MAYO CLINIC PROCEEDINGS:
INNOVATIONS, QUALITY & OUTCOMES

ORIGINAL ARTICLE

Vital Signs—Only Machine Learning Model for Acute Inpatient Deterioration: A Retrospective Multicenter Study

Santiago Romero-Brufau, MD, PhD; Radit Smunyahirun, PhD;
Timothée Filhol, MiM; Lucille Niederhauser, MS; Thanawin Trakoolwilaiwan, MS;
and Gurpreet Singh, PhD

Abstract

Objective: To develop predictive models that are compatible with vital signs monitoring devices to identify patients at risk of clinical deterioration, defined as requiring a rapid response team intervention or an unplanned intensive care unit transfer.

Patients and Methods: Targeted vital signs from 227,858 inpatients admitted to general care or telemetry beds at a multihospital health care institution between January 1, 2019, and July 31, 2023, were selected. After filtering for high-quality data, 30,118 patients were used to train a Light Gradient Boosting Machine, and 30,095 were reserved for blind validation. We developed a machine learning model designed to minimize false positives while maintaining clinical relevance in identifying low-prevalence clinical deterioration events.

Results: At a sensitivity of 73.4% (95% CI, 72.2%-74.4%), the model achieved a positive predictive value (PPV) of 30.4% (95% CI, 29.6%-31.3%), with a C-statistic of 0.874 (95% CI, 0.867-0.881), alert rate of 0.170 (95% CI, 0.167-0.173) per patient per day, and normalized alert rate of 2.41 (95% CI, 2.31-2.51). Stratified analysis by hospital revealed that PPV was highest at the Rochester site, reaching 54.9% (95% CI, 52.9%-57.0%) and outperforming the EPIC deterioration index by 46% or a factor of 6 (7.57%).

Conclusion: Achieving a high PPV is crucial because it ensures a larger proportion of alerts are true positives, reducing the burden of false alarms. The considerable improvement in results comes from the novel 2-window feature extraction method. This technique enables the model to capture both long-term trends and recent changes in patient status, enhancing predictive performance.

■ Mayo Clin Proc Inn Qual Out 2025;9(5):100663

Clinical deterioration refers to the worsening of a patient's physiological condition, often characterized by important deviations from baseline health status. This process is marked by declining vital signs, such as unstable blood pressure, elevated heart rate (HR), respiratory distress, or low oxygen saturation (SpO_2).[1-3] Left unrecognized or unmanaged, clinical deterioration can lead to severe outcomes, including organ failure, cardiac arrest, or even death.[4-6] In hospital settings, it remains a major cause of preventable mortality, especially in general care units. The timely identification of clinical deterioration is critical because delayed recognition and intervention can be associated with poor patient outcomes and increased health care costs.[7,8]

Vital signs serve as key indicators of a patient's health status and are often the first signals of clinical deterioration. Changes such as a rapid HR, irregular breathing patterns, or a drop in SpO_2 levels can indicate the onset of critical events.[9-13] These physiological changes typically occur hours before severe events, such as cardiac arrest or the need for mechanical ventilation, providing an opportunity for early detection. Research shows that early intervention based on these warning signs, such as activating a rapid response

From the Department of Otolaryngology, Head and Neck Surgery, SW Rochester, MN (S.R-B.); Department of Biostatistics, Harvard University, Boston, MA (S.R-B.); Respiree Private Ltd, Singapore, Singapore (R.S., T.F., T.T., G.S.); and École polytechnique fédérale de Lausanne (EPFL), Lausanne, Switzerland (L.N.).

team or transferring patients to a higher level of care, can significantly reduce mortality rates and improve patient outcomes.[14-19] Prompt recognition and timely action not only save lives but also prevent complications, shorten hospital stays, and enhance overall quality of care.

Numerous early warning systems, including widely used tools such as the Modified Early Warning Score, the National Early Warning Score, and the EPIC deterioration index have been developed to identify clinical deterioration.[20-40] Although these systems have improved the recognition of at risk patients, many experience a low positive predictive value (PPV), leading to frequent false alarms.[41-43] A low PPV can contribute to alarm fatigue among health care providers, with excessive alerts desensitizing staff and delaying responses to true emergencies.

Most early warning machine learning models can incorporate a broader range of inputs beyond just vital signs[22-24,31,35-38] that may not always be readily available at bedside. With an increasing adoption of vital sign monitoring devices and wearable sensors,[44-47] developing models that rely solely on vital signs can be advantageous, given that these inputs can be more consistently available.

To address these challenges, we hypothesized that an extreme gradient boosting—based model, using readily available vital signs from a large patient cohort, could achieve a high PPV while maintaining good sensitivity for detecting low-prevalence clinical deterioration events. To achieve that, our model incorporates 3 novel techniques: a segmented window approach for extracting selective features from longitudinal vital signs data, using only vital signs as inputs and a proxy for clinician concern. We validated the model across multiple hospital sites and diverse racial groups, confirming its generalizability. Furthermore, we developed and validated alternative models using a reduced set of vital signs, highlighting their potential applicability in real-world settings where complete vital sign data or laboratory reports may be limited.

PATIENTS AND METHODS

Data Set

Data from Mayo Clinic Platform_Accelerate was used for training and validating the models. The data sources were Mayo Clinic sites across Rochester, Florida, and Arizona and the Mayo Clinic Health System (MCHS). The MCHS includes 16 hospitals and 53 clinics in rural areas. Our development focused on adult patients (aged >18 years) who were admitted to general care or telemetry beds. Data were extracted from inpatient admissions between January 1, 2019, and July 31, 2023. COVID-19 patients were included during this period, with both rapid response team activations and vital signs measurement frequency decreasing by approximately 50% from March to May 2020, before returning to pre-COVID levels. Hospitalizations were excluded if they consisted entirely of intensive care unit stays or were primarily for research, rehabilitation, or psychiatric care. Longitudinal data from eligible patients was collected, including 4 vital signs, respiration rate, HR, SpO_2, and systolic blood pressure, as well as 2 demographic variables: age and sex. These data served as the raw input for the model.

We defined an episode as the continuous period a patient spends in a general care or telemetry bed. If a deterioration event occurs, it concludes the current episode. A new episode begins when the patient starts another continuous stay in a general care or telemetry bed. As a result, 1 hospitalization could give rise to multiple distinct episodes. In total, there were 227,858 unique patients with 444,582 unique episodes.

The demographic characteristics of patients with high-quality data used for training and validation are shown in Table 1. For model validation, approximately 50% (36,784) of all episodes were set aside and kept entirely unseen during training. The remaining 50% (36,806) was used for training the model and optimizing its hyperparameters. The data split was stratified and performed at the patient level to maintain a similar ratio of deterioration events and prevent any data leakage between both sets.

Model

Supplemental Figure 1 (available online at http://www.mcpiqojournal.org) depicts the workflow of the machine learning algorithm.

Data Quality Check. To ensure data quality, we excluded time points where at least 2 of

TABLE 1. Descriptive Statistics of the Train Set and the Validation Set

Variable		Train set (n = 36,806)	Validation set (n = 36,784)
Deterioration	Total	4083 (11.1)	4011 (10.9)
	RRT	4006 (10.9)	3940 (10.7)
	Unplanned ICU Transfer	77 (0.2)	71 (0.2)
No. of time points		890,265	869,527
Sampling time (min)	25th percentile	1	1
	50th percentile	15	17
	75th percentile	95	106
Length of stay (d:h)	25th percentile	1:18	1:18
	50th percentile	2:20	2:20
	75th percentile	5:09	5:09
Patients		30,118	30,095
Sex	Male	15,552 (51.6)	15,509 (51.5)
	Female	14,566 (48.4)	14,586 (48.5)
Age (y)	18-60	10,484 (34.8)	10,600 (35.2)
	60-80	14,957 (49.7)	14,901 (49.5)
	80 and above	4620 (15.3)	4536 (15.1)
	Unknown	57 (0.2)	58 (0.2)
Race	White	27,084 (90.0)	27,071 (90.0)
	Black	1301 (4.3)	1355 (4.5)
	Asian	617 (2.0)	607 (2.0)
	Native American	311 (1.0)	303 (1.0)
	Other	418 (1.4)	397 (1.3)
	Unknown	387 (1.3)	362 (1.2)
Location	Rochester	11,764 (39.1)	11,676 (38.8)
	Arizona	6173 (20.5)	6347 (21.1)
	Florida	6616 (22.0)	6567 (21.8)
	MCHS	5565 (18.5)	5505 (18.3)

Values are n (%) unless specified.

ICU, intensive care unit; MCHS, Mayo Clinic Health System; RRT, rapid response team.

the 4 vital signs were missing. Since we aimed to preserve the integrity of the data, we chose not to implement any imputation techniques to deal with missing values. Additionally, we removed episodes that contained time gaps of more than 6 hours between consecutive measurements. This ensured that at least 3 vital signs were recorded within every 6-hour period. Furthermore, in general, a patient who did not require vital sign monitoring for more than 6 hours could be assumed to be in a more stable condition.

Extract Features. Before feature extraction, any out-of-range vital sign values were replaced with missing values. These out-of-range values were likely the result of instrument or human errors rather than accurate measurements. This step was crucial to ensure the model did not consider these erroneous values because their inclusion could significantly impact its predictions. The acceptable ranges for each vital sign are shown in Supplemental Table 1 (available online at http://www.mcpiqojournal.org).

Moreover, the time duration between each recorded time point (sampling time) was also used for feature extraction. Features were extracted using 2 running windows: the current window and the baseline window, as illustrated in Supplemental Figure 2 (available online at http://www.mcpiqojournal.org). The current window was fixed with a duration of 1 hour, while the size of the baseline window varied depending on the time point.

To extract features at a specific time point *t*:

- The current window spanned from $t - 1$ (1 hour before) to *t* (the time point).
- The baseline window ended 24 hours before *t* ($t - 24$) and started at the beginning of the episode.

For each window, the median, minimum, and maximum of the vital signs and sampling time were extracted.

A complete feature vector for a given time point comprised the statistical features described, raw vital signs, age, and sex. In total there were 37 features, with some of the features having missing values. Missing values were left unaltered because tree-based models could naturally handle them during the training and prediction processes.

Model Training. The model used is a gradient boosting tree-based algorithm called Light Gradient Boosting Machine. For episodes with a deterioration event, the feature vector from the time point closest to the event was selected for training as a positive class. For episodes without a deterioration event, a feature vector was randomly selected from any time point within the episode to represent the negative class. Hyperparameter optimization for the model was performed using grid search with 10-fold cross-validation, aiming to maximize the average precision score. Penalty for misclassifying a positive instance was scaled to be 10 times greater than that for a negative instance. Initially, this penalty was set at 8 to reflect the ratio of negative samples over positive samples noted in Table 1, but we determined that a penalty of 10 produced better results. The penalty reflected clinical priorities, where missing a true deterioration event is considered significantly costlier than generating a false alert.

Statistical Analyses

The model served as a predictive tool, estimating the probability of a deterioration event occurring within the next 24 hours. Descriptive statistics for both the training and validation sets are presented in Table 1.

To validate the model in practical use, the alert generation algorithm was designed to notify caregivers of potential deterioration events based on the model's predictions. An alert is triggered when the predicted probability exceeds a predefined threshold. Once triggered, no further alerts are allowed within the prediction period, which is set to 24 hours. Using this algorithm, the model's performance can be evaluated on the validation set as follows:

- A true positive (TP) is an alert that is followed by a deterioration event within 24 hours.
- A false positive (FP) is an alert that is not followed by a deterioration event within 24 hours.
- A false negative (FN) is a deterioration event that is not preceded by an alert within the previous 24 hours.
- A true negative (TN) is a number of days (or fraction of 24 hours) with no deterioration events and no alerts triggered in the preceding 24 hours.

Every time point within the episodes was processed and assigned a probability by the model. Based on the alert generation algorithm, an alert was either triggered or not at each time point, depending on the model's predicted probability.

The 24-hour alert suppression effectively reduces alarm fatigue and prevents double-counting TPs. However, if a FP triggers the suppression, it may cause a FN. One could

TABLE 2. Performance of Proposed Model, EDI, and MC-EWS

Metric	Our model (Rochester)	Our model (all sites)	EDI	MC-EWS
Sensitivity	0.73	0.73	0.73	0.73
Specificity	0.94	0.88	0.34	0.94
PPV	0.55	0.30	0.08	0.12
Alert rate (alerts/d)	0.13	0.17	0.68	0.07
Normalized alert rate	1.3	2.4	9.1	6.1

EDI, EPIC deterioration index; MC-EWS, Mayo Clinic Early Warning Score; PPV, positive predictive value.

argue that FPs occurring just before the 24-hour window preceding a deterioration event might be considered TPs because patient deterioration could begin around that time. Nevertheless, the number of FNs does not significantly impact the final results presented in Table 2.

With the 4 metrics explained earlier (TP, FP, FN, and TN), the C-statistic, or the area under the receiver operating characteristic (ROC) curve, can be calculated. However, when the outcome prevalence is low or the data is imbalanced, the C-statistic alone may not provide sufficient insights, and multiple metrics should be considered for a comprehensive evaluation.[48] The metrics we calculated include sensitivity, specificity, PPV, alert rate, and normalized alert rate (detailed in Supplemental Appendix 1, available online at http://www.mcpiqojournal.org).

RESULTS

Using the previously outlined validation protocol, the model demonstrated strong performance. At a sensitivity of 73.4% (95% CI, 72.2%-74.4%), it achieved a specificity of 87.6% (95% CI, 87.3%-87.8%), a PPV of 30.4% (95% CI, 29.6%-31.3%), an area under the ROC curve of 0.874 (95% CI, 0.867-0.881) and an average precision of 0.478 (95% CI, 0.460-0.496). The model also produced an alert rate of 0.170 (95% CI, 0.167-0.173) alerts per day and a normalized alert rate of 2.41 (95% CI, 2.31-2.51). For benchmarking purposes, we extracted the EDI[37] from Mayo Clinic Platform_Accelerate data and compared the model's predictions against it. The ROC curve, precision-recall curve, and normalized alert rate are shown in Figure 1.

The model's performance was stratified by hospital site and patient race to assess potential effect measure modification (Figure 2). Among the hospital sites, the Rochester site demonstrated the highest PPV at a sensitivity of 0.73, achieving a PPV of 0.55, whereas the Florida site recorded the lowest PPV at 0.09. When stratified by race, the other category exhibited the highest PPV at a sensitivity of 0.73, with a PPV of 0.45, whereas the lowest PPV was observed in the Black patient group, at 0.19.

DISCUSSION

Several related studies have explored a variety of features and machine learning architectures to predict patient deterioration.[20-40] We compared our model to the Mayo Clinic Early Warning Score (MC-EWS)[35] because it was trained on a similar data set, consisting of Mayo Clinic Rochester and Arizona admissions between January 2010 and June 2015. At a sensitivity of 0.73, the results, summarized in Table 2, show that our model achieved a higher PPV with an 18% improvement (0.30). However, it had a 6% lower specificity (0.88), and the alert rate was higher by 0.1 alerts per day (0.17). The best performance was achieved when validated on Rochester data, with a PPV of 0.55. It is important to note that our model and MC-EWS were validated on different data sets, and 2 metrics, alert rate and specificity, are strongly influenced by the length of the data set. The alert rate is inversely proportional to the total duration of the time series in the data set, whereas specificity depends on TN counts, which is also influenced by data length under our validation protocol.

In contrast, PPV and sensitivity are more comparable metrics because they rely on TP, FP, and FN counts, which are unaffected by data set length. A normalized alert rate is also more comparable because it is adjusted by the ideal alert rate, making it independent of data set length. Additionally, the deterioration event ratio in our data set was 10.9%, compared with 12.1% in the MC-EWS data set, placing both within a similar range and supporting the validity of PPV comparisons.

We believe a likely key contributor to the improved performance achieved by our model is the feature extraction approach wherein 2 aggregation windows are used: the current window and the baseline window. The current window represents the period immediately preceding the prediction, whereas the baseline window covers the time from admission up to a specified period before the prediction time. In principle, one could extend this approach to include

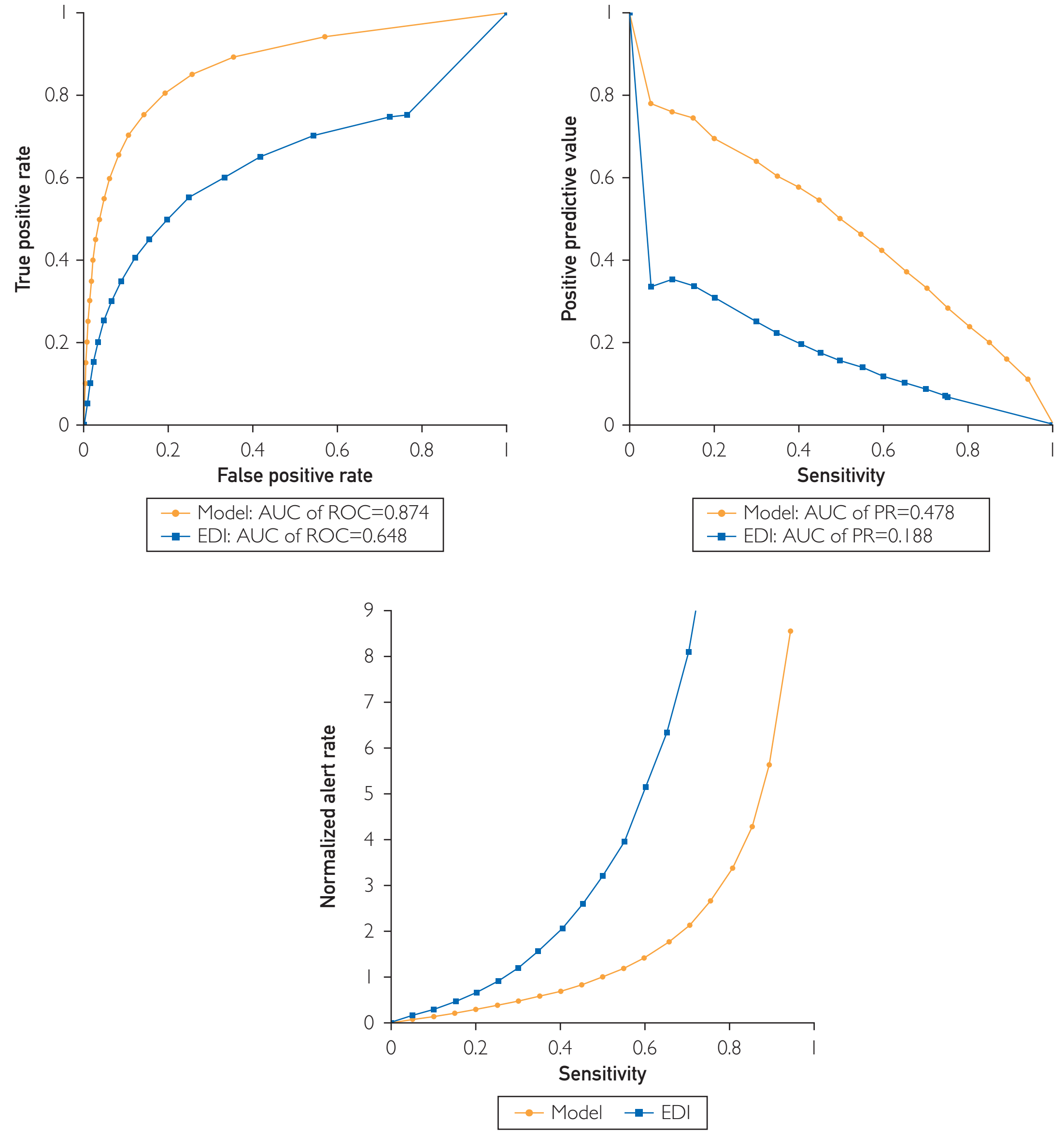

FIGURE 1. ROC curve, PR curve, and normalized alert rate of the model and EDI. AUC, area under the curve; EDI, EPIC deterioration index; PR, precision-recall; ROC, receiver operating characteristic.

multiple windows segmented at different time points and with different lengths.

The durations of these windows were tuned as hyperparameters during model optimization, with the optimal values determined to be 1 hour for the current window and 24 hours for the baseline window, as illustrated in Supplemental Figure 2. This approach suggests that the most informative features are derived from recent data within the past hour, compared against baseline trends over the preceding 24 hours. Furthermore, we observed that including raw vital signs at the current time was crucial to the model's performance.

An additional feature was sampling time, which significantly improved model performance. Sampling time can indirectly reflect a "worry factor" because shorter intervals

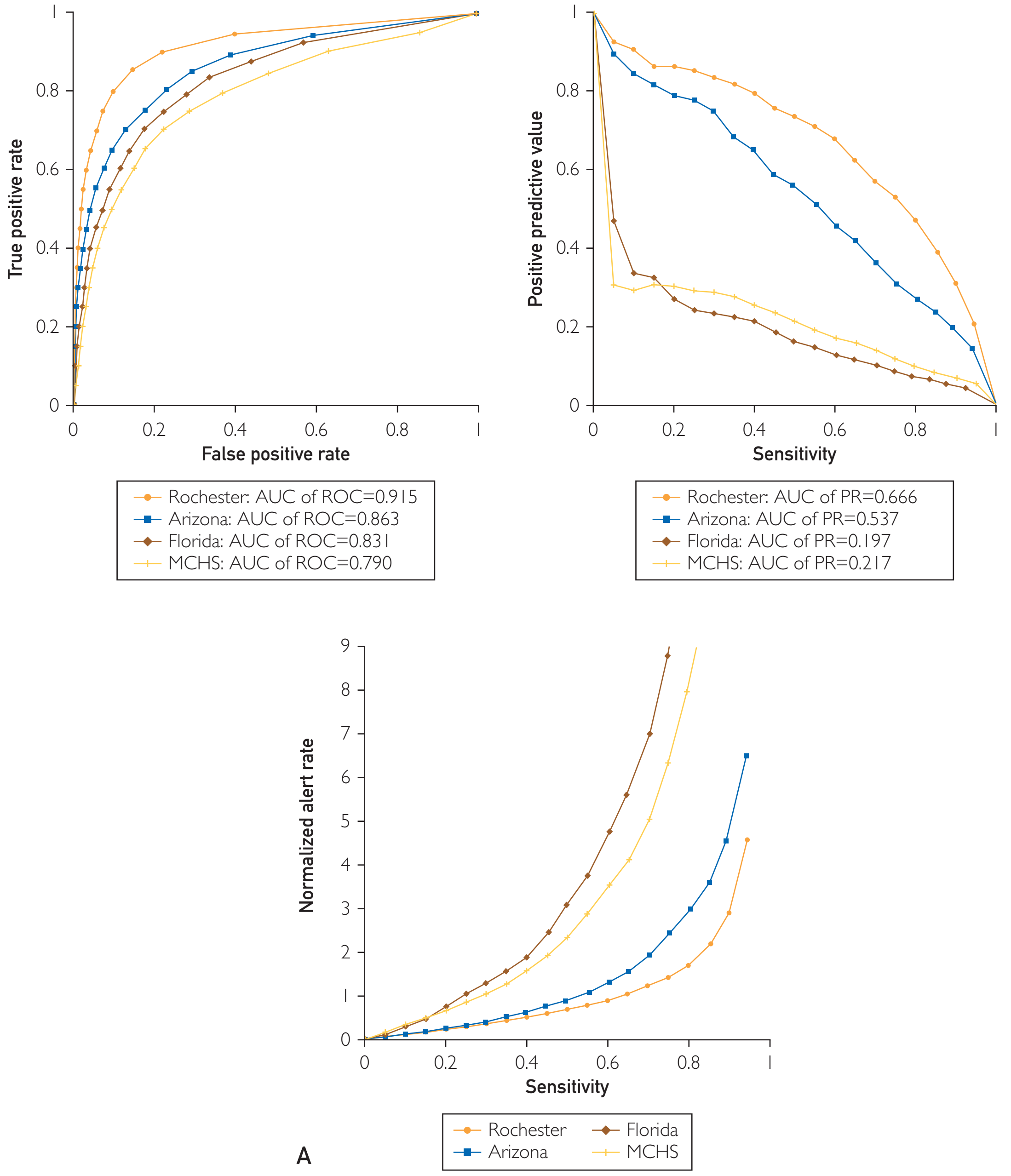

FIGURE 2. ROC curve, PR curve, and normalized alert rate of the model stratified by (A) hospital site and (B) race. AUC, area under the curve; MCHS, Mayo Clinic Health System; PR, precision-recall; ROC, receiver operating characteristic.

between measurements may indicate that nurses were more concerned about a particular patient, leading to more frequent vital sign monitoring. This worry factor has been suggested as an important predictor of clinical deterioration.[35,43]

The hospital-stratified results revealed notable performance differences across sites. The Rochester site outperformed the others, followed by Arizona, MCHS, and Florida. Rochester accounted for approximately 40% of the data (Table 1), making it possible for the model

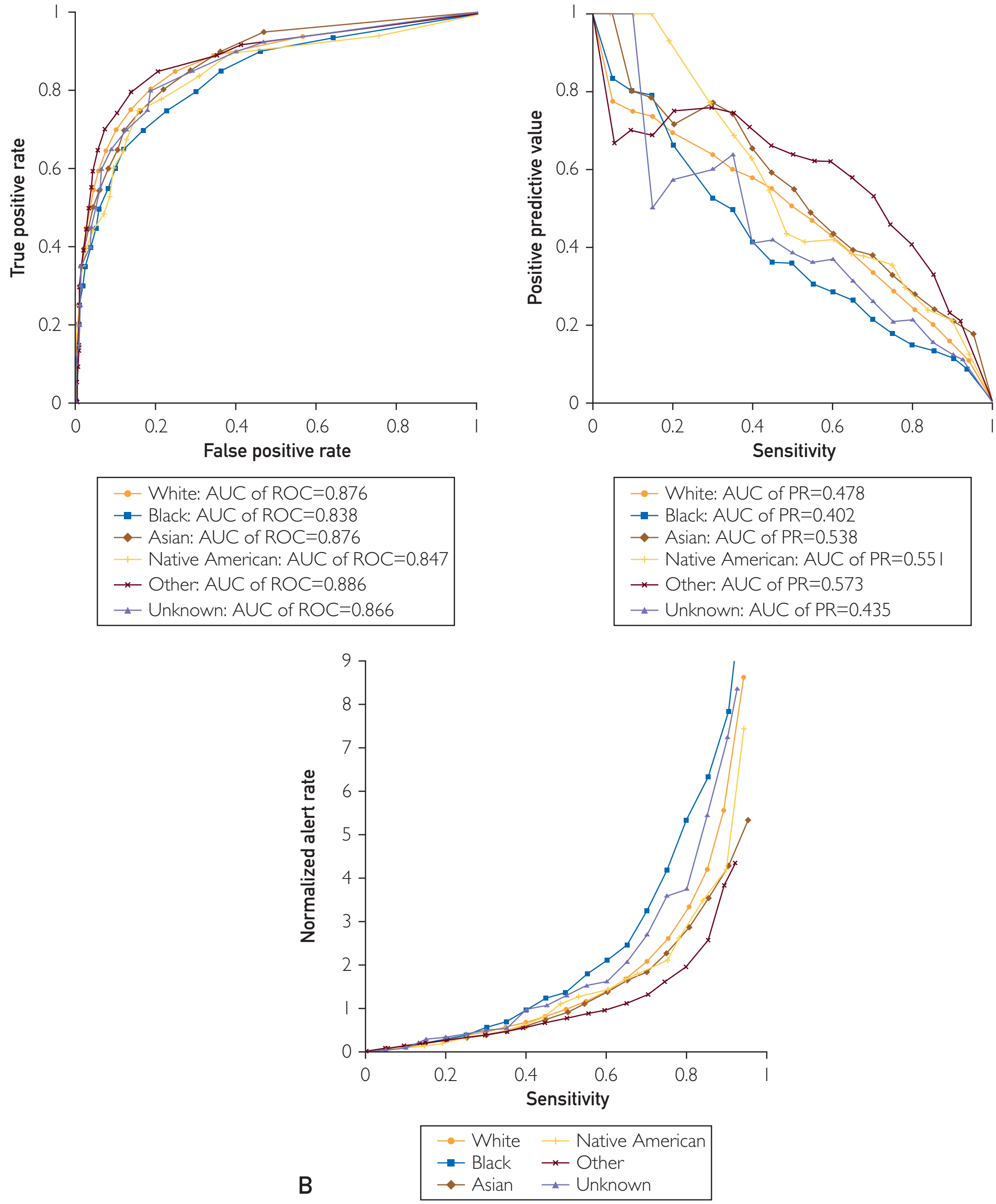

FIGURE 2. (*continued*).

to overfit to this site. However, the other 3 hospitals each contribute approximately 20% of the data, yet their performances vary widely. Arizona achieved a PPV of 0.32, whereas Florida and MCHS had PPVs of 0.09 and 0.13, respectively. This discrepancy suggests that the observed performance differences are not purely owing to overfitting; otherwise, the results across Arizona, MCHS, and Florida would be more consistent.

To better understand these disparities, we performed a statistical analysis of the data

TABLE 3. Descriptive Statistics of the Train Set and the Validation Set Stratified by Hospital

Variable		Rochester train set	Rochester validation set	Arizona train set	Arizona validation set	Florida train set	Florida validation set	MCHS train set	MCHS validation set
No. of patients		11,764	11,676	6173	6347	6616	6567	5565	5505
Age (y)	18-60	4774 (40.58)	4717 (40.40)	1948 (31.56)	2121 (33.42)	2138 (32.32)	2132 (32.47)	1624 (29.18)	1629 (29.59)
	60-80	5678 (48.27)	5709 (48.90)	3176 (51.45)	3149 (49.61)	3418 (51.66)	3356 (51.10)	2685 (48.25)	2687 (48.81)
	80 and above	1300 (11.05)	1237 (10.59)	1040 (16.85)	1066 (16.80)	1055 (15.95)	1070 (16.29)	1225 (22.01)	1163 (21.13)
	Unknown	12 (0.10)	13 (0.11)	9 (0.15)	11 (0.17)	5 (0.08)	9 (0.14)	31 (0.56)	26 (0.47)
No. of episodes		14,034	13,793	7877	8191	8425	8336	6470	6464
Deterioration	Total	1761 (12.55)	1753 (12.71)	1351 (17.15)	1358 (16.58)	524 (6.22)	459 (5.51)	447 (6.91)	441 (6.82)
	RRT	1757 (12.52)	1749 (12.68)	1332 (16.91)	1345 (16.42)	477 (5.66)	412 (4.94)	440 (6.80)	434 (6.71)
	Unplanned ICU transfer	4 (0.03)	4 (0.03)	19 (0.24)	13 (0.16)	47 (0.56)	47 (0.56)	7 (0.11)	7 (0.11)
No. of time points		327,885	301,682	173,447	183,083	267,803	267,753	121,130	117,009
Sampling time	25th percentile (min)	1	1	9	6	1	1	5	5
	50th percentile (min)	10	15	30	30	15	15	30	30
	75th percentile	57 min	1 h	3 h 10 min	3 h 6 min	2 h	2 h	2 h 42 min	2 h 51 min

ICU, intensive care unit; MCHS, Mayo Clinic Health System; RRT, rapid response team.

from each hospital (Table 3). Several key differences between hospitals were identified. First, the ratio of deterioration events varies substantially. Age distributions also differ across sites. Another notable factor is the difference in sampling time distributions, which reflect variations in standard operating procedures.

These findings underscore the complexity of hospital-level differences and their impact on model performance. Factors such as event ratios, patient demographic characteristics, and operational procedures all likely contribute to the observed variability.

Race-stratified analyses show that White patients account for approximately 90% of the dataset. At a sensitivity of 0.73, the PPV for White patients was 0.31, nearly identical to the overall model PPV of 0.30. This suggests that the model's performance is predominantly driven by White patients. Non-White racial groups each comprise approximately 2% of the data set, and their small sample sizes may lead to variability in performance metrics because the model primarily reflects characteristics of White patients de facto. However, the features used, vital signs, sex, and age, are primarily physiological and race agnostic, indicating a low risk of racial bias.[49-51]

Limitations

This study has several limitations. First, data points were considered valid only if at most 1 vital sign was missing at a time. This restriction may have led to the exclusion of potentially useful data. The use of wearable devices could help address this limitation by capturing continuous and more complete vital sign data. Nevertheless, most of the excluded data points were still accounted for in models that incorporated fewer vital signs. Second, episodes with time gaps of more than 6 hours between consecutive measurements were excluded from the analysis. Although this approach ensured data consistency, it may have led to the removal of relevant cases, potentially affecting the model's generalizability. However, we believe that this model will be suitable primarily for patients in acute care setting, hence justifying the minimum gap of 6 hours. Third, the variation in hospital-level results could be mitigated by adopting site-specific optimization strategies to enhance the generalizability and robustness of our model.

Future Research

To further improve our machine learning model, first, future work should focus on validation across broader hospital-wide populations, specifically building strategies around model configurations to fit specific hospital-level workflows. Second, one can consider integrating the model with wearable devices to better enable continuous monitoring, facilitating early detection of inpatient deterioration beyond traditional hospital environments. Third, real-time deployment within clinical workflows should be investigated to evaluate practical utility, incorporating clinician feedback to refine predictions and ensure seamless integration into decision-making processes. By embedding the model in hospital workflow systems, timely alerts can be delivered to clinicians, who can acknowledge, suppress noncritical notifications, or provide context-driven clinical advice. This approach, leveraging higher PPV and alert suppression mechanisms, aims to reduce alarm fatigue and enhance workflow efficiency, ultimately leading to better patient outcomes.

CONCLUSION

Our machine learning model, developed using 4 vital signs, respiration rate, HR, SpO_2, and systolic blood pressure, demonstrated a PPV of 0.55 for a sensitivity of 0.73, which was 6 times higher than the EPIC deterioration index (0.08) with the normalized alert rate of 1.3 on Rochester data. This improvement is attributed to the use of sampling time, which reflects the worry factor, and an advanced feature extraction approach that incorporates both a baseline window and a current window.

POTENTIAL COMPETING INTERESTS

The authors joined Mayo Clinic Platform_Accelerate to access data source and computational resources and report that part of the manuscript was submitted for allocation of date of filing to Intellectual Property Office of Singapore.

ETHICS STATEMENT

This study involved analysis of deidentified electronic health record (EHR) data via Mayo

Clinic Platform Discover. Data shown and reported in this manuscript has been extracted from the EHR using an established protocol for data extraction, aimed at preserving patient privacy. Data have been determined to be deidentified pursuant to an expert's evaluation, in accordance with the HIPAA Privacy Rule. Any data beyond what are reported in the article, including but not limited to the raw EHR data, cannot be shared or released due to the parameters of the expert determination to maintain the data deidentification. Contact corresponding authors for additional details regarding Mayo Clinic Platform Discover.

SUPPLEMENTAL ONLINE MATERIAL

Supplemental material can be found online at http://www.mcpiqojournal.org. Supplemental material attached to journal articles has not been edited, and the authors take responsibility for the accuracy of all data.

Abbreviations and Acronyms: **HR,** heart rate; **PPV,** positive predictive value; **ROC,** receiver operating characteristic; **SpO_2,** oxygen saturation

Data Previously Presented: These data were presented in the preprints with *The Lancet* on May 1st, 2025 (https://ssrn.com/abstract=5219319).

Correspondence: Address to Timothée Filhol, MiM, Respiree Private Ltd, 23 Cantonment Road, Singapore 089743, Singapore (timothee.filhol@respiree.com).

REFERENCES

1. Padilla RM, Mayo AM. Clinical deterioration: a concept analysis. *J Clin Nurs*. 2018;27(7-8):1360-1368. https://doi.org/10.1111/jocn.14238.
2. Lavoie P, Pepin J, Alderson M. Defining patient deterioration through acute care and intensive care nurses' perspectives. *Nurs Crit Care*. 2016;21(2):68-77. https://doi.org/10.1111/nicc.12114.
3. Jones D, Mitchell I, Hillman K, Story D. Defining clinical deterioration. *Resuscitation*. 2013;84(8):1029-1034. https://doi.org/10.1016/j.resuscitation.2013.01.013.
4. Buist M, Bernard S, Nguyen TV, Moore G, Anderson J. Association between clinically abnormal observations and subsequent in-hospital mortality: a prospective study. *Resuscitation*. 2004;62(2):137-141. https://doi.org/10.1016/j.resuscitation.2004.03.005.
5. Poloujadoff MP, Lapostolle F, Lockey D, et al. Survival of severely shocked patients who present with absent radial pulse and unrecordable blood pressure in the pre-hospital phase. *Resuscitation*. 2006;69(2):185-189. https://doi.org/10.1016/j.resuscitation.2005.08.016.
6. Kause J, Smith G, Prytherch D, et al. A comparison of antecedents to cardiac arrests, deaths and emergency intensive care admissions in Australia and New Zealand, and the United Kingdom—the ACADEMIA study. *Resuscitation*. 2004;62(3):275-282. https://doi.org/10.1016/j.resuscitation.2004.05.016.
7. Lourenço LBA, Meszaros MJ, Silva MFN, São-João TM. Nursing training for early clinical deterioration risk assessment: protocol for an implementation study. *JMIR Res Protoc*. 2023;12:e47293. https://doi.org/10.2196/47293.
8. Connell CJ, Endacott R, Jackman JA, Kiprillis NR, Sparkes LM, Cooper SJ. The effectiveness of education in the recognition and management of deteriorating patients: a systematic review. *Nurse Educ Today*. 2016;44:133-145. https://doi.org/10.1016/j.nedt.2016.06.001.
9. Bleyer AJ, Vidya S, Russell GB, et al. Longitudinal analysis of one million vital signs in patients in an academic medical center. *Resuscitation*. 2011;82(11):1387-1392. https://doi.org/10.1016/j.resuscitation.2011.06.033.
10. Storm-Versloot MN, Verweij L, Lucas C, et al. Clinical relevance of routinely measured vital signs in hospitalized patients: a systematic review. *J Nurs Scholarsh*. 2014;46(1):39-49. https://doi.org/10.1111/jnu.12048.
11. Gabayan GZ, Gould MK, Weiss RE, Derose SF, Chiu VY, Sarkisian CA. Emergency department vital signs and outcomes after discharge. *Acad Emerg Med*. 2017;24(7):846-854. https://doi.org/10.1111/acem.13194.
12. Winter J, Waxman MJ, Waterman G, et al. Pediatric patients discharged from the emergency department with abnormal vital signs. *West J Emerg Med*. 2017;18(5):878-883. https://doi.org/10.5811/westjem.2017.5.33000.
13. Lauteslager T, Dishakjian V, Watson L, Savarese J, Williams AJ, Leschziner GD. Detecting early signs of deterioration and preventing hospitalizations in skilled nursing facilities using remote respiratory monitoring. *Respir Med Case Rep*. 2024;50:102044. https://doi.org/10.1016/j.rmcr.2024.102044.
14. Jackson SA. Rapid response teams: what's the latest? *Nursing*. 2017;47(12):34-41. https://doi.org/10.1097/01.NURSE.0000526885.10306.21.
15. Lee A, Bishop G, Hillman KM, Daffurn K. The medical emergency team. *Anaesth Intensive Care*. 1995;23(2):183-186. https://doi.org/10.1177/0310057X9502300210.
16. DeVita MA, Smith GB, Adam SK, et al. "Identifying the hospitalised patient in crisis"—a consensus conference on the afferent limb of rapid response systems. *Resuscitation*. 2010;81(4):375-382. https://doi.org/10.1016/j.resuscitation.2009.12.008.
17. Tirkkonen J, Tamminen T, Skrifvars MB. Outcome of adult patients attended by rapid response teams: a systematic review of the literature. *Resuscitation*. 2017;112:43-52. https://doi.org/10.1016/j.resuscitation.2016.12.023.
18. Jung B, Daurat A, De Jong A, et al. Rapid response team and hospital mortality in hospitalized patients. *Intensive Care Med*. 2016;42(4):494-504. https://doi.org/10.1007/s00134-016-4254-2.
19. van Galen LS, Struik PW, Driesen BE, et al. Delayed recognition of deterioration of patients in general wards is mostly caused by human related monitoring failures: a root cause analysis of unplanned ICU admissions. *PLoS One*. 2016;11(8):e0161393. https://doi.org/10.1371/journal.pone.0161393.
20. Liu VX, Lu Y, Carey KA, et al. Comparison of early warning scoring systems for hospitalized patients with and without infection at risk for in-hospital mortality and transfer to the intensive care unit. *JAMA Netw Open*. 2020;3(5):e205191. https://doi.org/10.1001/jamanetworkopen.2020.5191.
21. Mann KD, Good NM, Fatehi F, et al. Predicting patient deterioration: a review of tools in the digital hospital setting. *J Med Internet Res*. 2021;23(9):e28209. https://doi.org/10.2196/28209.
22. Churpek MM, Yuen TC, Winslow C, Meltzer DO, Kattan MW, Edelson DP. Multicenter comparison of machine learning methods and conventional regression for predicting clinical deterioration on the wards. *Crit Care Med*. 2016;44(2):368-374. https://doi.org/10.1097/CCM.0000000000001571.
23. Wengerter BC, Pei KY, Asuzu D, Davis KA. Rothman Index variability predicts clinical deterioration and rapid response

activation. *Am J Surg*. 2018;215(1):37-41. https://doi.org/10.1016/j.amjsurg.2017.07.031.

24. Kipnis P, Turk BJ, Wulf DA, et al. Development and validation of an electronic medical record-based alert score for detection of inpatient deterioration outside the ICU. *J Biomed Inform*. 2016;64:10-19. https://doi.org/10.1016/j.jbi.2016.09.013.
25. Ludikhuize J, Smorenburg SM, de Rooij SE, de Jonge E. Identification of deteriorating patients on general wards; measurement of vital parameters and potential effectiveness of the Modified Early Warning Score. *J Crit Care*. 2012;27(4). 424.e7-424.e13, https://doi.org/10.1016/j.jcrc.2012.01.003.
26. Churpek MM, Adhikari R, Edelson DP. The value of vital sign trends for detecting clinical deterioration on the wards. *Resuscitation*. 2016;102:1-5. https://doi.org/10.1016/j.resuscitation.2016.02.005.
27. Byrd TF IV, Southwell B, Ravishankar A, et al. Validation of a proprietary deterioration index model and performance in hospitalized adults. *JAMA Netw Open*. 2023;6(7):e2324176. https://doi.org/10.1001/jamanetworkopen.2023.24176.
28. Brekke IJ, Puntervoll LH, Pedersen PB, Kellett J, Brabrand M. The value of vital sign trends in predicting and monitoring clinical deterioration: a systematic review. *PLoS One*. 2019;14(1):e0210875. https://doi.org/10.1371/journal.pone.0210875.
29. Hammond NE, Spooner AJ, Barnett AG, Corley A, Brown P, Fraser JF. The effect of implementing a modified early warning scoring (MEWS) system on the adequacy of vital sign documentation. *Aust Crit Care*. 2013;26(1):18-22. https://doi.org/10.1016/j.aucc.2012.05.001.
30. Churpek MM, Yuen TC, Edelson DP. Predicting clinical deterioration in the hospital: the impact of outcome selection. *Resuscitation*. 2013;84(5):564-568. https://doi.org/10.1016/j.resuscitation.2012.09.024.
31. Smith GB, Prytherch DR, Meredith P, Schmidt PE, Featherstone PI. The ability of the National Early Warning Score (NEWS) to discriminate patients at risk of early cardiac arrest, unanticipated intensive care unit admission, and death. *Resuscitation*. 2013;84(4):465-470. https://doi.org/10.1016/j.resuscitation.2012.12.016.
32. Verma AA. Toward the rigorous evaluation of early warning scores. *JAMA Netw Open*. 2024;7(10):e2438966. https://doi.org/10.1001/jamanetworkopen.2024.38966.
33. Kim SY, Kim S, Cho J, et al. A deep learning model for real-time mortality prediction in critically ill children. *Crit Care*. 2019;23(1):279. https://doi.org/10.1186/s13054-019-2561-z.
34. Smith GB, Prytherch DR, Schmidt PE, Featherstone PI. Review and performance evaluation of aggregate weighted 'track and trigger' systems. *Resuscitation*. 2008;77(2):170-179. https://doi.org/10.1016/j.resuscitation.2007.12.004.
35. Romero-Brufau S, Whitford D, Johnson MG, et al. Using machine learning to improve the accuracy of patient deterioration predictions: Mayo Clinic Early Warning Score (MC-EWS). *J Am Med Inform Assoc*. 2021;28(6):1207-1215. https://doi.org/10.1093/jamia/ocaa347.
36. Escobar GJ, Liu VX, Schuler A, Lawson B, Greene JD, Kipnis P. Automated identification of adults at risk for in-hospital clinical deterioration. *N Engl J Med*. 2020;383(20):1951-1960. https://doi.org/10.1056/NEJMsa2001090.
37. Weissman GE. A bold first toe into the uncharted waters of evaluating proprietary clinical prediction models. *Ann Am Thorac Soc*. 2021;18(7):1116-1117. https://doi.org/10.1513/AnnalsATS.202103-332ED.
38. Edelson DP, Churpek MM, Carey KA, et al. Early warning scores with and without artificial intelligence. *JAMA Netw Open*. 2024;7(10):e2438986. https://doi.org/10.1001/jamanetworkopen.2024.38986.
39. Romero-Brufau S, Huddleston JM, Naessens JM, et al. Widely used track and trigger scores: are they ready for automation in practice? *Resuscitation*. 2014;85(4):549-552. https://doi.org/10.1016/j.resuscitation.2013.12.017.
40. Rakers MM, van Buchem MM, Kucenko S, et al. Availability of evidence for predictive machine learning algorithms in primary care: a systematic review. *JAMA Netw Open*. 2024;7(9):e2432990. https://doi.org/10.1001/jamanetworkopen.2024.32990.
41. Patel MS, Jones MA, Jiggins M, Williams SC. Does the use of a "track and trigger" warning system reduce mortality in trauma patients? *Injury*. 2011;42(12):1455-1459. https://doi.org/10.1016/j.injury.2011.05.030.
42. Bedoya AD, Clement ME, Phelan M, Steorts RC, O'Brien C, Goldstein BA. Minimal impact of implemented early warning score and best practice alert for patient deterioration. *Crit Care Med*. 2019;47(1):49-55. https://doi.org/10.1097/CCM.0000000000003439.
43. Romero-Brufau S, Gaines K, Nicolas CT, Johnson MG, Hickman J, Huddleston JM. The fifth vital sign? Nurse worry predicts inpatient deterioration within 24 hours. *JAMIA Open*. 2019;2(4):465-470. https://doi.org/10.1093/jamiaopen/ooz033.
44. Weenk M, Koeneman M, van de Belt TH, Engelen LJLPG, van Goor H, Bredie SJH. Wireless and continuous monitoring of vital signs in patients at the general ward. *Resuscitation*. 2019;136:47-53. https://doi.org/10.1016/j.resuscitation.2019.01.017.
45. Leenen JPL, Leerentveld C, van Dijk JD, van Westreenen HL, Schoonhoven L, Patijn GA. Current evidence for continuous vital signs monitoring by wearable wireless devices in hospitalized adults: systematic review. *J Med Internet Res*. 2020;22(6):e18636. https://doi.org/10.2196/18636.
46. Areia C, Biggs C, Santos M, et al. The impact of wearable continuous vital sign monitoring on deterioration detection and clinical outcomes in hospitalised patients: a systematic review and meta-analysis. *Crit Care*. 2021;25(1):351. https://doi.org/10.1186/s13054-021-03766-4.
47. Eddahchouri Y, Peelen RV, Koeneman M, Touw HRW, van Goor H, Bredie SJH. Effect of continuous wireless vital sign monitoring on unplanned ICU admissions and rapid response team calls: a before-and-after study. *Br J Anaesth*. 2022;128(5):857-863. https://doi.org/10.1016/j.bja.2022.01.036.
48. Romero-Brufau S, Huddleston JM, Escobar GJ, Liebow M. Why the C-statistic is not informative to evaluate early warning scores and what metrics to use. *Crit Care*. 2015;19(1):285. https://doi.org/10.1186/s13054-015-0999-1.
49. Gathers CL, Morgan RW, Alvey JS, et al. Race is not associated with medical emergency team evaluation prior to in-hospital cardiac arrest. *Resuscitation*. 2025;211:110609. https://doi.org/10.1016/j.resuscitation.2025.110609.
50. Joseph L, Chan PS, Bradley SM, et al. Temporal changes in the racial gap in survival after in-hospital cardiac arrest. *JAMA Cardiol*. 2017;2(9):976-984. https://doi.org/10.1001/jamacardio.2017.2403.
51. Weisfeldt M, Becker L. Racial differences in in-hospital cardiac arrest: good news: cautious optimism is welcome. *JAMA Cardiol*. 2017;2(9):984-985. https://doi.org/10.1001/jamacardio.2017.2432.