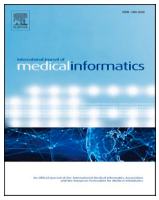
International Journal of Medical Informatics 219 (2026) 106560

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf



Original Research Article

Evaluation of the ROX index for predicting intubation in ICU patients

Alessandra Angelucci [a,*], Andrea Aliverti [a], Ilaria Pradella [a], Caterina Reni [a], Greta Roncalli [a], Maddalena Tacconelli [a], Maurizio Cecconi [b,c], Massimiliano Greco [b,c]

[a] Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milan, Italy
[b] Department of Biomedical Sciences, Humanitas University, Pieve Emanuele, Italy
[c] Department of Anesthesiology and Intensive Care, IRCCS Humanitas Research Hospital, Rozzano, Italy

ARTICLE INFO

Keywords:
ROX index
High-flow nasal cannula
Intubation
Intensive Care Unit
Critical care databases

ABSTRACT

Background: The ROX index was proposed as a decision-support tool to assess the effectiveness of high-flow oxygen therapy (HFOT) in patients with acute hypoxemic respiratory failure (AHRF) affected by pneumonia.
Objective: The purpose of this work was to assess the discriminative power of the ROX index across heterogeneous intensive care unit (ICU) populations. As a secondary, hypothesis-generating objective, we explored whether ROX-based risk stratification may provide a standardized reference for describing variability in observed intubation practices across datasets and centers.
Methods: Patients affected by AHRF and receiving oxygen support were identified from two large public ICU databases (MIMIC-IV and eICU). Oxygen support was stratified based on recorded flow rates, *i.e.*, $LPMO_2 \geq 6$ for conventional oxygen therapy (COT) and ≥ 30 for HFOT. All AHRF patients were initially considered, regardless of the underlying pathology, with a subgroup analysis performed in patients with pneumonia. ROX index predictions were compared with actual intubation rates in different datasets, and alternative thresholds were explored using Youden's method.
Results: In the primary three-category analysis, ROX risk strata produced only modest likelihood ratios (LRs) for observed intubation. In the merged cohorts, high-risk ROX categories showed LR values ranging from 1.36 to 2.06, whereas low-risk categories showed LR values close to 1, ranging from 0.85 to 0.90. Binary cut-off analyses confirmed limited discrimination, with AUROC values between 0.56 and 0.64.
Conclusions: When applied across heterogeneous real-world populations, the ROX index shows limited discriminative ability for predicting intubation and should not be used as a standalone decision tool. However, it may serve as a standardized reference to explore variability in intubation practices across centers, particularly in retrospective analyses.

1. Introduction

The decision to intubate a patient in the intensive care unit represents a particularly critical moment, as unnecessary intubation may entail risks and complications arising from mechanical ventilation itself, while delayed intubation increases the likelihood of death. For this reason, a tool that facilitates the early identification of patients who may benefit from mechanical ventilation would help prevent delayed intubation.

An objective indicator could also reduce the influence of the individual clinician's subjective judgment and serve as a common reference to support clinical decision-making and enable benchmarking, thereby improving the consistency of choices among different professionals. In this context, Roca et al. [1,2] introduced the ROX (Respiratory Rate-Oxygenation) index to predict the need for intubation in patients undergoing high-flow oxygen therapy (HFOT). This index was developed in 2016 based on a two-hospital cohort of 157 patients with pneumonia.

The ROX index ,computed as in Eq. (1), combines three clinical parameters that can be readily measured in a non-invasive manner: peripheral blood oxygen saturation (SpO_2), the inspiratory fraction of oxygen (FiO_2), and respiratory rate (RR).

$$ROX = \frac{SpO_2/FiO_2}{RR} \tag{1}$$

Based on the ROX index value twelve hours after the start of HFOT, Roca et al. divided patients into different risk categories, with those having a

* Corresponding author.
E-mail address: alessandra.angelucci@polimi.it (A. Angelucci).

https://doi.org/10.1016/j.ijmedinf.2026.106560
Received 29 January 2026; Received in revised form 19 June 2026; Accepted 21 June 2026

ROX index greater or equal to 4.88 being in the low-risk category and being unlikely to require intubation. Furthermore, they discussed how patients in a grey zone, i.e. with a ROX index between 3.85 and 4.88, might be treated [2]. Patients in this zone are considered to be at intermediate risk and those below 3.85 belong to the high-risk category and are more likely to require intubation. In addition to the studies focused on pneumonia [3], subsequent research has evaluated the ROX index in broader populations of patients with acute hypoxemic respiratory failure (AHRF) [4–6], with variable performance depending on population characteristics, definitions of outcomes, and clinical context.

This study investigates the discriminative power of the ROX index in patients with AHRF admitted to the intensive care unit (ICU) and treated with different levels of oxygen support. In addition, we explored the extent to which ROX-based risk stratification aligns with subsequent clinical decisions to proceed with invasive mechanical ventilation across datasets.

2. Materials and methods

The primary objective of this study was to assess the discriminative power of the ROX index in heterogeneous ICU populations with AHRF, including patients with and without pneumonia and patients receiving different levels of oxygen support.

Secondary objectives were to explore alternative ROX cutoffs across datasets using Youden's method and to evaluate, in a descriptive and hypothesis-generating manner, whether ROX-based risk stratification may provide a standardized reference for describing variability in observed intubation practices across databases and centers [7].

2.1. Data extraction from the databases

Several public databases, shown in Fig. 1, were taken into consideration, each presenting different internal structures and data classification methods.

Among the considered databases, the U.S. databases MIMIC-IV and eICU were chosen because they contain information on patients with all the characteristics of interest for our study. Other databases were excluded due to the lack or inconsistency of data, often regarding mechanical intubation or FiO_2.

The MIMIC-IV database [8,9] is a large, publicly available dataset comprising detailed clinical information from patients admitted to the Beth Israel Deaconess Medical Center in Boston, Massachusetts. The database is divided into two main modules: the hospital (*hosp*) module, which contains general demographic and administrative patient information, and the ICU module, which includes detailed data on more than 65,000 intensive care admissions.

The eICU database [10] collects data from 335 ICUs across 208 hospitals in the United States. The dataset covers hospitalizations that occurred between 2014 and 2015, including a population of 139,000 patients and a total of 200,000 ICU stays, accounting for repeated admissions of the same patient. The main strength of the eICU database lies in its ability to integrate data from multiple hospitals, facilitating comparative analyses across different healthcare institutions.

The variables extracted from each database included oxygen flow rate ($LPMO_2$, liters per minute of oxygen), FiO_2, SpO_2, RR, age, sex, pneumonia diagnosis, and records of invasive mechanical ventilation/intubation. AHRF was not available as a single explicitly coded variable in the databases; therefore, it was operationally identified from the combination of documented oxygen support and the availability of hypoxemia-related variables required to compute the ROX index. Because MIMIC-IV and eICU use different internal structures, these variables were identified through database-specific item codes and then harmonized across datasets before analysis. All variables and parameters that referred to a single hospital admission are identified by a specific ITEMID, while each patient is uniquely represented by a SUBJECTID. For each ITEMID, all corresponding records were retrieved, merged, and subsequently linked to the relevant patient.

2.2. Patient selection and ROX index computation

Patients were considered eligible for the AHRF cohort when they had documented oxygen support and available SpO_2, FiO_2, and RR measurements allowing computation of the ROX index within a predefined time window.

When multiple ICU admissions were available for the same patient, only the first hospitalization associated with a unique patient identifier was retained.

The term "all diagnoses" refers to the full AHRF cohort regardless of underlying etiology; pneumonia was then analyzed as a predefined subgroup.

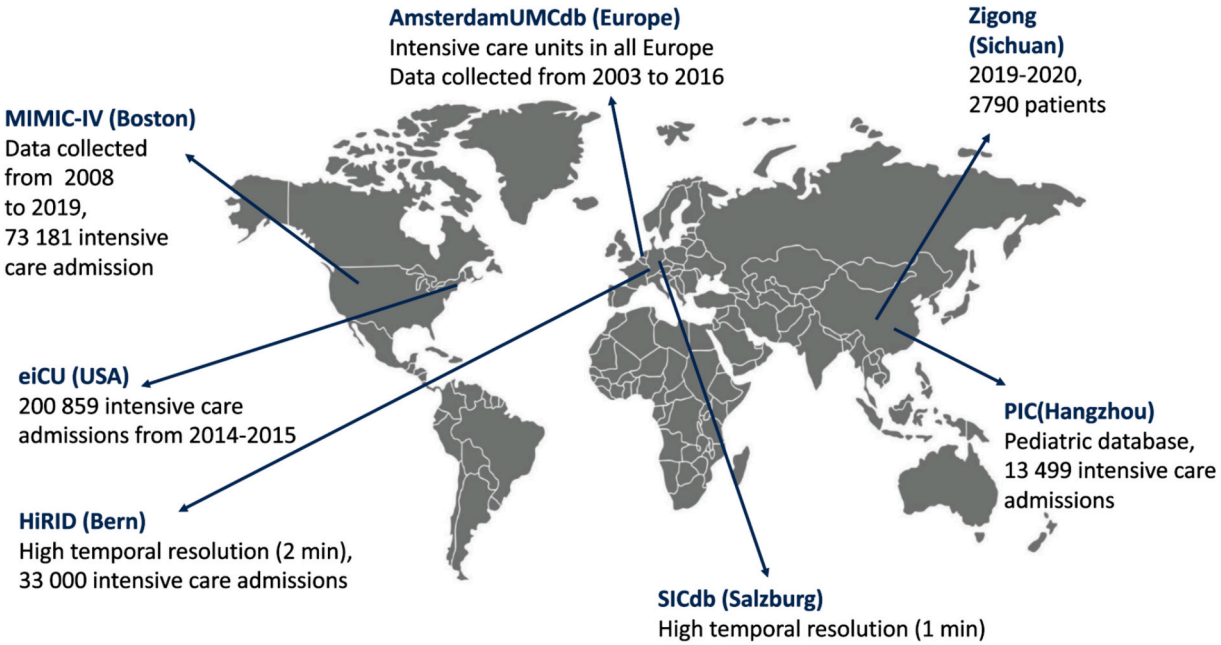


Fig. 1. Public databases initially considered for the analysis.

However, HFOT (or use of high-flow nasal cannula, HFNC) was not consistently labeled as a distinct therapy. Consequently, to distinguish HFOT from conventional oxygen therapy (COT), selection was based on the recorded $LPMO_2$. Two threshold values, derived from the literature, were adopted: $LPMO_2 \geq 6$ for COT as reported by Thille et al. [11], and $LPMO_2 \geq 30$ for HFOT [12]. Accordingly, two initial cohorts of patients were selected based on these two distinct $LPMO_2$ values.

From these cohorts, only patients for whom the ROX index could be computed approximately 12 h after the initiation of therapy (specifically between 11.5 and 14 h) were retained. This time window was selected according to the findings of Roca et al. [1,2], who demonstrated that the discriminative performance of the ROX index is maximal within this interval. The computation of the ROX index at this time point depended on the availability of the necessary parameters. For SpO_2 and RR, the first available measurements within the selected time window were used, whereas for FiO_2 the last value recorded immediately before those measurements was considered.

The subsequent step involved identifying patients who underwent mechanical intubation within a maximum of 72 h following the time point corresponding to the ROX index. This criterion ensured that the decision to intubate was temporally and clinically related to the ROX index value. Patients older than 80 years were excluded from the study, as their clinical management and general condition could be significantly influenced by age-related frailty.

From the two main cohorts defined by the $LPMO_2$ thresholds, additional subcohorts were extracted, including only patients diagnosed with pneumonia, identified by appropriate labels in the databases. This restriction was applied to maintain consistency with the original validation study of the ROX index [1]. The final analysis was therefore conducted on two primary samples and their corresponding pneumonia-only subgroups. The workflow of the study is shown in Fig. 2.

3. Statistical analysis

In the statistical analysis, we focused on the relationship between ROX-based risk stratification and observed clinical decisions to intubate, rather than on formal validation of predictive performance. Therefore, other outcomes such as death or recovery were not considered.

Patients were classified according to the thresholds previously introduced by Roca et al. [1,2], which were reported in the Introduction section. Three different risk categories for intubation were derived: the high-risk group included all patients with a ROX index <3.85, the intermediate risk group included patients with a ROX index between 3.85 and 4.88; and the low-risk group included patients with a ROX index >4.88 at 12 h after the start of therapy. The proportion of intubated and non-intubated patients was then evaluated within each category and each database to assess how ROX-based stratification aligned with clinical practice.

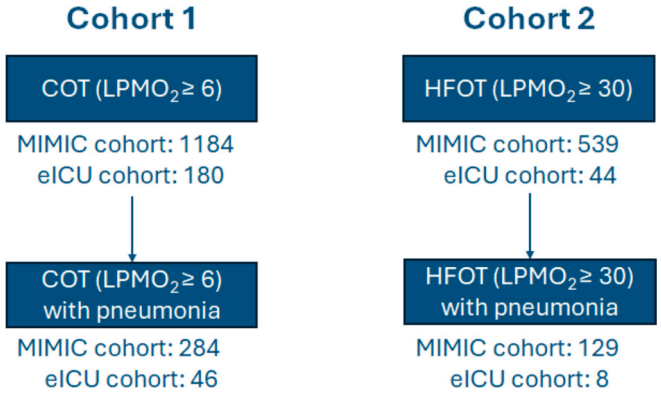


Fig. 2. Workflow of the study, with the total population divided using two thresholds ($LPMO_2 \geq 6$, and $LPMO_2 \geq 30$). The two cohorts are not mutually exclusive.

In this setting, patients intubated within 72 h after ROX assessment can be interpreted as cumulative cases for the event of interest, as described by Saha and Heagerty [13]. Regarding the clinical course within the 72-hour window, patients who died without being intubated or who were discharged due to clinical recovery were classified as 'non-events' for the primary outcome. This classification was adopted to define a fixed-window binary endpoint, namely observed intubation versus no observed intubation within 72 h after ROX assessment [13].

Since the classification based on ROX index defines an ordinal three-level diagnostic variable, the primary descriptive analysis was based on 3×2 contingency tables crossing ROX risk category with observed intubation within 72 h after ROX assessment. For each ROX category, we calculated category-specific likelihood ratios (LRs), defined as the proportion of intubated patients falling within a given ROX category divided by the proportion of non-intubated patients falling within the same category. Thus, for category k, LR_k was calculated as in (2):

$$LR_k = \frac{\left[\frac{n(intubatedincategoryk)}{N(intubated)}\right]}{\left[\frac{n(non-intubatedincategoryk)}{N(nonintubated)}\right]} \quad (2)$$

Values greater than 1 indicate that the corresponding ROX category is more frequently observed among intubated than among non-intubated patients, whereas values below 1 indicate that the category is more frequently observed among non-intubated patients. This approach was chosen because LRs are more appropriate than binary sensitivity and specificity when the test result is expressed on more than two ordered categories [14–16]. Patients who died without intubation or were discharged without intubation during the 72-hour window were included in the non-intubated group. This approach differs from a time-dependent competing-risk framework, in which competing events such as death or discharge would be explicitly modeled [13].

Additionally, Youden's index [17] was calculated for each of the obtained samples to identify a cut-off value that provides the best trade-off between specificity and sensitivity when applied to the same outcome, *i.e.*, the need for intubation. As a secondary analysis, the ROX index was dichotomized using the fixed high-risk threshold of 3.85, with ROX $<$ 3.85 considered test-positive and ROX $\geq$ 3.85 considered test-negative. This threshold was selected because it identifies the high-risk category discussed in Roca et al.'s work and corresponds to the clinically most conservative definition of increased intubation risk. In addition, sample-specific thresholds were derived using Youden's index. Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated only for these binary 2×2 analyses.

Both LRs and Youden's index analyses were carried out only on the two macro-samples obtained by merging the MIMIC-IV and eICU databases, namely $LPMO_2 \geq 6$ for COT, and $LPMO_2 \geq 30$ for HFOT. Two corresponding pneumonia subgroups were also analyzed.

4. Results

The demographic characteristics of the analyzed population are described in Table 1.

In Fig. 3, the results of each database analyzed individually are shown first, followed by the combined analysis of both.

Table 2 reports the category-specific LRs for observed intubation within 72 h according to ROX risk category in the merged MIMIC-IV + eICU cohorts.

In the primary 3×2 analysis, ROX risk categories showed only modest shifts in the observed probability of intubation. In the merged MIMIC-IV + eICU cohort with $LPMO_2 \geq 6$, the likelihood ratio was 2.05 for the high-risk category, 2.0 for the intermediate-risk category, and 0.86 for the low-risk category. Similar values were observed in the

Table 1
Demographic characteristics of each sample analyzed. SD = standard deviation.

Database	Oxygen therapy	Diagnosis	Men [%]	Women [%]	Age [years, mean ± SD]
MIMIC-IV	$LPMO_2 \geq 6$	All	55.7	44.3	62.20 ± 13.93
		Pneumonia	55.6	44.4	60.68 ± 13.58
	$LPMO_2 \geq 30$	All	58.6	41.4	62.13 ± 13.58
		Pneumonia	61.2	38.8	60.78 ± 13.05
eICU	$LPMO_2 \geq 6$	All	62.8	37.2	63.03 ± 12.43
		Pneumonia	67.4	32.6	63.40 ± 11.24
	$LPMO_2 \geq 30$	All	59.1	40.9	63.39 ± 11.95
		Pneumonia	50	50	67.75 ± 8.5
MIMIC-IV + eICU	$LPMO_2 \geq 6$	All	56.3	43.7	62.27 ± 13.81
		Pneumonia	57.3	42.7	61.15 ± 13.30
	$LPMO_2 \geq 30$	All	58.7	41.3	62.22 ± 13.46
		Pneumonia	60.6	39.4	61.19 ± 12.92

pneumonia subgroup, with LR values of 2.06, 1.84, and 0.85, respectively. In the merged $LPMO_2 \geq 30$ cohort, the corresponding LR values were 1.99, 1.62, and 0.85 in the overall population, and 1.36, 1.43, and 0.90 in the pneumonia subgroup.

These findings indicate that patients in the high-risk ROX category were more frequently represented among intubated than among non-intubated patients; however, the magnitude of this association was limited. Conversely, the low-risk category was only weakly associated

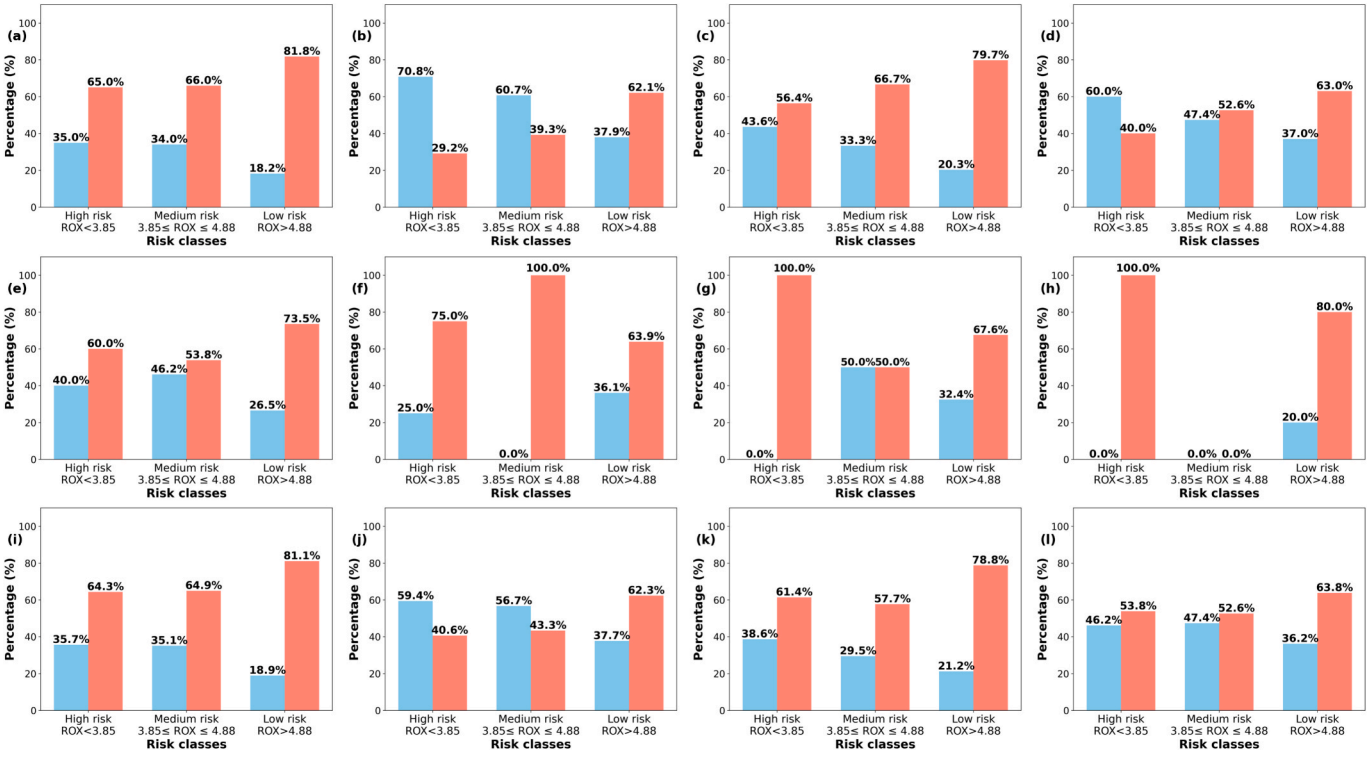


Fig. 3. Subplots showing, for each cohort, the percentage of intubated (blue) and non-intubated (orange) patients within each ROX risk category. Percentages are calculated using, as denominator, the total number of patients in the corresponding ROX risk category: a) MIMIC-IV, $LPMO_2 \geq 6$ (all diagnoses); b) MIMIC-IV, $LPMO_2 \geq 6$ (pneumonia); c) MIMIC-IV, $LPMO_2 \geq 30$ (all diagnoses); d) MIMIC-IV, $LPMO_2 \geq 30$ (pneumonia); e) eICU, $LPMO_2 \geq 6$ (all diagnoses); f) eICU, $LPMO_2 \geq 6$ (pneumonia); g) eICU, $LPMO_2 \geq 30$ (all diagnoses); h) eICU, $LPMO_2 \geq 30$ (pneumonia); i) MIMIC-IV + eICU, $LPMO_2 \geq 6$ (all diagnoses); j) MIMIC-IV + eICU, $LPMO_2 \geq 6$ (pneumonia); k) MIMIC-IV + eICU, $LPMO_2 \geq 30$ (all diagnoses); l) MIMIC-IV + eICU, $LPMO_2 \geq 30$ (pneumonia).

Table 2
Category-specific likelihood ratios for observed intubation within 72 h according to ROX risk category in the MIMIC-IV + eICU cohorts. High-risk was defined as ROX < 3.85, intermediate-risk as $3.85 \leq$ ROX ≤ 4.88, and low-risk as ROX > 4.88. LR values were calculated as the proportion of intubated patients in a given ROX category divided by the proportion of non-intubated patients in the same category. LR > 1 indicates that the category was more frequent among intubated patients, whereas LR < 1 indicates that the category was more frequent among non-intubated patients.

Database	Oxygen therapy	Diagnosis	LR high-risk	LR intermediate-risk	LR low-risk
MIMIC-IV + eICU	$LPMO_2 \geq 6$	All	2.05	2.00	0.86
		Pneumonia	2.06	1.84	0.85
	$LPMO_2 \geq 30$	All	1.99	1.62	0.85
		Pneumonia	1.36	1.43	0.90

with absence of intubation, with LR values close to 1 across cohorts.

Using the Youden method on the merged database, a cut-off of 6.99 was identified for the whole COT cohort (Fig. 4a) and 4.70 for the pneumonia COT subcohort (Fig. 4b). With HFOT, the optimal cut-off with the Youden method was 6.09 for the group with all diagnoses (Fig. 4c) and 8.61 for the pneumonia group (Fig. 4d). Table 3 reports the performance of binary ROX classifications in the merged cohorts. Results obtained using the fixed high-risk threshold cut-off of 3.85 are reported to allow comparison with a binary cut-off–based approach, whereas Youden-derived thresholds were estimated separately for each cohort.

In the non-pneumonia cohorts (both $LPMO_2 \geq 6$ and $LPMO_2 \geq 30$), the Youden-derived thresholds (6.99 and 6.09, respectively) showed relatively narrower confidence intervals and AUC values around 0.63–0.64, indicating a modest but more stable discriminative ability. Conversely, in the pneumonia subgroups, the optimal thresholds exhibited very wide confidence intervals (*e.g.*, 4.22–10.78 for $LPMO_2 \geq 30$ and 4.46–13.00 for $LPMO_2 \geq 6$), and both sensitivity and specificity showed marked variability.

5. Discussion

In this study we present an analysis on the MIMIC-IV and eICU databases assessing the relationship between ROX-based stratification and observed intubation practices, with the aim of exploring whether it may serve as a standardized reference for benchmarking intubation practices across different centers and studies.

In the primary 3 × 2 analysis, ROX-based risk categories were

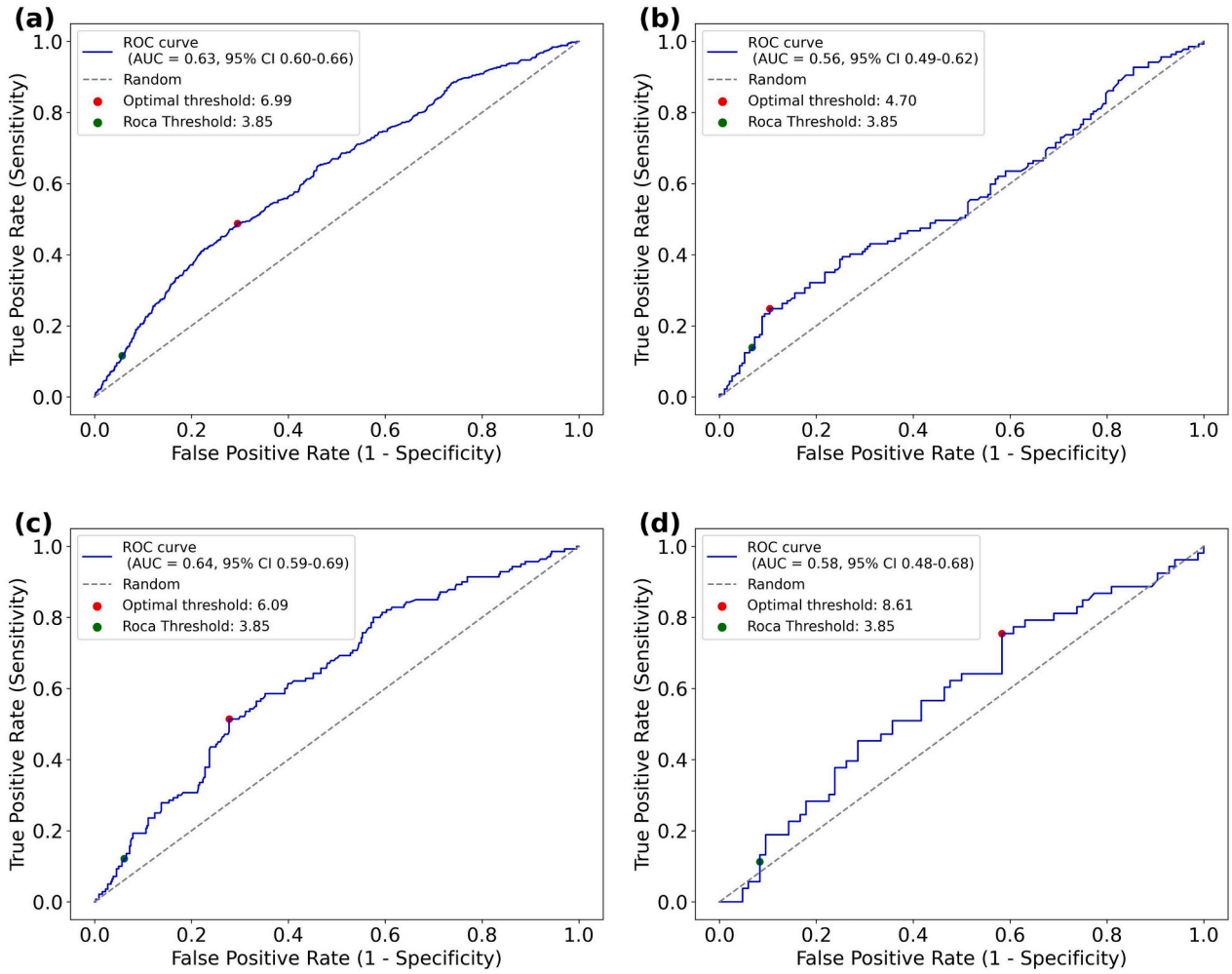


Fig. 4. Subplot showing, for each cohort, the ROC curve with the optimal threshold determined with Youden's method and fixed high-risk threshold of 3.85 (Roca Threshold in the subplots); a) MIMIC-IV + eICU, $LPMO_2 \geq 6$ (all diagnoses); b) MIMIC-IV + eICU, $LPMO_2 \geq 6$ (pneumonia); c) MIMIC-IV + eICU, $LPMO_2 \geq 30$ (all diagnoses); d) MIMIC-IV + eICU, $LPMO_2 \geq 30$ (pneumonia).

Table 3
Sensitivity, specificity, PPV and NPV of the analyzed samples (MIMIC-IV + eICU) with different cutoff values, i.e., the fixed high risk ROX threshold of 3.85 proposed by Roca et al., and a cutoff determined with Youden's method. Values for sensitivity, specificity, PPV and NPV obtained using Roca's thresholds are presented as estimate (95% CI) using the exact binomial confidence intervals (Clopper-Pearson). Values obtained with the Youden method are presented as estimates (95% CIs) derived from bootstrap internal validation.

Cutoff method	Oxygen therapy	Diagnosis	Cutoff value	Sensitivity [%]	Specificity [%]	PPV [%]	NPV [%]
Roca	$LPMO_2 \geq 6$	All	3.85	11.6 (8.8–15)	94.3 (93.1–95.4)	35.7 (27.8–44.1)	79.8 (77.9–81.6)
		Pneumonia		13.9 (8.6–20.8)	93.3 (88.8–96.4)	59.4 (40.6–76.3)	60.4 (54.6–66.0)
	$LPMO_2 \geq 30$	All		12.1 (7.2–18.7)	93.9 (91.3–95.9)	38.6 (24.4–54.5)	77.2 (73.4–80.7)
		Pneumonia		11.3 (4.3–23.0)	91.7 (83.6–96.6)	46.1 (19.2–74.9)	62.1 (52.9–70.7)
Youden	$LPMO_2 \geq 6$	All	6.99 (6.06–9.40)	48.7 (38.6–71.5)	70.5 (49.7–80.0)	30.8 (25.7–36.7)	83.5 (81.7–87)
		Pneumonia	4.70 (4.46–13.00)	24.8 (20.1–91.5)	89.6 (21.5–93.9)	63.0 (44.2–75.0)	62.7 (57.7–74.0)
	$LPMO_2 \geq 30$	All	6.09 (5.83–9.30)	51.4 (46.8–87.6)	72.2 (37.9–76.4)	36.6 (27.5–43.1)	82.3 (79.6–90.8)
		Pneumonia	8.61 (4.22–10.78)	75.5 (25.0–89.4)	41.7 (29.8–91.1)	44.9 (37.2–66.7)	72.9 (59.3–83.3)

associated with observed intubation, but the magnitude of this association was limited. Across the merged cohorts, the high-risk category showed LR values of approximately 2 in most analyses, indicating only a modest increase in the relative frequency of intubation. Conversely, the low-risk category showed LR values close to 1, suggesting that low ROX-risk classification only weakly reduced the likelihood of observed intubation. These findings indicate that the original ROX categories, when applied to heterogeneous ICU populations, provide limited incremental information for distinguishing patients who subsequently underwent intubation from those who did not. The comparison between cohorts defined by different oxygen flows reflects not only different oxygen delivery modalities but also inherently different patient severity profiles, which may intrinsically influence the performance of the ROX index and contribute to the observed differences between groups.

Several methodological factors may have influenced the observed performance of the ROX index in this study. First, the inclusion of a broader and heterogeneous population of patients with acute hypoxemic respiratory failure, beyond the original population in which the ROX index was developed, as expected, affected its discriminative ability. Second, the retrospective nature of the data limits standardization of oxygen delivery modalities and features, including the timing and clinical context of ROX index calculation, which may differ from other studies. Finally, the clinician-driven nature of intubation, varying across centers and practices, introduces additional variability that may weaken the association between ROX-based categories and observed intubation.

An additional methodological issue concerns the interpretation of intubation as an outcome. In the original studies by Roca et al. [1,2], escalation to invasive mechanical ventilation was considered within a structured clinical framework for HFOT failure. In contrast, in the present retrospective analysis, intubation was treated as an observed event recorded in two ICU databases, and may have been driven by factors different from those considered in the original studies by Roca et al. Therefore, differences in local protocols, treatment decisions, and practices may have influenced the observed relationship between ROX categories and intubation.

Nonetheless, the aim of this study was to explore the use of the ROX index across multiple centers as a standardized reference for benchmarking intubation practices in both high flow and conventional oxygen therapy; in this context, the differences observed between MIMIC-IV and eICU reflect variability in clinical practice rather than intrinsic limitations of the index alone.

An additional limitation concerns the different structure and granularity of the two main U.S. databases. Although eICU includes data from 208 hospitals over a two-year period, the number of patients meeting our inclusion criteria was relatively small compared with MIMIC-IV. This discrepancy should not be interpreted as reflecting the true incidence of AHRF in the participating hospitals. Rather, it likely reflects differences in data structure, variable availability, local documentation practices, and the way oxygen therapy, FiO_2, RR, SpO_2, and intubation events were recorded across databases. In particular, the strict requirement for all variables needed to compute the ROX index within the predefined 11.5–14 h window after oxygen-support initiation may have substantially reduced the number of eligible eICU records. This limitation should also be considered when interpreting the merged analysis, since the contribution of eICU is smaller and potentially affected by greater heterogeneity across centers.

Previous studies have evaluated the ROX index in relatively well-defined populations and settings. Across studies, both predictive performance and the number of included patients vary depending on the definitions of acute respiratory failure, HFNC, and pneumonia [18,19]. For example, Yu et al. validated a machine learning model against the ROX index in predefined cohorts, applying strict definitions of HFNC use within the MIMIC-IV and eICU databases [19,20]. The existing literature is also affected by heterogeneity in outcome definitions, including variable definitions of HFNC failure and the absence of standardized criteria for intubation. A meta-analysis by Junhai et al., including 9 studies (6 in pneumonia and 3 in AHRF), reported a pooled sensitivity of approximately 0.67, highlighting the variability in predictive performance. The same analysis also showed a wide range of ROX index cut-offs (2.7–9.2), with higher discriminatory accuracy observed in studies using thresholds > 5 compared to ≤ 5 [21].

In this context, the present study was designed not to further validate the ROX index, but to investigate its discriminative power across heterogeneous, real-world ICU populations with respiratory failure. By relating ROX-based stratification to observed intubation practices across different oxygen therapy levels, we aimed to explore its potential as a common reference framework to characterize variability in clinical decision-making across centers, rather than as a standalone predictor of treatment failure.

These aspects can be further explored through the analysis of the European database. As an example, in the HiRID dataset [8,22], which reflect a Swiss ICU setting as highlighted in Fig. 1, a higher proportion of intubated patients is observed in the high-risk ROX group, exceeding that of non-intubated patients. This reflects the different ICU admission criteria between Europe and the United States. In Europe, ICU bed capacity relative to total hospital beds is typically lower, and admissions are generally reserved for the most critically ill patients, often already requiring invasive mechanical ventilation. The good predictive power of the ROX index in the European context may therefore be attributed to the fact that Roca's original study was conducted in French and Spanish hospitals, although further analyses are needed to confirm this finding.

The study presented in this work has several limitations. First, its retrospective observational design limits the ability to infer causality, and the observed associations may be influenced by unmeasured confounding factors. The use of predefined oxygen flow rate thresholds in retrospective data to distinguish HFOT from conventional oxygen therapy may have led to some level of misclassification in the intermediate range (10–30 L/min), which may represent a transitional level of support or humidified oxygen therapy rather than true COT or HFOT. Clinical decisions such as initiation of HFOT and endotracheal

intubation were not protocolized and may reflect local practice and clinician experience, rather than physiological deterioration. As a result, intubation should be interpreted as a clinician-driven outcome rather than a definitive biological or clinical endpoint. We excluded patients over 80 years of age to reduce heterogeneity; however, this may limit the generalizability of the findings.

A further limitation concerns the handling of death and discharge within the 72-hour window. These events may act as competing events for intubation, since they can preclude or modify the subsequent observation of the event of interest. For this reason, our analysis should be interpreted as a descriptive assessment of ROX-based stratification with respect to observed intubation practice within a fixed time window, rather than as a formal time-to-event competing-risk analysis [13]. Accordingly, patients who died without intubation or were discharged without intubation were classified as non-intubated. This choice was consistent with the descriptive aim of relating ROX-based stratification to observed intubation practice, but it may affect the interpretation of both category-specific LRs and secondary binary performance metrics if some deaths represented unobserved treatment failures or if intubation was not performed because of treatment-limitation decisions. A multinomial or competing-risk framework distinguishing intubation, death without intubation, discharge/recovery without intubation, and remaining non-intubated patients would provide a more detailed characterization of these trajectories. However, such an analysis was beyond the scope of the present study and would be limited by the imbalance in sample size across datasets and subgroups, particularly in smaller eICU subcohorts. This should be addressed in future studies with larger and more balanced samples.

Although the ROX index was originally proposed as a bedside decision-support tool, in this analysis its performance appeared insufficient to reliably guide individual clinical decisions across heterogeneous real-world settings. By focusing on observed intubation events, the present study explores how ROX-based stratification relates to clinical practice, highlighting its potential role as a reference framework to characterize variability in intubation decisions rather than as a standalone tool for guiding escalation of care. Second, we could not control all factors which may be related with the decision to intubate, including hemodynamic and neurologic status, invasive procedures, transport, or other issues. Moreover, the decision to intubate a patient is assumed to be correct and is not validated, for example, by comparing it with the final clinical outcome or mortality. Nonetheless, the strength of this study lies in the use of multi-center databases: the patients analyzed come from different hospitals, each with distinct protocols and clinical decision-making processes. These differences were integrated within a single dataset, providing a more comprehensive and representative view of clinical reality.

In conclusion, when applied across a heterogeneous population of patients with acute hypoxemic respiratory failure, the ROX index showed limited ability to discriminate patients requiring intubation, both in the overall cohort and in the subgroup of patients with pneumonia. Its performance appears insufficient to support clinical decision-making across diverse real-world settings, particularly when applied beyond the populations in which it was originally developed. However, ROX-based risk stratification may still provide a descriptive reference to explore variability in observed intubation practices across datasets and centers, although formal benchmarking would require more balanced samples and dedicated between-center analyses. Its application should nonetheless be complemented, when possible, by additional clinical and physiological parameters, particularly in retrospective and heterogeneous analyses.

CRediT authorship contribution statement

Alessandra Angelucci: Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Investigation, Data curation, Conceptualization. **Andrea Aliverti:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization. **Ilaria Pradella:** Writing – original draft, Visualization, Software, Formal analysis, Data curation. **Caterina Reni:** Writing – original draft, Visualization, Software, Formal analysis, Data curation. **Greta Roncalli:** Writing – original draft, Visualization, Software, Formal analysis, Data curation. **Maddalena Tacconelli:** Writing – original draft, Visualization, Software, Formal analysis, Data curation. **Maurizio Cecconi:** Writing – review & editing, Supervision, Resources, Conceptualization. **Massimiliano Greco:** Writing – review & editing, Supervision, Methodology, Investigation, Data curation, Conceptualization.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT only to support English language editing and proofreading. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Funding: This work was partly funded by the National Plan for National Recovery and Resilience Plan (NRRP) Complementary Investments (PNC, established with the decree-law 6 May 2021, n. 59, converted by law n. 101 of 2021) in the call for the funding of the Research Initiatives for Technologies and Innovative Trajectories in the health and care sectors (Directorial Decree n. 931 of 06-06-2022)–AdvaNced Technologies for Human centEred Medicine (ANTHEM) under Project PNC0000003. This work reflects only the authors' views and opinions, neither the Ministry for University and Research nor the European Commission can be considered responsible for them.

References

[1] O. Roca, J. Messika, B. Caralt, M. García-de-Acilu, B. Sztrymf, J.D. Ricard, J. R. Masclans, Predicting success of high-flow nasal cannula in pneumonia patients with hypoxemic respiratory failure: the utility of the ROX index, J. Crit. Care 35 (2016) 200–205, https://doi.org/10.1016/j.jcrc.2016.05.022.

[2] O. Roca, B. Caralt, J. Messika, M. Samper, B. Sztrymf, G. Hernández, M. García-De-Acilu, J.P. Frat, J.R. Masclans, J.D. Ricard, An index combining respiratory rate and oxygenation to predict outcome of nasal high-flow therapy, Am. J. Respir. Crit. Care Med. 199 (2019) 1368–1376, https://doi.org/10.1164/RCCM.201803-0589OC.

[3] L.F. Reyes, A. Bastidas Goyes, E.A. Tuta Quintero, K.D. Pedreros, Y.F. Mantilla, M. Herrera, G.A. Carmona, L.D. Saza, L.E. Bello, C.A. Muñoz, J.C. Chaves, J. C. Arias, P.M. Alcaraz, M.D. Hernández, A.P. Nonzoque, N. Trujillo, A.F. Pineda, G. S. Montaño, Validity of the ROX index in predicting invasive mechanical ventilation requirement in pneumonia, BMJ Open Respir. Res. 9 (2022), https://doi.org/10.1136/BMJRESP-2022-001320.

[4] D. Basoulis, P. Avramopoulos, M. Aggelara, G. Karamanakos, P.M. Voutsinas, A. Karapanou, M. Psichogiou, M. Samarkos, F. Ntziora, N.V. Sipsas, Validation of sequential ROX-Index score beyond 12 hours in predicting treatment failure and mortality in COVID-19 patients receiving oxygen via high-flow nasal cannula, Can. Respir. J. 2023 (2023), https://doi.org/10.1155/2023/7474564.

[5] L.C. Myers, D. Mark, B. Ley, M. Guarnieri, M. Hofmeister, S. Paulson, G. Marelich, V.X. Liu, Validation of respiratory rate-oxygenation index in patients with COVID-19-related respiratory failure, Crit. Care Med. 50 (2022) E638–E642, https://doi.org/10.1097/CCM.0000000000005474.

[6] A. Li, M.E. Cove, J. Phua, S.H. Puah, V. Ng, A. Kansal, Q.L. Tan, J.T. Sahagun, J. Taculod, A.Y.H. Tan, A. Mukhopadhyay, C.K. Tay, K. Ramanathan, Y.W. Chia, D. W. Sewa, M. Chew, S.J.W. Lew, S. Goh, S. Dhanvijay, J.J.E. Tan, K.C. See, Expanding the utility of the ROX index among patients with acute hypoxemic respiratory failure, PLoS One 17 (2022), https://doi.org/10.1371/JOURNAL.PONE.0261234.

[7] A.R. Jagesar, T.A. Dam, T. Struja, C.M. Sauer, M. Otten, L.A. Biesheuvel, A.R. J. Girbes, L. Adhikari, Z. Zhang, M. Faltys, N. Rodemund, P.J. Thoral, L.A. Celi, P.

W.G. Elbers, Sharing is caring: a systematic review of publicly available intensive care data sets, J. Crit. Care 90 (2025) 155205, https://doi.org/10.1016/J.JCRC.2025.155205.

[8] A.L. Goldberger, L.A. Amaral, L. Glass, J.M. Hausdorff, P.C. Ivanov, R.G. Mark, J. E. Mietus, G.B. Moody, C.K. Peng, H.E. Stanley, PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals, Circulation 101 (2000), https://doi.org/10.1161/01.CIR.101.23.E215.

[9] A.E.W. Johnson, L. Bulgarelli, L. Shen, A. Gayles, A. Shammout, S. Horng, T. J. Pollard, B. Moody, B. Gow, L. Wei, H. Lehman, L.A. Celi, R.G. Mark, MIMIC-IV, a freely accessible electronic health record dataset, Scientific Data 1 (10) (2023), https://doi.org/10.1038/s41597-022-01899-x.

[10] T.J. Pollard, A.E.W. Johnson, J.D. Raffa, L.A. Celi, R.G. Mark, O. Badawi, The eICU collaborative research database, a freely available multi-center database for critical care research, Sci. Data 5 (2018) 180178, https://doi.org/10.1038/SDATA.2018.178;SUBJMETA.

[11] A.W. Thille, F. Balen, G. Carteaux, T. Chouihed, J.P. Frat, C. Girault, E. L'Her, N. Marjanovic, M.A. Nay, P. Ray, M. Reffienna, L. Retenauer, A. Roch, G. Thiery, J. Truchot, Oxygen therapy and noninvasive respiratory supports in acute hypoxemic respiratory failure: a narrative review, Ann. Intensive Care 14 (2024), https://doi.org/10.1186/S13613-024-01389-W.

[12] O. Roca, J. Li, T. Mauri, High-flow nasal cannula: evolving practices and novel clinical and physiological insights, Intensive Care Medicine 50 (5) (2024) 758–761, https://doi.org/10.1007/S00134-024-07386-8.

[13] P. Saha, P.J. Heagerty, Time-dependent predictive accuracy in the presence of competing risks, Biometrics 66 (2010) 999, https://doi.org/10.1111/J.1541-0420.2009.01375.X.

[14] G.M. Schuetz, P. Schlattmann, M. Dewey, Use of 3×2 tables with an intention to diagnose approach to assess clinical performance of diagnostic tests: meta-analytical evaluation of coronary CT angiography studies, The BMJ 345 (2012) e6717.

[15] D.L. Simel, J.R. Feussner, E.R. Delong, D.B. Matchar, Intermediate, indeterminate, and uninterpretable diagnostic test results, Med. Decis. Making 7 (1987) 107–114, https://doi.org/10.1177/0272989X8700700208.

[16] A.R. Feinstein, The inadequacy of binary models for the clinical reality of three-zone diagnostic decisions, J. Clin. Epidemiol. 43 (1990) 109–113, https://doi.org/10.1016/0895-4356(90)90064-V.

[17] R. Fluss, D. Faraggi, B. Reiser, Estimation of the Youden Index and its associated cutoff point, Biom. J. 47 (2005) 458–472, https://doi.org/10.1002/BIMJ.200410135;WGROUP:STRING:PUBLICATION.

[18] X. Li, C. Jiang, Q. Xie, H. Wang, J. Xu, G. Liu, P. Chang, G. Zhang, Real-time prediction of HFNC treatment failure in acute hypoxemic respiratory failure using machine learning, Scient. Rep. 1 (15) (2025) 30245, https://doi.org/10.1038/s41598-025-16061-x.

[19] H. Yu, S. Saffaran, R. Tonelli, J.G. Laffey, Q. Zhang, A.M. Esquinas, L.M. de Lima, L. Kawano-Dourado, I.S. Maia, A.B. Cavalcanti, E. Clini, D.G. Bates, Evaluating the effect of heart and respiratory rate measurement errors on the ability to predict the outcome of high flow nasal cannula therapy: a multi-centre study, Crit. Care 29 (2025) 535, https://doi.org/10.1186/S13054-025-05765-1.

[20] H. Yu, S. Saffaran, R. Tonelli, J.G. Laffey, A.M. Esquinas, L.M. de Lima, L. Kawano-Dourado, I.S. Maia, A.B. Cavalcanti, E. Clini, D.G. Bates, Machine learning models compared with current clinical indices to predict the outcome of high flow nasal cannula therapy in acute hypoxemic respiratory failure, Crit. Care 29 (2025), https://doi.org/10.1186/S13054-025-05336-4.

[21] Z. Junhai, Y. Jing, C. Beibei, L. Li, The value of ROX index in predicting the outcome of high flow nasal cannula: a systematic review and meta-analysis, Respir. Res. 1 (23) (2022) 33, https://doi.org/10.1186/S12931-022-01951-9.

[22] S.L. Hyland, M. Faltys, M. Hüser, X. Lyu, T. Gumbsch, C. Esteban, C. Bock, M. Horn, M. Moor, B. Rieck, M. Zimmermann, D. Bodenham, K. Borgwardt, G. Rätsch, T. M. Merz, Early prediction of circulatory failure in the intensive care unit using machine learning, Nat. Med. 3 (26) (2020) 364–373, https://doi.org/10.1038/s41591-020-0789-4.