Dynamically Personalized Detection of Hemorrhage

Chirag Nagpal[1], **Xinyu Li**[1], **Michael R. Pinsky**[2], **Artur Dubrawski**[1]
CHIRAGN@CS.CMU.EDU, XINYUL2@ANDREW.CMU.EDU, PINSKY@PITT.EDU, AWD@CS.CMU.EDU
[1] *Auton Lab, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA*
[2] *School of Medicine, University of Pittsburgh, Pittsburgh, PA, USA*

Abstract

Rapid detection of hemorrhage is of major interest to the critical care community, enabling clinicians to take swift actions to mitigate adverse outcomes. In this paper, we describe a model that allows rapid detection of the onset of hemorrhage by monitoring the Central Venous Pressure (CVP). As opposed to prior work in the domain, our model does not rely on prior availability of a stable physiology of a patient as a baseline of reference, and it makes generative assumptions on the monitored vital sign. This allows for rapid on-the-fly personalization to a previously unseen patient's physiology. This property makes the proposed approach particularly relevant to e.g. trauma care and other scenarios where reference hemodynamic data may not be readily available for any new patient. We compare our model against strong discriminative alternatives and demonstrate its potential utility through empirical evaluation.

1. Introduction

Hemorrhage, a critical condition associated with substantial blood loss, can lead to hemodynamic instability, multi-organ dysfunction, and death (Gutierrez et al., 2004). Early detection of hemorrhage is of significance to critical care medicine, as it can help prevent the loss of significant amounts of blood and subsequent hypovolemic shock, allowing timely application of treatment within the critical 'golden hours', and enabling favorable patient outcomes (Rivers et al., 2001). However, standard practice relies on monitoring macro-level indicators of hypovolemia, such as increased heart rate and decreased blood pressure, and yields noticeable indications only after a large amount of blood loss has already taken place (Bárdossy et al., 2011). In this study, we develop models aimed at reducing the time to detection of hemorrhage, hoping that they can guide clinicians in taking proactive steps to mitigate its adverse consequences in patients.

Technical Significance Prior approaches to the problem of hemorrhage detection in vital sign time series have predominantly relied upon supervised methods. They have been facing challenges stemming from a wide diversity of vital sign presentation between individual patients, typically addressed by normalizing currently monitored data to baselines of reference collected from the specific individuals when they were in stable condition (Wertz et al., 2019). Those approaches may work well in e.g. scheduled surgery scenarios, but do not scale well to e.g. trauma patients who are rushed to treatment without the luxury of having stable reference data available. We therefore propose an alternative autoregressive system for hemorrhage detection that does not depend on an individual's stable physiologi-

cal characteristics and instead relies on the instantaneous waveform data for inference. At the test time, our model allows rapid personalization to data of a previously unseen patient in order to discern between bleeding and non-bleeding phases of the vital sign trajectory.

Clinical Relevance Hemorrhage is a common experience of patients in trauma or certain types of surgeries. If not timely discovered and treated, it leads to a substantial risk of mortality and long-term morbidity in survivors (Kauvar and Wade, 2005; Hu, 2004; Mastracci et al., 2012). However, due to the limitations of the sensitivities of physiological vital signs used in current medical practice and their bed-side interpretation, the diagnosis of hemorrhage and resuscitation are often delayed (Bárdossy et al., 2011). We aim to facilitate early discovery of hemorrhage during the initial stage of bleeding using the hemodynamic vital signs collected in bed-side monitoring units. Our results will support more timely treatment of bleeding patients, guide allocation of relevant clinical resources, and result in improved patient outcomes.

2. Prior Work

Both the clinical and machine learning communities are gaining interest in the automatic detection of hemorrhage and subsequent hypovolemic shock. For instance, Falck et al. (2018) applied deep neural networks with Gated Recurrent Units (Cho et al., 2014) and dilated causal convolution (van den Oord et al., 2016) to the high-frequency waveform data of multiple vital signs after normalization with subject-specific baseline characteristics collected during stable state, and achieved Area under Receiver Operating Characteristic Curve (AUC) greater than 0.9 for 30 minutes into bleeding. In intensive care scenarios, this reliance on personal baseline normalization may be overly optimistic as we do not have the luxury of observing the stable physiological characteristics if a patient is already bleeding.

Romero Zambrano et al. (2015) trained Random Forest model with patterns that characterized the sequential changes in the shape of Arterial Blood Pressure (ABP) waveform, which achieved AUC 0.75 and 0.85 at 1 minute and 5 minute post bleeding onset respectively. Lei et al. (2017) classified the respiration cycles during non-bleeding and bleeding phases by learning representative clusters of the correlation structures between central venous blood pressure waveforms observed at the top of the inspiration and bottom of the expiration phases of each respiration cycle, separately for non-bleeding and bleeding phases. Their work achieved AUC 0.862 for 30 minutes into bleeding. Guillame-Bert et al. (2014) used random forest classifiers and regressors for early bleed detection and estimation of the amount of blood lost from a comprehensive set of multiple high-frequency vitals. They have demonstrated the ability to reliably detect hemorrhage within a couple of minutes from its onset. Of further relevance to our work, they found central venous pressure waveform to be the most impactful vital sign for the task, and emphasized the importance of measuring timeliness of detection as an operationally important performance metric, besides accuracy of the classification algorithms that can be quantified with e.g. AUC. In clinical practice, rapid detection with high recall rates allows bleeding to be identified and treated timely, while highly specific discrimination attainable at a later time may be less beneficial for effective clinical intervention. Based on that, we argue that generative or single-class models may be more naturally tailored to the task of fast detection of hemorrhage than more common supervised methods, unless the latter are optimized for minimum latency of detection and

maximum recall, instead of just AUC or accuracy of prediction. We further elaborate on the importance of rapid detection in Section 6.

Unsupervised methods have been studied and utilized by researchers in bed-side monitoring applications. For instance, Fiterau et al. (2011) proposed a probabilistic model to predict heart failure alarms by leveraging multiple vital signs in an unsupervised fashion, thus the model was able to adapt to a patient's real-time status. Our method is motivated by a similar idea: by utilizing only the instantaneous physiological state of a patient, the model should be able to adapt and capture the change to make rapid predictions immediately without requiring any baseline data from a stable state.

3. Data and Featurization

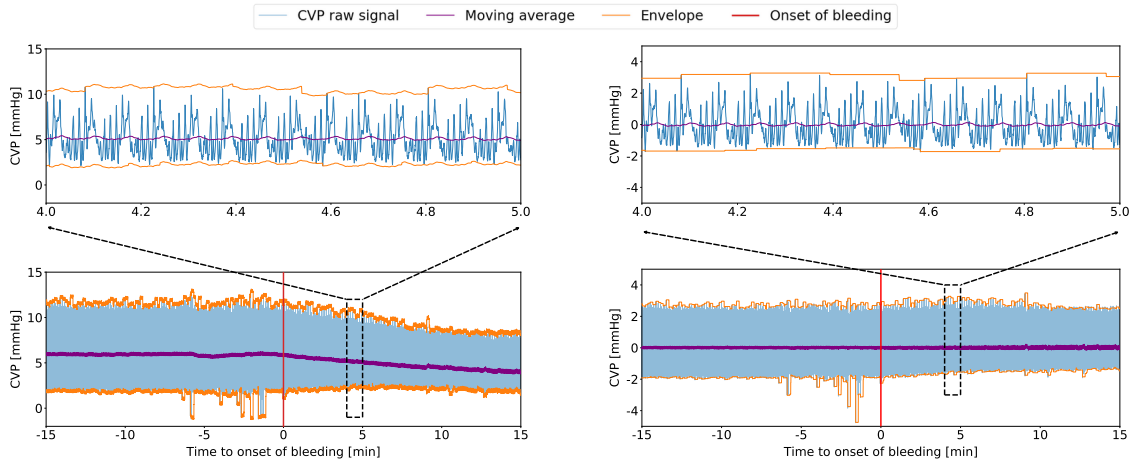
We utilize a real-world laboratory animal dataset. 21 healthy Yorkshire Pigs were sedated and surgically retrofitted to measure the Central Venous Pressure (CVP) and other vital signs. Following the surgery, the pigs were stabilized for a period of 30 minutes (referred to as the 'stable' phase in the paper). Following this, each individual pig was bled at a constant rate of 20mL/min until mean arterial pressure (MAP) dropped to 30 mmHg (referred to as the 'bleeding' phase). During both the stable phase and the bleeding phase, The CVP raw waveform data was collected at a frequency of 250Hz which we downsampled to 50Hz (uniform subsampling with a 20% duty cycle) for all subsequent experiments described in the paper. Fig 1 demonstrates the Central Venous Pressure Waveform for a single subject during both the stable and bleeding phase.

3.1. Featurization

- **Buffer Window ($\mathcal{X}_t$) :** At each time step, we use the observations over the past 10 second window of the raw signal to aggregate statistics required for computing the features for the model. These statistics include the signal moving average, m_t (mean of the raw signal over the entire Buffer Window) and the signal envelope, e_t (the maximum and minimum value of the signal over the buffer period.)

- **Regressive Feature Window (X_t):** At each instance, we normalize the signal by subtracting the statistics computed from the current Buffer Window. This involves subtracting the moving average from the raw signal and dividing the result by the signal envelope to obtain the standardized signal, $\bar{x}_t$. This standardized signal accumulated over a period of 1 second is used as the multivariate feature representation ($\bar{X}_t$) for the proposed model and the High Frequency (HF) baselines.

Note that since the standardization is conducted on-the-fly and uses only the data from the current Buffer Window, our approach does not need to rely on the baseline physiology from the stable non-bleeding phase. This deliberate design choice enables our model to be deployed at test time even if no information about an individual's baseline physiology is available as is the case in trauma care scenarios.



a) b)

Fig. 1: a) The raw CVP signal of a single subject from 15 minutes preceding bleeding onset to 15 minutes post bleeding onset, with its moving average and envelope. Notice how the moving average and the envelope changes post the onset of bleeding. We zoom into the waveform data during a 1-minute time window. b) The CVP signal in a) after normalization with the moving average and envelope of buffer window.

4. The Proposed Model for Rapid Detection Of Hemorrhage (RaDeOH)

4.1. Emission Model

For each patient in our Training Set, we construct two separate models corresponding to the the Stable and Bleeding Phase respectively. We call them emission models and train them in an autoregressive fashion to estimate $P(x_{t+1}|\tilde{X}_t, z_{t+1}, \Theta)$. Here, $\tilde{X}_t$ are the observations in the immediately preceding step and X_{t+1} is the predicted value of the signal at time step $t+1$. z_{t+1} is a binary variable corresponding to whether the model was fit on the bleeding or the non-bleeding phase and Θ corresponds to the learned model parameters.

Readers familiar with time series modelling would naturally consider the class of Autoregressive Moving Average (ARMA) family as an appropriate choice for the emission models. However, we would like to point out that the assumptions of ARMA may make the model somewhat sensitive to artifacts. In particular, these models assume $P(x_{t+1}|X_t, \Theta) \approx \mathrm{N}(f(\Theta^\top X_t), \sigma^2)$ with $f(.)$ typically being linear. Since the Normal distribution has unbounded support, for maximum a posteriori inference the output is unbounded. We restrict our model such that the output in expectation is bounded with respect to the immediate signal history and thus somewhat well behaved, while allowing for non-linear interactions within the co-variates with the use of multilayer perceptrons.

A question arises on the choice of parameterization for the emission models corresponding to each unique patient. Unified parameterizations (or architectures) for all the individuals may be too restrictive, and so at the training time we fit multiple alternative models and choose the parameterization which leads to empirically the best discriminative performance on a held out set of observations made for the same training individual.

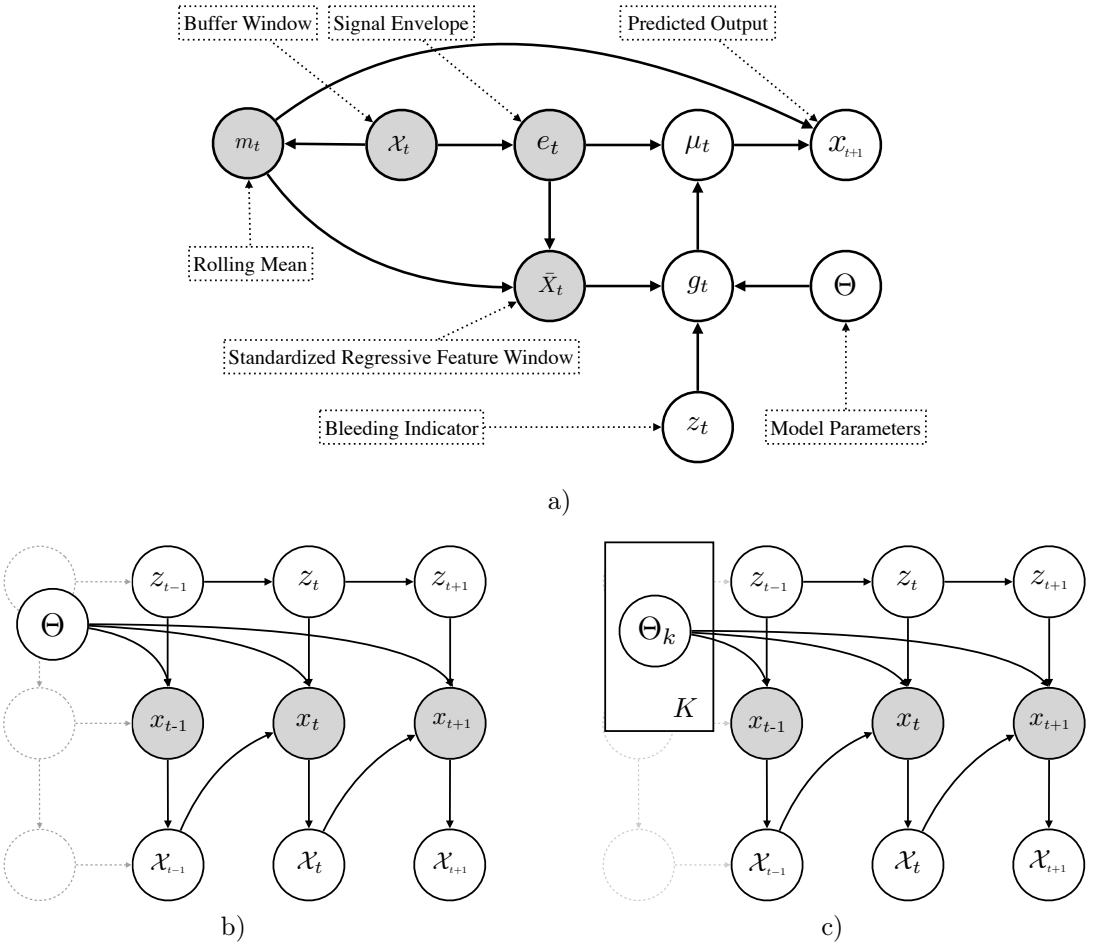


Fig. 2: a) The emission model for an individual subject. The observed variables include the raw signal over the buffer window, $\mathcal{X}_t$, the derived statistics, m_t, e_t, and the standardized signal $\bar{X}_t$. g_t is a Bernoulli output of the model with the parameters Θ. b) The model with latent temporal dependencies (emission specifics have been abstracted out.) c) The model with personalization through the 'Emission Mixture Model', at test time.

Table 1: The Generative Model

1. Transition to the latent state, z_{t+1}, conditioned on the previous state, z_t
   $z_{t+1}|z_t = T(z_{t+1}, z_t)$
2. Compute the envelope (e_t) and rolling mean (m_t) using previous $\mathcal{X}_t$
3. Compute the standardized signal $\bar{X}_t$ using X_t, e_t, m_t
4. Draw a Bernoulli g_t given $\bar{X}_t$ and Θ
   $g_t|\bar{X}_t, \Theta \sim \text{Bernoulli}(f(\bar{X}_t, \Theta))$
5. Select μ_t from g_t
   $\mu_t = e_t^{\uparrow}$, if $g_t = 1$, $\quad \mu_t = e_t^{\downarrow}$, if $g_t = 0$
6. Draw x_t, given μ_t as,
   $x_t|, \mu_t, \sigma \sim \text{Normal}(\mu_t, \sigma^2)$

From the directed acyclic assumptions in Fig. 2 and the generative process as in Table 1 we can write the Conditional Probability under the model as

$$\Pr(x_{t+1}|\tilde{X}_t, m_t, z_{t+1}, \Theta) = \sum_{g_t} \Pr(g_t|\tilde{X}_t, z_{t+1}, \Theta)\Pr(x_{t+1} + m_t|\mu_t, \sigma^2) \tag{1}$$

4.2. Personalization - An Emission Mixture Model

In the previous section we described a unique Emission Model for each of the K individuals in our training dataset. At test time however, we observe emissions from a previously unseen test patient. It is natural to assume that different patients have different underlying physiologies responsible for heterogeneous cardiovascular function and diverse appearance in the lens of vital signs. Thus a single individual emission model may not be generalizable enough to describe the trajectory of a previously unseen patient. We instead hypothesize that the new patient can be described as a mixture of the previously learnt subject-specific Emission Models, drawing from the database of historical patients.

$$\Pr(x_{t+1}|\tilde{X}_t, m_t, z_{t+1} = \mathbf{0}, \mathbf{\Theta}) = \sum_k \alpha_k \Pr^{(k)}(x_{t+1}|\tilde{X}_t, m_t, z_{t+1} = \mathbf{0}, \Theta^k) \tag{2}$$

$$\hat{\alpha} = \arg\max_\alpha \Big[\prod_{x_{t+1} \in \mathcal{X}_t} \Pr(x_{t+1}|\tilde{X}_t, m_t, z_{t+1} = \mathbf{0}, \mathbf{\Theta}) \Big] \quad \text{s.t.} \quad \sum_k \alpha = 1 \tag{3}$$

In order to obtain the counterfactual emission model, or the model corresponding to the bleeding phase, we can simply fix the mixture parameters and replace the Emission Models with their respective counterfactuals corresponding to the bleeding phase.

$$\Pr(x_{t+1}|\tilde{X}_t, m_t, z_t = \mathbf{1}, \mathbf{\Theta}) = \sum_k \hat{\alpha}_k \Pr^{(k)}(x_{t+1}|\tilde{X}_t, m_t, z_t = \mathbf{1}, \Theta^k) \tag{4}$$

Greedy Selection. To compare how personalization impacts performance of the model, we compare it to a greedy selection strategy where, instead of a weighted average of the individual emission models we instead select one model from the K trained models that best describes the held-out unseen subject at test time. Results corresponding to this strategy are hereafter referred to as RaDeOH-G.

4.3. Learning

Note there are two learning phases in our framework. The first involves learning the emission model parameters for each individual in the dataset. The second phase involves on-the-fly learning of the mixing components at test time.

Emission Model Parameters, Θ. We train a separate autoregressive models, $\Pr(x_{t+1}|\tilde{X}_t, m_t, z_t, \mathbf{\Theta})$ for each individual subject in the training dataset. Notice from Eq. 1 that the loss corresponding to the modeling assumption may be written as:

$$\ell\ell(\Theta; x_{t+1}) = \log \sum_{g_t \in \{0,1\}} \Pr(g_t|\tilde{X}_t, z_t)\Pr(x_{t+1}|\mu_t + e_t, \sigma^2, g_t) \tag{5}$$

For pragmatic reasons, we use the following form of the loss for a single observation:

$$\mathbf{NLL}(\Theta; x_{t+1}) = -\log \sum_{g_t \in \{0,1\}} \exp\left(\textsf{Log-Sigmoid}_{(g_t)}(\bar{X}_t, \Theta, \boldsymbol{f}) - \frac{||x_{t+1} - (m_t + e_t^{(k)})||_{\mathbf{2}}}{\sigma^2}\right), \sigma^2 = 1 \tag{6}$$

Here $\boldsymbol{f}$ is a potentially nonlinear function that operates on the input, $\hat{X}_t$, with the parameter set Θ. The summation outside is made numerically stable using the log-sum-exp trick. For the purposes of this paper, for $\boldsymbol{f}$ we experiment with both linear functions and 1- and 2-hidden-layer Multilayer Perceptrons with ReLU activations. We train these models via backpropagation using the popular first order optimizer, Adam (Kingma and Ba, 2014) over the entire dataset as $\mathbf{NLL}(\Theta; \mathcal{D}) = \sum_{x \in \mathcal{D}} \mathbf{NLL}(\Theta; x)$.

We fit separate models for the stable and bleeding phases, and choose the set which gives the best discriminative performance in terms of Area under Receiver Ooperating Characteristic Curve for a held out set of observations randomly sampled from each individual in the training set.

Mixing Parameters, $\boldsymbol{\alpha}$. The second set of parameters to be learnt are the mixture weights in order to personalize the model to the unseen subject at test time. As mentioned previously, this personalization is carried out only on the first 10 seconds of observations corresponding to the buffer window $\mathcal{X}_t$. The mixing parameters are a simplex that sum to one, a constraint we incorporate by using a softmax over the mixing weights. Thus the optimization objective for Eq. 3 can be written as:

$$\mathbf{P\text{-}Loss}(\boldsymbol{\alpha}; x_{t+1}) = -\log \sum_{k} \exp\left(\textsf{Log-Softmax}_{k}(\boldsymbol{\alpha}) - \mathbf{NLL}(\Theta; x_{t+1})\right) \tag{7}$$

Here, k represents an indiviudal subject in the dataset. We minimize this objective over the buffer window at test time, $\mathbf{P\text{-}Loss}(\boldsymbol{\alpha}; \mathcal{D}) = \sum_{x \in \mathcal{X}_t} \mathbf{P\text{-}Loss}(\boldsymbol{\alpha}; x)$. Similar to other parameters, we train these weights using Adam over a fixed number of epochs.

4.4. Inference

At test time, we are interested in inferring z_t given the past history of observations $\{x_1, x_2, ..., x_t\}$ and the learnt model parameters, Θ. In this section, we demonstrate that even with the conditionals on the observation on the previous timesteps, inference is similar to the filtering process in standard Hidden Markov Models or more commonly, the forwards algorithm.

For notational simplicity we denote the observed sequence as $x_{1:t}$. We emphasize that we are primarily interested in the conditional distribution $P(z_t|x_{1:t})$.

The joint $P(z_1, X_1|\Theta)$ is $P(X_1|z_1, \Theta)P(z_1|\Theta)$, but from the directed acyclic assumptions encoded in the model, $z_1 \perp\!\!\!\perp \Theta$, this simplifies to $\alpha(z_1) = P(X_1|z_1, \Theta)P(z_1)$. Now from Bayes Theorem, the posterior probability is given by,

$$\Pr(z_1|x_1, \Theta) = \frac{\Pr(x_1|z_1, \Theta)\Pr(z_1)}{\int_{z_1} \Pr(x_1|z_1, \Theta)\Pr(z_1)}. \tag{8}$$

Note that $\Pr(z_1)$ is a prior, and in the simplest case it can be assumed uninformative, i.e. uniform. $\Pr(X_1|z_1, \Theta)$ can be computed using the fitted models, $\Pr(x_1|z_1 = \mathbf{1}, \Theta)$ and $\Pr(x_1|z_1 = \mathbf{0}, \Theta)$.

Let us now consider the case, where we want to infer z_2 given $\{x_1, x_2\}$. Note that the joint distribution is:

$$\gamma(z_2) = \Pr(z_2, x_1, x_2|\Theta) = \Pr(x_2|z_2, x_1, \Theta)\Pr(z_2, x_1|\Theta). \tag{9}$$

We can rewrite $\Pr(z_2, x_1|\Theta)$ as, $\int_{z_1} \Pr(z_2, x_1|z_1, \Theta)\Pr(z_1|\Theta)$ Note that $z_2 \perp\!\!\!\perp x_1, |z_1, \Theta$, thus, Eq. 9 can be rewritten as:

$$\gamma(z_2) = \Pr(z_2, x_1, x_2|\Theta) = \Pr(x_2|z_2, x_1, \Theta) \int_{z_1} \Pr(z_2|z_1)\Pr(x_1|z_1, \Theta)dz_1. \tag{10}$$

Here, $P(z_2|z_1)$ are the transition probabilities. Note that from this we can derive the familiar recurrence relations corresponding to the forwards algorithm for HMMs as:

$$\gamma(z_t) = \Pr(z_t, x_{1:t}|\Theta) = \Pr(x_t|z_t, x_{t-1}, \Theta) \int_{z_{t-1}} \Pr(z_t|z_{t-1})\gamma(z_{t-1})dz_{t-1}. \tag{11}$$

Now, proceeding as in Eq. 8, we can get $\Pr(z_t|x_{1:t}, \Theta)$ using a definition of the conditional:

$$\Pr(z_t|x_{1:t}, \Theta) = \frac{\Pr(x_{1:t}, z_t|\Theta)}{\int_{z_t} \Pr(x_{1:t}, z_t|\Theta)dz_t}. \tag{12}$$

The transition matrix $\Pr(z_t|z_{t-1})$ that governs the probability of transition from one state to another is a 2×2 matrix. For practical implementation, we do not learn the transition matrix, but instead fix it to be symmetric, with the probability of transition from stable to bleeding being a hyperparameter, τ, and the probability of the reverse transition being $(1 - \tau)$ ensuring the row marginals of the transition matrix sum to one. τ is tuned during cross validation.

5. Baselines

We compare RaDeOH against two types of discriminative models that we describe as High Frequency and Low Frequency based on the temporal features used to train these baselines.

5.1. High Frequency Reference Models

The High Frequency (HF) models rely on the same set of autoregressive features as RaDeOH. Similar to RaDeOH, these models are trained on CVP data downsampled to 50Hz and standardized using the buffer window. We train a separate model for each of the subjects in our training data and tune hyper-parameters for each individual model on a held-out set of observations from the same subject. At the test time, we ensemble the outputs from each individual model to obtain predictions for the held-out subjects. We choose Logistic Regression (LR), Random Forest (RF), and Multilayer Perceptron (MLP) as the HF classifiers. The hyperparameters tuned for each of these model types are shown in the Appendix.

5.2. Low Frequency Reference Models

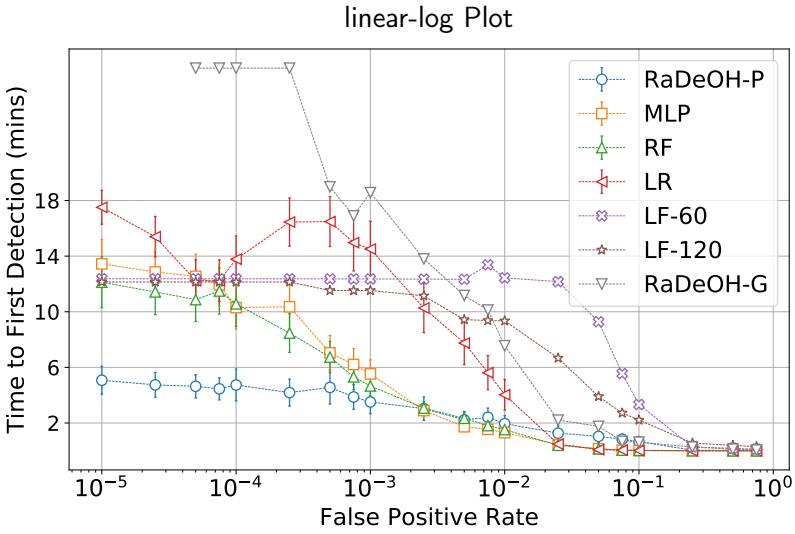
As mentioned in Section 1, the visible physiological indicators often used by clinicians to diagnose or monitor hemorrhage in current practice, such as decreased blood pressure, gradually emerge as the bleeding proceeds. Compared to the autoregressive features that can reflect the instantaneous changes in the waveforms used in HF models, these gradual trends change at much lower rates. We represent these low frequency (LF) trends of CVP with the moving linear slopes of the signal moving average and envelope, derived by fitting the least-squares linear regressions to the trailing time windows of 60 and 120 seconds at each time step. These are the features used to train Logistic Regression models, referred to as 'LF-60' and 'LF-120' in later sections, to mimic the performance attainable by human clinicians in current practice. For the LF models, we concatenate linear slopes of all subjects in the training set to train a single model, and use it to predict the held-out subject at the test time.

6. Results

We evaluate the models in a Leave-One-Subject-Out-Cross-Validation (LOSOCV) protocol. For RaDeOH and the HF baselines, we test the model on an individual held-out subject and average these predictions over all the models, and for LF baselines, we train a single model in each iteration in LOSOCV and obtain predictions for the held-out subject. We report the averaged Activity Monitoring Operator Characteristic (AMOC) (Fawcett et al., 1999; Jiang et al., 2009) that report the False Positive Rate (FPR) as a fucntion of the Time to First True Detection (T2FD), to demonstrate the tradeoffs between latency of detection and predictive error rates. We define the first 'True Detection', as the amount of time from the onset of hemorrhage that the model triggers alarms at the minimum frequency of 5Hz (10% duty cycle of the sampling frequency) sustained over a duration of at least 2 seconds (Fig. 3).

For each of the individual Activity Monitoring Operating Characteristic (AMOC) curves, we may sometimes run into a scenario that below a certain threshold of classifier sensitivity, the false positive rate (FPR) does not fall for higher times to detection, making averaging over these thresholds difficult. Since we perform filtering during inference, RaDeOH is more prone to this condition arising out of numerical challenges as opposed to the baselines. In order to mitigate these effects, for all such lower thresholds the FPR is extrapolated to the FPR at the next higher threshold at which the FPR is well defined. This strategy is pessimistic and we would like to point out that it favors the baseline models more than our proposed approach, ensuring a conservative comparison.

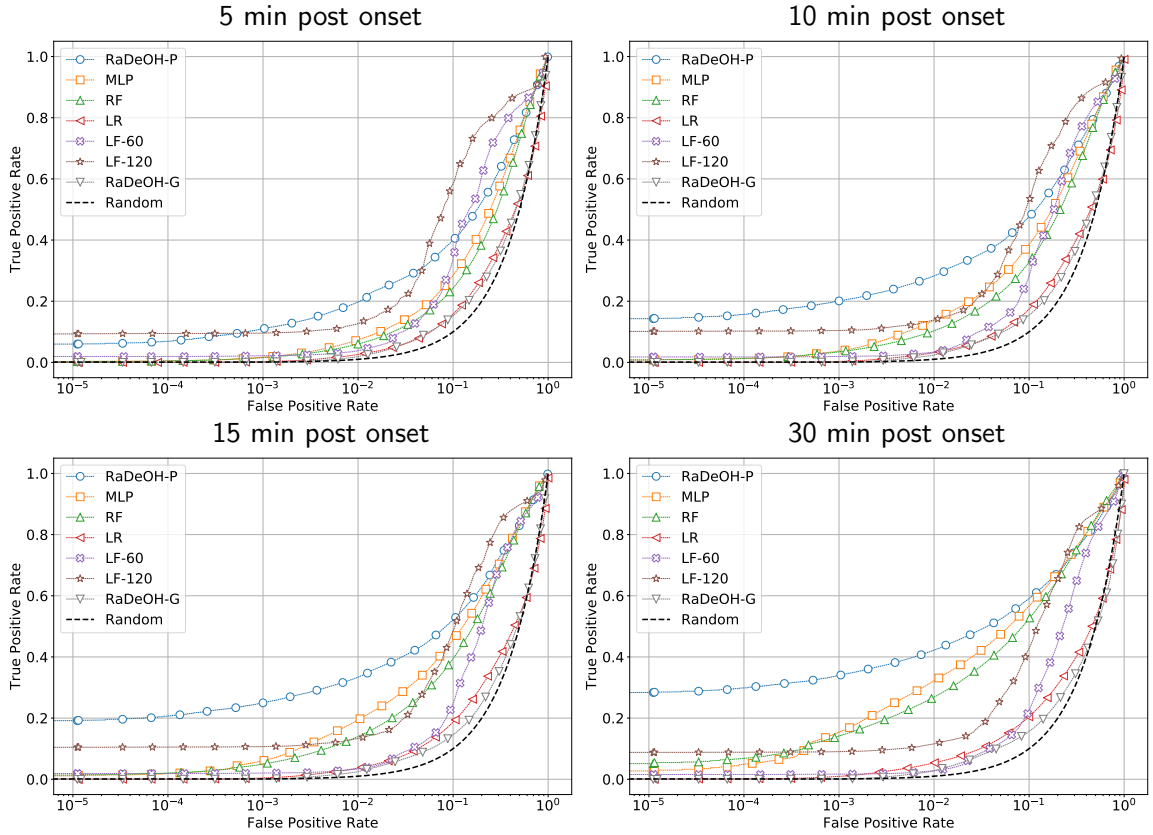
For completeness, we also report the True Positive Rates (TPR) at those thresholds and the Receiver Operating Characteristic (ROC) diagrams in Fig. 4 at different horizons of 5, 10, 15 and 30 minutes into bleeding. However, we maintain that the model selection for the applications where timeliness of detection is the key driving factor, the AMOC provides a more immediately relevant assessment than ROC. While the LF-60, LF-120 seem to have higher recall rates as compared to RaDeOH, high FPRs for these models reduce their practical utility.



Model	Time to First True Detection (sec)				
	FPR$=1\times10^{-5}$	**FPR**$=1\times10^{-4}$	**FPR**$=1\times10^{-3}$	**FPR**$=1\times10^{-2}$	**FPR**$=1\times10^{-1}$
LF-60	742.3 ± 126.3	742.2 ± 126.2	741.5 ± 126.2	746.2 ± 134.7	199.9 ± 85.1
LF-120	728.5 ± 139.0	728.4 ± 139.0	691.6 ± 143.1	560.7 ± 139.3	132.6 ± 72.2
LR	1050.6 ± 72.9	826.7 ± 100.3	870.6 ± 119.4	242.0 ± 65.9	1.4 ± 0.6
RF	727.3 ± 109.7	633.8 ± 97.2	279.8 ± 42.0	91.4 ± 20.00	1.6 ± 0.6
MLP	806.5 ± 104.8	617.9 ± 92.1	331.9 ± 60.6	78.3 ± 16.7	1.0 ± 0.4
RaDeOH-G	No Detection	1651.2 ± 0.0	1114.1 ± 146.9	453.6 ± 106.0	37.2 ± 21.41
RaDeOH-P	304.5 ± 60.1	284.0 ± 68.9	210.7 ± 50.4	117.2 ± 34.3	39.5 ± 18.9

Model	True Positive Rates				
	FPR$=1\times10^{-5}$	**FPR**$=1\times10^{-4}$	**FPR**$=1\times10^{-3}$	**FPR**$=1\times10^{-2}$	**FPR**$=1\times10^{-1}$
LF-60	$1.54\times10^{-2.0}\pm0.0$	$1.56\times10^{-2.0}\pm0.0$	$1.68\times10^{-2.0}\pm0.01$	$2.82\times10^{-2.0}\pm0.01$	$2.25\times10^{-1.0}\pm0.04$
LF-120	$8.78\times10^{-2.0}\pm0.04$	$8.84\times10^{-2.0}\pm0.04$	$9.03\times10^{-2.0}\pm0.04$	$1.16\times10^{-1.0}\pm0.05$	$4.29\times10^{-1.0}\pm0.07$
LR	$7.54\times10^{-4.0}\pm0.0$	$1.61\times10^{-3.0}\pm0.0$	$9.49\times10^{-3.0}\pm0.0$	$5.37\times10^{-2.0}\pm0.01$	$2.05\times10^{-1.0}\pm0.01$
RF	$5.07\times10^{-2.0}\pm0.01$	$6.84\times10^{-2.0}\pm0.02$	$1.41\times10^{-1.0}\pm0.02$	$2.69\times10^{-1.0}\pm0.03$	$5.25\times10^{-1.0}\pm0.03$
MLP	$2.69\times10^{-2.0}\pm0.01$	$4.98\times10^{-2.0}\pm0.01$	$1.54\times10^{-1.0}\pm0.02$	$3.22\times10^{-1.0}\pm0.03$	$5.67\times10^{-1.0}\pm0.03$
RaDeOH-G	$2.49\times10^{-5.0}\pm0.0$	$1.53\times10^{-4.0}\pm0.0$	$1.08\times10^{-3.0}\pm0.0$	$2.77\times10^{-2.0}\pm0.02$	$1.57\times10^{-1.0}\pm0.05$
RaDeOH-P	$2.82\times10^{-1.0}\pm0.07$	$2.99\times10^{-1.0}\pm0.07$	$3.4\times10^{-1.0}\pm0.07$	$4.22\times10^{-1.0}\pm0.07$	$5.91\times10^{-1.0}\pm0.07$

Fig. 3: The Activity Monitoring Operating Characteristic for RaDeOH compared to the baselines. RaDeOH considerably outperforms the baselines within the first 5 minutes from the onset of bleeding. False positive rates decrease with the amount of time allowed to make the determination, and for RaDeOH, they are below 1% within the first two minutes and drop to 0.1% within about another minute from the onset of bleeding.



Model	TPR at FPR $= 1 \times 10^{-3}$			
	5 min post onset	10 min post onset	15 min post onset	30 min post onset
LF60	0.022 ± 0.02	0.020 ± 0.01	0.020 ± 0.01	0.017 ± 0.01
LF120	0.096 ± 0.05	0.104 ± 0.05	0.107 ± 0.05	0.090 ± 0.04
LR	0.002 ± 0.00	0.003 ± 0.00	0.004 ± 0.00	0.009 ± 0.00
RF	0.016 ± 0.00	0.035 ± 0.01	0.050 ± 0.01	0.141 ± 0.02
MLP	0.017 ± 0.01	0.037 ± 0.01	0.061 ± 0.01	0.154 ± 0.02
RaDeOH-P	0.110 ± 0.04	0.201 ± 0.06	0.250 ± 0.07	0.340 ± 0.07

Model	TPR at FPR $= 1 \times 10^{-2}$			
	5 min post onset	10 min post onset	15 min post onset	30 min post onset
LF-60	0.038 ± 0.03	0.034 ± 0.02	0.034 ± 0.01	0.028 ± 0.01
LF-120	0.127 ± 0.06	0.139 ± 0.06	0.133 ± 0.05	0.116 ± 0.05
LR	0.026 ± 0.00	0.032 ± 0.00	0.037 ± 0.00	0.054 ± 0.01
RF	0.060 ± 0.01	0.103 ± 0.02	0.140 ± 0.02	0.269 ± 0.03
MLP	0.074 ± 0.02	0.138 ± 0.02	0.194 ± 0.02	0.322 ± 0.03
RaDeOH-P	0.197 ± 0.06	0.283 ± 0.07	0.334 ± 0.07	0.422 ± 0.07

Model	Area Under ROC Curve			
	5 min post onset	10 min post onset	15 min post onset	30 min post onset
LF-60	0.739 ± 0.02	0.725 ± 0.02	0.700 ± 0.02	0.749 ± 0.03
LF-120	0.821 ± 0.02	0.813 ± 0.02	0.788 ± 0.02	0.815 ± 0.03
LR	0.528 ± 0.00	0.530 ± 0.00	0.530 ± 0.00	0.524 ± 0.00
RF	0.717 ± 0.02	0.752 ± 0.02	0.797 ± 0.02	0.661 ± 0.02
MLP	0.741 ± 0.02	0.773 ± 0.02	0.803 ± 0.02	0.688 ± 0.02
RaDeOH-P	0.765 ± 0.03	0.787 ± 0.03	0.806 ± 0.04	0.721 ± 0.04

Fig. 4: ROC curves comparing the performance of RaDeOH versus the HF and LF Baselines at different horizons of the onset of bleeding.

7. Conclusion

We proposed a novel approach for rapid detection of hemorrhage using Central Venous Pressure waveform data that leverages only the instantaneous physiological characteristics of patients. Compared to previous work in hemorrhage detection, our method does not require normalization with the patients' stable status data as personal reference, or require a period long enough for the commonly used physiological indicators of hemorrhage to become humanly noticeable for assessment. With only short time window of CVP observation, our method outperforms popular discriminative classifiers including Logistic Regression, Random Forest and Multilayer Perceptron trained on the same set of autoregressive features, in terms of time to detection (as shown in AMOC data) and overall discriminative power (as shown in ROC data) during early stages of bleeding, i.e. the initial few minutes from the onset of hemorrhage. By alleviating the need for stable reference data of each subject, our method is particularly applicable to monitoring emergency and trauma patients or patients undergoing unscheduled surgeries for whom such reference data may not be immediately available. These characteristics of RADEOH could improve clinical practice by timely and reliably informing medical personnel of emerging health crises that can be discerned from routinely collected high frequency bedside monitoring data, in all monitored patients.

Our current work has several limitations that can be overcome with additional research. Currently, we only utilize the Central Venous Pressure (CVP) waveform data. One obvious next step is to incorporate other vital signs available in clinical or field care settings, including both non-invasively collected data such as Plethysmography, and the invasive measurements such as Arterial Blood Pressure. Our prior research into parsimony of hemodynamic data indicated potential benefits from including such additional data whenever available (Wertz et al., 2019). For practical applicability, it would be desirable to devise models that could pragmatically use whatever data that is being instantaneously available, even a small subset of generally measurable vital signs, e.g. due to limited access to some means of hemodynamic monitoring in the field, or due to limitations of the patient's health status. The proposed graphical model formulation could allow incorporating incomplete data with minor extensions, making the method applicable in a wider range of health care settings. We also consider extending the model to identify other physiological states of clinical importance such as the emergence of sepsis or hemorrhagic shock (in our data, some subjects developed hemorrhagic shock in the course of bleeding and required aggressive resuscitation). Another direction of possible future development is to enable simultaneous prediction of multiple types of conditions. Such multi-tasking approach would be particularly useful for rapid disambiguation among conditions that may initially present with similar symptoms but require different therapies, such as e.g. hypotension due to hypovolemia vs. endotoxic shock.

Acknowledgments

This work has been partially supported by the US Army Medical Research and Material Command (W81XWH19C0101, W81XWH19C0083), DARPA (FA87501720130), and NIH (R01GM117622).

References

Gergely Bárdossy, Gábor Halász, and Tibor Gondos. The diagnosis of hypovolemia using advanced statistical methods. *Computers in Biology and Medicine*, 41(11):1022–1032, 2011. doi: 10.1016/j.compbiomed.2011.09.002.

Kyunghyun Cho, Bart van Merrienboer, ÃĞaglar GülÃğehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. In *EMNLP*, 2014.

Fabian Falck, Michael R. Pinsky, and Artur Dubrawski. Deep sequence modeling for hemorrhage diagnosis. In *Proceedings of the Neural Information Processing Systems Machine Learning for Health (ML4H) Workshop 2018*, 2018.

Tom Fawcett, Foster J Provost, et al. Activity monitoring: Noticing interesting changes in behavior. In *KDD*, volume 99, pages 53–62. Citeseer, 1999.

Madalina Fiterau, Artur Dubrawski, and Can Ye. Real-time adaptive monitoring of vital signs for clinical alarm preemption. *Emerging Health Threats Journal*, 4(0):11146, 2011. doi: 10.3402/ehtj.v4i0.11146.

Mathieu Guillame-Bert, Artur Dubrawski, Lujie Chen, Andre Holder, Michael R. Pinsky, and Gilles Clermont. Utility of empirical models of hemorrhage in detecting and quantifying bleeding. *Critical Care Medicine*, 40, 2014. doi: 10.1007/s00134-014-3451-0.

Guillermo Gutierrez, H David Reines, and Marian E Wulf-Gutierrez. Clinical review: hemorrhagic shock. *Critical Care*, 8(5):373, 2004. doi: 10.1186/cc2851.

Serena S. Hu. Blood loss in adult spinal surgery. *European Spine Journal*, 13(S01):S3–S5, 2004. doi: 10.1007/s00586-004-0753-x.

Xia Jiang, Gregory F Cooper, and Daniel B Neill. Generalized AMOC curves for evaluation and improvement of event surveillance. In *AMIA Annual Symposium Proceedings*, volume 2009, page 281. American Medical Informatics Association, 2009.

David S Kauvar and Charles E Wade. The epidemiology and modern management of traumatic hemorrhage: Us and international perspectives. *Critical Care*, 9(Suppl 5):S1âĂŞS9, 2005. doi: 10.1186/cc3779.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Eric Lei, Kyle Miller, and Artur Dubrawski. Learning mixtures of multi-output regression models by correlation clustering for multi-view data. *CoRR*, abs/1709.05602, 2017. URL http://arxiv.org/abs/1709.05602.

Tara Mastracci, Mohit Bhandari, Raman Mundi, Sandro Rizoli, Bartolomeu Nascimento, and Martin Schreiber. Operative blood loss, blood transfusion and 30-day mortality in older patients after major noncardiac surgery. *Canadian Journal of Surgery*, 55(6):426–428, 2012. doi: 10.1503/cjs.025212.

Emanuel Rivers, Bryant Nguyen, Suzanne Havstad, Julie Ressler, Alexandria Muzzin, Bernhard Knoblich, Edward Peterson, and Michael Tomlanovich. Early goal-directed therapy in the treatment of severe sepsis and septic shock. *New England Journal of Medicine*, 345 (19):1368–1377, 2001. doi: 10.1056/nejmoa010307.

Sergio Romero Zambrano, Mathieu Guillame-Bert, Artur Dubrawski, Gilles Clermont, and Michael R. Pinsky. Detection of hemorrhage by analyzing shapes of the arterial blood pressure waveforms. *Intensive Care Medicine Experimental*, 3(S1):A589, 2015. doi: 10.1186/2197-425x-3-s1-a589.

Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alexander Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. Wavenet: A generative model for raw audio. In *Arxiv*, 2016. URL https://arxiv.org/abs/1609.03499.

Anthony Wertz, Gilles Clermont, Artur Dubrawski, and Michael R. Pinsky. Hemodynamic monitoring parsimony: Minimal information for rapid hemorrhage detection. *Intensive Care Medicine Experimental*, 7, 2019.

Appendix A

The hyperparameters tuned for each of the High Frequency (HF) baseline models are listed in the table below.

Table 2: Each classifier is tuned with all combinations of hyperparameters, and the set of hyperparameters that achieves the best AUC on the held-out set of each subject is chosen to fit the best model.

Model	parameters	values
LR	Penalty	l1, l2
	C (inverse of regularization strength)	$10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^{2}$
RF	Number of trees	$10, 20, 50$
	Maximum depth of the trees	$3, 5, 10$
MLP	Number of hidden layers	$1, 2$
	Number of nodes in hidden layer 1	30
	Number of nodes in hidden layer 2	20