Parsimony of hemodynamic monitoring data sufficient for the detection of hemorrhage

Michael R. Pinsky, MD[a], **Anthony Wertz, MS**[b], **Gilles Clermont, MD**[a], **Artur Dubrawski, PhD**[b]

[a]University of Pittsburgh, School of Medicine

[b]Carnegie Mellon University, School of Computer Science, Auton Lab

Abstract

Background: Individualized hemodynamic monitoring approaches are not well validated. Thus, we evaluated of the discriminative performance improvement that might occur when moving from non-invasive monitoring (NIM) to invasive monitoring and with increasing levels of featurization associated with increasing sampling frequency and referencing to a stable baseline to identify bleeding during surgery in a porcine model.

Methods: We collected physiologic waveform data (250Hz) from NIM, central venous (CVC), arterial (ART) and pulmonary arterial (PAC) catheters, plus mixed venous O_2 saturation and cardiac output from 38 anesthetized Yorkshire pigs bled at 20 mL/min until a mean arterial pressure of 30 mmHg following a 30-min baseline period. Pre-bleed physiologic data defined a personal stable baseline for each subject independently. Nested models were evaluated using simple hemodynamic metrics (SM) averaged over 20 second windows and sampled every minute, beat-to-beat (B2B) and waveform data (WF) using random forest classification models to identify bleeding with or without normalization to personal stable baseline, using a leave-one-pig-out cross-validation to minimize model overfitting. Model hyperparameters were tuned to detect stable or bleeding states. Bleeding models were compared use both each subject's personal baseline and a grouped average (universal) baseline. Timeliness of bleed onset detection was evaluated by comparing the tradeoff between a low false positive rate (FPR) and shortest time to bleed detection. Predictive performance was evaluated using a variant of the Receiver Operating Characteristic focusing on minimizing FPR and false negative rates (FNR) for true positive and true negative rates, respectively.

Results: In general, referencing models to a personal baseline resulted in better bleed detection performance for all catheters than using universal baselined data. Increasing granularity from SM to B2B and WF progressively improved bleeding detection. All invasive monitoring out-performed NIM for both time to bleeding detection and low FPR and FNR. In that regard, when referenced to

Address correspondence to: Michael R. Pinsky, MD, 1215.4 Kaufmann Medical Building, 3471 Fifth Avenue, Pittsburgh, PA 15213, Tel: 412-647-8109, Fax: 412-647-8060, pinsky@pitt.edu.

Conflicts of Interest: Michael R. Pinsky is consultant to Edwards LifeSciences and LiDCO

personal baseline with SM analysis, PAC and ART+PAC performed best; for B2B CVC, PAC and ART+PAC performed best; and for WF PAC, CVC, ART+CVC and ART+PAC performed equally well and better than other monitoring approaches. Without personal baseline NIM performed poorly at all levels, while all catheters performed similarly for SM, with B2B PAC and ART+PAC were best, and for WF PAC, ART, ART+CVC and ART+PAC performed equally well and better than the other monitoring approaches.

Conclusions: Increasing hemodynamic monitoring featurization by increasing sampling frequency and referencing to personal baseline markedly improves the ability of invasive monitoring to detect bleed.

Keywords

animal model; hemodynamic monitoring; circulatory shock; machine learning

Introduction

Hemodynamic monitoring is often done to identify the onset of cardiovascular insufficiency, its progression and response to treatment. However, hemodynamic monitoring is not performed using a single device, measure or sampling rate. It reflects a variety of invasively and non-invasively physiologic sensors collected both intermittently and continuously, and often processed to create fused parameters. No systematic evaluation of their utility has been done, either alone or in combination, at varying levels of featurization and relative to a stable baseline to identify cardiovascular insufficiency. Hypovolemia is the most common cause of cardiovascular decompensation in the operating room. Its magnitude and duration predict acute kidney injury, prolonged length of stay and mortality.[1] Hypotension is often due to hypovolemia. Overt signs of progressive hypovolemia can be obscure. Yet delayed treatment of hypovolemia is associated with poor outcomes.[1,2] Since clinicians tend to progress from non-invasive to more invasive monitoring and at greater measurement frequencies if they feel the patient is at increased risk of cardiovascular insufficiency, we examined the discriminative impact of such a strategy in our animal model of hemorrhage.

This study addresses monitoring knowledge gaps of surgical patients at risk of hypovolemia. Most patients have continuous non-invasive electrocardiogram (ECG) and pulse oximetry monitoring, while invasive monitoring is less common and inconsistently applied[3] or used.[4,5] Central venous catheters (CVCs) are commonly placed and measure central venous pressure, often used to identify volume status and guide fluid therapy,[4] though it poorly defines volume status or volume responsiveness.[6] Arterial catheters (ART) are used less often[3,4] and primarily to assess mean arterial pressure (MAP) as a surrogate for organ perfusion pressure. ART pressure waveform beat-to-beat (B2B) parameters, like pulse pressure variation (PPV), cardiac output (CO) and stroke volume variation (SVV) predict volume responsiveness,[7] and when coupled with resuscitation protocols reduce perioperative complications[8] and mortality.[9] Finally, the pulmonary artery catheter (PAC) continuously report central venous and pulmonary artery pressures, CO, and mixed venous O_2 saturation (SvO_2). Although useful in the diagnosis and management of heart failure and pulmonary hypertension,[10] its value in the assessment of hemorrhagic shock is undefined. Finally, invasiveness delays surgery and carries risk of the procedures, making their universal use

unwise. We hypothesized that for each hemodynamic monitoring modality, the more closely we look at in terms of sampling frequency, data featurization and subject-specific baseline, the more accurately we may discriminate between stable and hemorrhage states.

Methods

Overview:

The experiment was designed to study the discriminative impact of three factors: data normalization, granularity, and sensing modalities, on our ability to detect hemorrhage. Our design moved through three stages: data collection and preprocessing; model training; and evaluation. Physiologic data were collected from pig models of hypovolemia both while the pig was hemodynamically stable then bled at a constant rate. Hemodynamic data were filtered for artifacts and then multiple derived metrics (referred to as *features*) were computed. Subsets of these features were generated by grouping on normalization (either using a universal baseline or personal subject-specific baselines), granularity (simple metrics (SM) only, beat-to-beat (B2B) metrics, or waveform (WF)-based metrics), and available sensing modalities (non-invasive only (NIM), adding CVC, PAC, and/or ART). For each subset of features, a classification model (referred to here as a *classifier*) to discriminate between the stable and bleeding periods was trained and validated in a leave-one-pig-out cross-validation (LOOCV) framework to minimize model overfitting. Normal ROC area under the curve (AUC) only describes overall model sensitivity and specificity, it does not highlight model performance at both very low false positive (FPR) and false negative rate (FNR), two characteristics important in an alert monitor. Thus, the results of this evaluation are presented through modified Receiver Operating Characteristic (ROC) curves highlighting positive and negative predictive performance, and Activity Monitoring Operating Characteristic (AMOC) curves to present the tradeoff between time to detection (TTD) (how fast bleeding can be detected after starting) and false alarm rate.[11]

Experimental model and data collection:

The protocol was approved by the University of Pittsburgh IACUC and previously described.[12] Briefly, 46 female Yorkshire pigs (wt. 21.3 ± 0.65 kg) were anesthetized (ketamine, xylazine, and telazol) for induction, intubated and ventilated (8 mL/kg tidal volume, FiO_2 0.4, 3 cmH_2O PEEP) on maintenance anesthesia (2.0–3.0% isoflurane). 0.9% NaCl solution was infused at 1 mL/kg/h prior to the study. A PAC (Edwards LifeSciences, Irvine, CA) equipped with distal fiberoptics to continuously measure SvO_2 was inserted via internal jugular vein and central venous and pulmonary artery pressures measured. A triple lumen 18-ga femoral artery catheter was placed to measure arterial pressure and an 8 Fr femoral vein introducer placed for blood removal. The arterial pressure signal was simultaneously recorded on a LiDCO*plus*™ (LiDCO Ltd., London, UK) monitor. Triplicate bolus thermodilution cardiac output calibrated the PAC continuous cardiac output and LiDCO monitors. A pulse oximeter (Masimo, Irvine CA) was placed on the tail and the pulse oximeter O_2 saturation and plethysmographic (pleth) waveform collected. Following surgery, the pigs were rested 30 minutes without further manipulation to establish a baseline, then bled using a roller pump (Masterflex L/S easy-load II, Cole-Parmer; Vernon Hills, IL)

at 20 mL/min until the MAP decreased to 30 mmHg. Data were collected during a stabilization period to the end of the bleed.

Data acquisition:

All monitored outputs were collected at 250 Hz. B2B heart rate, systolic and diastolic arterial and pulmonary arterial pressures, MAP and mean pulmonary artery pressures were calculated and SvO_2 was recorded. B2B CO, stroke volume (SV), PPV and SVV were recorded (LiDCO*plus*). Twelve additional arterial pressure features including heart rate variability (HRV) parameters, such as R-R interval standard deviation, entropy, Fournier analysis, etc., were obtained on B2B data. For comparison between modalities, we assumed that SvO_2 approximated central venous O_2 saturation, and included PAC SvO_2 in the CVC features. NIM included all data from ECG and pulse oximetry, whereas invasive monitoring was as defined by parameters available from the specific catheter plus NIM. Data are reported as mean±SD.

Nested models:

We started with one model including only NIM measurements and progressively added additional signals to compare performance. NIM comprised ECG, HR, Pleth, HRV and SpO_2. All invasive models included NIM data. CVC comprised central venous pressure and SvO_2; ART comprised arterial pressures and CO; and PAC comprised central venous and pulmonary artery pressures, CO and SvO_2, with ART+CVC and ART+PAC each comprising a union of the two catheter-related inputs.

Features by data density:

To assess the impact of data granularity on our ability to detect bleed onset, we trained multivariable models using data at three levels of granularity: twenty second averages of simple hemodynamic metrics (SM) sampled every minute, B2B, and waveform (WF), with each level conveying increasing richness of data features, with each higher granularity set including lower granularity features.

Personal baseline normalization or universal normalization:

We trained models which incorporated features normalized either by all subjects (universal) or by each subject individually (personal). The clinical interpretation of the different normalization schemes contrasts scenarios where knowledge of the patient's stable vitals (personal) is available versus when it is not (universal). Normalization was performed for each feature independently by centering on the mean and scaling by the SD in the same way a Z-score would be computed. The mean±SD used were computed using ten minutes of data captured in the stable (pre-bleed) period. For the universal baseline, a single mean±SD was computed for each feature across all pigs in the training set. For the personal baseline the parameters were computed for each pig individually.

Evaluation methods:

Each data model configuration contained several time series (e.g. for the ART+PAC model with personalized baseline: SM=26, B2B=48 and WF=354), reflecting increasing extents of

data featurization. We used machine learning methodology to automatically define optimal bleeding alert criteria for each group. Features were computed in rolling time windows and updated at 2 Hz. Classifiers were trained separately for each feature set configuration to predict bleeding, with data prior to the start of bleed labeled as "Stable" and after as "Bleeding". Although we evaluated several machine learning algorithms, we used only the random forest classifiers as the evaluation platform because they have been shown to perform consistently well across the range of considered feature sets compared to other classifiers[11,13] and allowed us to focus on analyzing relative value of those different featurizations. The model predictions were validated using a LOOCV to mitigate effects of over-fitting: similar to splitting data into training and testing sets repeated on multiple splits (i.e. *folds*) of the data to control for the uncertainty in training and testing set selection. In our LOOCV procedure, the models generalize their utility to never before seen individuals. We use a rolling window approach to compute features in the data. For example, we compute a metric multiple times at different points in the data using samples in a specific interval. The size of the window used depends on the feature (e.g. a mean maybe computed using thirty seconds of data and another mean two minutes of data). All test results were agglomerated into overall performance metrics for each model configuration. In order to quantify relative importance of individual features of data, we quantified the amount of information they provide when test data is being assessed using information gain, and accumulated these gains per feature across all samples of the test data used in LOOCV (results reported as IGmax in Table 1).

Model performance:

First, we used a variant of the ROC curves plotting either true positive rate (TPR) as a function of FPR, or true negative rate (TNR) as a function of FNR. The TPR versus FPR is the traditional ROC approach. Since precise monitoring aims to minimize FPR to reduce inappropriate false alerts and low FNR so that if the monitor reports stability, stability is present, we displayed the ROC curves in two simultaneous back-to-back views (referred to here as the *ROCOR*) with the *x*-axes scaled logarithmically to focus on model performance at very low FPR and FNR, respectively. Second, with AMOC curves we plotted FPR as a function of average time to detection (TTD) from the onset of bleeding given a specific minimum detection rate. The AMOC presents the tradeoff between detection latency and FPR, both of which we wanted to minimize, but as the detection threshold increases (i.e. fewer false positives) so does the time it takes to identify bleeding. Sensitivity of detection was defined as the earliest time of class separation (non-bleeding v. bleeding) with the lowest FPR. In the case of the AMOC and the ROCOR plots (ribbons depict the 95% confidence intervals). We used a threshold of 80% detection (TPR) rate for reference purposes between models. ROCOR curves expressed much of the same relations as the AMOC curves, except they addressed the tradeoff between prediction performance (bleed or no bleed) and error (or alarm) rates, whereas the AMOCs reported the tradeoff between timeliness of detection of bleeding and false alarm rate. Because our objective in this analysis was not to develop models that present the probability of bleeding, but rather present trade-offs between prediction performance, detection latency, and FPR, model calibration was not required as it would not affect the ordering of comparative results.

Model optimization:

The random forest model used allowed for the specification of a few different parameters (which we refer to as *hyperparameters*), for example the number of trees and maximum tree depth. To select the optimal hyperparameters for each data configuration (i.e. modality, granularity, and baseline), a grid-search hyperparameter selection was conducted to select optimal hyperparameters for each model. Each hyperparameter selection was evaluated using LOOCV. For any configuration, two models were selected: one that optimizes performance of positive (bleed) detection, which was selected based on AMOC and ROC performance; and another that optimized performance of negative (stable) detection, selected based on ROC performance alone. The simultaneous use of two such models enables confident identification of both test subjects who bleed and those who do not. This symmetric assessment capability is useful to guide clinical resource allocation based on confident adjudication of both the positive and negative cases (Figure 1E). As illustrated, most models performed similarly down to FPR of one in fifty, but degraded thereafter. Similarly, most models had a short TTD at a high FPR but at a TTD <4 minutes at a FPR $<10^{-3}$ marked performance differences occurred. Positive prediction models were selected to minimize TTD at low error rates while not sacrificing TPR (an indication of overfitting in this selection process). Negative prediction models were selected using a similar process, but only TNR at low error rates (FNR) was evaluated.

Results

Baseline hemodynamics prior to bleed reflected normal post-induction values for pigs with a MAP 68±9.3 mmHg CVP 3.4±1.6 mmHg, PAP 13.8±1.9 mmHg, HR 91.3±10.9 1/min, CO 4.24±1.46 L/min, and SvO_2 76.6±6.4%. These baseline values of individual hemodynamic measures using threshold values not referenced to a personalized baseline did not identify the onset of hypovolemia until approximately 10–15 minutes after the start of bleeding (Figure 1). Using personal baseline references improved detection for arterial pressure-related parameters. The multivariable B2B and WF models performed better than raw data threshold alert at time to detection for less than one in one hundred (10^{-2}) error rate.

Nested Model optimization and performance:

The classification performance for the SM, B2B and WF models with and without personalized baseline normalization is shown in Figures 2 and 3, respectively. We defined better model performance on the ROCOR as higher TNR or TPR for the lower FNR or FPR, respectively. Area Under the ROC Curve (AUC) for each model is presented in Table 1. The AMOC curves for the SM, B2B and WF models with and without personal baseline normalization are shown in Figures 4 and 5, respectively. We defined better model performance as a lower FPR for the shorter time to detection. ROCOR and AMOC curves with and without personal baseline normalization, as data granularity is varied for each monitoring modality separately, are reported in Figures E2–E15 and the specific features used for each model that impacted performance are reported in Table E1.

Universal baseline normalization analyses:

Under most conditions, invasive monitoring outperformed NIM. Performance at predicting absence of bleeding was difficult with SM (Figure 2, left) improving with B2B and WF for all but the NIM model, which itself did not improve until WF granularity was used. For identifying bleeding (Figure 2, right), NIM performed worse than invasive monitoring. With SM models, CVC did not perform better than the NIM with minimal separation across monitoring inputs, but in B2B and WF PAC and ART+PAC performed best, with ART and ART+CVC becoming less different with WF analyses. While B2B provided a boost from ART to ART+PAC models, minimal boost in performance was seen with the ART to ART +CVC or PAC to ART+PAC models. Thus, there appears to be minimal benefit from adding universally baselined CVP to ART. SM inputs did not differ across monitoring modalities TTD and FPR (Figure 3). However, with both B2B and WF, PAC, ART+PAC and ART markedly improved performance. At higher FPR the models performed similarly, but as the FPR decreases in the B2B granularity PAC and ART+PAC improved substantially, allowing a short TTD of ~7 minutes at 10^{-3} FPR.

Personal baseline normalization analyses:

Personal baseline normalization for ROCOR and AMOC analyses Figures 4 and 5, respectively), improved the performance of all models compared to universally baselined data. Despite CVC performing poorly when using the universal baseline normalization, when personalized CVC outperformed all universally normalized models at a 10^{-2} FPR, even at the SM granularity, underscoring the need for personalized baseline referencing prior to monitoring for potential deterioration, independent of monitoring modality. Similarly, although the performance of the NIM SM and B2B models remained poor, NIM WF markedly improved its positive predictive performance and equaled ART for negative predictive performance (Figure 3C). Although PAC, ART+PAC and ART+CVC demonstrated better performance at all levels of granularity, their positive predictive performance separation across modalities was minimal, unlike the negative predictive performance. At SM ART+PAC and PAC, TTD was shorter at lower FPR than all other modalities (Figure 5). At B2B ART+PAC markedly reduced its FPR while CVP and ART-CVP also improved performance to similar TTD. With WF granularity, ART+PAC, PAC, CVC, and ART+CVC all TTD <5 minutes with extremely low FPRs.

Discussion

This study had four fundamental findings. First, increasing sampling frequency by allowing for more extensive data featurization improves model performance for all modalities. The marked improvement of the CVC model when moving from SM to either B2B or WF when using personal baseline normalization underscores the power of this analysis to use commonly available physiologic data streams in novel ways to accurately define stability and bleed onset. This is in contradistinction to raw CVP values which are not useful in defining intravascular volume status.[4] At a FPR of 10^{-2}, the CVC B2B model can detect bleed in >80% of pigs in <4 minutes, and the FPR can be decreased to 10^{-3} with an additional minute of detection latency. The NIM model also performed well when WF featurizations were used, detecting bleed onset in ~7 minutes at 10^{-2} FPR. The PAC

predictive performance deserves special mention. Prior clinical literature failed to demonstrate the clinical usefulness of PAC-related monitoring.[14] However, the data used from such monitoring was not analyzed on featurized data or with B2B or WF granularity. Considering the utility of CVP, pulmonary vascular pressures, CO and SvO_2 in defining cardiovascular stability (Table E1) it is not surprising that PAC out-performed other individual monitoring devices improving slightly by adding ART.

Second, referencing hemodynamic data to personal baselines markedly improves bleeding detection for all devices. This is a clinically reasonable scenario for elective surgery where stable post-induction baseline values usually exist. Third, combining data from different devices differently improves performance depending on sampling frequency and availability of personalized baseline references. Thus, there is no one single monitoring approach that will be effective in identifying hypovolemia across all surgical conditions, but for similar conditions a common approach is possible. Fourth, performance of models varies depending on how the performance is measured. If the goal is to minimize FNR avoiding reporting a subject as stable when they are not, versus minimizing the latency of detection of bleeding with few false alerts (low FPR), then the choice of the optimal monitoring configuration may differ. Importantly, as the model granularity and the type of baseline used change, the most useful features in driving model performance also vary. Thus, one model alone cannot perform optimally to address both detection of stability (TNR) and bleeding (TPR) at low error rates as the monitoring device, granularity and access to personal baselines vary.

Model performance assessment methods:

We assessed model performance in identifying active bleeding using ROCOR curves that display performance simultaneously for true positives (real bleeding) and true negatives (not bleeding). Usually, such analyses report only AUC as a measure of model performance. But AUC metric alone is inadequate to assess model performance in specific operating conditions because it reports expected model performance across all operational settings, whereas in clinical practice only one or two decision thresholds (one for positive and one for negative case determination, at low error rates) would be used. Since a major monitoring fear is in reporting no bleed in a bleeding patient (i.e. false negative), it is important to make the FNR low. Our ROCOR display, by expanding the FNR and FPR into log scale, graphically magnifies the performance differences across models at the operationally relevant low FNR and FPR settings. These analyses demonstrate that increasing data granularity and referencing to a stable personal baseline minimize both FNR and FPR for the same probability of detection. ROC analysis does not address the timeliness of alerts. AMOC analysis compares TTD versus FPR, with FPR plotted in log scale to highlight operationally relevant performance ranges. There are tradeoffs between timeliness of detection and the number of false alerts. When clinicians are continually at the bedside or if the event if unattended would rapidly lead to catastrophic consequences, a higher FPR may be tolerated to enable faster detection. Whereas for remote monitoring with limited clinical resources and a slower expected pathologic decline rates and lesser immediacy of demise, a delayed detection time but at a lower FPR may be preferred. Thus, the operationally optimal settings of each model, as defined by the AMOC and ROCOR, will be condition- and situation-specific. However, in our study, the normalized WF models performed excellently

with a very low FPR and rapid detection of >80% of subjects, suggesting that at some point very sensitive and specific models may create a universal alert algorithm across clinical environments with a limited device monitoring dataset.

Limitations:

This study has several limitations. First, we used an anesthetized porcine model at a fixed bleeding rate from an initial stable cardiovascular state. Surgical bleeding and hypovolemia, if present, is rarely at a fixed rate and a baseline is often unavailable for emergency surgery. However, these baseline conditions closely mimic routine surgical patients and would be applicable to identifying when stability was lost, independent of the bleeding rate. Second, we used machine learning approaches to plumb the physiologic signatures and manually tuned our models for optimal performance for both TNR and TPR for the lowest FNR and FPR. These models will be dependent of the quality of the physiologic data available and clinicians may accept different tradeoffs of precision and accuracy. Still, this approach represents a robust and easily duplicated method of making such assessments of information utility. As described above there is no one level of stability or instability that would to optimal for all patients, care environments and conditions. Thus, the levels of performance and utility of any predictive model will be patient, procedure, location and care specific. We specifically chose models with extremely low FPR and FNR which needed to be associated with slightly prolonged TTD. To the extent that excessive false alerts do not promote alarm fatigue, and if early detection of hypovolemia is required, then the faster detecting models would still be operationally useful. Thus, this study provides a roadmap to tailor models specifically to the clinician's individual monitoring needs. Third, we had only 38 animals during a post-induction pre-bleed interval to define a universally-baselined normalization. If a much larger clinical database of stable post-induction data were available and able to be stratified by patient, surgery and co-morbidities, universally-baselined normalization performance may improve significantly. This is one of the promises of big data and data mining in healthcare what will need to be studied. Next, we used machine learning approaches to create unique models from our existing database. We do not report the specific output code making its direct reapplication limited. However, we used standard modeling methodologies and model types. Given that machine learning is a mature scientific field and readily available through computer science resources, reproducibility of our results should be straightforward.

Finally, for any smart alarm-based monitoring to be useful, it must be linked to actionable events whose reversal improves outcome. Recent large observational data analyses underscore the clinical significance of increasing levels of arterial hypotension both in magnitude and duration.[1,2] Futier et al.[15] randomized 298 intraoperative patients into tight blood pressure control using SM of arterial pressure monitored continuously and found a decreased incidence of composite complications. As shown for AMOC curve analyses, our SM models when normalized detected >80% of bleeds at a mean TTD in <8 minutes for invasive monitoring, decreasing TTD even further with B2B and WF analysis. Thus, these models report clinically relevant alerts superior to simple threshold monitoring even in the setting of pre-existing mild hypotension. Although we chose bleeding as it is a common cause of intraoperative cardiovascular instability, hypotension also occurs because of loss of

vasomotor tone, vascular obstruction and primary cardiac dysfunction. Thus, any alerting system needs to be used within the clinical context to maximize its usefulness. Our results consistently demonstrate that as data granularity increases, more invasive monitoring data is added, and a personal baseline is used, the models progressively and markedly improved their performance.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Glossary of Abbreviations

AMOC	Activity monitoring operating characteristic
AUC	Area under the curve
ART	Arterial catheter
B2B	Beat-to-beat
CVC	Central venous catheter
CO	Cardiac output
ECG	Electrocardiogram
FNR	False negative rate
FPR	False positive rate
FiO$_2$	Fractional concentration of oxygen
HRV	Heart rate variability
LOOCV	Leave-one-pig out cross-validation
MAP	Mean arterial pressure
SvO$_2$	Mixed venous oxygen saturation
NIM	Non-invasive hemodynamic monitoring
pleth	Plethysmographic
PEEP	Positive end-expiratory pressure
PAC	Pulmonary artery catheter
PPV	Pulse pressure variation
ROC	Receiver operator characteristic

ROCOR	Receiver operator characteristic combining FPR/TPR+FNR/TNR
SM	Simple metrics
SVV	Stroke volume variation
TTD	Time to detection
TNR	True negative rate
TPR	True positive rate
WF	Waveform

References:

1. Welsh M, Devereaux PJ, Garg AX, Kurz A, Turan A, Rodseth, Cywinski J, Thabane L, Sessler DI. Relationship between intraoperative mean arterial pressure and clinical outcomes after noncardiac surgery: toward an empirical definition of hypotension. Anesthesiology 119:507–15, 2013. [PubMed: 23835589]

2. Maheshwari K, Nathanson BH, Munson SH, Khangulov V, Stevens M, Badani H, Khanna AK, Sessler DI. The relationship between ICU hypotension and ibn-hospital mortality and morbidity in septic patients. Intensive Care Med 44:857–67, 2018 [PubMed: 29872882]

3. Cannesson M, Pastel G, Ricks C, Hoeft A, Perel A. Hemodynamic monitoring and management on patients undergoing high risk surgery: a survey among North American and European anesthesiologists. Crit Care 15: R197, 2011. [PubMed: 21843353]

4. Cecconi M, Hofer C, Teboul JL, Pettila V, Wilkman E, Molnar Z, Della Rocca G, Aldecoa C, Artigas A, Jog S, Sander M, Spies C, Lefant J-Y, DeBacker D, on behalf of the FENICE investigators and the ESICM Trial Group. Fluid challenges in intensive care: the FENICE study. Intensive Care Med 41:1529–37, 2015. [PubMed: 26162676]

5. Cannesson M, Desebbe O, Rosamel P, Delannoy B, Robin J, Bastien O, Lehot J-J. Pleth variability index to monitor the respiratory variations in the pulse oximeter plethysmographic waveform amplitude and predict fluid responsiveness in the operating theatre. BJA 101:200–6, 2008. [PubMed: 18522935]

6. Marik PE, Cavallazzi R. Does central venous pressure predict fluid responsiveness? An updated meta-analysis and a plea for common sense. Crit Care Med 41:1774–81, 2013. [PubMed: 23774337]

7. Marik PE, Cavallazzi R, Vasu T, Hirani A. Dynamic changes in arterial waveform derived variables and fluid responsivness in mechanically ventilated patients: a systematic review of the literature. Crit Care Med 37:2642–55, 2009. [PubMed: 19602972]

8. Goepfert MS, Richter HP, zu Eulenburg C, Gruetzmacher J, Rafflenbeul E, Roeher K, von Sandersleben A, Diedrichs S, Reichenspurner H, Goetz A, Reuter DA. Individually optimized hemodynamic therapy reduces complications and length of stay in the intensive care unit. Anesthesiology 119:824–36, 2013. [PubMed: 23732173]

9. Bednarczyk JM, Fridfinnson JA, Kumar A, Blanchard L, Rabbani R, Bell D, Funk D, Turgeon AF, Abou-Setta AM, Zarychanski R. Incorporating dynamic assessment of fluid responsiveness into Goal-Directed Therapy: A systematic review and meta-analysis. Crit Care Med 45:1538–45, 2017. [PubMed: 28817481]

10. Vanderpool RR, Pinsky MR, Naeije R, Deible C, Kosaraju V, Bunner C, Mathier MA, Lacomis J, Champion HC, Simon MA. RV-pulmonary arterial coupling predicts outcome in patients referred for pulmonary hypertension. Heart 101:37–43, 2015. [PubMed: 25214501]

11. Fawcett T and Provost F, Activity monitoring: noticing interesting changes in behavior, in Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining ACM: San Diego, California, United States (1999)

12. Gomez H, Mesquida J, Hermus L, Polanco P, Kim HK, Zenker S, Torres A, Namas R, Vodovotz Y, Clermont G, Puyana JC, Pinsky MR. (2012) Physiologic responses to severe hemorrhagic shock and the genesis of cardiovascular collapse: Can irreversibility be anticipated? J Surg Res 178:358–69. [PubMed: 22475354]

13. Lonkar R, Dubrawski A, Fiterau M, Garnett R. Mining intensive care vitals for leading indicators of adverse health events. Emerging Health Threats Journal 2011, 4: 11073 10.3402/ehtj.v4i0.11073

14. Harvey S, Harrison DA, Singer M, Ashcroft J, Jones CM, Elbourne D, Brampton W, Williams D, Young D, Rowan K. Assessment of the clinical effectiveness of pulmonary artery catheters in the management of patients in intensive care (PAC-Man): a randomized controlled trial. Lancet 366:472–7, 2005. [PubMed: 16084255]

15. Futier E, Lefrant JY, Guinot PG, Godet T, Lorne E, Cuvillon P, Bertran S, Leone M, Pastene B, Piriou V, Molliex S, Albanese J, Julia JM, Tavernier B, Imhoff E, Bazin JE, Constantin JM, Pereira B, Jaber S, for the INPRESS Study Group. Effect of individualized vs standard blood pressure management strategies on postoperative organ dysfunction among high-risk patients undergoing major surgery: A randomized clinical trial. JAMA 318:1346–57, 2017. [PubMed: 28973220]

Key Points

Question;

Are all monitoring modalities equally effective at identifying bleeding onset during surgery?

Findings:

Invasive pulmonary artery and arterial catheter parameters performed best with central venous catheter data improving with greater featurization, whereas referencing to a stable baseline markedly improve d the performance of all monitoring devices.

Meaning:

Increasing featurization of pressure waveform data and referencing it to an individual baseline markedly improves detection of bleeding at a low false positive rate across all monitoring devices.
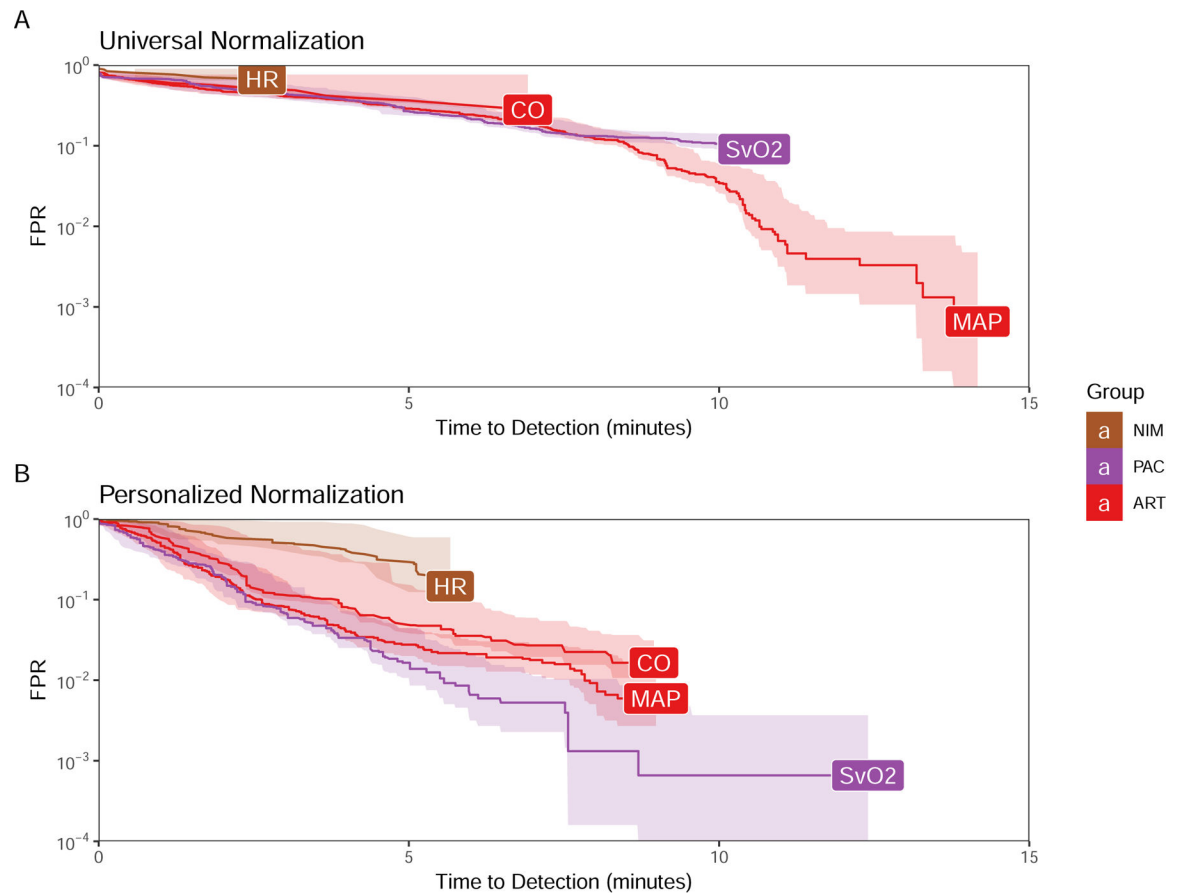
Figure 1.
AMOC curves generated using primary hemodynamic parameters as the predictors using either universal (A) or personal (B) baseline normalization.
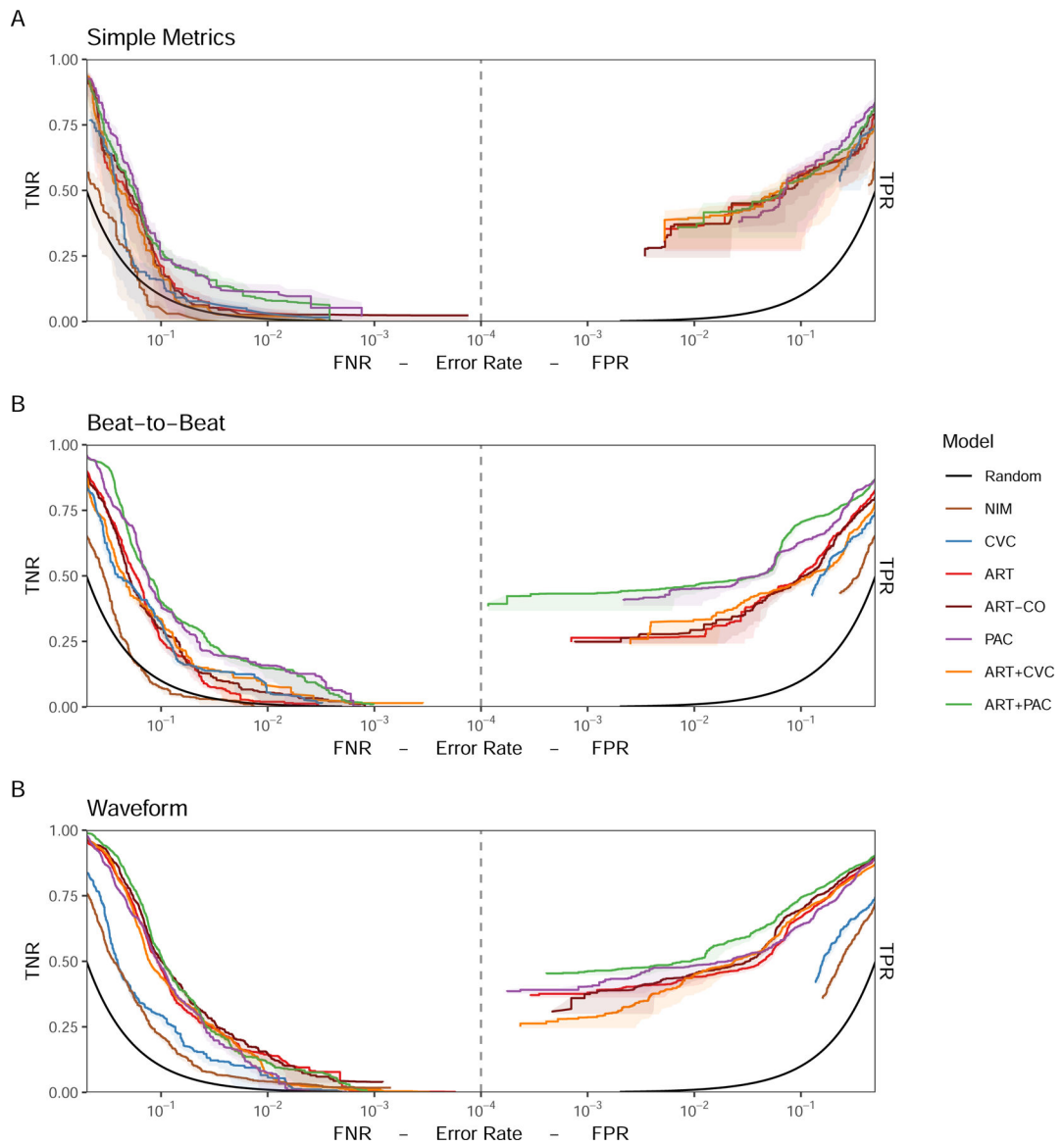
Figure 2.
ROCOR showing (A) Simple Metrics, (B) Beat-to-Beat, and (C) Waveform model classification performance for each monitoring modality or combination of modalities optimized for detection of stability (TNR vs FNR, left) and detection of bleeding (TPR vs FPR, right) for models using universal baseline normalization.
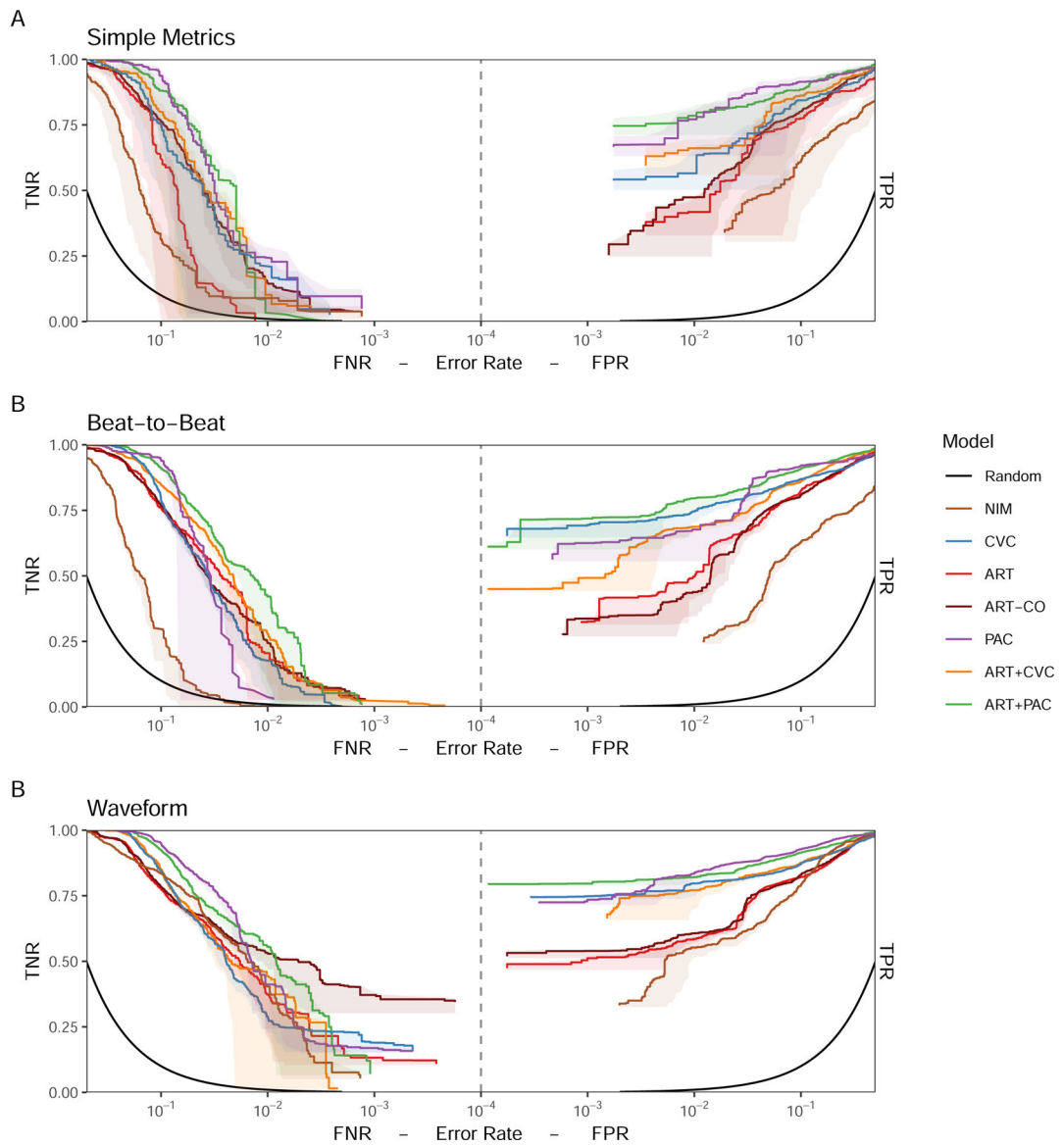
Figure 3.
AMOC showing (A) Simple Metrics, (B) Beat-to-Beat, and (C) Waveform model detection performance for each monitoring modality or combination of modalities optimized to detect bleeding onset with the lowest FPR for universal baseline normalization.
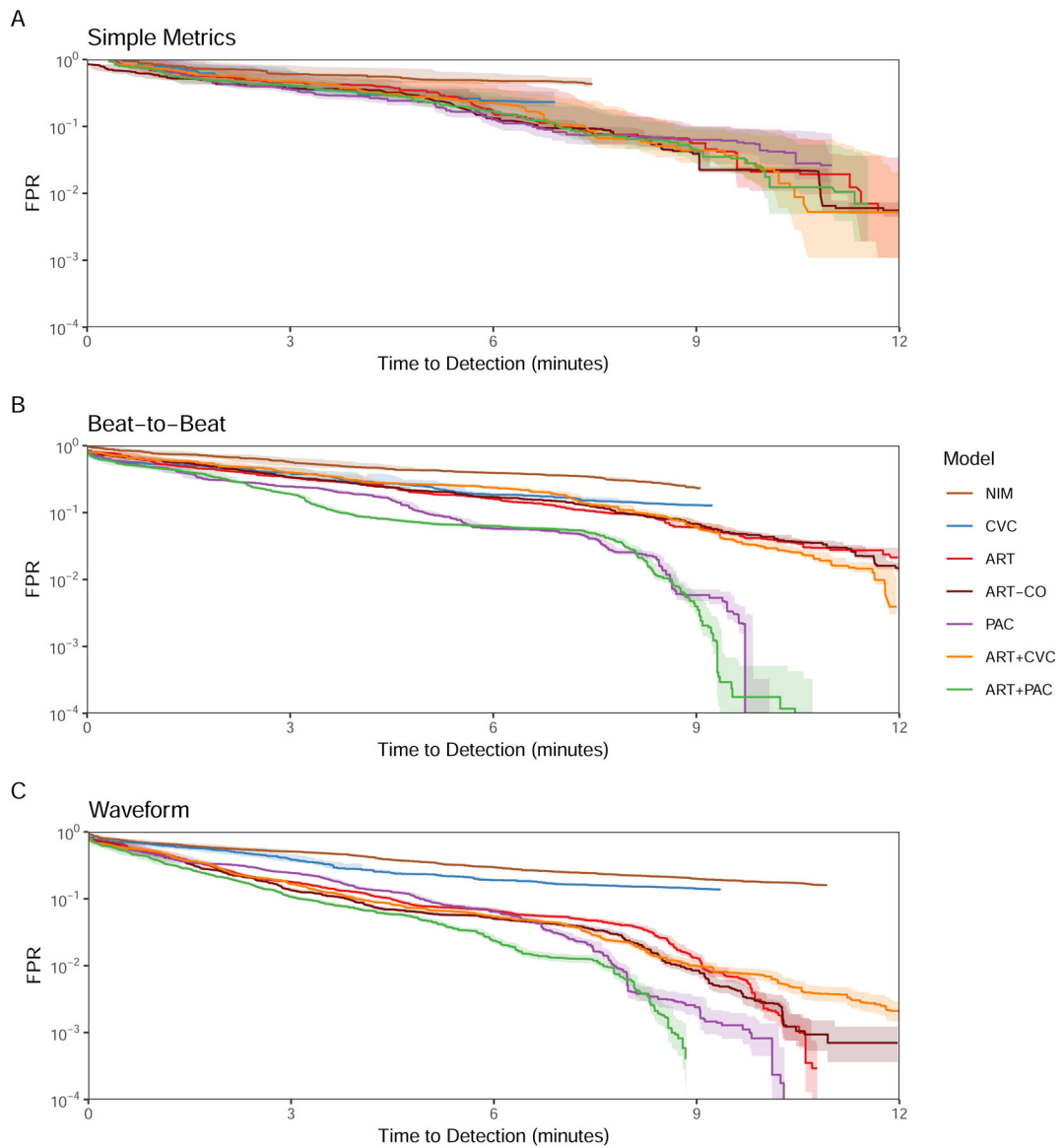
Figure 4.
ROCOR showing (A) Simple Metrics, (B) Beat-to-Beat, and (C) Waveform model classification performance for each monitoring modality or combination of modalities optimized for detection of stability (TNR vs FNR, left) and detection of bleeding (TPR vs FPR, right) for models using personal baseline normalization.
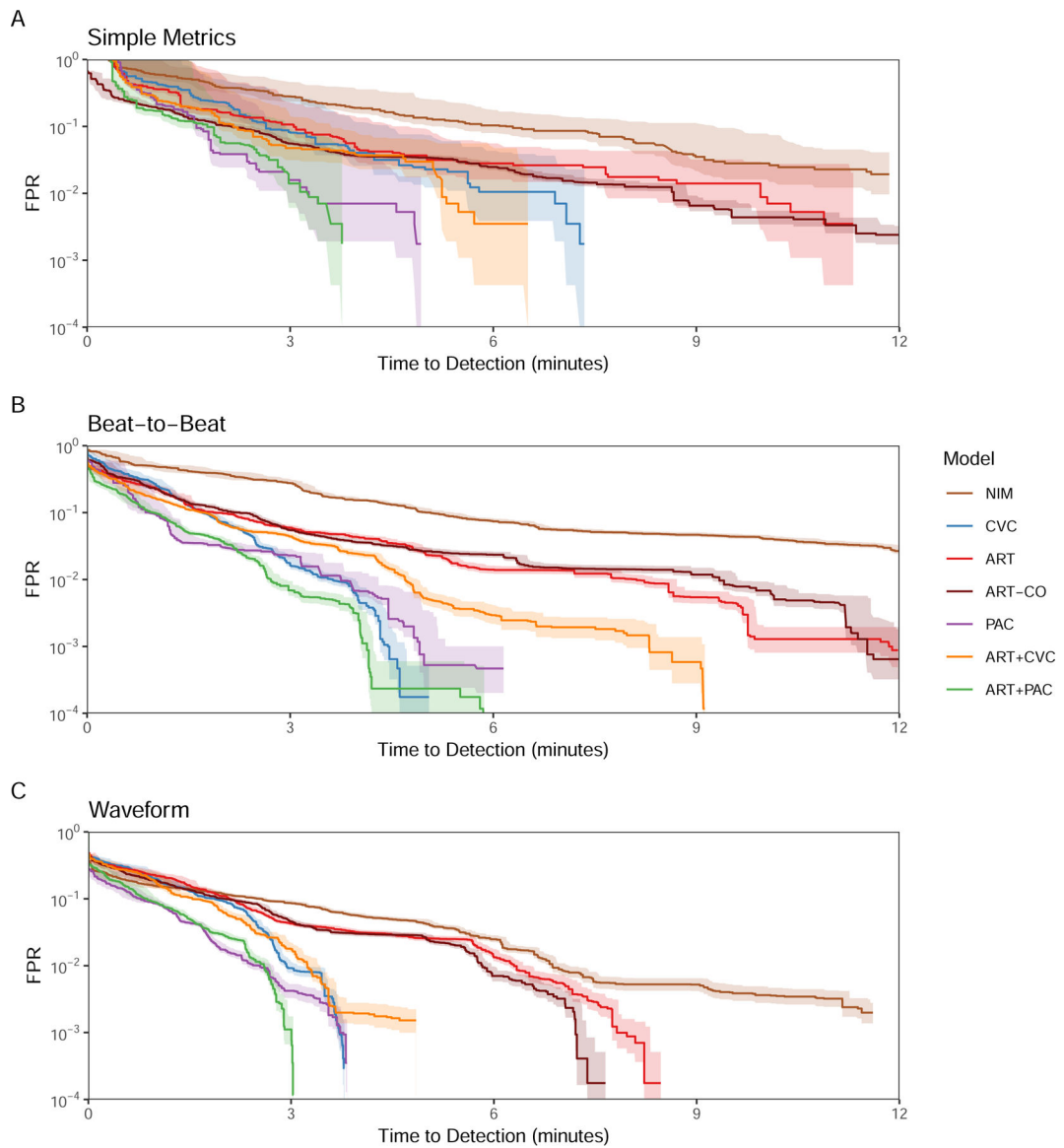
Figure 5:
AMOC showing (A) Simple Metrics, (B) Beat-to-Beat, and (C) Waveform model detection performance for each monitoring modality or combination of modalities optimized to detect bleeding onset with the lowest FPR for personal normalization.

Table 1

Area under the Receiver Operator Characteristic (ROC) curve (AUC) and 95% Confidence Interval (CI) for each model

Simple Metrics

Universal Normalization

Model	AUC	AUC 95% CI	
NIM	0.415	0.317	0.450
CVC	0.542	0.391	0.577
ART	0.566	0.419	0.602
ART-CO	0.574	0.508	0.581
PAC	0.638	0.502	0.672
ART+CVC	0.554	0.399	0.589
ART+PAC	0.605	0.486	0.641

Personalized Normalization

Model	AUC	AUC 95% CI	
NIM	0.681	0.551	0.713
CVC	0.861	0.748	0.883
ART	0.824	0.698	0.849
ART-CO	0.858	0.841	0.862
PAC	0.905	0.790	0.922
ART+CVC	0.874	0.755	0.896
ART+PAC	0.885	0.743	0.904

Beat-to-Beat

Universal Normalization

Model	AUC	AUC 95% CI	
NIM	0.429	0.410	0.435
CVC	0.571	0.546	0.578
ART	0.608	0.588	0.614
ART-CO	0.597	0.575	0.603
PAC	0.698	0.682	0.704
ART+CVC	0.559	0.533	0.566
ART+PAC	0.719	0.707	0.724

Personalized Normalization

Model	AUC	AUC 95% CI	
NIM	0.664	0.634	0.670
CVC	0.869	0.856	0.873
ART	0.871	0.855	0.875
ART-CO	0.865	0.852	0.869
PAC	0.913	0.865	0.916
ART+CVC	0.906	0.897	0.909
ART+PAC	0.923	0.912	0.926

Waveform

Universal Normalization

Model	AUC	AUC 95% CI	
NIM	0.511	0.496	0.518
CVC	0.573	0.553	0.579
ART	0.735	0.725	0.741
ART-CO	0.759	0.748	0.765
PAC	0.725	0.712	0.730
ART+CVC	0.733	0.722	0.738
ART+PAC	0.782	0.772	0.787

Personalized Normalization

Model	AUC	AUC 95% CI	
NIM	0.884	0.873	0.888
CVC	0.868	0.857	0.872
ART	0.884	0.874	0.888
ART-CO	0.891	0.882	0.894
PAC	0.931	0.922	0.934
ART+CVC	0.896	0.874	0.900
ART+PAC	0.904	0.888	0.907