Valerie Pfannschmidt*, Yaren Aydin, Mark Schoberer, and André Stollenwerk

Detection of Negative Arterial to End-Tidal CO_2 Gradients in Mechanically Ventilated Neonatal Lambs

https://doi.org/10.1515/cebme-2025-0190

Abstract: The end-tidal partial pressure of CO_2 ($PetCO_2$) is widely used as a noninvasive and continuously measurable surrogate for the arterial partial pressure of CO_2 ($PaCO_2$). The arterial to end-tidal CO_2 gradient (ΔpCO_2(a-et)), which is usually positive, changes over time and may become negative. Hence, estimating $PaCO_2$ from $PetCO_2$ is challenging. Yet, knowledge on the prevalence of negative ΔpCO_2(a-et) is limited. We analyze a dataset of 63 mechanically ventilated preterm lambs on the occurrence of negative ΔpCO_2(a-et). Among 1,182 paired measurements of $PaCO_2$ and $PetCO_2$, 185 instances of negative ΔpCO_2(a-et) are found, which are distributed among 36 subjects. Further, we present three classifiers for detecting negative ΔpCO_2(a-et) on a balanced subdataset based on noninvasive measurements of routine monitoring. The classifiers achieve promising results and when validated on human data can contribute to robustly estimating $PaCO_2$.

Keywords: Binary Classification, Machine Learning, Noninvasive Monitoring, Neonates

1 Introduction

The arterial partial pressure of CO_2 ($PaCO_2$) is the core parameter to assess adequacy of CO_2 elimination under mechanical ventilation. Invasive arterial blood gas analysis (ABG) is the 'gold standard' to determine $PaCO_2$. In preterm neonates $PaCO_2$ is of particular importance, because fluctuations and extremes of $PaCO_2$ are associated with severe intraventricular hemorrhage [1]. However, arterial blood gas samples allows only intermittent assessment of $PaCO_2$. As a continuous and noninvasive surrogate for $PaCO_2$ the end-tidal partial pressure of CO_2 ($PetCO_2$) can be used. $PetCO_2$ corresponds to the partial pressure of CO_2 in the breathing gas at the end of expiration and can be derived on a breath-by-breath basis from continuous capnometry. $PetCO_2$ has been found to have good correlation with $PaCO_2$ in neonates with healthy lungs but the correlation weakens with lung diseases [2]. Having available a noninvasive and reliable surrogate for $PaCO_2$ would be of great value for individual clinical monitoring and adjustment of mechanical ventilation not only in neonates, but in all patient groups. Further, this would contribute to robust physiological closed-loop control of mechanical ventilation.

$PetCO_2$ is correlating with $PaCO_2$ because the expired air at the end of expiration is mostly consisting of alveolar air, which is assumed to be equilibrated with the arterial blood. Yet, there remains a difference between $PaCO_2$ and $PetCO_2$, often called the arterial to end-tidal CO_2 gradient, which will be denoted with ΔpCO_2(a-et) in this paper. Typically, ΔpCO_2(a-et) is expected to be positive, i.e., $PaCO_2$ being higher than $PetCO_2$, due to air from deadspace regions of the lungs mixing with the alveolar air. However, ΔpCO_2(a-et) can also be negative, i.e., measured $PaCO_2$ is found to be lower than measured $PetCO_2$. Observations of negative ΔpCO_2(a-et) have been reported in healthy adults, especially during exercise and pregnancy, and in children and neonates [3]. Knowledge of changes in the sign of ΔpCO_2(a-et) would be a helpful for achieving robust $PetCO_2$ based estimation of $PaCO_2$. For preterm neonates, this would directly contribute to detecting unsafe values of $PaCO_2$. Still, knowledge about the prevalence of negative ΔpCO_2(a-et) in mechanically ventilated (preterm) neonates is currently limited to few studies [2, 4–6] and challenged by the usually small number of subjects in clinical studies involving preterm neonates.

This work contributes to solving the above mentioned challenges in two ways. First, we present additional knowledge on the prevalence of negative ΔpCO_2(a-et) by analyzing data from mechanically ventilated preterm lambs, an established model for preterm neonates [7]. Second, we propose three classifiers for detecting negative ΔpCO_2(a-et) from data of noninvasive routine monitoring without knowledge of $PaCO_2$.

***Corresponding author: Valerie Pfannschmidt,** Chair of Embedded Software (Informatik 11), RWTH Aachen University, Aachen, Germany, e-mail: pfannschmidt@embedded.rwth-aachen.de.de
Yaren Aydin, André Stollenwerk, Chair of Embedded Software (Informatik 11), RWTH Aachen University
Mark Schoberer, Neonatology Section of the Department of Paediatric and Adolescent Medicine, RWTH Aachen University Hospital, Aachen, Germany

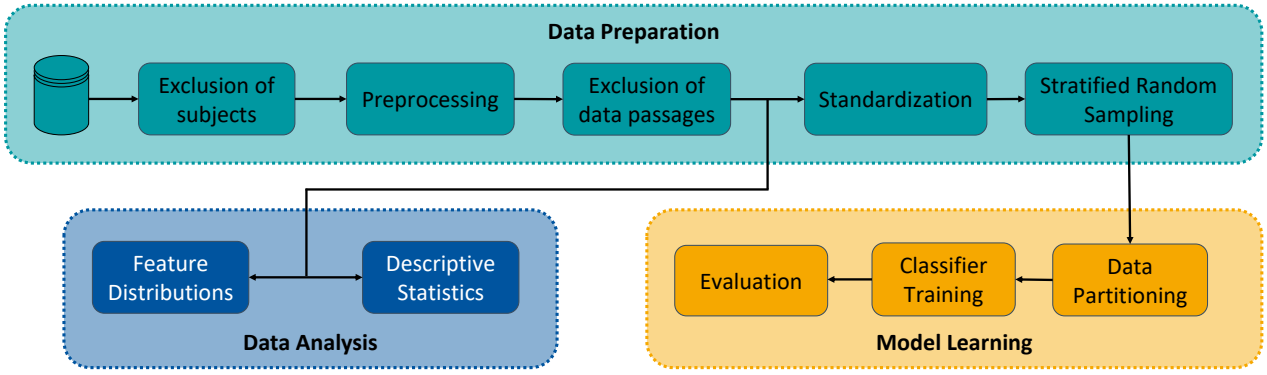

Fig. 1: Data processing pipeline used in this work. The turquoise shaded steps involve data preparation, while the blue shaded steps represent general data analysis, and the orange shaded steps correspond to classifier learning and evaluation.

2 Materials and Methods

2.1 Data Base

The data base includes 650 hours of ventilation data recorded from 63 preterm lambs on the first day of life. The primary study involved evaluation of a $PaCO_2$ controller for neonates and parts of the data have been published in our corresponding work [8, 9]. Mechanical ventilation was adjusted to the subjects either by neonatologists or an automatic ventilation controller. Maneuvers for evaluating the controller involved for example the induction of hyper- and hypoventilation as shown in [8]. These passages were purposefully kept in this work to include a wide range of $PaCO_2$ values.

2.2 Ventilation Monitoring and Data Collection

Ventilation was provided by a LeoniPlus ventilator (Löwenstein Medical SE & Co. KG, Bad Ems, Germany) with integrated SpO_2 monitoring. Ventilation monitoring data including respiratory rate (RR), expiratory tidal volume (V_{te}), inspiratory fraction of oxygen (FiO_2), positive end-expiratory pressure (PEEP), and mean airway pressure (P_{mean}) were provided by the ventilator. The fraction of CO_2 in the breathing gas was measured with a MASIMO IRMA CO2 (Masimo Corporation, Irvine, USA) proximal mainstream sensor at 50 Hz. $PetCO_2$ was determined from the raw data as shown in [9]. Blood samples for ABGs were taken manually from an umbilical arterial catheter and analyzed with an i-Stat 1 (Abbott Laboratories, Chicago, USA) handheld bloodgas analyzer. The ABGs were documented manually, whereas all other data were automatically logged by a custom application [10]. ABGs were made on a regular basis every 30 minutes, which allows to closely evaluated the evolution of ΔpCO_2(a-et) on the first day of life.

2.3 Data Preparation and Analysis

The data processing pipeline used in this work is shown in Figure 1. Of the 63 lambs in the database, three lambs were excluded because of incomplete data following technical problems. Further, two lambs had to be excluded because of severe health complications. The data of the resulting 58 lambs were then preprocessed which involved smoothing and filling of missing data with a moving mean over 60 seconds. Remaining data gaps of up to five minutes were filled with forward filling. In the next step, data gaps unable to be filled (i.e., gaps of larger time intervals than 5 minutes) were excluded. In addition, data passages were excluded if $V_{te}$was below 12 ml as this was identified as a lower threshold for reliable CO_2 measurement in our ventilation setup. $V_{te}$was then normalized by the individual animal bodyweight per kilogram. The resulting dataset contained 1,182 ABGs and will be denoted as the *main dataset*. The main dataset was used for general data analysis and feature engineering presented in Section 3.1.

As the main dataset was found to be strongly imbalanced a balanced dataset with the same number of observations of positive and negative ΔpCO_2(a-et) was created for classifier training and evaluation. First, the main dataset was standardized for improved classifier learning. Then, stratified random sampling from the main dataset was performed. The resulting dataset includes 370 ABGs with corresponding noninvasive data, i.e., 185 cases of positive and negative ΔpCO_2(a-et) each. This smaller data will further be denoted as the *balanced dataset*. Because of the small size of the balanced dataset stratified 10-fold cross validation was used for model training and validation to reduce the influence of selection bias.

As performance metrics accuracy, precision, recall, and the F1 Score are used. Accuracy describes the proportion of correct classifications made by a model among all instance to be classified. Precision measures the proportion of true pos-

itive cases among all positively classified instances. Recall, also named sensitivity or true positive rate, denotes the proportion of true positive classifications among all actual positive instances. Finally, the F1 Score is the harmonic mean of precision and recall and thus helps, to get an impression of the balance between both metrics. All four metrics take values between zero (worst) and one (best).

3 Results

3.1 Data Analysis

In the main dataset of 1,182 paired measurements of $PaCO_2$ from ABGs and $PetCO_2$ 185 instances of negative ΔpCO_2(a-et) were found, corresponding to 15 % of measurement pairs. Negative ΔpCO_2(a-et) were found in 37 out of 58 animals (64 % of subjects), of which 2 animals accounted for more than 10 % of these events. In 36 of 58 animals both positive and negative ΔpCO_2(a-et) were observed, and in one animal only negative ΔpCO_2(a-et) occurred. In 35 animals at least two changes in the sign of ΔpCO_2(a-et) were found. The sign of ΔpCO_2(a-et) changed twice in 15 of these animals, 3-4 times in 10 of these animals and between 5 and 13 times in the other 10 animals. Among all measurements the mean ΔpCO_2(a-et) was 5.7 $\pm$ 7.3 mmHg. Among the cases of negativ ΔpCO_2(a-et) the mean ΔpCO_2(a-et) was -4.1 $\pm$ 5.0 mmHg. Fig. 2 shows the feature distribution of four selected features, showing that negative ΔpCO_2(a-et) was found more often at high $PetCO_2$, low RR and low P_{mean}. Although these trends are qualitatively visible, the Pearson correlation for all of these features is weak with r_{PetCO2} = 0.22, r_{RR} = -0.32, and r_{Pmean} = -0.26. We did not find a correlation between negative ΔpCO_2(a-et) and the magnitude of V_{te}normalized to body weight or absolute expired or inspired tidal volume.

3.2 Classification

As shown in the previous section, negative ΔpCO_2(a-et) and several changes between positive and negative ΔpCO_2(a-et) are no seldom phenomenon in our main data set. Therefore, detecting negative ΔpCO_2(a-et) on noninvasive measurements of routine monitoring may be helpful for individual patient monitoring and robust estimation of $PaCO_2$ from $PetCO_2$. We present results of three classifiers for binary classification of ΔpCO_2(a-et) . A logistic regression, random forest and gradient boosting model were trained to detect negative ΔpCO_2(a-et) in our balanced dataset. As feature, i.a., $PetCO_2$, RR, V_{te}/bodyweight and P_{mean} were used. The classification performance on the validation data of the balanced dataset from 10-fold cross validation is shown in Fig. 3. The three models show a relatively balanced and comparable performance on all

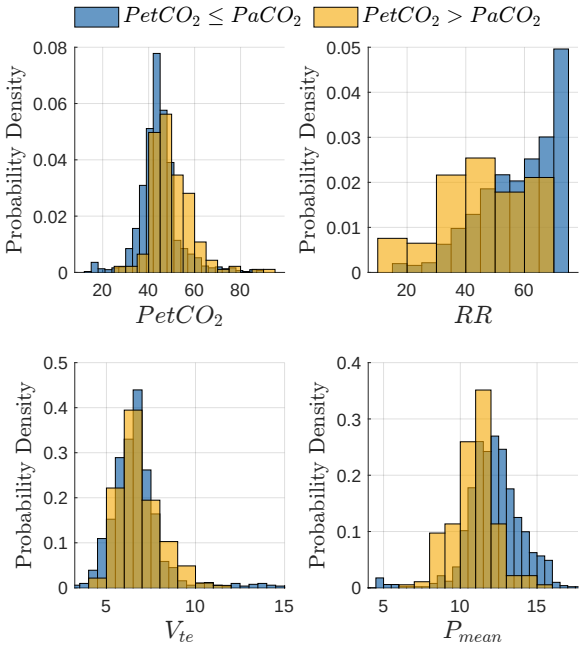

Fig. 2: Feature distribution of four selected features. $PetCO_2$ is given in mmHg, RR in 1/min, V_{te}in ml/kg bodyweight, and P_{mean} in mbar.

metrics. Whereas the Logistic Regression and Gradient Boosting Model have a mean value between 0.78 - 0.82 among all metrics, the Random Forest model has superior precision with P_{RF} = 0.86, but weaker recall with R_{RF} = 0.77. The Random Forest model further has the best accuracy of 0.83. Among all metrices, the variability of performance between the different runs of cross validation is a lot larger than the performance differences between the single models.

4 Discussion

In our main dataset the instances of negative ΔpCO_2(a-et) accounted for 15 % of all 1,182 measurements, which were distributed among 64 % of subjects. [5] report a prevalence of negative ΔpCO_2(a-et) of 11.8 % among 754 measurements from 32 preterm neonates of extremely low birth weight (< 1000 g). [6] report 6 % of negative ΔpCO_2(a-et) in 143 measurements of 45 preterm neonates of at least very low birth weight (< 1,500 g). [2] found negative ΔpCO_2(a-et) in 13.5 % of 133 paired measurements from 32 ventilated neonates of varying birth weight (840 - 3,500 g). In all three studies, blood sampling was done through an arterial catheter and $PetCO_2$ taken using mainstream capnography, as it was in our study. Our findings align with these known work, although being at the upper end of reported negative ΔpCO_2(a-et) prevalence.

Whereas our dataset is of reasonable size in terms of a clinical data set, it is small in terms of a base for training ma-

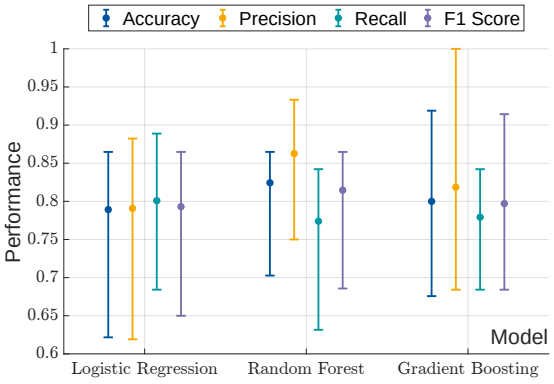

Fig. 3: Classification results from 10-fold cross validation. The span of performances among the 10 different evaluation sets is represented with the whiskers, while each mean is marked with a circle.

chine learning algorithms. Still, all three presented classifiers achieved a good mean accuracy of around 0.8. The most important metric for our application is recall, as the aim of our classification is to detect a specific condition, to alert a physician (who could then potentially take a blood probe for confirmation). With a recall of around 0.78 the classifiers are able to detect a substantial amount of negative ΔpCO_2(a-et) cases. However, around 20 % of instances of negative ΔpCO_2(a-et) remain undetected. As the three models do not differ much in their performance, we believe, the remaining performance gap can rather be attributed to availability of underlying information in the dataset, then the type of model. Unfortunately, we did not find a single feature with a moderate or strong correlation with negative ΔpCO_2(a-et) among the measurements of routine monitoring. Future work will therefore involve further feature analysis to determine, if this finding is due to physiological complexity or can be solved by advanced feature engineering and data processing.

Although the preterm lamb model is well-established for preterm pulmonary pathology and treatment [7], validation of our results on human data is needed.

5 Conclusion

This work indicates that negative ΔpCO_2(a-et) may be a common phenomenon in mechanically ventilated neonates, which should be considered when estimating $PaCO_2$ from $PetCO_2$. Further studies on data of human subjects are needed to confirm our findings on the prevalence of negative ΔpCO_2(a-et) in preterm neonates and how these reflect in monitoring data. The presented classification models can successfully detect the occurrence of negative ΔpCO_2(a-et) on data from noninvasive routine monitoring. In the future, the accuracy of the models may be further enhanced by advancing feature engineering. When validated on human data the presented classifiers can contribute to robustly estimating $PaCO_2$ noninvesively.

Author Statement

Research funding and ethical approval: The animal trials which were retrospectively analyzed in this work were supported by the German Federal Ministry of Education and Research (BMBF) (Grant 13GW0292C) and ethically approved by the Dutch Central Authority for Scientific Procedures on Animals (AVD10700202010347). Conflict of interest: the authors state not conflict of interest.

References

[1] J. Fabres, W. A. Carlo, V. Phillips, *et al.*, "Both extremes of arterial carbon dioxide pressure and the magnitude of fluctuations in arterial carbon dioxide pressure are associated with severe intraventricular hemorrhage in preterm infants," *Pediatrics*, vol. 119, no. 2, pp. 299–305, 2007.

[2] Y. R. Bhat and N. Abhishek, "Mainstream end-tidal carbon dioxide monitoring in ventilated neonates," *Singapore medical journal*, vol. 49, no. 3, pp. 199–203, 2008.

[3] K. Bhavani-Shankar, "Negative arterial to end-tidal co2 gradients in children," *Canadian Journal of Anaesthesia*, vol. 41, no. 11, pp. 1125–1126, 1994.

[4] G. F. Rich and J. M. Sconzo, "Continuous end-tidal co2 sampling within the proximal endotracheal tube estimates arterial co2 tension in infants," *Canadian Journal of Anaesthesia*, vol. 38, no. 2, pp. 201–203, 1991.

[5] S. Amuchou Singh and N. Singhal, "Dose end-tidal carbon dioxide measurement correlate with arterial carbon dioxide in extremely low birth weight infants in the first week of life?," *Indian pediatrics*, vol. 43, no. 1, pp. 20–25, 2006.

[6] D. Trevisanuto, S. Giuliotto, F. Cavallin, N. Doglioni, S. Toniazzo, and V. Zanardo, "End-tidal carbon dioxide monitoring in very low birth weight infants: correlation and agreement with arterial carbon dioxide," *Pediatric pulmonology*, vol. 47, no. 4, pp. 367–372, 2012.

[7] R. de Matteo, N. Blasch, V. Stokes, P. Davis, and R. Harding, "Induced preterm birth in sheep: a suitable model for studying the developmental effects of moderately preterm birth," *Reproductive Sciences*, vol. 17, no. 8, pp. 724–733, 2010.

[8] M. Buglowski, V. Pfannschmidt, S. Becker, *et al.*, "Closed-loop control of arterial co2 in mechanical ventilation of neonates," in *2022 44th Annual Intl Conf of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2022.

[9] V. Pfannschmidt, M. Buglowski, Hütten, *et al.*, "Closed-loop control of arterial co2 for neonatal mechanical ventilation: In-vivo interaction with spontaneous breathing," *IFAC-PapersOnLine*, vol. 58, no. 24, pp. 281–286, 2024.

[10] V. Pfannschmidt, M. Buglowski, M. Grüne, *et al.*, "An interconnected modular setup for in-vivo evaluation of automated mechanical ventilation," in *AUTOMED 2024 Proceedings, 29-30, 2024*.