2023 IEEE International Conference on Digital Health (ICDH) | 979-8-3503-4103-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICDH60066.2023.00045

Predicting Adverse Neonatal Outcomes for Preterm Neonates with Multi-Task Learning

Jingyang Lin[1] Junyu Chen[1] Hanjia Lyu[1] Igor Khodak[2] Divya Chhabra[2]
Colby L Day Richardson[2] Irina Prelipcean[2] Andrew M Dylag[2] Jiebo Luo[1]
[1]University of Rochester [2]University of Rochester Medical Center
{jlin81,jchen175,hlyu5}@ur.rochester.edu jluo@cs.rochester.edu
{igor_khodak,divya_chhabra,colby_richardson,irina_prelipcean,andrew_dylag}@urmc.rochester.edu

***Abstract*—Diagnosis of adverse neonatal outcomes is crucial for preterm survival since it enables doctors to provide timely treatment. Machine learning (ML) algorithms have been demonstrated to be effective in predicting adverse neonatal outcomes. However, most previous ML-based methods have only focused on predicting a single outcome, ignoring the potential correlations between different outcomes, and potentially leading to suboptimal results and overfitting issues. In this work, we first analyze the correlations between three adverse neonatal outcomes and then formulate the diagnosis of multiple neonatal outcomes as a multi-task learning (MTL) problem. We then propose an MTL framework to jointly predict multiple adverse neonatal outcomes. In particular, the MTL framework contains shared hidden layers and multiple task-specific branches. Extensive experiments have been conducted using Electronic Health Records (EHRs) from 121 preterm neonates. Empirical results demonstrate the effectiveness of the MTL framework.**

***Index Terms*—adverse neonatal outcome, preterm neonate, machine learning, multi-task learning, model interpretability**

I. Introduction

The neonatal period, the first four weeks of an infant's life, is a critical stage with high mortality rates [11]. Predictions indicate that approximately 26 million newborns will die between 2019 and 2030 [14], making the issue of high neonatal mortality a global concern. In 2021, one-tenth of US babies were born prematurely[1]. Therefore, prioritizing attention to Extremely Low Gestational Age Newborns (ELGANs) is essential, particularly those born before 28 weeks of gestation. Early diagnosis of *adverse neonatal outcomes* is crucial for survival and treatment preparation. It also helps identify risk factors contributing to high neonatal mortality rates.

Machine learning (ML) is a branch of AI that learns patterns from data without explicit programming [10]. ML model can make predictions on unseen data based on these learned patterns [1]. Innovations [7], [12], [16] have been proposed to predict neonatal outcomes using ML techniques. However, most of them focus on individual outcomes, ignoring correlations among these outcomes, which leads to inefficient usage of data [17]. Therefore, this work aims to **jointly diagnose multiple neonatal outcomes** to improve data efficiency by leveraging outcome correlations.

[1]https://www.cdc.gov/reproductivehealth/features/premature-birth

In this work, we first collect a dataset including 121 preterm neonates from two medical centers and focus on three adverse neonatal outcomes, including severe bronchopulmonary dysplasia (BPD) [9], pulmonary hypertension (PH) diagnosis [6], and discharge weight. The analysis of data correlations revealed the relevance of these outcomes. By considering the correlations across neonatal outcomes, we first formulate the diagnosis of multiple neonatal outcomes as a multi-task learning (MTL) problem and propose an MTL framework. Technically, our proposed framework is based on a Neural Network (NN) model, consisting of shared hidden layers and task-specific hidden layers. The shared layers capture **correlated yet often hidden knowledge among all neonatal outcomes**, while the task-specific layers focus on **unique features of each neonatal outcome**. This approach enables joint learning of multiple outcomes to improve data efficiency.

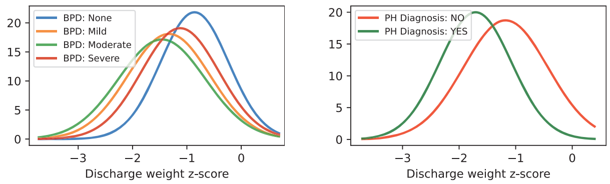
Adverse neonatal outcome prediction experiments are conducted using 121 preterm neonates. Each Electronic Health Record (EHR) consists of 69 input attributes and three primary neonatal outcomes. Five traditional machine learning algorithms are compared. The F1 score and the AUC are used to evaluate algorithms for the two categorical outcomes (severe BPD and PH diagnosis). The MSE is used for the continuous outcome (discharge weight). The results show that each task-specific MTL method outperforms its base single-task model.

Overall, our main contributions are: 1) this work focuses on preterm neonates instead of newborns in general, 2) the proposed MTL framework for jointly predicting multiple adverse neonatal outcomes using limited annotated data, and 3) empirical results verify the effectiveness of the MLT framework.

II. Related Work

Prediction of neonatal outcomes with machine learning models. Mangold *et al.* [12] conducts a systematic review that analyzes existing ML models for predicting neonatal outcomes. Sheikhtaheri *et al.* [16] consider a more practical setting where the experiments are only performed on newborns in neonatal intensive care units (NICUs). Hsu *et al.* [7] leverage ML models to estimate the neonatal mortality rate with respiratory failure. In contrast to prior research [7], [12], [16], we intend to investigate correlations between neonatal outcomes and leverage them to further improve ML models.

979-8-3503-4103-4/23/$31.00 ©2023 IEEE
DOI 10.1109/ICDH60066.2023.00045



(a) BPD vs Discharge Weight (b) PH Diag. vs Discharge Weight

Fig. 1: Correlation among different adverse neonatal outcomes.

Multi-task learning. MTL involves training a single model to perform multiple tasks at the same time [15], [19]. The joint training schema learns shared features that are useful for all tasks, promoting its generalization ability. In the clinical scenario, prior research [4], [8] has explored the effectiveness of multi-task learning. Hu *et al.* [8] employ an MTL framework to screen commercially available and effective inhibitors against SARS-CoV-2. Harutyunyan *et al.* [4] empirically demonstrates that multi-task training acts as a regularizer for almost all tasks. In this work, we first explore the use of MTL in jointly predicting multiple neonatal outcomes for preterm infants.

III. Data Collection and Preprocessing

Data Collection. This study is performed using two combined ELGANs cohorts ($N = 184$) from the PROP and PRISM studies across two medical centers (University of Rochester and University at Buffalo). PROP and PRISM were studies performed under Institutional Review Board (IRB) approval (ClinicalTrials.gov: NCT01435187 and ClinicalTrials.gov: NCT01789268). Patients who died before 36 weeks' Corrected Gestational Age (CGA), have incomplete nutrition or respiratory profiles, or are enrolled in a blinded nutrition study are excluded from analyses. Demographic data is collected to summarize maternal exposures, partner exposures, and other important pregnancy outcomes such as the presence of chorioamnionitis. Resuscitation data is collected from interventions necessary after the infant is born in the delivery room. Once admitted to the neonatal intensive care unit (NICU), nutrition and respiratory variables are collected every day. In summary, we collect Electronic Health Records (EHRs) from 184 preterm neonates. Each neonatal record consists of 431 input attributes and three primary outcomes (i.e., severe BPD, PH diagnosis, and discharge weight).

Data Preprocessing. In our study, we preprocess electronic health records (EHRs) from 184 preterm infants, resulting in 431 input attributes for each neonate. First, **data cleaning** involves removing duplicates, dropping attributes with single or excessive missing values, and filling missing values using the MICE [18]. Next, **data transformation** encompasses converting nominal features to numeric variables, aggregating time-series features, encoding categorical variables, and applying z-score normalization. After preprocessing, the data contain 69 input attributes of 121 neonatal samples. The preprocessed data is then used for further analysis to train and evaluate ML methods. Due to the limited available labeled data in this study, it is critical to investigate techniques such as MTL to improve the data efficiency.

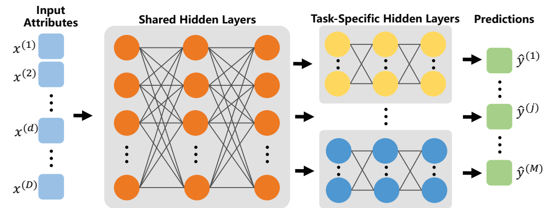


Fig. 2: Overview of multi-task learning framework.

Correlation analysis on adverse neonatal outcomes. MTL allows the model to learn shared features relevant to all tasks, which could improve the performance of each individual task. Therefore, it is necessary to assess the relevance of three neonatal outcomes in our dataset. Specifically, Fig.1 presents the correlations among the outcomes. Fig.1a reveals a negative correlation between **BPD and discharge weight**, indicating that the discharge weights tend to decrease as the severity of BPD increases. Similarly, Fig. 1b illustrates the negative correlation between **PH diagnosis and discharge weight**.

IV. Methodology

Problem Formulation. We formulate the prediction of adverse neonatal outcomes as a multi-task learning problem. The input feature $\boldsymbol{x}_i \in \mathbb{R}^{1\times D}$ consists of information about the parents, delivery room, nutrition, and so on. D denotes the number of input features, and $\boldsymbol{x}_i$ denotes the i-th sample in the dataset. $\boldsymbol{y} = \{\boldsymbol{y}^{(j)}\}_{j=1}^{M}$ refers to the ground truth of different adverse neonatal outcomes, where M indicates the total number of adverse neonatal outcomes and $\boldsymbol{y}^{(j)}$ is the j-th outcome. For the MTL framework, we define the shared hidden layers as $f(\cdot;\theta)$ transferring input features $\boldsymbol{x}$ into latent features $f(\boldsymbol{x};\theta)$, where θ denotes the learnable parameters. Following the shared hidden layers, M task-specific branches $\{g^{(j)}(\cdot;\theta^{(j)})\}_{j=1}^{M}$ transfer the shared latent features into the final predictions $\hat{\boldsymbol{y}}_i$ for different neonatal outcomes:

$$\hat{\boldsymbol{y}}_i = \{g^{(j)}(f(\boldsymbol{x}_i;\theta);\theta^{(j)})\}_{j=1}^{M}, \tag{1}$$

where $\hat{\boldsymbol{y}}_i$ denotes the prediction of the i-th sample, and $\theta^{(j)}$ represents the parameters of the j-th task-specific branch.

Single-Task Learning Baseline. We employ a Neural Network with multiple hidden layers as our backbone. For the classification task, we obtain the prediction scores using a softmax function and then use the cross-entropy loss to optimize the classifier. For the regression task, we use the MSE loss to optimize the regression model.

Multi-Task Learning Framework. Single-task learning framework might ignore the potential correlations between different outcomes, leading to suboptimal results. As shown in Fig. 2, we propose a novel multi-task learning framework to leverage the potential correlation between various adverse neonatal outcomes. Technically, the proposed multi-task learning framework consists of shared hidden layers and multiple task-specific branches. Each task-specific branch contains several hidden layers and a prediction layer. Overall, the proposed MTL framework can exploit the correlations among neonatal outcomes since it is better equipped to capture the general patterns relevant to all tasks.

TABLE I: Overall performance on prediction of adverse neonatal outcomes. The best score for each task is highlighted in bold. The improvements in task-specific MTL methods are shown in red. The results of BPD and PH Diagnosis are in %.

Methods	**BPD** (Task 1)		**PH Diagnosis** (Task 2)		**Discharge Weight** (Task 3)
	F1 ↑	AUC ↑	F1 ↑	AUC ↑	MSE ↓
Logistic Regression [3]	40.2	69.0	28.0	66.4	-
Random Forest [5]	42.4	81.2	44.3	75.3	0.266
XGBoost [2]	31.8	72.4	39.8	**77.2**	0.296
ElasticNet [20]	-	-	-	-	0.267
Neural Network [13] (base model)	43.1	81.2	32.3	75.1	0.279
MLT (tuned for BPD)	**48.0** (↑4.9)	**83.2** (↑2.0)	39.7	68.2	0.288
MLT (tuned for PH Diagnosis)	35.8	65.7	**46.6** (↑14.3)	76.5 (↑1.4)	0.330
MLT (tuned for Discharge Weight)	43.2	79.8	42.0	71.8	**0.257** (↓0.022)

V. Experiment

Dataset. Adverse neonatal outcome prediction experiments are conducted using the data of 121 premature infants. Each sample has 69 input features and three adverse neonatal outcomes. We use 5-fold cross-validation to obtain a more reliable estimate of the ML models' performance.

Evaluation metric. We use the F1 score and Area under the ROC Curve (AUC) of the positive samples to evaluate the performance of the BPD and PH diagnosis classification tasks. We use the mean squared error (MSE) to estimate the discharged weight prediction, which is a regression task.

Comparison with Other Traditional ML Techniques. Table I summarizes the quantitative results on three tasks about the adverse neonatal outcomes. We compare our proposed multi-task learning framework with several traditional machine learning frameworks. Overall, the results across different tasks show that our proposed MTL framework outperforms other traditional machine learning methods. Note that Neural Network (NN) is a degraded version of our proposed multi-task learning framework. The comparison between the NN and different task-specific MTL methods suggests that MTL methods can significantly outperform their base model. The improvements indicate that MTL can boost the performance of the primary task by leveraging the other relevant auxiliary tasks with extra information for learning the primary task. Therefore, data efficiency highlights the necessity of exploring MTL techniques, especially for small datasets.

VI. Conclusion

This paper investigates multi-task learning for predicting multiple adverse neonatal outcomes. Unlike previous ML-based methods that formulate this problem as a single-task problem, our proposed MTL framework jointly learns several relevant tasks to capture the shared general patterns across various tasks. Our proposed framework incorporates shared hidden layers and task-specific branches, effectively integrating general patterns into ML models. Empirical results verify the effectiveness of our MTL method with limited data. We hope our study will draw the community's attention to the diagnosis of multiple neonatal outcomes for premature infants.

Acknowledgments

The study was supported in part by K08HL155491 (AM Dylag), The Prematurity and Respiratory Outcomes Program (PROP): National Institutes of Health, NHLBI and NICHD through U01 HL101813 to the University of Rochester and University at Buffalo (GS Pryhuber). NIH/NIAID HHSN272201200005C University of Rochester Respiratory Pathogens Research Center (URRPRC).

References

[1] Harith Al-Sahaf, Ying Bi, Qi Chen, et al. A survey on evolutionary machine learning. *Journal of the Royal Society of New Zealand*, 2019.
[2] Tianqi Chen and Carlos Guestrin. XGBoost: A scalable tree boosting system. In *KDD*, 2016.
[3] David R Cox. The regression analysis of binary sequences. *Journal of the Royal Statistical Society: Series B (Methodological)*, 1958.
[4] Hrayr Harutyunyan, Hrant Khachatrian, David C Kale, et al. Multitask learning and benchmarking with clinical time series data. *Scientific data*, 2019.
[5] Tin Kam Ho. Random decision forests. In *ICDAR*, 1995.
[6] Marius M Hoeper, Harm Jan Bogaard, Robin Condliffe, et al. Definitions and diagnosis of pulmonary hypertension. *Journal of the American College of Cardiology*, 2013.
[7] Jen-Fu Hsu, Chi Yang, Chun-Yuan Lin, et al. Machine learning algorithms to predict mortality of neonates on mechanical intubation for respiratory failure. *Biomedicines*, 2021.
[8] Fan Hu, Jiaxin Jiang, and Peng Yin. Prediction of potential commercially available inhibitors against sars-cov-2 by multi-task deep learning model. *Biomolecules*, 2022.
[9] Alan H Jobe and Eduardo Bancalari. Bronchopulmonary dysplasia. *American journal of respiratory and critical care medicine*, pages 1723–1729, 2001.
[10] Michael I Jordan and Tom M Mitchell. Machine learning: Trends, perspectives, and prospects. *Science*, 2015.
[11] Parviz Karimi, Leily Mahmudi, Milad Azami, and Gholamreza Badfar. Mortality in neonatal intensive care units in iran: a systematic review and meta-analysis. *Iranian Journal of Neonatology IJN*, 2019.
[12] Cheyenne Mangold, Sarah Zoretic, Keerthi Thallapureddy, et al. Machine learning models for predicting neonatal mortality: a systematic review. *Neonatology*, 2021.
[13] Warren S McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 1943.
[14] Aramesh Rezaeian, Marzieh Rezaeian, Seyede Fatemeh Khatami, et al. Prediction of mortality of premature neonates using neural network and logistic regression. *Journal of Ambient Intelligence and Humanized Computing*, 2020.
[15] Sebastian Ruder. An overview of multi-task learning in deep neural networks. *arXiv:1706.05098*, 2017.
[16] Abbas Sheikhtaheri, Mohammad Reza Zarkesh, Raheleh Moradi, and Farzaneh Kermani. Prediction of neonatal deaths in nicus: development and validation of machine learning models. *BMC medical informatics and decision making*, 2021.
[17] Trevor Standley, Amir Zamir, Dawn Chen, Leonidas Guibas, Jitendra Malik, and Silvio Savarese. Which tasks should be learned together in multi-task learning? In *ICML*, 2020.
[18] Stef Van Buuren and Karin Groothuis-Oudshoorn. mice: Multivariate imputation by chained equations in r. *Journal of statistical software*, 2011.
[19] Yu Zhang and Qiang Yang. A survey on multi-task learning. *IEEE Transactions on Knowledge and Data Engineering*, 2021.
[20] Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, 2005.