

Perspective

Intelligent Clinical Decision Support

Michael R. Pinsky ^{1,*} , Artur Dubrawski ²  and Gilles Clermont ¹

¹ Department of Critical Care Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA 15261, USA; cler@pitt.edu

² Auton Laboratory, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA; awd@cs.cmu.edu

* Correspondence: pinsky@pitt.edu

Abstract: Early recognition of pathologic cardiorespiratory stress and forecasting cardiorespiratory decompensation in the critically ill is difficult even in highly monitored patients in the Intensive Care Unit (ICU). Instability can be intuitively defined as the overt manifestation of the failure of the host to adequately respond to cardiorespiratory stress. The enormous volume of patient data available in ICU environments, both of high-frequency numeric and waveform data accessible from bedside monitors, plus Electronic Health Record (EHR) data, presents a platform ripe for Artificial Intelligence (AI) approaches for the detection and forecasting of instability, and data-driven intelligent clinical decision support (CDS). Building unbiased, reliable, and usable AI-based systems across health care sites is rapidly becoming a high priority, specifically as these systems relate to diagnostics, forecasting, and bedside clinical decision support. The ICU environment is particularly well-positioned to demonstrate the value of AI in saving lives. The goal is to create AI models embedded in a real-time CDS for forecasting and mitigation of critical instability in ICU patients of sufficient readiness to be deployed at the bedside. Such a system must leverage multi-source patient data, machine learning, systems engineering, and human action expertise, the latter being key to successful CDS implementation in the clinical workflow and evaluation of bias. We present one approach to create an operationally relevant AI-based forecasting CDS system.

Keywords: database; machine learning; hemodynamic monitoring; predictive analytics



Citation: Pinsky, M.R.; Dubrawski, A.; Clermont, G. Intelligent Clinical Decision Support. *Sensors* **2022**, *22*, 1408. <https://doi.org/10.3390/s22041408>

Academic Editors: Yangquan Chen, Nunzio Cennamo, M. Jamal Deen, Subhas Mukhopadhyay, Simone Morais and Junseop Lee

Received: 27 January 2022

Accepted: 6 February 2022

Published: 12 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

There is a need for Artificial Intelligence (AI)-based tools in acute care environments to aid in the early detection and assessment of cardiorespiratory insufficiency (CRI) because the amount of information exceeds human capacity to process it, internalize the extracted knowledge, and then act upon it consistently and appropriately. The new onset of CRI is common in acutely ill hospitalized patients. Misdiagnosis and/or delayed treatment leads to increased morbidity, mortality, and cost of care [1]. Instability manifests through acute but frequently subtle changes in vital signs (VS) trends indicative of attempted compensation and evolving decompensation. Decompensation and instability occur even in highly monitored patient care environments such as in the Intensive Care Unit, and the longer a patient is in such a decompensated state the more difficult it is to mitigate or reverse resultant damage at the organ and cellular levels [2,3]. Being able to detect impending instability, rather than detecting its presence at a late stage, could permit timely stabilization, thereby reducing morbidity, mortality, and resource use. However, the technology needed to forecast impending instability is not well developed. Many early alerting approaches, once embedded in Electronic Health Record (EHR) systems, have subsequently been rolled back for a variety of reasons, including unacceptable performance, lack of perceived clinical usefulness, interference with existing workflows, and increased clinician documentation burden [4,5]. There is a clear need for better performing, trustworthy models that can be effectively melded into the clinical workflow. We and others have used machine learning

(ML) to detect patterns predictive of impending instability before overt manifestations occur from real-time physiologic monitoring, often linked to EHR data. Such advanced intelligent clinical decision support (CDS) is only the first step towards developing AI-based systems that provide trustworthy, personalized predictions and recommendations to preemptively mitigate instability [6,7], hopefully leading to improved patient-centered outcomes.

Importantly, CRI usually develops over time and, therefore, it can potentially be predicted. Many researchers have demonstrated the feasibility of forecasting its overt onset. For instance, we have shown in animal models, step-down units, and ICUs that, typically, hemodynamic and respiratory instability develops over a time scale where clinical mitigations could be initiated in advance of the overt manifestations of instability [8–20]. However, creating AI tools with good performance is not enough. This perspective paper describes one approach to creating an operationally relevant AI-based forecasting CDS system.

2. Challenges of AI-Based CDS

AI-based tools must be deemed useful and trustworthy by end-users. There are two primary and very different challenges associated with using AI-based CDS tools at the bedside. The first is a lack of clinician trust in the system, a subject we and others have studied [21]. The other, less well-recognized challenge relates to the over-reliance on AI-based decision support [22–24]. CDS tools should be developed using sound human factor engineering principles to minimize information overload, to operate cohesively within the clinical workflow, and in support of an a priori expectation of incremental usefulness if validated.

Regrettably, AI-based tools may be biased. Bias in AI algorithms has been known to plague business applications for years. More recently, much emphasis has been put on exploring sources of biases in healthcare AI, including algorithms based on EHR [25–29]. Several landmark papers demonstrated that the naïve application of ML models to data may lead to erroneous conclusions or lack robustness across populations and subgroups of interest [30]. Statisticians have developed sophisticated methods to deal with observed and unobserved confounding, such as propensity-based methods [31,32]. Moreover, the issue of fairness in machine learning has recently become a very active research area, especially in healthcare applications [33]. Future predictive models should go beyond a simple examination of the impact of sex, race, and age as biological variables on model predictions. They should systematically explore data, algorithms, and results for the presence of bias. Potentially, models drawn from high-frequency data (i.e., 1 Hz and 250 Hz), when available, may be less exposed to certain forms of bias. However, this assumption needs to be validated across different clinical datasets to ensure their broad and consistent applicability [34].

Hence, AI-based tools need to be robust across environments. External validation of prediction algorithms and CDS tools is essential for the scalable impact of AI-based solutions. Creating generalizable tools is a difficult problem, for which there exist several non-mutually exclusive approaches. These approaches include: (1) external testing and transfer learning; (2) learning from multi-site broadly representative datasets; (3) federated approaches to learning. There are challenges and limitations to each of these methods. The current cybersecurity and privacy protection climate around health care data severely impedes the sharing of large datasets, including patient-level data. There is also a limit as to the notion of the generalizability of models. After all, predictions need to be accurate in a local environment, as trust is developed locally. When and how generally robust models can enhance local performance at sites where they were not developed remains to be explored [35].

Performing AI-based CDS in real time poses additional challenges. Although there are a growing number of tools for real-time decision support, most are rule-based and focused on detection (e.g., bedside alarms), rather than on personalized forecasting and therapeutics. There are some recent inspiring examples of promising work towards automated EHR-

based early warning systems [36–38] and the forecasting of hypotension [39], which have impacted clinical workflow. However, current state of the art applications do not draw from multi-resolution, multi-domain data, including unstructured EHR data and monitor-derived waveform data. Furthermore, system architecture to build real-time AI-based systems needs to be not only developed locally, but also to maintain inter-institutional applicability. Such system requirements are not trivial to satisfy in practice.

3. Examples of Forecasting and Phenotyping Instability in the ICU

We have developed predictive models using controlled animal laboratory data and historical ICU data that demonstrate good performance in predicting clinically relevant tachycardia, hypotension, and bleeding. We also demonstrated the incremental benefit of high-frequency data in increasing the reliability of these prediction models [8,17,39–41]. For instance, advanced signal processing predicted clinically relevant tachycardia and hypotension in ICU patients [42,43]. To determine a clinically relevant tachycardia target in these studies, we first examined tachycardia's impact upon outcomes from the MIMIC-III database. We found that although increasing HR > 100/min was associated with progressively increasing vasopressor use, morbidity, and mortality, a clear step-up in the length of stay and mortality occurred with HR > 130/min. Those tachycardia patients with a HR > 130/min had increased vasopressor support, longer ICU stays, and increased ICU mortality. Thus, we defined clinically relevant tachycardia as HR > 130/min, lasting ≥ 5 min with >10% density of occurrence over this time interval. Using data sampled at 1 Hz from 2809 subjects, classifiers were trained to create a risk score for future tachycardia [44]. Risk trajectory was generated from time windows moving away from the tachycardia event at 1-min increments. The classifiers performed generally well. The area under receiver operating characteristic curve (AUROC) score ranged from 0.842 when a regularized logistic regression model was used to 0.921 when a random forest (RF) classifier was used. Risk trajectory analysis showed average risks for the tachycardia group of 0.78 just before the tachycardia episodes, while control group risks remained <0.3 and with significant separation between subsequent tachycardia and control stable patients at ~75 min before the initial tachycardia event (Figure 1).

We also applied AI tools to predict hypotension, defined as systolic blood pressure < 90 mmHg and a mean arterial pressure <60 mmHg for ≥ 5 measurements within 10 min [13]. We used an RF classifier to predict hypotension and performance was measured by AUROC and the area under the precision-recall curve (AUPRC). We identified 1307 cases and matched them to 1619 non-hypotensive controls. The RF model showed AUROC of 0.93, 0.91, and 0.88 at 15, 30, and 60 min, respectively, before hypotension and AUPRC of 0.77 at 60 min before hypotension. Mean risk trajectory showed a clear separation from mean control risk trajectory >15 min before hypotension in 80% of cases. Since alerts predicting impending hypotension may also cause alarm fatigue if they are presented to the bedside clinicians too often, a second-level RF model analyzed the recent shape of the risk trajectory, combined with the existence of prior alerts, to generate potential alerts. We then imposed a lockout period of 15 min, where the system would not re-alert, even if alert conditions persisted. The resulting alert system produced on average 0.79 alerts/subject/hour, with a positive predictive value (probability of developing hypotension) of 65% and sensitivity of 92.4%. Thus, using this strategy to minimize alarm events and alarm fatigue did not materially impede the performance of this model.

We formally evaluated improvements in the performance of an instability model attainable when moving from non-invasive monitoring (NIM) to adding central venous, pulmonary artery, and arterial catheterization-derived variables and, as their sampling frequency was increased, from simple metrics (SM), computed every minute, to heart-beat-to-beat (B2B), to waveform (WF) with and without personalized stable baseline reference in our porcine 20 mL/min bleed cohort [41]. RF classification was used to identify the onset of bleeding. Model performance was evaluated using the AUROC curve, as well as the activity monitoring operating characteristic (AMOC) curve. The AMOC curve displays the

tradeoffs between an earlier time to detection (i.e., how early bleeding can be detected after its onset) and increased false alarm rate (FPR). Referencing models to a personal stable baseline before a bleed improved bleed detection, as did an increase in data granularity from SM to B2B to WF. All invasive monitoring out-performed NIM (Figure 2). Thus, these data demonstrate that using more invasive monitoring-related data, increasing sampling frequency, and referencing to a personal baseline, cumulatively improve the detection of bleeding onset.

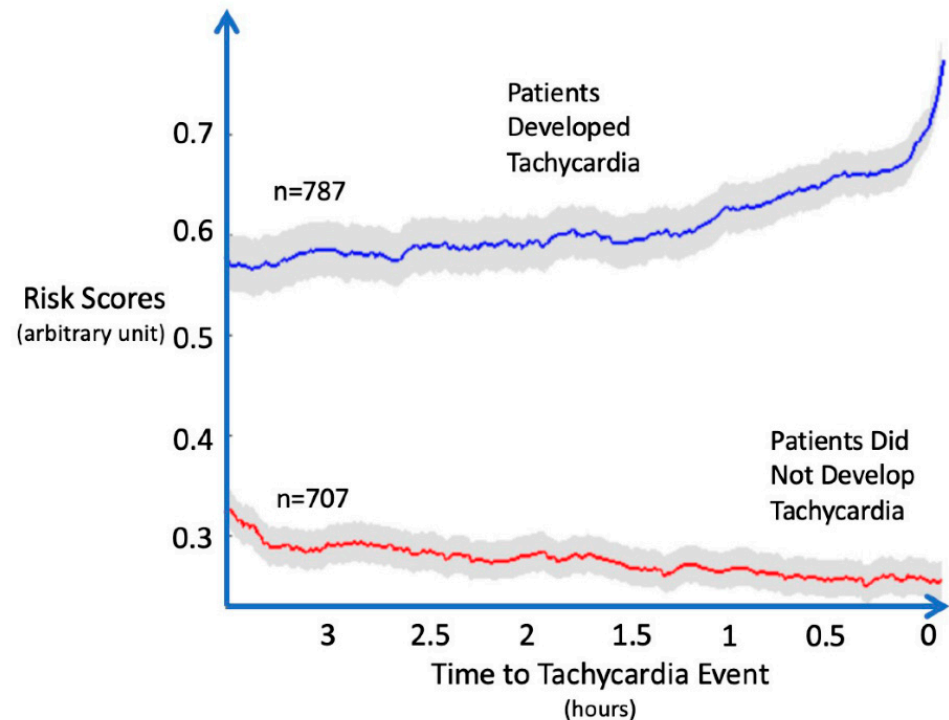


Figure 1. Model prediction of initial tachycardic episode using external control data matched for every episode of tachycardia. Comparison of a model trained on MIMIC-II data to identify an initial episode of tachycardia (heart rate (HR) > 130/min) in an external validation cohort from that same database. Results are shown as risk score changes over time as the future tachycardic and non-future tachycardic (control) groups move toward the event. The control group’s time series data were time-matched to correspond to the future tachycardic group’s time in the ICU. Data derived from Yoon et al. [43].

The limited availability of sufficiently large collections of clinically assessed reference data is one of the major bottlenecks preventing the wider development and adoption of robust AI-based CDS in practice. However, reviewing large amounts of clinical time series data to identify CRI can be both time-consuming and fraught with inter-rater variability of labeled events. To address these issues, we developed and applied an efficient protocol for a multi-expert, multi-tier ground truth elicitation framework with application to artifact classification for predicting patient instability [44], efficiently utilizing precious human expertise, and yielding accurate downstream models with one-quarter of the amount of time needed if conducted by the content experts manually. We also developed an active learning algorithm that prioritized which instances of data should be labeled by humans to maximally boost the eventual performance models, further reducing by half the need for direct visualization reviews while maintaining human interpretability of resulting model predictions [45,46]. In addition to being labor-intensive, the clinical expert data annotation process is often prone to error and uncertainty, especially if the cases are to be assessed individually. Our studies showed that more reliable labels of real versus artifact, or minor versus clinically relevant instability, can be collected by asking other kinds of questions. For instance, supplementing direct labeling of each data instance with qualitative comparisons

such as: “Comparing patient A and B, who appears healthier?” [47,48]. Answering such questions is often easier, yields more reliable labels, and requires less annotation effort to achieve equivalent performance. We have also demonstrated how to completely avoid the laborious process of pointillistic labeling of reference data for clinical applications of AI using weak supervision [49]. By harvesting multiple labeling functions that a human expert would use in their mind to assess each case at hand, one can automate the process of data annotation. This is particularly appealing when facing large amounts of clinical data needing annotation. In one such exercise, we have shown that a handful of labeling functions derived from basic clinical knowledge can eliminate the need for manual data annotation and yield predictive models of performance comparable to the equivalent models trained on data point-by-point labeled by expert clinicians when evaluated on the task of detecting arrhythmia in ECG signals [49].

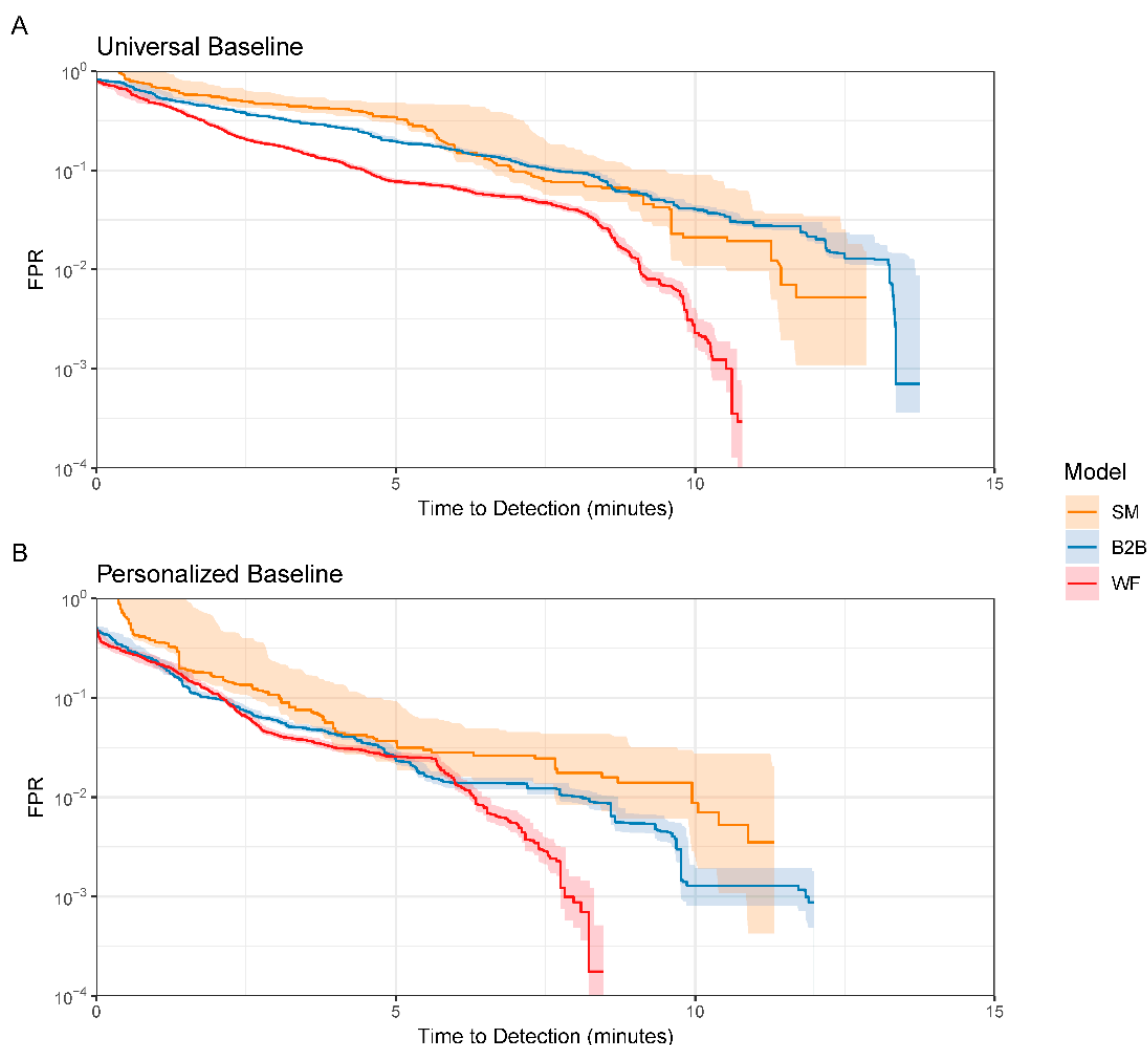


Figure 2. Activity monitoring operating characteristic analysis of models developed with increasingly granular arterial pressure physiologic data, for models developed using a universal baseline (A), and models developed using a personalized baseline (B) (see text for details). Displayed as the time to detection of bleeding versus false-positive rate for arterial catheter data only for increasing granularity levels: simple metrics (SM), beat-to-beat (B2B), and waveform (WF). Data displayed with shading equal to 95% confidence range. Data derived from Pinsky et al. [41].

Finally, the importance of a usable and informative human–computer interface contributes to user trust in a major way and is often overlooked. Unless a CDS works autonomously, it requires a machine–human interface to allow the clinicians to both oper-

ationalize the alerts and corresponding information and to audit its performance. Traditionally, such interfaces included a visual display from an electronic monitor, either a bedside monitor or handheld device such as a tablet or smartphone. We and others have proactively involved the end-users in graphical user interface (GUI) development [50]. We demonstrated that clinical end-users are usually eager to engage in GUI development and readily provide useful feedback on issues related to both the GUI design and the foundational CDS architecture. This feedback revealed specific items that bedside clinician users found important, including a better trend evolution display and context of alert relative to overall care. Unfortunately, bypassing input from clinical experts during the early-stage development of a GUI CDS system is quite common [51,52].

4. Conclusions

AI-based CDS processes need to be highly iterative using multiple pathways and forms of feedback involving both the modelers and the target clinicians, linked by *in silico* trials of effectiveness and acceptance by end-users and using clear metrics of success. One can never fully eliminate bias, but newer AI approaches may be able to mitigate several of the sources of bias CDS systems contend with, yielding diagnostic, predictive, or prescriptive tools that optimize accuracy and preserve the fairness of the recommended decisions across subpopulations. Any AI-based CDS will have a finite lifecycle and will require periodic reevaluation and refinement to adapt to changing patient demographics, data ecosystems, bedside workflows, and evolution in clinical practices if they are to sustain effectiveness and trustworthiness. Some of these needs can be automated or semi-automated with the use of AI. For instance, efficient methods for acquiring training information for models will likely be able to facilitate such adaptivity at operationally feasible costs. End-user involvement in the design and evaluation of CDS systems across their lifecycle will also promote trustworthiness, adoption into practice, and sustainability.

Future research needs to optimally leverage multi-institutional datasets to not only develop clinically relevant predictive models, but also to establish effective and efficient methodologies to plumb these databases within the context of health information security and democratization. Such broad-based efforts are underway [53] and, hopefully, will lead to novel and insightful ways of joining models and applications across groups of investigators and across healthcare facilities, once developed.

Author Contributions: Conceptualization, all authors; methodology, all authors; software, A.D.; validation, all authors.; formal analysis, A.D.; investigation, M.R.P. and G.C.; resources, M.R.P. and G.C.; data curation, G.C. and A.D.; writing—original draft preparation, M.R.P.; writing—review and editing, all authors.; visualization, A.D.; supervision, M.R.P.; project administration, M.R.P.; funding acquisition, all authors. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by NIH R0-GM117622 and R01-NR013912, DoD award W81XWH-19-C-0101, and DARPA award FA8750-17-2-0130.

Institutional Review Board Statement: The study was approved by the Institutional Animal Care and Use Committee Review Board of the University of Pittsburgh Protocol number 19014099 PHS Assurance Number: D16-00118, for studies involving animals.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are available upon request. Contact pinskyme@pitt.edu to initiate such a request.

Conflicts of Interest: Michael Pinsky is a founding member of Intelomed, Inc. Artur Dubrawski is a founding member of Auton Systems LLC, and Marinus Analytics LLC. Gilles Clermont is Chief Medical Officer of NOMA AI, Inc.

References

1. Galhotra, S.; Scholle, C.C.; Dew, M.A.; Mininni, N.C.; Clermont, G.; DeVita, M.A. Medical emergency teams: A strategy for improving patient care and nursing work environments. *J. Adv. Nurs.* **2006**, *55*, 180–187. [[CrossRef](#)] [[PubMed](#)]
2. Burke, J.R.; Downey, C.; Almoudaris, A.M. Failure to Rescue Deteriorating Patients: A Systematic Review of Root Causes and Improvement Strategies. *J. Patient Saf.* **2022**, *18*, e140–e155. [[CrossRef](#)]
3. Hall, K.K.; Lim, A.; Gale, B. The Use of Rapid Response Teams to Reduce Failure to Rescue Events: A Systematic Review. *J. Patient Saf.* **2020**, *16*, S3–S7. [[CrossRef](#)]
4. Ginestra, J.C.; Giannini, H.; Schweickert, W.D.; Meadows, L.; Lynch, M.J.; Pavan, K.; Chivers, C.J.; Draugelis, M.; Donnelly, P.J.; Fuchs, B.D.; et al. Clinician Perception of a Machine Learning–Based Early Warning System Designed to Predict Severe Sepsis and Septic Shock. *Crit. Care Med.* **2019**, *47*, 1477–1484. [[CrossRef](#)] [[PubMed](#)]
5. Wong, A.; Otles, E.; Donnelly, J.P.; Krumm, A.; McCullough, J.; DeTroyer-Cooley, O.; Pestue, J.; Phillips, M.; Konye, J.; Penzoza, C.; et al. External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients. *JAMA Intern. Med.* **2021**, *181*, 1065–1070. [[CrossRef](#)] [[PubMed](#)]
6. Pinsky, M.R.; Payen, D. Functional hemodynamic monitoring. *Crit. Care.* **2005**, *9*, 566–572. [[CrossRef](#)]
7. Pinsky, M.R.; Dubrawski, A. Gleaning Knowledge from Data in the Intensive Care Unit. *Am. J. Respir. Crit. Care Med.* **2014**, *190*, 606–610. [[CrossRef](#)] [[PubMed](#)]
8. Wertz, A.; Holder, A.L.; Guillame-Bert, M.; Clermont, G.; Dubrawski, A.; Pinsky, M.R. Increasing Cardiovascular Data Sampling Frequency and Referencing It to Baseline Improve Hemorrhage Detection. *Crit. Care Explor.* **2019**, *1*, e0058. [[CrossRef](#)]
9. Hravnak, M.; DeVita, M.A.; Clontz, A.; Edwards, L.; Valenta, C.; Pinsky, M.R. Cardiorespiratory instability before and after implementing an integrated monitoring system. *Crit. Care Med.* **2011**, *39*, 65–72. [[CrossRef](#)]
10. Cancio, L.C.; Batchinsky, A.I.; Salinas, J.; Kuusela, T.; Convertino, V.A.; Wade, C.E.; Holcomb, J.B. Heart-Rate Complexity for Prediction of Prehospital Lifesaving Interventions in Trauma Patients. *J. Trauma Inj. Infect. Crit. Care* **2008**, *65*, 813–819. [[CrossRef](#)]
11. Batchinsky, A.I.; Cancio, L.C.; Salinas, J.; Kuusela, T.; Cooke, W.H.; Wang, J.J.; Boehme, M.; Convertino, V.A.; Holcomb, J.B. Prehospital Loss of R-to-R Interval Complexity is Associated with Mortality in Trauma Patients. *J. Trauma Inj. Infect. Crit. Care* **2007**, *63*, 512–518. [[CrossRef](#)]
12. Hu, X. An algorithm strategy for precise patient monitoring in a connected healthcare enterprise. *NPJ Digit. Med.* **2019**, *2*, 30. [[CrossRef](#)] [[PubMed](#)]
13. Holder, A.L.; Clermont, G. Using what you get: Dynamic physiologic signatures of critical illness. *Crit. Care Clin.* **2015**, *31*, 133–164. [[CrossRef](#)] [[PubMed](#)]
14. Guillame-Bert, M.; Dubrawski, A.; Wang, D.; Hravnak, M.; Clermont, G.; Pinsky, M.R. Learning temporal rules to forecast instability in continuously monitored patients. *J. Am. Med. Inform. Assoc.* **2016**, *24*, 47–53. [[CrossRef](#)] [[PubMed](#)]
15. Tarassenko, L.; Clifton, D.A.; Pinsky, M.R.; Hravnak, M.T.; Woods, J.R.; Watkinson, P. Centile-based early warning scores derived from statistical distributions of vital signs. *Resuscitation* **2011**, *82*, 1013–1018. [[CrossRef](#)]
16. Yoon, J.H.; Jeanselme, V.; Dubrawski, A.W.; Hravnak, M.; Pinsky, M.R.; Clermont, G. Prediction of hypotension events with physiologic vital sign signatures in the intensive care unit. *Crit. Care.* **2020**, *24*, 661. [[CrossRef](#)]
17. Chen, L.; Dubrawski, A.; Hravnak, M.; Clermont, G.; Pinsky, M.R. Modelling Risk of Cardio-Respiratory Instability as a Heterogeneous Process. *Annu. Symp. Am. Med. Inform. Assoc.* **2015**, *2015*, 1841–1850.
18. Chen, L.; Ogundele, O.; Clermont, G.; Hravnak, M.; Pinsky, M.R.; Dubrawski, A.W. Dynamic and Personalized Risk Forecast in Step-Down Units. Implications for Monitoring Paradigms. *Ann. Am. Thorac. Soc.* **2017**, *14*, 384–391. [[CrossRef](#)]
19. Ruminski, C.M.; Clark, M.T.; Lake, U.E.; Kitzmiller, R.R.; Keim-Malpass, J.; Robertson, M.P.; Simons, T.R.; Moorman, J.R.; Calland, J.F. Impact of predictive analytics based on continuous cardiorespiratory monitoring in a surgical and trauma intensive care unit. *Int. J. Clin. Monit. Comput.* **2018**, *33*, 703–711. [[CrossRef](#)]
20. Fairchild, K.D.; Schelonka, R.L.; Kaufman, D.A.; Carlo, W.A.; Kattwinkel, J.; Porcelli, P.J.; Navarrete, C.T.; Bancalari, E.; Aschner, J.L.; Walker, M.; et al. Septicemia mortality reduction in neonates in a heart rate characteristics monitoring trial. *Pediatr. Res.* **2013**, *74*, 570–575. [[CrossRef](#)]
21. Gisolfi, N. Model-Centric Verification of Artificial Intelligence. Ph.D. Thesis, The Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, USA, December 2021.
22. Goddard, K.; Roudsari, A.V.; Wyatt, J. Automation bias: A systematic review of frequency, effect mediators, and mitigators. *J. Am. Med. Inform. Assoc.* **2012**, *19*, 121–127. [[CrossRef](#)]
23. Lebiere, C.; Blaha, L.M.; Fallon, C.K.; Jefferson, B. Adaptive Cognitive Mechanisms to Maintain Calibrated Trust and Reliance in Automation. *Front. Robot. AI* **2021**, *8*, 652776. [[CrossRef](#)] [[PubMed](#)]
24. Calzoni, L.; Clermont, G.; Cooper, G.F.; Visweswaran, S.; Hochheiser, H. Graphical Presentations of Clinical Data in a Learning Electronic Medical Record. *Appl. Clin. Inform.* **2020**, *11*, 680–691. [[CrossRef](#)] [[PubMed](#)]
25. Gianfrancesco, M.A.; Tamang, S.; Yazdany, J.; Schmajuk, G. Potential Biases in Machine Learning Algorithms Using Electronic Health Record Data. *JAMA Intern. Med.* **2018**, *178*, 1544–1547. [[CrossRef](#)] [[PubMed](#)]
26. Agniel, D.; Kohane, I.S.; Weber, G.M. Biases in electronic health record data due to processes within the healthcare system: Retrospective observational study. *BMJ* **2018**, *361*, 1479. [[CrossRef](#)]

27. Bellamy, R.K.E.; Mojsilovic, A.; Nagar, S.; Ramamurthy, K.N.; Richards, J.; Saha, D.; Sattigeri, P.; Singh, M.; Varshney, K.R.; Zhang, Y.; et al. AI Fairness 360: An extensible toolkit for detecting and mitigating algorithmic bias. *IBM J. Res. Dev.* **2019**, *63*, 4:1–4:15. [[CrossRef](#)]
28. Wolff, R.F.; Moons, K.G.; Riley, R.D.; Whiting, P.F.; Westwood, M.; Collins, G.S.; Reitsma, J.B.; Kleijnen, J.; Mallett, S.; Altman, D.; et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Ann. Intern. Med.* **2019**, *170*, 51–58. [[CrossRef](#)] [[PubMed](#)]
29. Saleiro, P.; Kuester, B.; Hinkson, L.; London, J.; Stevens, A.; Anisfeld, A.; Rodolfa, K.T.; Ghani, R. Aequitas: A Bias and Fairness Audit Toolkit. *arXiv* **2018**, arXiv:1811.05577.
30. Parikh, R.B.; Teeple, S.; Navathe, A.S. Addressing Bias in Artificial Intelligence in Health Care. *JAMA J. Am. Med. Assoc.* **2019**, *322*, 2377. [[CrossRef](#)] [[PubMed](#)]
31. Lederer, D.J.; Bell, S.C.; Branson, R.; Chalmers, J.D.; Marshall, R.; Maslove, D.M.; Ost, D.E.; Punjabi, N.M.; Schatz, M.; Smyth, A.R.; et al. Control of Confounding and Reporting of Results in Causal Inference Studies. Guidance for Authors from Editors of Respiratory, Sleep, and Critical Care Journals. *Ann. Am. Thorac. Soc.* **2019**, *16*, 22–28. [[CrossRef](#)]
32. Maslove, D.M.; Leisman, D.E. Causal Inference from Observational Data: New Guidance from Pulmonary, Critical Care, and Sleep Journals. *Crit. Care Med.* **2019**, *47*, 1–2. [[CrossRef](#)] [[PubMed](#)]
33. Jeanselme, V.; De-Arteaga, M.; Elmer, J.; Perman, S.M.; Dubrawski, A. Sex differences in post cardiac arrest discharge locations. *Resusc. Plus* **2021**, *8*, 100185. [[CrossRef](#)] [[PubMed](#)]
34. Maslove, D.M.; Elbers, P.W.G.; Clermont, G. Artificial intelligence in telemetry: What clinicians should know. *Intensiv. Care Med.* **2021**, *47*, 150–153. [[CrossRef](#)] [[PubMed](#)]
35. Escobar, G.J.; Liu, V.X.; Schuler, A.; Lawson, B.; Greene, J.D.; Kipnis, P. Automated Identification of Adults at Risk for In-Hospital Clinical Deterioration. *N. Engl. J. Med.* **2020**, *383*, 1951–1960. [[CrossRef](#)]
36. Churpek, M.M.; Yuen, T.C.; Winslow, C.; Robicsek, A.A.; Meltzer, D.O.; Gibbons, R.D.; Edelson, D.P. Multicenter Development and Validation of a Risk Stratification Tool for Ward Patients. *Am. J. Respir. Crit. Care Med.* **2014**, *190*, 649–655. [[CrossRef](#)] [[PubMed](#)]
37. Bartkowiak, B.; Snyder, A.M.; Benjamin, A.; Schneider, A.; Twu, N.M.; Churpek, M.M.; Roggin, K.K.; Edelson, D.P. Validating the Electronic Cardiac Arrest Risk Triage (eCART) Score for Risk Stratification of Surgical Inpatients in the Postoperative Setting: Retrospective Cohort Study. *Ann. Surg.* **2019**, *269*, 1059–1063. [[CrossRef](#)] [[PubMed](#)]
38. Wijnberge, M.; Geerts, B.F.; Hol, L.; Lemmers, N.; Mulder, M.P.; Berge, P.; Schenk, J.; Terwindt, L.E.; Hollmann, M.W.; Vlaar, A.P.; et al. Effect of a Machine Learning-Derived Early Warning System for Intraoperative Hypotension vs. Standard Care on Depth and Duration of Intraoperative Hypotension during Elective Noncardiac Surgery: The HYPE Randomized Clinical Trial. *JAMA* **2020**, *323*, 1052–1060. [[CrossRef](#)]
39. Chen, Y.; Yoon, J.H.; Pinsky, M.R.; Ma, T.; Clermont, G. Development of hemorrhage identification model using non-invasive vital signs. *Physiol. Meas.* **2020**, *41*, 055010. [[CrossRef](#)]
40. Chen, Y.; Hong, C.; Pinsky, M.R.; Ma, T.; Clermont, G. Estimating Surgical Blood Loss Volume Using Continuously Monitored Vital Signs. *Sensors* **2020**, *20*, 6558. [[CrossRef](#)]
41. Pinsky, M.R.; Wertz, A.; Clermont, G.; Dubrawski, A. Parsimony of Hemodynamic Monitoring Data Sufficient for the Detection of Hemorrhage. *Anesth. Analg.* **2020**, *130*, 1176–1187. [[CrossRef](#)]
42. Johnson, A.E.W.; Pollard, T.J.; Shen, L.; Lehman, L.-W.H.; Feng, M.; Ghassemi, M.; Moody, B.; Szolovits, P.; Celi, L.A.; Mark, R.G. MIMIC-III, a freely accessible critical care database. *Sci. Data* **2016**, *3*, 160035. [[CrossRef](#)] [[PubMed](#)]
43. Yoon, J.H.; Mu, L.; Chen, L.; Dubrawski, A.; Hravnak, M.; Pinsky, M.R.; Clermont, G. Predicting tachycardia as a surrogate for instability in the intensive care unit. *Int. J. Clin. Monit. Comput.* **2019**, *33*, 973–985. [[CrossRef](#)] [[PubMed](#)]
44. Wang, D.; Chen, L.; Fiterau, M.; Dubrawski, A.; Hravnak, M.; Bose, E.; Wallace, D.; Kaynar, M.; Clermont, G.; Pinsky, M.R. Multi-Tier ground truth Elicitation Framework with Application to Artifact Classification for Predicting Patient Instability. *Intensive Care Med.* **2014**, *40*, S389.
45. Fiterau, M.; Dubrawski, A.; Wang, D.; Chen, L.; Guillaume-Bert, M.; Hravnak, M.; Clermont, G.; Bose, E.; Holder, A.; Kaynar, A.M.; et al. Semi automated adjudication of vital sign alerts in step-down units. *Intensive Care Med. Exp.* **2015**, *3*, A769. [[CrossRef](#)]
46. Fiterau, M.; Dubrawski, A.; Chen, L.; Hravnak, M.; Clermont, G.; Bose, E.; Guillaume-Bert, M.; Pinsky, M.R. Artifact adjudication for vital sign step-down unit data can be improved using Active Learning with low-dimensional models. *Intensive Care Med.* **2014**, *40*, S289.
47. Xu, Y.; Zhang, H.; Miller, K.; Singh, A.; Dubrawski, A. Noise-Tolerant Interactive Learning Using Pairwise Comparisons. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017.
48. Sheng, J.; Chen, L.; Xu, Y.; Pinsky, M.R.; Hravnak, M.; Dubrawski, A. Using Comparisons to Reduce Cost of Data Annotation Required to Train Models for Bedside Monitoring. *Crit. Care Med.* **2019**, *47*, 606. [[CrossRef](#)]
49. Goswami, M.; Boecking, B.; Dubrawski, A. Weak Supervision for Affordable Modeling of Electrocardiogram Data. *arXiv* **2021**, arXiv:2201.02936.
50. Helman, S.; Terry, M.A.; Pellathy, T.; Williams, A.; Dubrawski, A.; Clermont, G.; Pinsky, M.R.; Al-Zaiti, S.; Hravnak, M. Engaging Clinicians Early During the Development of a Graphical User Display of An Intelligent Alerting System at the Bedside. *Internat. J. Med. Inform.* **2022**, *159*, 104643. [[CrossRef](#)]

51. Kelly, C.J.; Karthikesalingam, A.; Suleyman, M.; Corrado, G.; King, D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med.* **2019**, *17*, 1–9. [[CrossRef](#)]
52. Shah, N.H.; Milstein, A.; Bagley, S.C. Making Machine Learning Models Clinically Useful. *JAMA J. Am. Med. Assoc.* **2019**, *322*, 1351. [[CrossRef](#)]
53. Laird, P.; Wertz, A.; Welter, G.; Maslove, D.; Hamilton, A.; Yoon, J.H.; Lake, D.E.; Zimmet, A.E.; Bobko, R.; Moorman, J.R.; et al. The critical care data exchange format: A proposed flexible data standard for combining clinical and high-frequency physiologic data in critical care. *Physiol. Meas.* **2021**, *42*, 065002. [[CrossRef](#)] [[PubMed](#)]