BOOST ENSEMBLE LEARNING FOR CLASSIFICATION OF CTG SIGNALS

Marzieh Ajirak,[†] *Cassandra Heiselman,*[◇] *J. Gerald Quirk,*[◇] *Petar M. Djurić*[†]

[†]Department of Electrical and Computer Engineering
[◇]Department of Obstetrics, Gynecology and Reproductive Medicine
Stony Brook University
Stony Brook, NY 11794, USA

ABSTRACT

During the process of childbirth, fetal distress caused by hypoxia can lead to various abnormalities. Cardiotocography (CTG), which consists of continuous recording of the fetal heart rate (FHR) and uterine contractions (UC), is routinely used for classifying the fetuses as hypoxic or non-hypoxic. In practice, we face highly imbalanced data, where the hypoxic fetuses are significantly underrepresented. We propose to address this problem by boost ensemble learning, where for learning, we use the distribution of classification error over the dataset. We then iteratively select the most informative majority data samples according to this distribution. In our work, in addition to addressing the imbalanced problem, we also experimented with features that are not commonly used in obstetrics. We extracted a large number of statistical features of fetal heart tracings and uterine activity signals and used only the most informative ones. For classification, we implemented several methods: Random Forest, AdaBoost, k-Nearest Neighbors, Support Vector Machine, and Decision Trees. The paper provides a comparison in the performance of these methods on fetal heart rate tracings available from a public database. Our results on the publicly available Czech database show that most applied methods improved their performances considerably when boost ensemble was used.

Index Terms— Boost ensemble learning, imbalanced learning, cardiotocography

1. INTRODUCTION

During the process of childbirth, fetal distress caused by hypoxia can lead to various abnormalities. Since the brain of a neonate is easily influenced by oxygen supply, a lack of oxygen can cause serious damage to the brain [19]. Therefore, continuous monitoring of the fetal state during labor is of great importance. Cardiotocography (CTG), which consists of continuous recording of the fetal heart rate (FHR) measured in beats per minute (bpm), and uterine contractions (UC) measured in mmHg, is routinely adopted by doctors to monitor and assess the fetal state during childbirth [16]. CTG recordings provide continuous information about the state of the fetus during delivery, which is used to screen for signs of fetal distress [1, 21]. This allows obstetricians to intervene in risky situations in order to prevent potential neurological adverse outcomes. Figure 1 shows an example of such recordings.

In clinical practice, the interpretation of CTG recordings is generally performed visually by obstetricians and field experts. Unfortunately, due to the complexity of these signals, the visual analysis of CTG signals using common guidelines leads to diagnostic errors [9].

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

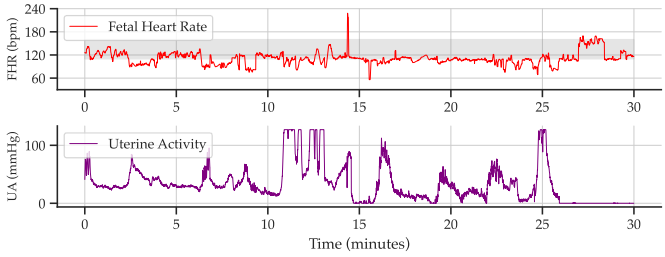


Fig. 1. Fetal heart rate tracing with its associated uterine activity recording.

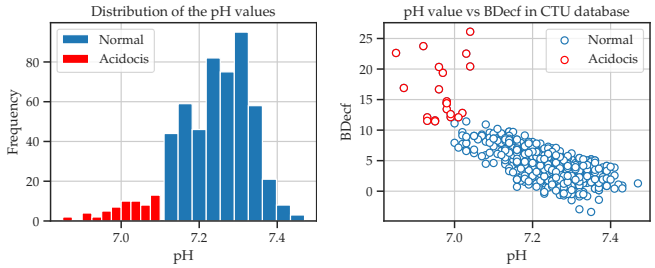


Fig. 2. Highly imbalanced data in Czech database.

In addressing the subjectivity of CTG interpretation, computer-aided diagnosis systems based on advanced machine learning (ML) have recently been developed to assist obstetricians in making objective medical decisions [23, 24, 25]. These data-driven approaches use signal processing and ML methods, including preprocessing, feature extraction, and classifiers to discriminate between two or more classes. These classes are often formed using the umbilical cord artery pH values, the base deficit of extracellular fluid (BDecf), or Apgar scores measured after delivery. The classes distinguish between hypoxic and non-hypoxic fetuses.

A large body of work on this matter notwithstanding, computational methods have not been widely adopted by obstetricians [7]. One reason is that traditional classification algorithms most often demonstrate unsatisfactory performance when they operate on highly imbalanced data, which is the case in CTG signal classification, where the number of fetuses with hypoxia is much lower than non-hypoxic fetuses. The reason for the poor performance is that most classifier algorithms wrongly assume a relatively balanced distribution between positive and negative samples [13]. Therefore, the minority class is usually ignored due to the overwhelming

number of majority instances. Clearly, detecting the minority class, i.e., the hypoxic fetuses, is of greater interest than fetuses who are non-hypoxic (the majority class). Figure 2 shows the imbalance ratio of an open-access CTG database, which we later use in our experiments [5].

To overcome this issue, various imbalanced learning methods have been proposed. These algorithms attempt to eliminate the bias toward the majority class through data resampling (undersampling or oversampling) [8, 10, 12, 20]. The most successful methodology to tackle the imbalanced learning problem has been the one based on ensemble learning [22, 11]. In ensemble learning, we undersample various subsets from the majority class and produce multiple classifiers using the combination of each subset with the minority class data. However, the primary deficiency of most existing undersampling methods is that their data sampling strategies are heuristic-based and that they usually throw away potentially valuable samples [13].

On the other hand, it is reported that the highest accuracy for CTG recording classification is achieved with deep neural networks [24, 17]. The problem with deep learning is that the feature extraction is transferred to the machine with the caveat that the produced features will be complicated if not impossible to interpret [15]. Hence, there is a need for more informative and interpretable features [6]. It is proved to be efficient to apply comprehensive feature extraction algorithms and then filter the respective features by an efficient feature selection algorithm [2, 6].

In this paper, instead of solely balancing the hypoxic and nonhypoxic fetuses, we use the distribution of classification error over the dataset and iteratively select the most informative majority data samples according to this distribution. We did not rely on features provided by obstetricians and field experts; instead, we extracted a large number of time series features (of the FHR tracings and UC signals) and kept the most informative ones. Subsequently, we used the selected features for classification by Random Forest, Naïve Bayes, AdaBoost, k-Nearest Neighbors, Support Vector Machine, and Logistic Regression. We provide a detailed comparison of these methods on CTG signals available from a public database [5].

The remainder of this paper is organized as follows. In the next section, we present the proposed methodology, and in Section, 3 we describe the database used for testing the methodology with various ML methods. In Section 4, we explain how we conducted our experiments and provide detailed results of the ML methods used in the study. Section 5 presents concluding remarks about our method.

2. THE METHODOLOGY

2.1. Boost Ensemble Imbalanced Learning

Let $\mathbf{x}$ be the input features and y their corresponding labels. We denote one instance of such pair by $(\mathbf{x}, y)$. Given an imbalanced dataset $\mathcal{D} = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \cdots, (\mathbf{x}_N, y_N)\}$, we define the minority (positive) set as

$$\mathcal{P} = \{(\mathbf{x}, y) : y = 1, (\mathbf{x}, y) \in \mathcal{D}\}, \tag{1}$$

and the majority (negative) set as

$$\mathcal{N} = \{(\mathbf{x}, y) : y = 0, (\mathbf{x}, y) \in \mathcal{D}\}. \tag{2}$$

For highly imbalanced data, we have $|\mathcal{N}| \gg |\mathcal{P}|$. We use $f : \mathbf{x} \to [0, 1]$ to denote a single classifier and $F_k : \mathbf{x} \to [0, 1]$ to denote an ensemble of classifiers that is formed by k base classifiers.

Table 1. Extracted Features

FHR FEATURES	
BL	mean of baseline
STD	standard deviation
STV	short-term variability
LTV	long-term variability
STI	short-term irregularity
LTI	long-term irregularity
PS	Power Spectrum, VLF [0.00-0.03] Hz, LF [0.03-0.15] Hz, MF [0.15-0.5] Hz, HF [0.5-1] Hz
SKE	sample skewness
KUR	sample kurtosis
ENT	approximate entropy
ACO	lagged autocorrelation
QT	change of quantile of the distribution
FFT	Fast Fourier Transformation Coefficients
aboveL	length of intervals above Mean/Median
belowL	length of intervals below Mean/Median
aggLTN	aggregation of linear least-squares regression[4].
UA FEATURES	
nCont	number of detected contractions
ICI	mean of inter-contraction intervals
CLINICAL FEATURES	
Gestational Weeks, Sex, Age, Parity, Diabetes Hypertension, Preeclampsia	

Formally, given a data instance $(\mathbf{x}, y)$ and an ensemble classifier F_k, the ensemble training error is defined as the absolute difference between the predicted probability of $\mathbf{x}$ being positive and the ground truth label y, i.e., $|F_k(\mathbf{x}) - y|$. Suppose next that the error histogram on the dataset $\mathcal{D}$ is given by a vector $\boldsymbol{e}_{\mathcal{D}} \in \mathbb{R}^b$, where b is the number of bins in the histogram. Specifically, the i-th component of the vector $\boldsymbol{e}_{\mathcal{D}}$ can be computed as follows:

$$\boldsymbol{e}_{\mathcal{D}}(i) = \frac{\left|\left\{(\mathbf{x}, y) \mid \frac{i-1}{b} \leq \operatorname{abs}\left(F_k(\mathbf{x}) - y\right) < \frac{i}{b}, (\mathbf{x}, y) \in \mathcal{D}\right\}\right|}{|\mathcal{D}|}, \tag{3}$$

where, $1 \leq i \leq b$. Intuitively, the error vector $\boldsymbol{e}_{\mathcal{D}}$ in (3) shows how well a given classifier fits the dataset $\mathcal{D}$. When $b = 2$, the histogram reports the accuracy score by $\boldsymbol{e}_{\mathcal{D}}(1)$ and the misclassification rate in $\boldsymbol{e}_{\mathcal{D}}(2)$ (classification threshold is 0.5). With $b > 2$, the histogram shows the distribution of "unimportant" samples (with errors close to 0) and "important" samples (with errors close to 1).

We construct a similar histogram on the validating test. Then we concatenate the error distribution vectors of the training and validation sets, and obtain what we refer to a state, i.e,

$$\boldsymbol{e} = [\boldsymbol{e}_{\mathcal{D}_{train}} : \boldsymbol{e}_{\mathcal{D}_{test}}] \in \mathbb{R}^{2b}. \tag{4}$$

Next, we want to use the state $\boldsymbol{e}$ to select the next set of samples that are the most informative majority representations. Data samples with more significant errors are more likely to be selected for training in the next iteration. The process continues iteratively until the predefined iterations are completed. More precisely, we implement the method as follows.

We take advantage of the meta-training in MESA algorithm in [14]. Let g be a distribution described by the probabilities $\mu \in [0, 1]$

Algorithm 1 Boost Ensemble Imbalance Learning

Input: $\mathcal{D}_{train}, \mathcal{D}_{test}, g, f, b, k$
Initialize: Random undersample $|\mathcal{N}_0'|$ and $|\mathcal{P}|$ samples from majority class and train the first classifier f_0, where $|\mathcal{N}_0'| = |\mathcal{P}|$
For $t = 1, \ldots k-1$
  $F_t(x) = \frac{1}{t}\sum_{i=1}^{t} f_i(x)$
  Compute $\boldsymbol{e}_{\mathcal{D}_{train}}$ and $\boldsymbol{e}_{\mathcal{D}_{test}}$
  $\boldsymbol{e}_t = [\boldsymbol{e}_{\mathcal{D}_{train}} : \boldsymbol{e}_{\mathcal{D}_{test}}]$
  $\mu_t \sim g\left(\mu_t \mid s_t\right)$
  $\mathcal{D}'_{t+1,train} = Newset\left(\mathcal{D}_{train}; F_t, \mu_t,\right)$
  Train new classifier $f_{t+1}(x)$ with $\mathcal{D}'_{t+1,train}$
End
Return $F_k(x) = \frac{1}{k}\sum_{i=1}^{k} f_i(x)$

based on the errors e. These probabilities quantify how likely a sample will be selected for training in the next iteration. We jitter the probabilities μ by sampling from Gaussians centered at the μs and with variance σ^2. The newly selected μs after normalization form the distribution g that is used for generating data for the next training iteration,

Given the distribution $g : \mathbb{R}^{2b} \rightarrow [0, 1]$, we obtain the samples for training the new base classifiers. At the t-th iteration, having the current ensemble $F_t(\cdot)$, we can obtain $\boldsymbol{e}_{\mathcal{D}_{train}}, \boldsymbol{e}_{\mathcal{D}_{test}}$ and meta-state $\boldsymbol{e}_t$ by applying (3) and (4). Then a new base classifier $f_{t+1}(\cdot)$ is trained with the subset $\mathcal{D}'_{t+1,train} = Newset\left(\mathcal{D}_{train}; F_t, \mu_t\right)$, where $\mu_t \sim g\left(. \mid e_t\right)$ and $\mathcal{D}_{train}$ is the original training set. Note that $f_1(\cdot)$ was trained on a random balanced subset, as there is no trained classifier in the first iteration.

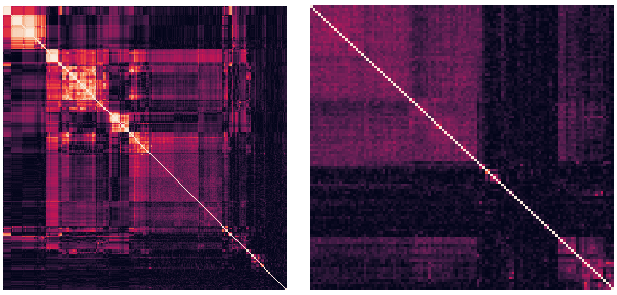
As described above, the sampler g is trained to optimize the generalized performance of the ensemble classifier by iteratively selecting the training data. It takes the current error of the training system as input and then outputs the parameter μ of a Gaussian function to decide each instance's sampling probability. This sampler is expected to learn and adapt its policy in order to decrease the classification error. This optimization problem can be defined as a Markov Decision Process (MDP) [18]. More specifically, in each iteration, we repeatedly train k base classifiers $f(\cdot)$ and form an ensemble classifier $F_k(\cdot)$. In each iteration, the ensemble provides the state of the MDP, i.e., the errors of training and test sets, $\boldsymbol{e}_t = [\boldsymbol{e}_{\mathcal{D}_{train}} : \boldsymbol{e}_{\mathcal{D}_{test}}]$, and then the action μ_t is selected by $\mu_t \sim g\left(. \mid \boldsymbol{e}_t\right)$. A new base classifier $f_{t+1}(\cdot)$ is trained using the subset,

$$\mathcal{D}'_{t+1,train} = Newset\left(\mathcal{D}_{train}; F_t, \mu_t\right). \tag{5}$$

After adding $f_{t+1}(\cdot)$ into the ensemble, the new state $\boldsymbol{e}_{t+1}$ is sampled according to $\boldsymbol{e}_{t+1} \sim p\left(\boldsymbol{e}_{t+1}; \boldsymbol{e}_t, \mu_t\right)$. The reward r can be defined as the difference in the final testing error before and after the ensemble update.

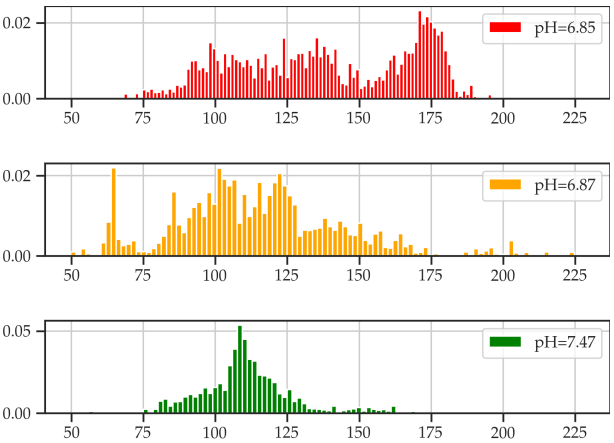
3. OPEN ACCESS INTRAPARTUM CTG DATABASE

We use the open-access intrapartum CTG database [5]. The database contains 552 CTG recordings that were acquired at the obstetrics ward of the University Hospital in Brno, Czech Republic. The clinical data of mothers and biomarkers of babies such as pH and BDecf values measured at birth were also collected. In this work, we use a bag of FHR features. The number of detected contractions and the average contraction times along with the clinical information, including Gestational Weeks, sex, age, parity, diabetes, hypertension, and preeclampsia are also used. All the features are summarized in Table 1. A problem of this filtering method is the redundancy in the



(a) Feature correlation matrix: presence of many redundant features.

(b) Feature correlation matrix after removing redundant features.



(c) Density of FHR values associated with different pH values

Fig. 3. CTG Features

feature selection. As long as the considered features are associated with the target, they will all be selected by the filter even though many of them are highly correlated with each other. We removed redundant features by dropping highly correlated features and keeping only one of them as a representative. Figures 3(a) and 3(b) show the correlation of selected features before and after the redundant features were removed, respectively. It is worth noting that the most informative feature is the standard deviation of the FHR signal in the last 30 minutes of the recording, Fig, 3(c).

4. EXPERIMENTS AND RESULTS

From the CTG recordings in the database, we extracted 706 basic features that are defined by varying various acquisition parameters. We used [4] for the task of rigorously extracting features from time-series data. Then we dropped the features that have close to zero correlation with the pH values. In order to reduce the dimension of the features effectively, we dropped the features that are highly correlated with each other and kept only one of them.

From the 552 recordings, 44 of them have a pH value lower or equal to 7.05, which is the threshold selected for defining the two classes in this study. Therefore, these 44 cases constituted the abnormal class while the rest 508 represented the normal class, making an imbalance ratio of 508:44. In order to compute the error histogram of the data, we used Leave-Two-Out cross-validation. The performance was evaluated using Area Under the ROC Curve (AUC), Classification Accuracy (CA), F1 score, Precision, and Recall. Classification

Accuracy is the fraction of predictions that were correct, i.e.,

$$\text{CA} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{6}$$

where TP stands for True Positive, TN for True Negative, FP for False Positive, and FN for False Negative predictions. Further, we used the following metrics:

$$\text{Recall} = \frac{TP}{TP + FN}, \tag{7}$$

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{8}$$

$$\text{F1} = 2 \times \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}. \tag{9}$$

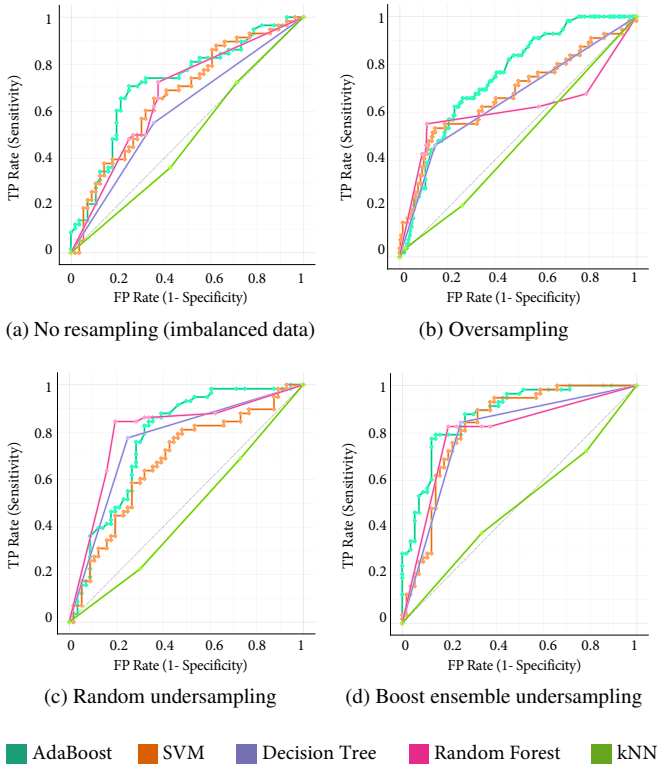


Fig. 4. ROC plots of various classifiers implemented with different schemes for data balancing.

We compared different ensemble random under-resampling with optimized boost-based undersampling and contrasted them with oversampling and no balancing at all. We used five different classifiers, i.e., Random Forest, AdaBoost, k-Nearest Neighbours (kNN), Decision Tree, and Support Vector Machine (SVM) in evaluating the different balancing techniques on CTG data. Table 2 provides the experimental results. We show that by learning a boost-based optimized resampling for balancing the data, many classifiers improve their performance. Since the dataset is highly imbalanced, oversampling the minority class deteriorates the overall classification because the minority class is poorly represented and lacks a clear structure. Thus, oversampling methods like SMOTE [3] have poor performance in classifying such data. We note that since the test data are balanced, the AUC is a reasonable metric for imbalanced learning evaluation. The ROC plots of all the classifiers under different resampling regimes are depicted in Fig 4. The proposed boost ensemble undersampling clearly shows the best performance for all the applied methods.

Table 2. Comparison of resampling methods

	AUC	CA	F1	Precision	Recall
No resampling					
SVM	0.684	0.837	0.374	0.486	0.304
Random forest	0.753	0.831	0.033	0.200	0.018
Decision Tree	0.626	0.819	0.452	0.441	0.464
AdaBoost	0.657	0.788	0.413	0.371	0.464
kNN	0.477	0.650	0.164	0.133	0.214
Oversampling (SMOTE)					
Random forest	0.783	0.781	0.780	0.781	0.781
Decision Tree	0.761	0.675	0.675	0.676	0.675
AdaBoost	0.640	0.640	0.640	0.640	0.640
SVM	0.601	0.614	0.614	0.615	0.614
kNN	0.438	0.465	0.451	0.458	0.465
Ensemble random undersampling					
Random Forest	0.760	0.763	0.763	0.766	0.763
Decision Tree	0.797	0.737	0.737	0.737	0.737
AdaBoost	0.712	0.711	0.710	0.714	0.711
SVM	0.641	0.623	0.623	0.623	0.623
kNN	0.487	0.500	0.565	0.493	0.661
Boost ensemble undersampling					
AdaBoost	0.796	0.798	0.785	0.824	0.750
Decision Tree	0.799	0.798	0.789	0.811	0.768
Random Forest	0.887	0.789	0.774	0.820	0.732
SVM	0.826	0.772	0.750	0.812	0.696
kNN	0.494	0.518	0.574	0.507	0.661

5. CONCLUSION

In this paper we studied the classification of CTG signals into hypoxic and normal classes. The main challenge of this task is that the data in practice are highly imbalanced, and straightforward application of ML methods on such data will yield poor performance. To address this issue, we adopted a method known as boot ensemble learning. The approach iteratively allows ML methods to retrain by using in every iteration more informative data samples. We compared the results of different resampling methods for imbalanced data learning. We also showed that by simply random undersampling we might lose potentially informative samples. Instead, by optimizing the undersampling in choosing the balanced training sets, we have much improved performance.

6. REFERENCES

[1] I. Amer-Wåhlin and K. Maršál. St analysis of fetal electrocardiography in labor. *Seminars in Fetal and Neonatal Medicine*, 16(1):29–35, 2011. Fetal and Neonatal Haemodynamics.

[2] V. Bolón-Canedo, N. Sánchez-Maroño, and A. Alonso-

Betanzos. *Feature selection for high-dimensional data*. Springer, 2015.

[3] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002.

[4] M. Christ, N. Braun, J. Neuffer, and A. W. Kempa-Liehr. Time series feature extraction on basis of scalable hypothesis tests (tsfresh – a python package). *Neurocomputing*, 307:72–77, 2018.

[5] V. Chudáček, J. Spilka, M. Burša, P. Janku, L. Hruban, M. Huptych, and L. Lhotská. Open access intrapartum ctg database. *BMC pregnancy and childbirth*, 14(1):1–12, 2014.

[6] B. D. Fulcher and N. S. Jones. Highly comparative feature-based time-series classification. *IEEE Transactions on Knowledge and Data Engineering*, 26(12):3026–3037, 2014.

[7] G. Georgoulas, P. Karvelis, J. Spilka, V. Chudáček, C. D. Stylios, and L. Lhotska. Investigating ph based evaluation of fetal heart rate (fhr) recordings. *Health and technology*, 7(2):241–254, 2017.

[8] 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*, 73:220–239, 2017.

[9] L. Hruban, J. Spilka, V. Chudáček, P. Janku, M. Huptych, M. Burša, A. Hudec, M. Kacerovskỳ, M. Kouckỳ, M. Procházka, et al. Agreement on intrapartum cardiotocogram recordings between expert obstetricians. *Journal of evaluation in clinical practice*, 21(4):694–702, 2015.

[10] 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*, pages 2980–2988, 2017.

[11] X.-Y. Liu, J. Wu, and Z.-H. Zhou. Exploratory undersampling for class-imbalance learning. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 39(2):539–550, 2008.

[12] 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)*, pages 970–974. IEEE, 2006.

[13] 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)*, pages 841–852. IEEE, 2020.

[14] Z. Liu, P. Wei, J. Jiang, W. Cao, J. Bian, and Y. Chang. Mesa: Boost ensemble imbalanced learning with meta-sampler. *Advances in Neural Information Processing Systems*, 33, 2020.

[15] R. Miotto, L. Li, B. A. Kidd, and J. T. Dudley. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. *Scientific reports*, 6(1):1–10, 2016.

[16] M. P. Nageotte. Fetal heart rate monitoring. *Seminars in Fetal and Neonatal Medicine*, 20(3):144–148, 2015.

[17] 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)*, pages 5866–5869. IEEE, 2018.

[18] M. L. Puterman. *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014.

[19] S. Raicevic, D. Cubrilo, S. Arsenijevic, G. Vukcevic, V. Živkovic, M. Vuletic, N. Barudžic, N. Andjelkovic, O. Antonovic, and V. Jakovljevic. Oxidative stress in fetal distress: potential prospects for diagnosis. *Oxidative medicine and cellular longevity*, 3(3):214–218, 2010.

[20] 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.

[21] M. J. Stout and A. G. Cahill. Electronic fetal monitoring: past, present, and future. *Clinics in perinatology*, 38(1):127–142, 2011.

[22] S. Wang and X. Yao. Diversity analysis on imbalanced data sets by using ensemble models. In *2009 IEEE symposium on computational intelligence and data mining*, pages 324–331. IEEE, 2009.

[23] L. Yang, M. Ajirak, C. Heiselman, J. G. Quirk, and P. M. Djurić. Unsupervised detection of anomalies in fetal heart rate tracings using phase space reconstruction. In *2021 29th European Signal Processing Conference (EUSIPCO)*, pages 1321–1325. IEEE, 2021.

[24] 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*, 19(1):1–15, 2019.

[25] Z. Zhao, Y. Zhang, and Y. Deng. A comprehensive feature analysis of the fetal heart rate signal for the intelligent assessment of fetal state. *Journal of clinical medicine*, 7(8):223, 2018.