Intensive Care Med
https://doi.org/10.1007/s00134-025-08025-6

SCIENTIFIC LETTER

Early prediction of non-invasive ventilation outcome using the TabPFN machine learning model: a multi-centre validation study

Hang Yu[1], Sina Saffaran[1], Israel S. Maia[2], Enrico Clini[3,4*] and Declan G. Bates[1] on behalf of the NIVPredict study group

Dear Editor,

Patients with acute hypoxemic respiratory failure (AHRF) who fail non-invasive ventilation (NIV) and require intubation have higher risk of mortality [1, 2]. Identifying these patients early is challenging, as no formal guidelines are currently available. Thresholds based on clinical scores or indices have been proposed to help identify patients who are at higher risk of failure, but uncertainty exists regarding their optimal cut-off values and their effectiveness in prompting treatment escalation is low [3].

Machine learning (ML) models could help to find more complex thresholds to identify earlier those patients who may require closer monitoring or further interventions. However, databases on patients with AHRF often lack granularity or scale in the case of NIV [2]. This complicates the deployment of standard ML models which have been developed for datasets involving many thousands of datapoints. We developed (Fig.E1, SM) a model to predict the outcome of NIV soon after treatment initiation, using a recently proposed ML model called tabular prior-data fitted network (TabPFN) [4]. TabPFN couples a unique pre-training strategy with in-context learning during inference for rapid, accurate predictions without hyperparameter tuning. In-context learning is particularly suitable for small datasets, as pre-training facilitates effective generalization from limited examples. This approach underlies the unprecedented performance of large-language models, where inclusion of a few examples can dramatically improve performance on unseen tasks.

We trained the model using routine measurements made before (T0) and one hour after (T1) initiation of NIV treatment in 624 AHRF patients from the RENOVATE trial in Brazil [5]. External validation was performed on a dataset on 368 AHRF patients from Italy, Spain and the US—see Table E1 for a complete description.

The TabPFN model achieved 74% predictive accuracy, 72% sensitivity, 73% specificity, 68% PPV, 76% NPV, and 0.78 AUC in repeated five-fold cross-validation on the training dataset (Table 1). On the external validation dataset it achieved 71% accuracy, 73% sensitivity, 69% specificity, 68% PPV, 74% NPV, and 0.76 AUC. A decision curve analysis (Fig. E5) showed that treatment escalation decisions guided by TabPFN provided a greater net benefit than default strategies across a wide range of decision thresholds. Calibration curves (Fig. E5) indicate that TabPFN was well-calibrated, with predicted risks closely matching actual outcomes.

The best performing standard ML model in external validation was XGBoost (68% accuracy, 66% sensitivity, 70% specificity and 0.71 AUC) while clinical indices such as HACOR were less accurate and more unbalanced predictors (60% accuracy, 77% sensitivity, 46% specificity, and 0.67 AUC)–Table 1.

The TabPFN model used the following measurements: PaO_2/FiO_2(T1), RR(T1), PaO_2/FiO_2(T0), pH(ΔT0-T1), FiO_2(ΔT0-T1), SAPSII(T0), $PaCO_2$(ΔT0-T1), PaO_2/FiO_2(ΔT0-T1), PEEP + PSV(T1), PEEP(T1), RR(ΔT0-T1). PSV was not used as including PEEP and PEEP + PSV

*Correspondence: enrico.clini@unimore.it
[3] Department of Medical and Surgical Sciences of Adult and Mother-Child SMECHIMAI, University of Modena Reggio-Emilia, Modena, Italy
Full author information is available at the end of the article

Table 1 Predictive performance metrics of ML models and clinical indices across datasets and time points

Model/indices	AUC	Accuracy	Sensitivity	Specificity	PPV	NPV
ML Models						
TabPFN (Training)	0.78 [0.71–0.83]	74% [69–77%]	72% [65–82%]	73% [67–78%]	68% [61%, 75%]	76% [68–84%]
TabPFN (Validation)	0.76	71%	73%	69%	68%	74%
XGboost (Training)	0.71 [0.63–0.79]	70% [64–77%]	62% [53–72%]	75% [64–83%]	56% [44–68%]	75% [71–80%]
XGBoost (Validation)	0.71	68%	66%	70%	65%	70%
Integrative Clinical Scores						
HACOR (T0) (Training)	0.62 [0.47–0.75]	60% [47–71%]	66% [35–90%]	55% [36–77%]	49% [37–62%]	72% [59–88%]
HACOR (T0) (Validation)	0.57	56%	65%	49%	51%	64%
HACOR (T1) (Training)	0.69 [0.53–0.82]	66% [54–78%]	71% [50–90%]	62% [45–77%]	55% [44–70%]	77% [64–91%]
HACOR (T1) (Validation)	0.67	60%	77%	46%	53%	71%
SAPSII (Training)	0.61 [0.47–0.72]	61% [49–68%]	55% [46–68%]	61% [47–74%]	51% [39–61%]	68% [59–78%]
SAPSII (Validation)	0.66	65%	52%	76%	64%	66%
Clinical Parameters						
PaO_2/FiO_2 (T0) (Training)	0.62 [0.47–0.76]	59% [45–70%]	68% [45–90%]	52% [36–68%]	48% [37–59%]	72% [58–87%]
PaO_2/FiO_2 (T0) (Validation)	0.60	50%	92%	12%	48%	64%
PaO_2/FiO_2 (T1) (Training)	0.67 [0.53–0.81]	61% [49–74%]	74% [55–90%]	52% [36–70%]	50% [40–62%]	76% [62–91%]
PaO_2/FiO_2 (T1) (Validation)	0.69	60%	81%	42%	55%	72%
RR (T0) (Training)	0.54 [0.40–0.69]	54% [41–67%]	52% [35–70%]	55% [39–71%]	43% [30–57%]	64% [52–75%]
RR (T0) (Validation)	0.49	48.6%	60%	39%	46%	52%
RR (T1) (Training)	0.61 [0.45–0.75]	60% [47–73%]	55% [35–75%]	63% [48–77%]	50% [35–65%]	69% [57–82%]
RR (T1) (Validation)	0.56	54%	52%	56%	51%	57%
ΔT0-T1 pH (Training)	0.58 [0.43–0.74]	57% [43–70%]	49% [25–85%]	62% [29–81%]	45% [29–63%]	66% [54–80%]
ΔT0-T1 pH (Validation)	0.54	50%	45%	54%	46%	53%

Internal Training Cohort: 624 patients (NIV success: 378, NIV failure: 246). External Validation Cohort: 368 patients (NIV success: 196, NIV failure: 172). For the training cohort, we performed repeated fivefold cross-validation. Results are reported as the mean and 95% confidence intervals, calculated over 200 iterations. Accuracy = # correctly predicted NIV outcomes / # patients. Sensitivity = # correctly predicted NIV failures / # NIV failures. Specificity = # correctly predicted NIV successes / # NIV successes. Area Under the Curve (AUC) = area under the Receiver Operating Characteristic (ROC) curve. Positive predictive value (PPV) = # correctly predicted NIV failures / # predicted NIV failures. Negative predictive value (NPV) = # correctly predicted NIV successes / # predicted NIV successes

Abbreviations: *T0* Baseline measurements (within 6 h prior to NIV initiation), *T1* Measurements taken 1–2 h after NIV initiation, *SAPSII* Simplified Acute Physiology Score II, *HACOR* A clinical index incorporating heart rate, acidosis, consciousness level, oxygenation, and respiratory rate, *PaO_2/FiO_2* ratio of arterial oxygen partial pressure to inspired oxygen fraction, *RR* respiratory rate, *ΔT0–T1 pH* change in pH between T0 and T1

as features allowed the model to capture both ventilatory support levels. PaO_2/FiO_2 was uniformly the most important feature in making accurate predictions (Figs. E6/E7). The increases in FiO_2 and low PaO_2/FiO_2 after 1–2 h were associated with predictions of NIV failure (Fig. E6(a)). For (many) more details, see the SM.

The performance gap between the TabPFN model and standard ML models and clinical indices is striking and suggests that with further prospective validation this new ML approach could help clinicians promptly identify patients at risk of failing NIV.

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1007/s00134-025-08025-6.

Author details
[1] School of Engineering, University of Warwick, Coventry, UK. [2] Hcor Research Institute, Hcor Hospital, Rua Desembargador Eliseu Guilherme, São Paulo, Brazil. [3] Department of Medical and Surgical Sciences of Adult and Mother-Child SMECHIMAI, University of Modena Reggio-Emilia, Modena, Italy. [4] Respiratory Diseases Unit, University Hospital of Modena Policlinico, Modena, Italy.

Acknowledgements
NIVPredict study group: Liam Weaver, PhD, School of Engineering, University of Warwick, Coventry, UK; Roberto Tonelli, MD, Department of Medical and Surgical Sciences of Adult and Mother-Child SMECHIMAI, University of Modena Reggio-Emilia, Modena, Italy, University Hospital of Modena Policlinico, Respiratory Diseases Unit, Modena, Italy; Luca S. Menga, MD, Department of Emergency, Intensive Care Medicine and Anesthesia, Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy, Istituto di Anestesiologia e Rianimazione, Universita Cattolica del Sacro Cuore, Rome, Italy, Keenan Research Centre, Li Ka Shing Knowledge Institute, St Michael's Hospital, Unity Health Toronto, Toronto, Canada, Division of Critical Care Medicine, University of Toronto, Toronto, Canada; Qingchen Zhang, PhD, School of Computer Science and Technology, Hainan University, Haikou, China; Moein Einollahzadeh Samadi, M.Sc, Institute for Computational Biomedicine, University Hospital RWTH Aachen, Germany; Andreas Schuppert, PhD, Institute for Computational Biomedicine, University Hospital RWTH Aachen, Germany; John G. Laffey, MD, Anaesthesia and Intensive Care Medicine, Galway University Hospitals, Galway, Ireland, Anaesthesia and Intensive Care Medicine, School of Medicine, University of Galway, Galway, Ireland; Luigi Camporota, MD, Intensive Care Medicine, Guy's and St Thomas' NHS Foundation Trust, London, UK, Division of Asthma Allergy and Lung Biology, King's College London, London, UK; Timothy E. Scott, MD, NIV Critical Care & Regional Weaning Centre, University Hospital

North Midlands NHS Trust, Stoke-on-Trent, UK; Abdisamad Ali, MD, NIV Critical Care & Regional Weaning Centre, University Hospital North Midlands NHS Trust, Stoke-on-Trent, UK; Antonio M. Esquinas, MD, Intensive Care Unit, Hospital Morales Meseguer, Murcia, Spain; Domenico L. Grieco, MD, Department of Emergency, Intensive Care Medicine and Anesthesia, Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy, Istituto di Anestesiologia e Rianimazione, Universita Cattolica del Sacro Cuore, Rome, Italy; Massimo Antonelli, MD, Department of Emergency, Intensive Care Medicine and Anesthesia, Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy, Istituto di Anestesiologia e Rianimazione, Universita Cattolica del Sacro Cuore, Rome, Italy; Lucas Martins de Lima, MD, Hcor Research Institute, Hcor Hospital, Rua Desembargador Eliseu Guilherme, São Paulo, Brazil; Letícia Kawano-Dourado, MD, Hcor Research Institute, Hcor Hospital, Rua Desembargador Eliseu Guilherme, São Paulo, Brazil; Alexandre Biasi Cavalcanti, MD, Hcor Research Institute, Hcor Hospital, Rua Desembargador Eliseu Guilherme, São Paulo, Brazil.

Funding
This letter was supported by the UKRI Engineering and Physical Sciences Research Council (Refs. EP/W000490/1, EP/Y003527/1) and The Royal Academy of Engineering (Ref. RF2122-21-258).

Data availability
Patient data used for internal validation and machine learning model development are available upon request to bona fide researchers for specific scientific purposes, subject to approval by the RENOVATE investigators. Data used for external validation include both publicly accessible datasets and datasets that are available upon request for specified scientific purposes from the corresponding author. Publicly available data include deidentified patient records from the Medical Information Mart for Intensive Care IV (MIMIC-IV) v2.2 database, accessible at https://physionet.org/content/mimiciv/2.2/.

Declarations

Conflicts of interest
All authors declare that they have no competing interests.

Publisher's Note

Accepted: 28 June 2025
Published online: 10 July 2025

References

1. Ferreyro BL, Angriman F, Munshi L, Del Sorbo L, Ferguson ND, Rochwerg B, Ryu MJ, Saskin R, Wunsch H, da Costa BR, Scales DC (2020) Association of noninvasive oxygenation strategies with all-cause mortality in adults with acute hypoxemic respiratory failure: a systematic review and meta-analysis. JAMA 324(1):57–67
2. Bellani G, Laffey JG, Pham T, Madotto F, Fan E, Brochard L, Esteban A, Gattinoni L, Bumbasirevic V, Piquilloud L, van Haren F, Larsson A, McAuley DF, Bauer PR, Arabi YM, Ranieri M, Antonelli M, Rubenfeld GD, Thompson BT, Wrigge H, Slutsky AS, Pesenti A (2017) Noninvasive ventilation of patients with acute respiratory distress syndrome: insights from the LUNG SAFE study. Am J Respir Crit Care Med 195(1):67–77
3. Yarnell CJ, Johnson A, Dam T, Jonkman A, Liu K, Wunsch H, Brochard L, Celi LA, De Grooth HJ, Elbers P, Mehta S, Munshi L, Fowler RA, Sung L, Tomlinson G (2023) Do thresholds for invasive ventilation in hypoxemic respiratory failure exist? A cohort study. Am J Respir Crit Care Med 207(3):271–282
4. Hollmann N, Müller S, Purucker L, Krishnakumar A, Körfer M, Hoo SB, Schirrmeister RT, Hutter F (2025) Accurate predictions on small data with a tabular foundation model. Nature 637(8045):319–326
5. RENOVATE Investigators and the BRICNet Authors (2025) High-flow nasal oxygen vs noninvasive ventilation in patients with acute respiratory failure: the RENOVATE randomized clinical trial. JAMA 333(10):875–890