Prediction of major complications affecting very low birth weight infants

Olli-Pekka Rinta-Koski*, Simo Särkkä†, Jaakko Hollmén*,
Markus Leskinen‡, Krista Rantakari‡ and Sture Andersson‡

*Aalto University, Department of Computer Science, Espoo, Finland

†Aalto University, Department of Electrical Engineering and Automation, Espoo, Finland

‡Children's Hospital, Pediatric Research Center, University of Helsinki, and Helsinki University Hospital, Helsinki, Finland

***Abstract*— Bronchopulmonary dysplasia (BPD), necrotizing enterocolitis (NEC), and retinopathy of prematurity (ROP) are severe complications affecting Very Low Birth Weight (VLBW) infants. Our findings show that data gathered in the intensive care unit during the first 24 or 72 hours of care can be used to predict whether a VLBW infant is at risk of developing BPD. Using Gaussian process classification, we achieved classification results with areas under the receiver operator characteristic curve of 0.85 (standard error (SE) 0.05) for 24h and 0.87 (SE 0.06) for 72h BPD data. This compares favourably with results achieved using the clinical standard SNAP-II and SNAPPE-II scores. Sensitivity for BPD was 0.52 (SE 0.06). Sensitivity for NEC and ROP was close to zero, suggesting that NEC and ROP can not be reliably predicted with this approach from our data set.**

***Index Terms*— biomedical time series analysis, Gaussian process classification, bronchopulmonary dysplasia, necrotizing enterocolitis, retinopathy of prematurity, neonatal intensive care, very low birth weight infants**

I. Introduction

Very Low Birth Weight (VLBW) infants are born with a birth weight less than 1500 g. VLBW infants require critical care in a neonatal intensive care unit (NICU) and are at a high risk of developing both acute and later serious health issues. Many of their later problems are thought to originate from early care phases.

Three major complications affecting VLBW infants are bronchopulmonary dysplasia (BPD) [1], necrotizing enterocolitis (NEC) [2], and retinopathy of prematurity (ROP) [3]. The aim of the present study was to develop machine learning tools for early prediction of BPD, NEC, and ROP in VLBW infants using time series data. The contribution of this paper is to show that Gaussian process (GP) classification [4] can be used to predict whether a patient is in danger of developing BPD using time series data from the first 24 or 72 hours after a VLBW infant's admission to the NICU. Prediction of NEC and ROP using data from the same time frame gives results with close to zero sensitivity, which suggests that this approach does not work for these diagnoses.

In our previous work, NICU data from initial 24 hours has been used to predict in-hospital mortality [5] with GP classification. Prior to this, logistic regression [6], [7], [8], [9] and classifiers [9], [10] based on support vector machines (SVM) [11] have been used for BPD. An algorithm has been developed for predicting NEC using proteins in urine [12]. The prediction of sepsis, NEC, and in-hospital mortality using biosignals have also been considered [13], [14], [15].

II. Background

BPD, NEC, and ROP all manifest from days to weeks postnatally, but the development of all these diseases starts perinatally or during the early postnatal period [16]. Early prediction and identification of these disorders is a primary target when aiming at preventing or having early treatments for these conditions and consequently improving the prognosis of VLBW infants.

BPD is a severe chronic pulmonary complication of preterm birth. This most common lung disease in preterm infants can be diagnosed starting from the age of four weeks (28 days) [16]. Despite many advances in neonatal medicine, the incidence of BPD has not declined [17]. The majority of infants developing BPD have a birth weight less than 1250 g and they have postnatally required ventilator treatment for respiratory distress [16]. With the survivors, BPD is a major cause of long-term lung dysfunction causing a heavy burden on health care services and medical resources throughout childhood.

NEC is a critical illness in which segments of intestine undergo necrosis (tissue death). Severe NEC typically manifests during the first several weeks after birth, but its origins are thought be in the early phases of care. It is a life-threatening condition that often requires surgery and increases the risks for long term consequences, such as malnutrition, growth failure, BPD, ROP, and neurodevelopmental problems. Current treatments of NEC are not always effective [2].

Over 30% of VLBW infants weighing less than 1250 g develop ROP that can lead to severe vision problems and blindness in one or both eyes. Prematurity, low birth weight, and inappropriate oxygen levels are known risk factors of ROP [18], [19]. The incidence of ROP peaks postnatally at adjusted gestational age of 36–38 weeks [20]. Despite advances in treatments, ROP continues to be a significant problem in VLBW patients.

978-1-5386-1030-5/17/$31.00

SNAP-II and SNAPPE-II scores measure neonatal illness severity and predict risk for mortality. SNAP-II represents mortality risk from physiological problems, to which SNAPPE-II adds supplemental risk factors, such as birth weight and growth restriction [21]. They are calculated with data collected during the first 12 hours after admission to the NICU using variables such as blood pressure, body temperature, serum pH, birth weight, gestational age, and so on. Infants with a high SNAPPE-II score have been found to have significantly higher rates of NEC and BPD [22].

III. Methods

A. Data

Our data set contains data collected from 2059 VLBW infants treated between 1999 and 2013 in the NICU of Helsinki University Hospital's Children's Hospital. There are 416 (≈20%) patients diagnosed with BPD, 65 (≈3%) patients diagnosed with NEC, and 153 (≈7%) patients diagnosed with ROP.

For each patient, there are static values (SNAP-II and SNAPPE-II score, birth weight, and gestational age at birth) and time series data (systolic, mean, and diastolic arterial blood pressure, heart rate, and oxygen saturation). Time series data is averaged over 2 minute intervals.

B. Classifier

The patients were classified into two classes (likely/unlikely to get diagnosis) $y_i \in \{-1, 1\}$ using GP classification with a probit measurement model (Eq. 1)

$$\begin{aligned} f(\mathbf{x}) &\sim \mathcal{GP}(0, k(\mathbf{x}, \mathbf{x}')) \\ p(y_i \mid f(\mathbf{x}_i)) &= \int_{-\infty}^{y_i f(\mathbf{x}_i)} \mathrm{N}(z \mid 0, 1)\, dz \end{aligned} \tag{1}$$

and a kernel constructed as a linear combination of squared exponential, linear, and constant kernels (Eq. 2). This classifier has been previously used for VLBW infant in-hospital mortality prediction [5].

$$\begin{aligned} k(\mathbf{x}, \mathbf{x}') &= \sigma_{\mathrm{se}}^2 \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{x}')^{\mathsf{T}} \Lambda^{-1} (\mathbf{x} - \mathbf{x}')\right) \\ &\quad + \mathbf{x}^{\mathsf{T}} \Sigma \mathbf{x}' + \sigma^2 \\ \Lambda &= \mathrm{diag}(l_1^2, \ldots, l_d^2), \Sigma = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_d^2) \\ l_i &= \text{the lengthscale parameters} \\ \sigma_i^2 &= \text{the standard deviations of the Gaussians} \end{aligned} \tag{2}$$

For training the classifier we used the GPstuff Toolbox [23] with Laplace approximation on the latent variables and circular composite design integration over the hyperparameters.

Gestational age and birth weight were used as static features. In addition, we used time series data for the following five variables: systolic, mean, and diastolic arterial blood pressure, ECG heart rate, and oxygen saturation. These parameters were chosen because of their clinical and scientific importance. In the case of time series data, availability was also a consideration; for the chosen variables, reasonably complete time series data for the 24h and 72h periods being analyzed was available for >1000 patients.

The classification results were validated by stratified 5-fold crossvalidation which takes the class priors into account when forming the partitions.

For comparison purposes, patients were classified using thresholding with SNAP-II and SNAPPE-II scores only. A reference classifier which assumes that no patient will get the diagnosis in question was also used to quantify the classification results.

IV. Results

We have compared the classification results using the area (AUC) under the receiver operator characteristic curve (ROC) [24]. Tables I (BPD), II (NEC), and III (ROP) show the classification result using all available features and using only time series variables (i.e. without BW/GA) for 24h and 72h data, as well as thresholding results for SNAP-II and SNAPPE-II and the reference classifier. The results are shown in descending order by AUC. Standard error (SE) is shown in parentheses for all results. Figures 1 (BPD), 2 (NEC), and 3 (ROP) show the ROC curves for the best 24h and 72h GP classifiers, and SNAP-II/SNAPPE-II thresholding.

A. Bronchopulmonary dysplasia

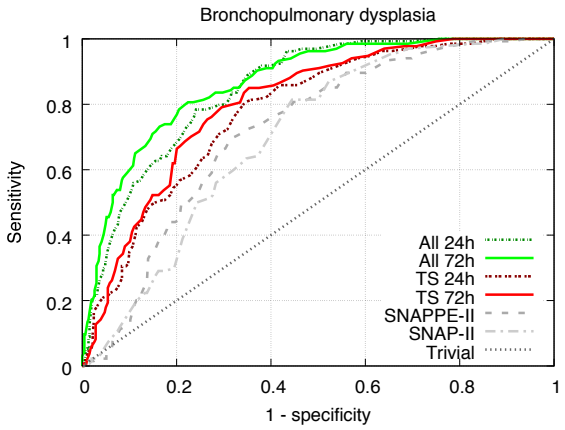

Fig. 1. Receiver operator characteristic curves for bronchopulmonary dysplasia results. False positive rate (1−specificity) on the X axis, true positive rate (sensitivity) on the Y axis.

The best AUC (0.87, SE 0.06) was achieved with 72h data using all available features. 24h data performed slightly worse (best AUC 0.85, SE 0.05), but still better than SNAP-II (AUC 0.70, SE 0.03) and SNAPPE-II (AUC 0.72, SE 0.04).

Best AUC achieved without either gestational age or birth weight was 0.81 (SE 0.05). While markedly worse than the best overall AUC, even this result surpassed that of SNAP-II and SNAPPE-II.

B. Necrotizing enterocolitis

The best AUC 0.74 (SE 0.02) was achieved with 72h data using all available features. However, sensitivity was close

TABLE I

CLASSIFICATION RESULTS FOR BRONCHOPULMONARY DYSPLASIA.

Variables	Acc	PPV	Sens	Spec	AUC
All 72h	**0.84** (0.01)	0.67 (0.04)	**0.52** (0.06)	0.93 (0.01)	**0.87** (0.06)
All 24h	0.81 (0.01)	0.65 (0.01)	0.42 (0.04)	0.93 (0.01)	0.85 (0.05)
TS 72h	0.78 (0.01)	0.54 (0.04)	0.31 (0.07)	0.91 (0.03)	0.80 (0.06)
TS 24h	0.77 (0.02)	0.51 (0.05)	0.42 (0.06)	0.88 (0.02)	0.78 (0.06)
SNAPPE-II	0.77 (0.00)	0.60 (0.24)	0.00 (0.00)	0.99 (0.00)	0.72 (0.04)
SNAP-II	0.77 (0.00)	0.80 (0.20)	0.00 (0.00)	**1.00** (0.00)	0.70 (0.03)
Reference	0.77 (0.00)	**1.00** (0.00)	0.00 (0.00)	**1.00** (0.00)	– (–)

All = all variables (gestational age, birth weight, and time series variables), TS = time series variables only (systolic/mean/diastolic arterial blood pressure, ECG heart rate, oxygen saturation), Acc = accuracy, PPV = positive predictive value, Sens = sensitivity, Spec = specificity, AUC = area under the receiver operator characteristic curve. Standard error is indicated in parentheses. Best values shown in bold.

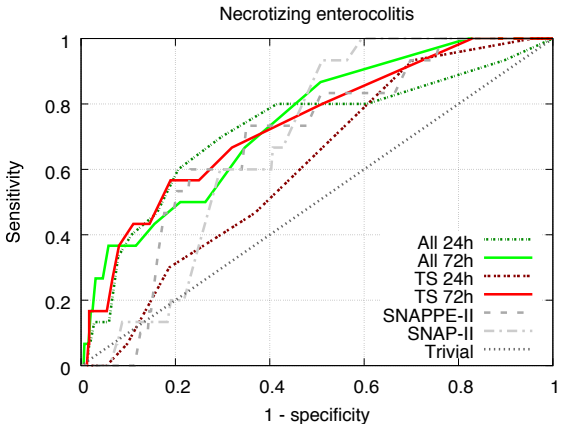

Fig. 2. ROC curves for necrotizing enterocolitis.

TABLE II

CLASSIFICATION RESULTS FOR NECROTIZING ENTEROCOLITIS.

Variables	Acc	PPV	Sens	Spec	AUC
All 72h	0.97 (0.00)	0.11 (0.07)	0.17 (0.11)	0.98 (0.00)	**0.74** (0.02)
TS 72h	0.96 (0.01)	0.20 (0.20)	0.00 (0.00)	0.98 (0.01)	**0.74** (0.02)
All 24h	0.97 (0.00)	0.40 (0.24)	0.00 (0.00)	0.99 (0.00)	0.72 (0.01)
SNAPPE-II	**0.98** (0.00)	0.80 (0.20)	0.00 (0.00)	**1.00** (0.00)	0.69 (0.04)
SNAP-II	**0.98** (0.00)	0.80 (0.20)	0.00 (0.00)	**1.00** (0.00)	0.68 (0.03)
TS 24h	0.97 (0.00)	0.60 (0.24)	0.00 (0.00)	**1.00** (0.00)	0.61 (0.01)
Reference	**0.98** (0.00)	**1.00** (0.00)	0.00 (0.00)	**1.00** (0.00)	– (–)

to zero at 0.17 (SE 0.11). For all other classification results, including the benchmarks, sensitivity was equal to zero.

C. Retinopathy of prematurity

TABLE III

CLASSIFICATION RESULTS FOR RETINOPATHY OF PREMATURITY.

Variables	Acc	PPV	Sens	Spec	AUC
All 72h	0.92 (0.01)	0.50 (0.22)	**0.05** (0.03)	0.99 (0.01)	**0.84** (0.03)
All 24h	0.92 (0.01)	0.40 (0.24)	0.00 (0.00)	0.99 (0.01)	0.80 (0.02)
TS 72h	0.92 (0.01)	0.60 (0.24)	0.00 (0.00)	0.99 (0.01)	0.74 (0.03)
SNAPPE-II	**0.93** (0.00)	0.80 (0.20)	0.00 (0.00)	**1.00** (0.00)	0.72 (0.06)
TS 24h	0.92 (0.01)	0.47 (0.23)	0.03 (0.03)	0.99 (0.01)	0.69 (0.02)
SNAP-II	**0.93** (0.00)	0.80 (0.20)	0.00 (0.00)	**1.00** (0.00)	0.67 (0.03)
Reference	**0.93** (0.00)	**1.00** (0.00)	0.00 (0.00)	**1.00** (0.00)	– (–)

As with BPD and NEC, the best AUC 0.83 (SE 0.03) was achieved with 72h data using all available features. Sensitivity was close to zero for all classification results. No classifier had a higher sensitivity than 0.05 (SE 0.03).

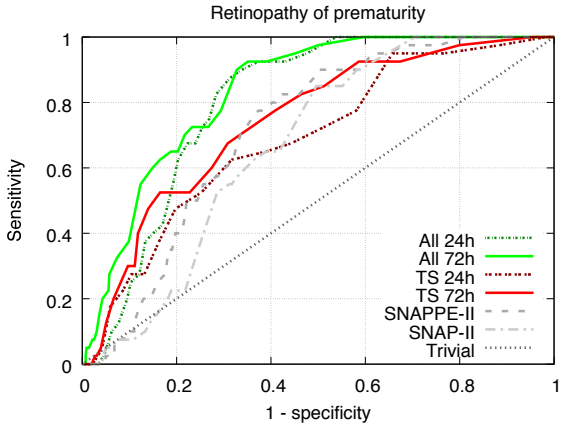

Fig. 3. ROC curves for retinopathy of prematurity.

V. DISCUSSION

The best predictive classification results were achieved for BPD, with AUC 0.85 (SE 0.05) for 24h and 0.87 (SE 0.06) for 72h data. This was in excess of the results for SNAP-II (AUC 0.70, SE 0.03) and SNAPPE-II (AUC 0.72, SE 0.04), indicating that GP classification using time series data is a better predictor of a patient's likelihood of developing BPD than either of these medical standard scores.

ROP classification results were better than SNAP-II and SNAPPE-II thresholding, however, sensitivity (true positive rate) in the best predictions was close to zero. Factors predisposing to ROP may emerge later than the (up to) 3-day period of this study, and so its reliable prediction may require data from a longer time period. It is also quite possible that the variables chosen for this study are not optimal for ROP prediction.

In the prediction of NEC, AUC for 24h data (0.72) was marginally better than SNAPPE-II (0.69). 72h data gave better results, with AUC 0.74 (SE 0.02). SNAP-II has not been found to predict later development of NEC [13], and our AUC 0.68 (SE 0.03) is in agreement. Heart rate characteristics, such as heart rate variability (HRV), are known to correlate with sepsis and NEC [14]. However, our data set is averaged over 2 minute intervals and does not have beat-to-beat information, so we could not use HRV as a feature.

A SVM-based approach has been used to achieve an accuracy of 0.832 for BPD prediction [10]. Our best GP accuracy result was 0.81 (SE 0.01) for 24h and 0.84 (SE 0.01) for 72h data. These accuracy values are not directly comparable, due to the difference in sample sizes (a cohort of 109 patients for [10] vs. 2059 for our study) and class sizes (≈42% positive for [10] vs. ≈20% positive for our study).

In another study [15], heart rate (one of our time series features) measured during the first 7 days was found to correlate with BPD, but no such correlation was found for

NEC and ROP. SNAP-II was also found to correlate with BPD, but not with NEC and ROP. This agrees with our findings.

There is a direct correlation between prediction accuracy and relative class size. Only 65 out of 2059 were diagnosed with NEC, making the negative class more than 30 times as large as the positive class.

Gestational age and birth weight have been found earlier to be important variables in predicting BPD [6]. There is a marked correlation with these two features and the general well-being of a VLBW infant, but they can not be used to achieve a good classification result by themselves.

The classifier, as presented, can be used to predict BPD but would have to be revised for NEC and ROP, perhaps by using a sliding time window tracking recent changes in patient state instead of only data from the early stages of care. Feature selection is also an important factor. The data set used in this paper contains time series data for blood pressure, heart rate, and oxygen saturation. Using supplementary oxygen, its effect on oxygen saturation, and the rates of change of both could shed more light on these, especially in the case of ROP.

VI. Conclusions

Time series data from the initial hours of a VLBW infant's life can be used to predict the infant's susceptibility to major complications. These predictions will in general be more accurate than just using the medical standard SNAP-II or SNAPPE-II scores, which are established with data available from the first 12 hours in the NICU.

In this study we looked at the predictive power of 24h and 72h data. As can be expected, classification results were improved when more data was available. However, 24h data already gives a good prediction of an infant's likelihood of developing BPD.

In contrast with BPD, our findings show that GP classification can not reliably predict NEC nor ROP using early time series data for blood pressure, heart rate, and oxygen saturation.

A classification tool based on this approach could assist care personnel in following the most important parameters in order to identify and predict patients most likely to develop complications and subsequently to develop personalized care for these patients at risk.

Ethics approval: The study was approved by the Helsinki University Central Hospital Ethics Committee, decision number 115/13/03/00/14 dated 8 April 2014.

References

[1] J. Gien and J. P. Kinsella, "Pathogenesis and treatment of bronchopulmonary dysplasia," *Current Opinion in Pediatrics*, vol. 23, no. 3, pp. 305–313, Jun. 2011.

[2] J. Neu and W. A. Walker, "Necrotizing Enterocolitis," *New England Journal of Medicine*, vol. 364, no. 3, pp. 255–264, Jan. 2011.

[3] B. W. Fleck and B. J. Stenson, "Retinopathy of Prematurity and the Oxygen Conundrum," *Clinics in Perinatology*, vol. 40, no. 2, pp. 229–240, Jun. 2013.

[4] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. The MIT Press, 2006.

[5] O.-P. Rinta-Koski, S. Särkkä, J. Hollmén, M. Leskinen, and S. Andersson, "Prediction of preterm infant mortality with Gaussian process classification," in *Proceedings of the 25th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, Bruges, Belgium, 26-28 April 2017, pp. 193–198.

[6] V. K. Bhutani and S. Abbasi, "Relative likelihood of bronchopulmonary dysplasia based on pulmonary mechanics measured in preterm neonates during the first week of life," *The Journal of pediatrics*, vol. 120, no. 4, pp. 605–613, 1992.

[7] C. A. Bhering, C. C. Mochdece, M. E. L. Moreira, J. R. Rocco, and G. M. Sant'Anna, "Bronchopulmonary dysplasia prediction model for 7-day-old infants," *Jornal de Pediatria*, vol. 83, no. 2, pp. 163–170, Mar. 2007.

[8] T. Gursoy, M. Hayran, H. Derin, and F. Ovali, "A Clinical Scoring System to Predict the Development of Bronchopulmonary Dysplasia," *American Journal of Perinatology*, vol. 32, no. 07, pp. 659–666, Oct. 2014.

[9] M. Ochab and W. Wajs, "Bronchopulmonary Dysplasia Prediction Using Support Vector Machine and Logit Regression," in *Information Technologies in Biomedicine, Volume 4*, ser. Advances in Intelligent Systems and Computing. Cham: Springer International Publishing, 2014, vol. 284, pp. 365–374.

[10] ——, "Expert system supporting an early prediction of the bronchopulmonary dysplasia," *Computers in Biology and Medicine*, vol. 69, pp. 236–244, Feb. 2016.

[11] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[12] K. G. Sylvester, X. B. Ling, G. Y. Liu, Z. J. Kastenberg, J. Ji, Z. Hu, S. Peng, K. Lau, F. Abdullah, M. L. Brandt, R. A. Ehrenkranz, M. C. Harris, T. C. Lee, J. Simpson, C. Bowers, and R. L. Moss, "A novel urine peptide biomarker-based algorithm for the prognosis of necrotising enterocolitis in human infants," *Gut*, vol. 63, no. 8, pp. 1284–1292, Aug. 2014.

[13] B. A. Sullivan and K. D. Fairchild, "Predictive monitoring for sepsis and necrotizing enterocolitis to prevent shock," *Seminars in Fetal and Neonatal Medicine*, vol. 20, no. 4, pp. 255–261, Aug. 2015.

[14] M. P. Griffin, D. E. Lake, T. M. O'Shea, and J. R. Moorman, "Heart Rate Characteristics and Clinical Signs in Neonatal Sepsis," *Pediatric Research*, vol. 61, no. 2, pp. 222–227, Feb. 2007.

[15] B. A. Sullivan, C. McClure, J. Hicks, D. E. Lake, J. R. Moorman, and K. D. Fairchild, "Early Heart Rate Characteristics Predict Death and Morbidities in Preterm Infants," *The Journal of Pediatrics*, vol. 174, pp. 57–62, Jul. 2016.

[16] E. Baraldi and M. Filippone, "Chronic lung disease after premature birth," *New England Journal of Medicine*, vol. 357, no. 19, pp. 1946–1955, 2007.

[17] V. C. Smith, J. A. F. Zupancic, M. C. McCormick, L. A. Croen, J. Greene, G. J. Escobar, and D. K. Richardson, "Trends in severe bronchopulmonary dysplasia rates between 1994 and 2002," *The Journal of Pediatrics*, vol. 146, no. 4, pp. 469–473, Apr. 2005.

[18] B. W. Fleck and N. McIntosh, "Pathogenesis of retinopathy of prematurity and possible preventive strategies," *Early Human Development*, vol. 84, no. 2, pp. 83–88, Feb. 2008.

[19] M. E. Hartnett and J. S. Penn, "Mechanisms and Management of Retinopathy of Prematurity," *New England Journal of Medicine*, vol. 367, no. 26, pp. 2515–2526, Dec. 2012.

[20] E. A. Palmer, J. T. Flynn, R. J. Hardy, D. L. Phelps, C. L. Phillips, D. B. Schaffer, B. Tung, F. J. Elsas, J. M. Botsford, K. W. Braune, and others, "Incidence and Early Course of Retlnopathy of Prematurity," *Ophthalmology*, vol. 98, no. 11, pp. 1628–1640, 1991.

[21] D. K. Richardson, J. D. Corcoran, G. J. Escobar, and S. K. Lee, "SNAP-II and SNAPPE-II: Simplified newborn illness severity and mortality risk scores," *The Journal of Pediatrics*, vol. 138, no. 1, pp. 92–100, Jan. 2001.

[22] S. Ucar, M. Varma, M. I. Ethemoglu, and N. K. Acar, "The Efficacy of SNAPPE-II in Predicting Morbidity and Mortality in Extremely Low Birth Weight Infants," *Archives of Disease in Childhood*, vol. 99, no. Suppl 2, p. A468, Oct. 2014.

[23] J. Vanhatalo, J. Riihimäki, J. Hartikainen, P. Jylänki, V. Tolvanen, and A. Vehtari, "GPstuff: Bayesian modeling with Gaussian processes," *Journal of Machine Learning Research*, vol. 14, no. Apr, pp. 1175–1179, 2013.

[24] J. Davis and M. Goadrich, "The relationship between Precision-Recall and ROC curves," in *Proceedings of the 23rd International Conference on Machine Learning*. ACM, 2006, pp. 233–240.