Fetal Heart Rate Analysis from a Multi-task Learning Perspective with Gaussian Processes

Tong Chen,* *Guanchao Feng*,* *Cassandra Heiselman*,† *J. Gerald Quirk*,† and *Petar M. Djurić**
* Department of Electrical and Computer Engineering
† Department of Obstetrics/Gynecology, Renaissance School of Medicine
Stony Brook University
Stony Brook, NY 11794, USA
Email: *{Tong.Chen, Guanchao.Feng, Petar.Djuric}@stonybrook.edu,
†{Cassandra.Heiselman, J.Gerald.Quirk}@stonybrookmedicine.edu

***Abstract*—Assessments of fetal heart rate tracings by obstetricians suffer from inter- and intra-observer variability whereas computerized fetal heart rate analysis lacks consensus on labels that have diagnostic capability. There are different measurements that carry important information about fetal well-being, although in the literature the most adopted one has been the umbilical cord blood pH value at birth. In this paper, instead of relying on pH-based labeling only, we propose Gaussian process-based multi-task learning that is able to learn multiple fetal well-being measurements simultaneously by explicitly modeling similarity between the tasks. We tested the proposed approach with different intrapartum databases on both regression and classification tasks. Our experimental results show that the proposed approach achieves superior performance compared to popular single-task learning models for fetal heart rate analysis.**

***Index Terms*—multi-task learning, fetal heart rate, Gaussian processes, Bayesian nonparametric methods, transfer learning**

I. Introduction

Electronic fetal monitoring is the most widely accepted method for intrapartum monitoring since it allows continuous fetal surveillance during labor to prevent adverse outcomes due to fetal hypoxia and ischemia. It is usually performed using cardiotocography (CTG), which is a simultaneous recording of the fetal heart rate (FHR) and uterine activity (UA). Although great effort has been made to establish various clinical guidelines for CTG interpretation and classification [1], there are high inter- and intra-variabilities in the obstetrician's visual interpretations of the signal patterns [2]. In computerized FHR analysis, which aims at supporting and improving CTG interpretations, FHR recordings are usually labeled by pH values of umbilical cord blood at birth [3]. The prevalence of umbilical cord blood pH-based labeling notwithstanding, there has been considerable controversy surrounding the reliability of pH and its threshold for classification of healthy and pathological delivery outcomes [4]. On the other hand, well-adopted clinical metrics/labels such as newborn Apgar scores and cord blood gas-based surrogates other than pH have rarely been used in computerized FHR analysis.

Multi-task learning (MTL) aims at learning multiple related tasks simultaneously by leveraging the shared information across tasks. Different from single-task learning (STL), which learns the tasks in an isolated manner, MTL uses data across different tasks to learn shared representations that are useful to improve performance on all the tasks. MTL has been applied to many biomedical and healthcare applications, for example, seizure electroencephalogram (EEG) signals recognition [5] and patient healthcare data analysis [6]. Besides, great research efforts have been devoted to adopting various frameworks for MTL. Recently, Gaussian processes (GPs) have been adopted for MTL since they provide a powerful and flexible framework for modeling unknown functions [7]–[9]. More importantly, unlike neural network-based and deep learning models, which require massive data to learn a large number of parameters, GPs are a nonparametric Bayesian data-efficient methodology because it allows for expressing the prior belief about the unknown functions.

In this work, we introduce the concept of MTL to FHR analysis for the first time and demonstrate its effectiveness using a GP-based MTL model. We tested it in regression and classification tasks using a widely-used open-access intrapartum CTG database and a new in-house CTG database, respectively. The experimental results show that GP-based MTL models outperformed the single-task learning GPs and other benchmark models in both regression and classification tasks. This indicates that for FHR analysis, MTL models can achieve better performance than STL models.

II. Background

A. Computerized Fetal Heart Rate Analysis

The ultimate goal of computerized FHR analysis is to estimate fetal wellbeings from acquired intrapartum signals as well as other available clinical data. In practice, the fetal wellbeing is measured or characterized by different fetal outcomes and metrics. Therefore, estimating or predicting these fetal outcomes and metrics are of great importance. Various computerized FHR analysis and evaluation approaches have been proposed in the literature. For example, a sparse support vector machine (sparse SVM) was adopted in [10] for the selection of informative features and FHR classification for fetal acidosis detection. In [11], the authors proposed a deep GP framework to provide informative latent space that can be used to generate FHR signals and improve the performance

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

of FHR classification. In [12], 8 machine learning models, including SVM and random forest, were implemented and evaluated for classification of CTG recordings.

However, most of the studies in FHR analysis learn only one task, which is to classify CTG recordings usually labeled by the umbilical cord blood pH value of the fetus at birth. Despite the wide use of such labeling approach, there is no consensus on the validity of adopting pH value for labeling as well as the cut-off value that should be used for diagnosis [4]. On the other hand, it has been shown that other fetal outcome measures are also able to provide valuable information on identifying adverse outcomes [13], [14]. Although some studies combined multiple fetal outcome measures, such as pH value and base deficit, to define abnormality for labeling [15], the models can only learn from the information based on the fetus classes that these surrogates define jointly. These studies, however, are still of STL nature. By contrast, within the MTL framework, different measurements can be accommodated directly. MTL models are able to learn both from the information based on the fetus classes that are defined by a specific label and from the underlying information shared by the classes.

B. Multi-task Learning

MTL is related to many active fields of research in machine learning, including transfer learning and multi-modal learning. Both MTL and transfer learning allow knowledge transfer across different tasks. However, transfer learning aims to enhance the performance of a target task using information learned from a source task while the objective of MTL is to improve the performance on all tasks. Furthermore, both multi-task and multi-modal learning improve generalization of models, while multi-modal learning solves tasks across multiple domains, where each domain has different types of data inputs, for example video, text and audio [16].

In the literature, there are three categories of MTL approaches, and they are based on the strategy of modeling the relatedness of the tasks. The first category is feature-based where one assumes that all related tasks share some feature representations. The second category is parameter-based where the task similarities are encoded by placing prior or constraints on model parameters. The last category is instance-based where models identify the instances in one task that will help learning other tasks. The GP-based MTL model belongs to the second category as it additionally introduces a similarity matrix to explicitly model the relatedness between tasks. In [17], MTL GP was formulated from a perspective of the relationship between linear models and GPs. Further, a sparse approximation for inference of GP-based multi-task approach was proposed in [18]. Recently, the authors in [19] introduced a continual learning method to model sequential observations.

In this work, we adopted GP-based MTL for computerized FHR analysis for the following reasons. Firstly, GP is a powerful Bayesian machinery that is inherently connected with many machine learning models, e.g., SVMs and neural networks. More importantly, data scarcity is one of the major challenges in computerized FHR analysis, for instance, the largest open-access intrapartum CTG database contains only 552 labeled CTG recordings. The Bayesian nature of the GPs enables them for data-efficient learning, which makes them suitable for computerized FHR analysis. Finally, another merit of the GPs is their capability for accurate quantification of uncertainties within a probabilistic framework. In turn, this is of great importance for decision making and risk quantification.

III. Model Description and Inference

A. Gaussian Processes

A GP is a collection of random variables such that every finite subset of them has a multivariate normal distribution. If $\mathbf{x}$ denotes an input, a GP can be seen as a distribution of a real-valued function $f(\mathbf{x})$, and it is fully specified by its mean function $m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})]$ and covariance function $k(\mathbf{x}, \mathbf{x}') = \mathbb{E}\left[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))\right]$. For simplicity, the mean is usually assumed to be zero.

The core of a GP is the covariance function $k(\mathbf{x}, \mathbf{x}')$ because it maps similarities or dependencies between inputs, e.g., $\mathbf{x}$ and $\mathbf{x}'$ to the covariance between outputs $f(\mathbf{x})$ and $f(\mathbf{x}')$. One of the most commonly used covariance function is the squared exponential,

$$k(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp\left(-\frac{||\mathbf{x} - \mathbf{x}'||^2}{2\ell}\right), \tag{1}$$

where σ^2 and $\ell > 0$ are its hyperparameters that are learned from training data by maximizing the log-likelihood.

B. Gaussian Process-based Multi-task Learning

When there are multiple different but related tasks that we want to learn simultaneously, multi-task learning is usually adopted as information learned from each task can be shared across all tasks to improve the model performance. And this can be viewed as modeling multiple related functions simultaneously. Compared to GP-based STL, where the covariance of outputs is mapped by the covariance function given the similarity between inputs, the GP-based MTL approach additionally introduces an inter-task similarity matrix that models the covariances between different latent functions belonging to different tasks.

Given inputs $\mathbf{X} = \{\mathbf{x}_n\}_{n=1}^N$, the complete set of corresponding labels for M tasks can be defined as $\mathbf{Y} \in \mathbb{R}^{N \times M}$, where Y_{nm} is the entry of the nth row and mth column of $\mathbf{Y}$, which represents the label of the mth task for the nth input $\mathbf{x}_n$. Let $\mathbf{y}$ be the vectorization of $\mathbf{Y}$ and $\mathbf{y}_o \subset \mathbf{y}$ be a set of observed labels. We place a zero-mean GP prior over the latent functions $\{f_m\}_{m=1}^M$. Then the covariance between outputs of the mth task and the m'th task can be modeled as

$$\langle f_m(\mathbf{x}), f_{m'}(\mathbf{x}')\rangle = K^f_{mm'} k^x(\mathbf{x}, \mathbf{x}'), \tag{2}$$

where $K^f_{mm'}$ is the (m, m')th entry of a positive semi-definite matrix $\mathbf{K}^f \in \mathbb{R}^{M \times M}$ that describes the inter-task dependencies and k^x is the covariance function over the inputs.

For regression tasks, we assume that the observed output of the mth task on the nth input $\mathbf{x}_n$ is a function of $\mathbf{x}_n$ and

$$Y_{nm} = f_m(\mathbf{x}_n) + \epsilon_m, \tag{3}$$

where $\epsilon_m \sim \mathcal{N}(0, \sigma_m^2)$ is additive Gaussian noise of the mth task. Given a set of labels $\mathbf{y}$, one can show that the likelihood of $\boldsymbol{\theta}$, where $\boldsymbol{\theta}$ is the set of parameters of k^x and the inter-task dependencies matrix $\mathbf{K}^f$, is given by

$$\mathbf{y}|\mathbf{X}, \boldsymbol{\theta} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{K}^f \otimes k^x(\mathbf{X}, \mathbf{X}) + \mathbf{D} \otimes \mathbf{I}\right), \quad (4)$$

where $\otimes$ is the Kronecker product, $\mathbf{D} \in \mathbb{R}^{M \times M}$ is a diagonal matrix whose (m, m)th entry is σ_m^2, and $\mathbf{I} \in \mathbb{R}^{N \times N}$ is the identity matrix. Therefore, the parameters of k^x and $\mathbf{K}^f$ can be learned by maximizing the marginal likelihood in (4).

To make predictions for a new input $\mathbf{x}_*$ for the mth tasks, one can use the fact that the predictive distribution $p(\mathbf{y}_*|\mathbf{y}, \mathbf{x}_*, \mathbf{X})$ is also Gaussian with mean

$$\mu_* = \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{y}, \quad (5)$$

and variance

$$\sigma_* = k^x(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{c}, \quad (6)$$

where $\mathbf{c} = \mathbf{k}_{:m}^f \otimes k^x(\mathbf{x}_*, \mathbf{X})$, $\mathbf{k}_{:m}^f$ is the mth column of $\mathbf{K}^f$, and $\boldsymbol{\Sigma} = \mathbf{K}^f \otimes k^x(\mathbf{X}, \mathbf{X}) + \mathbf{D} \otimes \mathbf{I}$. More details are available in [7], [8].

For classification tasks, the likelihood is

$$\pi_m(\mathbf{x}) = \lambda(f_m(\mathbf{x})), \quad (7)$$

where $\lambda(z) = (1 + \exp(-z))^{-1}$. Unlike in the regression case where the posterior of the latent functions and likelihood can be computed analytically, in classification, due to the non-Gaussianity introduced by the nonlinear mapping λ, expectation propagation (EP) is often used for inference.

IV. Experiments and Results

To demonstrate the usefulness of MTL in FHR analysis, we tested the GP-based MTL approach in regression and classification tasks using the CTU-UHB CTG database [20] and the Stony Brook University (SBU) CTG database, respectively. As mentioned previously, the CTU-UHB CTG database has only 552 CTG recordings, and due to its relatively small size, implementing deep learning-based approaches as benchmarks is impractical. These approaches often require large number of data points for training due to their number of parameters. Instead, we implemented an STL (regular) GP model and widely used baseline models in the literature of computerized FHR analysis as our benchmark models.

In our experiments, we adopted 15 well-accepted FHR features [11]. The features can be divided into three categories: time domain, frequency domain, and nonlinear features. Time domain features include the mean and standard deviation of the FHR signals, short-term variability (STV), long-term variability (LTV), short-term irregularity (STI), and long-term irregularity (LTI) [21]. Frequency domain features consist of the energies in four frequency bands: very low frequency (VLF: 0–0.03 Hz), low frequency (LF: 0.03–0.15 Hz), mild frequency (MF: 0.15–0.5 Hz), and high frequency (HF: 0.5–1 Hz), as well as the LF/(MF + HF) ratio [22]. Nonlinear features include approximate entropy (ApEn), sample entropy (SampEn), and two measures of the Poincaré plot, SD1 and SD2 [23]. The complete list of features is shown in Table I.

Category	Features
Time Domain	Mean, Standard deviation, STV, LTV, STI, LTI
Frequency Domain	VLF, LF, MF, HF, ratio
Nonlinear	ApEn, SampEn, SD1, SD2

TABLE I: Table of all features.

In our experiments, we used the squared exponential covariance function k^x with variance equal to one to reduce the number of parameters for estimation. Additionally, to increase the model expressiveness, we assigned a lengthscale ℓ for each input dimension in k^x, which is known as automatic relevance determination (ARD) [24].

A. CTU-UHB CTG Database

We first implemented MTL regression of six fetal outcome measures, including umbilical cord blood pH values, pCO2, base excess, base deficit, and the Apgar score at 1 minute and at 5 minute [20]. In this work, we included FHR recordings that had at least 60 minutes in length and had no missing fetal outcome measures. In total, there were 350 suitable FHR recordings with corresponding outcome measures.

In computerized FHR analysis, FHR recordings are usually labeled by umbilical arterial pH when detecting neonatal acidosis. However, there is no consensus on the validity of using pH as an indicator for labeling [25], while other blood gas-based measures and Apgar scores have also been shown to have valuable information regarding the status of the infant at birth [13], [14]. Therefore, in addition to the pH values, we also included other cord blood gas results, i.e., pCO2, base excess, and base deficit as well as the Apgar score at 1 minute and at 5 minute after delivery, and the goal was to predict these fetal outcome measures jointly.

First, we extracted the previously mentioned features from the last 60 minutes of FHR signals. Since the above labels, i.e., target variables are continuous, we implemented the GP-based multi-task regression. We also implemented the regular GP regression, linear regression, and regression tree for benchmark purposes. As a performance metric, we used the mean absolute error (MAE), which measures the difference between the actual and predicted values. We performed 5-fold cross-validation (CV) and the experiments were repeated 10 times. The results were averaged over the 10 runs, and they are shown in Fig. 1. It can be seen that the GP-based multi-task approach outperformed the STL-GP and the other benchmark models.

B. SBU CTG Database

We also implemented MTL classification of four binary neonatal indicators on a new in-house CTG database. This database consists of 5429 intrapartum CTG recordings and corresponding clinical data. The data were acquired between January 2018 and December 2020 at Stony Brook Gynecology and Obstetrics of the University Hospital of SBU. The data used in this work were a subset of the database. More

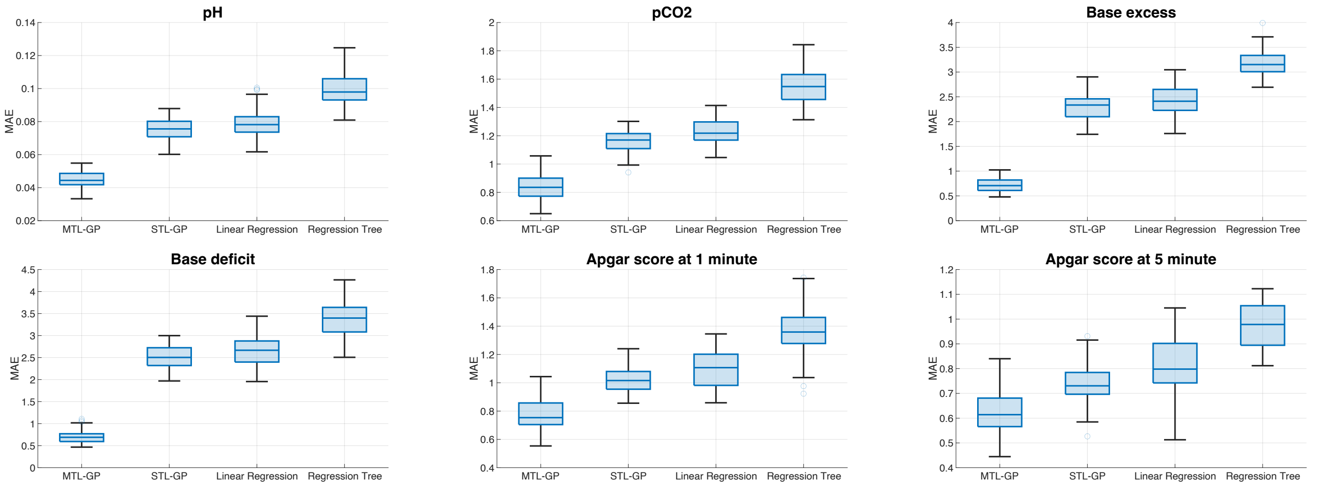

Fig. 1: Performance of GP-based multi-task learning approach (MTL-GP), GP regression (STL-GP), linear regression, and regression tree models. The tasks are regression of six fetal outcome measures, including umbilical cord blood pH values at birth, pCO2, base excess, base deficit and the Apgar score at 1 minute and at 5 minute, as shown in title of each plot.

specifically, we selected to work with 365 recordings with a missing data of less than 5% with a length of 60 minutes before delivery.

One main attraction of this database is the availability of information on multiple neonatal conditions, specifically, if the newborn 1) is diagnosed as having respiratory distress; 2) is diagnosed as having ischemia; 3) is diagnosed as having delayed transition; and 4) is admitted to the neonatal intensive care unit (NICU). In our experiments, we used these conditions as labels for detecting neonatal acidosis as they are more practical and can be used as alternatives to pH, since a newborn with a low pH does not necessarily mean that any pathological conditions will be developed in the future. As the above labels are all binary, we cast it as a multi-task classification problem where the goal is to predict if a newborn has a specific neonatal condition.

We implemented the GP-based multi-task classification and additionally included a GP classification model, logistic regression, and random forest as benchmark models. The performance was measured by the area under the receiver operating characteristic curve (AUC-ROC) which measures the diagnostic ability of a classification model as its discrimination threshold is varied. Similarly to the regression experiment in Section IV-A, we implemented a 5-fold CV and the experiments were repeated 10 times. The experimental results obtained by averaging over 10 runs are shown in Fig. 2. The results show that the GP-based multi-task approach outperformed the regular GP and the other benchmark classification models as it achieved the highest AUC-ROC in all tasks. This indicates that better diagnostic ability can be achieved by using MTL.

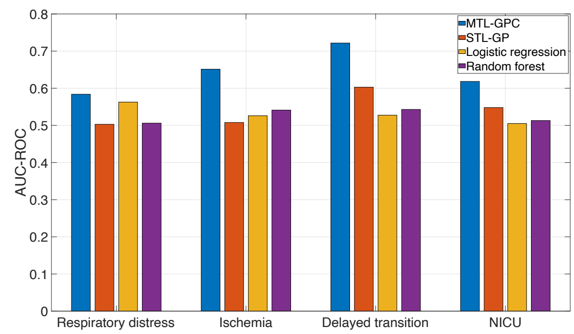

Fig. 2: Performance of GP-based multi-task learning classification model (MTL-GPC), GP classification (STL-GP), logistic regression, and random forest. The tasks are classification of four binary neonatal indicators, as shown on the x-axis.

V. Conclusions

In the paper, we cast FHR analysis tasks into a GP-based MTL framework that enables modeling of multiple fetal outcome measures and neonatal conditions simultaneously. Despite their usefulness in practice, these measurements are rarely exploited in computerized FHR analysis. We then demonstrated the effectiveness of the proposed approach with two different databases on various regression and classification tasks. Our results show that the GP-based MTL model can achieve better performance than benchmark STL models in all the tasks of FHR analysis. The GP-based MTL approach can readily be applied in the analysis of other biomedical domains where data are limited and multiple diagnoses are desired.

REFERENCES

[1] Diogo Ayres-de Campos, Catherine Y Spong, Edwin Chandraharan, and FIGO Intrapartum Fetal Monitoring Expert Consensus Panel, “FIGO consensus guidelines on intrapartum fetal monitoring: Cardiotocography,” *International Journal of Gynecology & Obstetrics*, vol. 131, no. 1, pp. 13–24, 2015.

[2] Lukáš Hruban, Jiří Spilka, Václav Chudáček, Petr Janků, Michal Huptych, Miroslav Burša, Adam Hudec, Marian Kacerovskỳ, Michal Kouckỳ, Martin Procházka, et al., “Agreement on intrapartum cardiotocogram recordings between expert obstetricians,” *Journal of Evaluation in Clinical Practice*, vol. 21, no. 4, pp. 694–702, 2015.

[3] Antoniya Georgieva, Patrice Abry, Václav Chudáček, Petar M Djurić, Martin G Frasch, René Kok, Christopher A Lear, Sebastiaan N Lemmens, Inês Nunes, Aris 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.

[4] Bruce E Josten, Timothy RB Johnson, and John 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.

[5] Yizhang Jiang, Jing Xue, Rong Wang, Kaijian Xia, Xiaoqing Gu, Jiaqi Zhu, Li Liu, and Pengjiang Qian, “Seizure recognition using a novel multitask radial basis function neural network,” *Journal of Medical Imaging and Health Informatics*, vol. 9, no. 9, pp. 1865–1870, 2019.

[6] Yuqi Si and Kirk Roberts, “Deep patient representation of clinical notes via multi-task learning for mortality prediction,” *AMIA Summits on Translational Science Proceedings*, vol. 2019, pp. 779, 2019.

[7] Edwin V Bonilla, Kian Chai, and Christopher Williams, “Multi-task Gaussian process prediction,” *Advances in Neural Information Processing Systems*, vol. 20, 2007.

[8] Gayle Leen, Jaakko Peltonen, and Samuel Kaski, “Focused multi-task learning using Gaussian processes,” in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2011, pp. 310–325.

[9] Mauricio Alvarez and Neil Lawrence, “Sparse convolved Gaussian processes for multi-output regression,” *Advances in Neural Information Processing Systems*, vol. 21, 2008.

[10] Jiří Spilka, Jordan Frecon, Roberto Leonarduzzi, Nelly Pustelnik, Patrice Abry, and Muriel Doret, “Sparse support vector machine for intrapartum fetal heart rate classification,” *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 3, pp. 664–671, 2016.

[11] Guanchao Feng, J Gerald Quirk, and Petar 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.

[12] Hakan Sahin and Abdulhamit Subasi, “Classification of the cardiotocogram data for anticipation of fetal risks using machine learning techniques,” *Applied Soft Computing*, vol. 33, pp. 231–238, 2015.

[13] Michael G Ross and Rageev Gala, “Use of umbilical artery base excess: Algorithm for the timing of hypoxic injury,” *American Journal of Obstetrics and Gynecology*, vol. 187, no. 1, pp. 1–9, 2002.

[14] Rahi Victory, Deborah Penava, Orlando Da Silva, Renato Natale, and Bryan Richardson, “Umbilical cord pH and base excess values in relation to adverse outcome events for infants delivering at term,” *American Journal of Obstetrics and Gynecology*, vol. 191, no. 6, pp. 2021–2028, 2004.

[15] Cristian Rotariu, Alexandru Pasarica, Hariton Costin, and Dragos Nemescu, “Spectral analysis of fetal heart rate variability associated with fetal acidosis and base deficit values,” in *2014 International Conference on Development and Application Systems (DAS)*. IEEE, 2014, pp. 210–213.

[16] Lukasz Kaiser, Aidan N Gomez, Noam Shazeer, Ashish Vaswani, Niki Parmar, Llion Jones, and Jakob Uszkoreit, “One model to learn them all,” *arXiv preprint arXiv:1706.05137*, 2017.

[17] Kai Yu, Volker Tresp, and Anton Schwaighofer, “Learning Gaussian processes from multiple tasks,” in *Proceedings of the 22nd International Conference on Machine Learning*, 2005, pp. 1012–1019.

[18] Trung V. Nguyen and Edwin V. Bonilla, “Collaborative multi-output Gaussian processes,” in *Proceedings of the Thirtieth Conference on Uncertainty in Artificial Intelligence*. 2014, UAI’14, p. 643–652, AUAI Press.

[19] Pablo Moreno-Muñoz, Antonio Artés-Rodríguez, and Mauricio A Álvarez, “Continual multi-task Gaussian processes,” *arXiv preprint arXiv:1911.00002*, 2019.

[20] Václav Chudáček, Jiří Spilka, Miroslav Burša, Petr Janků, Lukáš Hruban, Michal Huptych, and Lenka Lhotská, “Open access intrapartum CTG database,” *BMC Pregnancy and Childbirth*, vol. 14, no. 1, pp. 1–12, 2014.

[21] Jelte De Haan, Jan H Van Bemmel, B Versteeg, Anton FL Veth, Lodewijk AM Stolte, J Janssens, and Thomas KAB Eskes, “Quantitative evaluation of fetal heart rate patterns: I. processing methods,” *European Journal of Obstetrics & Gynecology*, vol. 1, no. 3, pp. 95–102, 1971.

[22] Maria G Signorini, Giovanni Magenes, Sergio Cerutti, and Domenico Arduini, “Linear and nonlinear parameters for the analysisof fetal heart rate signal from cardiotocographic recordings,” *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 3, pp. 365–374, 2003.

[23] Paul Fergus, Abir Hussain, Dhiya Al-Jumeily, De-Shuang Huang, and Nizar Bouguila, “Classification of caesarean section and normal vaginal deliveries using foetal heart rate signals and advanced machine learning algorithms,” *Biomedical Engineering Online*, vol. 16, no. 1, pp. 1–26, 2017.

[24] Christopher KI Williams and Carl Edward Rasmussen, *Gaussian Processes for Machine Learning*, vol. 2, MIT Press Cambridge, MA, 2006.

[25] Julian T Parer, Tekoa King, S Flanders, M Fox, and SJ Kilpatrick, “Fetal acidemia and electronic fetal heart rate patterns: Is there evidence of an association?,” *The Journal of Maternal-Fetal & Neonatal Medicine*, vol. 19, no. 5, pp. 289–294, 2006.