Machine learning prediction of non-invasive ventilation (NIV) failure in COPD: A TRIPOD-aligned real-world analysis

MSR270

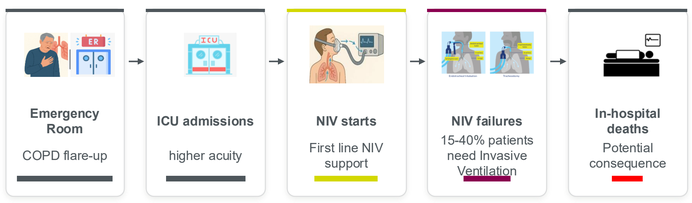
Komaleshwari Rani, MSc[1], Tuli De, PhD[2], Jackie Vanderpuye-Orgle, PhD[2]

[1]Parexel International, United Kingdom; [2]Parexel International, USA

Background

Non-invasive ventilation (NIV) fails in 15-40% of COPD patients treated in intensive care. NIV failure is associated with substantially higher mortality than NIV success (25-30% vs 5-10%) and prolonged ICU stay.[1,2]

Existing tools such as heart rate, acidosis, consciousness, oxygenation and respiratory rate (HACOR) combine a limited set of bedside variables assessed at specified time points after NIV initiation.[3,4] However, routinely collected physiological observations capture evolving patient status across the first 24 hours after NIV initiation.



Objectives

Clinical question: Can routinely collected physiological trajectories from the first 24 hours of ICU care identify patients at increased risk of NIV failure before escalation to invasive ventilation is required?

The objective of the study was to develop and evaluate a machine-learning model using routinely collected first-24-hour physiological data that predicts NIV failure in patients with COPD receiving intensive care.

Methods

Study population: Admissions in MIMIC-IV involving patients with COPD who received NIV.[5]

Outcome: Escalation to invasive mechanical ventilation.

Clinical data: Physiological observations recorded during the first 24 hours, divided into four sequential 6-hour windows. Predictors included vital signs, oxygen-delivery variables and GCS components.

Prediction approach: Temporal features were derived from physiological observations. Logistic regression and random forest models were evaluated on a held-out test set.

Pre-specified performance criteria: AUC ≥0.80, Sensitivity ≥0.75, Specificity ≥0.70

Model development and reporting were informed by TRIPOD guidance.[6]

Results

- Random Forest achieved AUC 0.804 (95% CI 0.724-0.879) and met the predefined discrimination target.
- The model met 2 of 3 pre-specified success criteria, achieving the ROC AUC and specificity targets but not the sensitivity target.
- On the independent test set, 22 of 35 NIV failures were correctly identified, while 13 failures were missed. There were 24 false-positive alerts among patients with NIV success.
- Random Forest demonstrated higher discrimination than Logistic Regression (AUC 0.804 vs 0.771). Five-fold cross-validation supported model stability (AUC 0.775 ± 0.009).

Table 1: Validation-set performance at selected threshold

Selected Threshold	Sensitivity	Specificity
0.6	0.68	0.83

Table 2: Test-set performance against pre-specified targets

Measure	Target	LogReg	Random Forest (95% CI)
ROC AUC	≥0.80	0.771	**0.804 (0.724-0.879)**
Sensitivity (failure)	≥0.75	0.600	**0.629 (0.475-0.788)**
Specificity (success)	≥0.70	0.818	**0.818 (0.744-0.880)**

Figure 1: Validation threshold sweep: Sensitivity and specificity

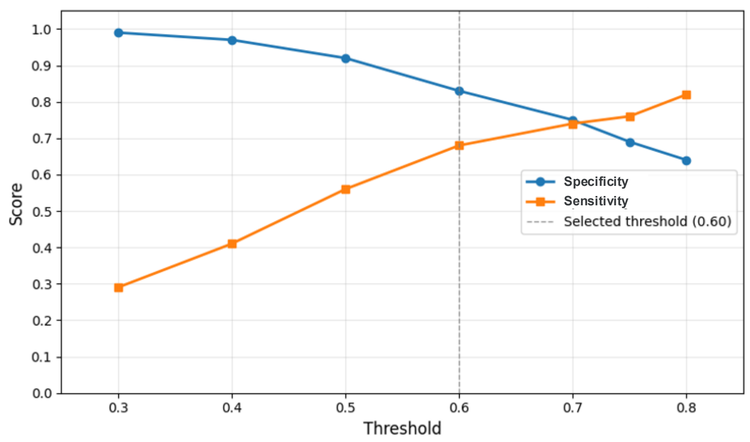


Figure 2: ROC curves for logistic regression and Random Forest on Held-out Test Data.

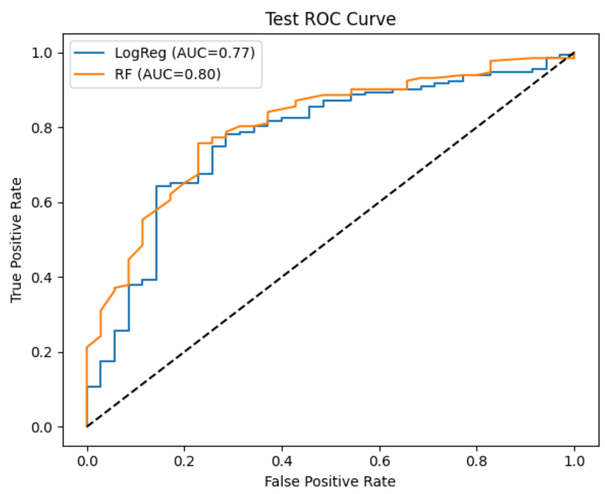


Figure 3: Test-set confusion matrix at selected operating point.

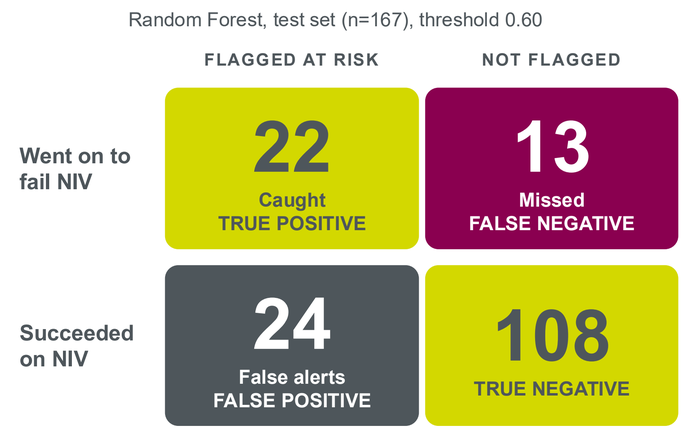
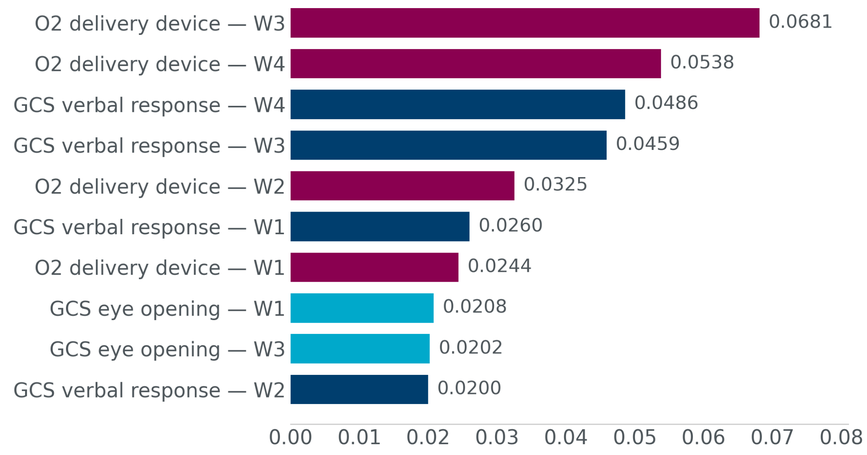


Figure 4: Top 10 predictors of NIV failure



Conclusions

- Routinely collected first-24-hour ICU data supported moderate-to-good discrimination of NIV failure in COPD patients (AUC 0.804).
- Although the ROC AUC and specificity targets were achieved, the pre-specified sensitivity target was not met (0.629 vs 0.75), indicating that a substantial proportion of future NIV failures remained undetected.
- The model is therefore best positioned for further evaluation as an assistive decision-support approach to prioritize clinical review, monitoring and escalation planning, rather than as a screening or autonomous decision system..
- The strongest predictors, oxygen-delivery variables and GCS verbal-response measures, were clinically plausible and aligned with recognised indicators of respiratory deterioration.
- Further development should focus on external validation and calibration assessment before any clinical implementation.

Abbreviations: AUC, area under the receiver operating characteristic curve; CI, confidence interval; COPD, chronic obstructive pulmonary disease; GCS, Glasgow Coma Scale; ICU, intensive care unit; LogReg, Logistic regression; MIMIC-IV, Medical Information Mart for Intensive Care; NIV, non-invasive ventilation; ROC, receiver operating characteristic; TRIPOD, Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis.

References

1. Chandra D, et al. Am J Respir Crit Care Med. 2012;185(2):152–159.
2. Osadnik CR, et al. Cochrane Database Syst Rev. 2017;7:CD004104.
3. Duan J, et al. Intensive Care Med. 2017;43(2):192–199.
4. Duan J, et al. Crit Care. 2022;26(1):196.
5. Johnson AEW, et al. Sci Data. 2023;10:1.
6. Collins GS, et al. Ann Intern Med. 2015;162(1):55–63.

Contact information

Komaleshwari Rani,
Senior Consultant, RWE Data Science, Advance Analytics, Access Consulting, Parexel International

Komaleshwari.rani@parexel.com

© 2026 Parexel International (MA) Corporation

parexel®

parexel.com