, S. N. Lemmens, I. Nunes, A. T. Papageorghiou *et al.*, "Computer-based intrapartum fetal monitoring and beyond: A review of the 2nd workshop on signal processing and monitoring in labor (october 2017, oxford, uk)," *Acta Obstetricia et Gynecologica Scandinavica*, vol. 98, no. 9, pp. 1207–1217, 2019.

[22] B. E. Josten, T. R. B. Johnson, and J. P. Nelson, "Umbilical cord blood ph and apgar scores as an index of neonatal health," *American Journal of Obstetrics and Gynecology*, vol. 157, no. 4, pp. 843–848, 1987.

[23] J. De Haan, J. Van Bemmel, B. Versteeg, A. Veth, L. Stolte, J. Janssens, and T. Eskes, "Quantitative evaluation of fetal heart rate patterns: I. processing methods," *European Journal of Obstetrics & Gynecology*, vol. 1, no. 3, pp. 95–102, 1971.

[24] M. G. Signorini, G. Magenes, S. Cerutti, and D. Arduini, "Linear and nonlinear parameters for the analysis of fetal heart rate signal from cardiotocographic recordings," *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 3, pp. 365–374, 2003.

[25] M. B. Tayel and E. I. AlSaba, "Poincaré plot for heart rate variability," *World Academy of Science, Engineering and Technology, International Journal of Medical, Health, Biomedical, Bioengineering and Pharmaceutical Engineering*, vol. 9, no. 9, pp. 708–711, 2015.

[26] G. Feng, J. G. Quirk, and P. M. Djurić, "Extracting interpretable features for fetal heart rate recordings with gaussian processes," in *International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, 2019, pp. 381–385.

[27] C. E. Rasmusen and C. Williams, *Gaussian Processes for Machine Learning*. MIT Press, 2006.

[28] G. Feng, J. G. Quirk, and P. M. Djurić, "Supervised and unsupervised learning of fetal heart rate tracings with deep gaussian processes," in *2018 14th Symposium on Neural Networks and Applications (NEUREL)*. IEEE, 2018, pp. 1–6.

[29] C. K. Williams and M. Seeger, "Using the nyström method to speed up kernel machines," *Advances in neural information processing systems*, vol. 13, pp. 682–688, 2001.

[30] E. Snelson and Z. Ghahramani, "Sparse gaussian processes using pseudo-inputs," in *Advances in neural information processing systems*, 2006, pp. 1257–1264.

[31] L. Csató and M. Opper, "Sparse online gaussian processes," *Neural computation*, vol. 14, no. 3, pp. 641–668, 2002.

[32] T. D. Bui, C. V. Nguyen, and R. E. Turner, "Streaming sparse gaussian process approximations," in *Advances in Neural Information Processing Systems*, 2017, pp. 3299–3307.

[33] R. M. Neal, *Bayesian Learning for Neural Networks*. Springer, 2012.

[34] Stony Brook University, "PreAna Fetal Heart Rate Database," https://preana-fo.ece.stonybrook.edu/database.html, accessed: 2025-02-17.

[35] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2015, published as a conference paper at the 3rd International Conference for Learning Representations, San Diego, 2015. Available at: https://arxiv.org/abs/1412.6980.

[36] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Computation*, vol. 13, no. 7, pp. 1443–1471, 2001.