A Composite Multi-Attention Framework for Intraoperative Hypotension Early Warning

Feng Lu[*1], **Wei Li**[*2], **Zhiqiang Zhou**[*3], **Cheng Song**[1], **Yifei Sun**[1], **Yuwei Zhang**[1], **Yufei Ren**[3],
Xiaofei Liao[1], **Hai Jin**[1], **Ailin Luo**[3], **Albert Y. Zomaya**[2]

[1] National Engineering Research Center for Big Data Technology and System, Services Computing Technology and System
Lab, Cluster and Grid Computing Lab, School of Computer Science and Technology, Huazhong University of Science and
Technology, China
[2] The Australia-China Joint Research Centre for Energy Informatics and Demand Response Technologies, Centre for
Distributed and High Performance Computing, School of Computer Science, The University of Sydney, Australia
[3] Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, China
{lufeng,songch,zhangyuwei,sunyifei,xfliao,hjin}@hust.edu.cn, {weiwilson.li,albert.zomaya}@sydney.edu.au,
{zqzhou,yfren,alluo}@tjh.tjmu.edu.cn

Abstract

Intraoperative hypotension (IOH) events warning plays a crucial role in preventing postoperative complications, such as postoperative delirium and mortality. Despite significant efforts, two fundamental problems limit its wide clinical use. The well-established IOH event warning systems are often built on proprietary medical devices that may not be available in all hospitals. The warnings are also triggered mainly through a predefined IOH event that might not be suitable for all patients. This work proposes a *composite multi-attention* (CMA) framework to tackle these problems by conducting short-term predictions on user-definable IOH events using vital signals in a low sampling rate with demographic characteristics. Our framework leverages a multi-modal fusion network to make four vital signals and three demographic characteristics as input modalities. For each modality, a multi-attention mechanism is used for feature extraction for better model training. Experiments on two large-scale real-world datasets show that our method can achieve up to 94.1% accuracy on IOH events early warning while the signals sampling rate is reduced by 3000 times. Our proposal CMA can achieve a mean absolute error of 4.48 mm Hg in the most challenging 15-minute mean arterial pressure prediction task and the error reduction by 42.9% compared to existing solutions.

Introduction

Intraoperative hypotension (IOH) events are closely associated with many postoperative complications, such as a sharp increase in mortality, myocardial injury, acute kidney damage, postoperative delirium, and other neurological complications (Salmasi et al. 2017). Early warning of IOH events throughout the perioperative period is an effective means to prevent postoperative complications. It enables clinicians to act proactively from the usual reactive responses to hypotensive events before the consequences are seen (Davies et al. 2020; Asai et al. 2019).

---

[*]These authors contributed equally.

Many efforts have been devoted to developing artificial intelligence solutions to provide IOH warnings throughout the perioperative period. In (Lee et al. 2021), the authors studied the performance of deep learning models on arterial pressure prediction and IOH early warning using invasive and non-invasive blood pressure readings. The results suggest that invasive blood pressure readings better predict IOH events via gaining continuous arterial pressure values (a.k.a. beat-to-beat blood pressure) from arterial waveform analysis. Invasive blood pressure monitoring is a well-developed and widely used technology in the intensive care unit and the operating rooms. The invasive monitoring involves the insertion of a catheter into a suitable artery and then displaying the numerical readings of systolic, diastolic, and *mean arterial pressure* (MAP) on a monitor. To our knowledge, (Hatib et al. 2018; Wijnberge et al. 2020) developed the hypotension prediction index algorithm using invasive arterial waveform and some other features to yield the best-known IOH prediction results. It is a shame that multiple factors in practice still limit the clinical use of invasive blood pressure monitoring. The typical medical concerns of invasive blood pressure monitoring include potential infection risk, local thrombosis, and drugs entering the arterial line to form crystals and thus causing catastrophic limb ischemia. In addition, the high cost of the monitoring equipment and its replacement parts, and stable electrical supply further hamper its appeal in many settings. Non-invasive and minimally invasive blood pressure monitoring techniques have gained increasing attention these days to serve as a viable alternative for patients. Limited by the techniques, the quality of the blood pressure readings from these alternatives is not yet comparable to the invasive solution.

Another challenge is that most existing early warning studies on intraoperative hypotension are designed for a specific definition of IOH. Once the IOH is defined, the data will be used to train the model according to this specific definition. When a new IOH definition needs to be used, the adoption cost is exceedingly high as the model is required to be reconstructed again from scratch. In (Inada et al. 2021),

the authors studied intraoperative hypotension using the definition of systolic blood pressure below 80 mm Hg. They trained a random forest and three boosting models using preoperative lifestyle questionnaires, patient demographics, and preoperative examination information. The random forest model achieved the best accuracy of 71.1% in the study. In (Hatib et al. 2018), the authors used a machine learning method to construct an intraoperative hypotension index system with a defined mean blood pressure represented by mean arterial pressure below 65 mm Hg for at least 1 minute. The accuracy of predicting hypotension 15 minutes ahead was 87.5%. In (Kendale et al. 2018), the authors compared the prediction models built by eight machine learning algorithms using the definition of mean blood pressure below 55 mm Hg within 10 minutes of anesthetization. The gradient boosting machine algorithm performed best in all models with the AUROC of 0.76. The systematic review (Wesselink et al. 2018) further shows that IOH has multiple definitions according to MAP and the minimum duration of hypotension. Not a single definition was found to be generally accepted in the field. The elevated risks of using different IOH definitions were also reported. The end-organ injury was reported for a more than 10 minutes exposure to MAP<80 mm Hg or shorter durations with MAP<70 mm Hg. More health risks were reported with increased durations for MAP<65–60 mm Hg or any exposure with MAP<55–50 mm Hg.

We propose a *composite multi-attention* (CMA) framework to address the mentioned challenges with a divide-and-conquer metaphor for enabling short-term IOH events prediction. The core idea of CMA is to use the machine learning model to predict short-term MAP instead of IOH events for better personalization of care. CMA consists of two main components, blood pressure prediction and IOH detection. In the blood pressure prediction component, we use multiple hypotension biomarkers as a whole to reveal the variation of MAP on time. This strategy enables our framework to no longer solely rely on the collected quality data from arterial pressure monitoring and acquire the ability to work with a wide range of blood pressure monitors. In our proposal, the use of the IOH definition is only involved in the IOH detection component, not the entire framework. The IOH definition is thus not fully embedded in the trained model but serves as a screening condition for event detection. This strategy creates a user-friendly interface for anesthesiologists to customize intraoperative hypotension treatment regimes based on the patient background that has been demonstrated with superior clinical effectiveness (Inada et al. 2021). In summary, our contributions are as follows:

- A composite multi-attention framework is proposed to issue early warnings for intraoperative hypotension by using the latest monitoring of vital signs in low sampling rates and demographic characteristics. Our proposal is the first work adaptive to the low sampling rate on vital signs and shows strong potential incorporated with a wide range of blood pressure monitors.

- The proposed framework fuses the low sampling se-

quences of the vital signals and demographic characteristics into the four modalities to enrich the representations of patient conditions. The multi-attention mechanism is also employed to extract salient features from the modalities for achieving high MAP prediction accuracy.

- Extensive experiments and ablation studies were conducted on two real-world surgery datasets to demonstrate the effectiveness of the proposed framework. One of the datasets we used in the experiment is the world's largest gathering of surgical records, with 387,291 patients from August 2015 to April 2021, which is at least 60 times larger than any available ones.

Preliminaries

Static Indicator: Static indicators are those demographic characteristics (Shimasaki 2020), such as age, sex, height, and weight, to help determine the IOH risks for patients.

Dynamic Indicator or Sequence: The vital signals, mean arterial pressure, diastolic blood pressure, body temperature, and heart rate are referred to as dynamic indicators in this work. These dynamic indicators have been instrumentally recorded during the surgery and retrieved at a fixed sampling rate to form dynamic sequences for close observation of the patient conditions. A dynamic sequence $\overrightarrow{D_i}$ is a series of readings of a dynamic indicator generated from a monitor from the beginning to the end of the surgery. A demonstrative example is given in Figure 1.

Surgery Case: A patient can be admitted to the hospital multiple times and has one or more surgeries in one admission. We denote a surgery case c by a tuple (p, v, o) for a patient p, at the admission v to have the surgery o.

Observation Window: The observation window represents the duration of the collected vital signs required to predict the arterial pressure in the following minutes. Please note that it is a fixed value within the lifetime of the surgery once it is decided. A demonstrative sample is given in Figure 1. As seen, we extract the latest sequence segment within the length of the observation window. It can be mathematically described as $\overrightarrow{D_o} = \{v_t, v_{t+1}, \cdots, v_{t+n}, 0 \le t < n\}$, where v_t represents a reading of the vital signal at time t. The observation window can move forward on a user-defined step.

MAP Short-term Prediction: Short-term prediction often refers to predicting a variable for 0 to 6 hours. In this work, we mean the prediction period is in between 0 to 15 minutes. As shown in Figure 1, MAP short-term prediction means using the latest sequence segments of the vital signals from the observation window to predict the following arterial pressure sequence segment. The predicted sequence segment has the same length as the observation window and contains the same amount of readings. This window refers to as the prediction window.

IOH Event: An IOH event is defined as the MAP value is lower than p mm Hg and lasts for at least d minutes. It can be denoted as
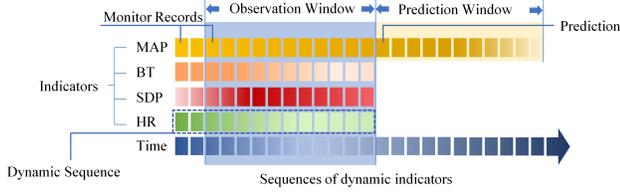
Figure 1: MAP (*mean arterial pressure*), SDP (*diastolic blood pressure*), BT (*body temperature*), and HR (*heart rate*) are four dynamic indicators used in our framework. Each square block represents a numeric reading of a dynamic indicator at a selected time point. The numeric readings obtained from the monitors of these indicators constitute the dynamic sequences. The observation window is used to continuously extract an equal-length latest reading from the dynamic sequence during the surgery for further data processing. The prediction window indicates the period of the dynamic sequence formed by the predicted values from our model, and its length is equivalent to the observation window.

$$e = \begin{cases} 1, \forall i \in S < p \\ 0, \exists i \in S \geq p \end{cases} \tag{1}$$

where S is a vector representing a partial sequence of the predicted MAP values. The size of S is determined by the number of samples found in d minutes.

User-definable IOH Event Warning: In clinical practice, the warning is expected to be triggered at least 2 minutes (Wijnberge et al. 2020) before IOH occurrence to create chances for anesthesiologists to intervene as early as possible. As mentioned, our framework can incorporate different IOH definitions. No matter what definition is used, it will be used as the selection criteria to detect IOH events in the prediction window. Supposing an IOH event will occur at t_e minute, our early warning should be triggered no later than the $t_e - 2$ minute. Please note that the early warning function becomes available only when $t_e > w_o + d_r$, where w_o represents the size of the observation window and d_r represents the reserved duration for the reaction of anesthetists with a general setting of at least 2 minutes.

Methodology

In this section, we introduce the technical details of our proposed CMA framework to issue an early warning for user-definable IOH events during surgery. The overview of our framework design is shown in Figure 2. We first select three static indicators, age, sex, and BMI (*body mass index*), and four dynamic indicators, MAP (*mean arterial pressure*), SDP (*diastolic blood pressure*), BT (*body temperature*), and HR (*heart rate*) as the raw inputs for our framework. We then develop a multi-modal representation module to combine those static and dynamic indicators into four modalities to enrich the representation of the raw inputs. Each modality is formed by one vital sign concatenated with three demographic characteristics. A multi-attention mechanism is

employed to extract salient features and their intertwined dependencies in each modality. After feeding the extracted features into the convolutional neural network, we can predict the MAP short-term results as time goes on. Combined with the user-definable IOH, the framework can thus provide the warning before the IOH event happens.

Composite Multi-modal Representation

As mentioned, we denote the dynamic sequence as a vector $\overrightarrow{D_i}$. Each vector is concatenated with all m static indicators, in the form of vectors $\overrightarrow{S_{i,j}}$, into a matrix to be a modal representation. The values in the matrix are normalized by Z-score (Friedman and Komogortsev 2019) or one-hot encoding based on whether the indicator is a continuous variable. Once completed, a composite modality M_i is thus formed with a dynamic sequence $\overrightarrow{D_i}$ and three sequences containing the repeated static indicators $\overrightarrow{S_{i,0}} \cdots \overrightarrow{S_{i,j}}$. All n modalities $M_0 \cdots M_i, i \in \{0 \cdots n\}$ are turned into the below tensor for further processing.

$$\begin{bmatrix} M_0 \\ \cdots \\ M_i \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} \overrightarrow{D_0} & \overrightarrow{S_{0,0}} & \overrightarrow{S_{...}} & \overrightarrow{S_{0,j}} \end{bmatrix}^T \\ \cdots \\ \begin{bmatrix} \overrightarrow{D_i} & \overrightarrow{S_{i,0}} & \overrightarrow{S_{...}} & \overrightarrow{S_{i,j}} \end{bmatrix}^T \end{bmatrix} \tag{2}$$

Modality Feature Extraction and Fusion

This section describes the multi-attention mechanism used in our framework, which has shown high capability in extracting the features and dependencies from modalities (Zhang et al. 2022; Wang et al. 2020; Song et al. 2022).

Positional Encoding The multi-attention mechanism is treated as a feed-forward layer that reads the sequence in the observation window at once. Since attention is computed independently on every data point in the sequence, the order between data points is ignored. The value embedded in the order could be further ignored in the later modules if we do not explicitly reveal it. Inspired by the work of (Vaswani et al. 2017), we adopt and apply the idea of positional encoding in our design to address this issue. We position the data points in the sequence and record their orders explicitly as a vector. More specifically, let t be the relative order of a data point in a sequence, $\overrightarrow{P_t} \in \mathbb{R}^d$ will be its corresponding encoding. It is defined as follow:

$$\overrightarrow{P_t^{(i)}} = f(t)^{(i)} := \begin{cases} sin(\omega_k \cdot t), & \text{if } i = 2k \\ cos(\omega_k \cdot t), & \text{if } i = 2k+1 \end{cases} \tag{3}$$

ω_k can be further defined as

$$\omega_k = \frac{1}{10000^{2k/d}} \tag{4}$$

where d represents the encoding dimension (in here, $d \equiv 7$). The positional encoding $\overrightarrow{P_t}$ in the form of a vector containing multiple pairs of sines and cosines for the sequence is then in conjunction with the sequence itself for the following modules for further processing.
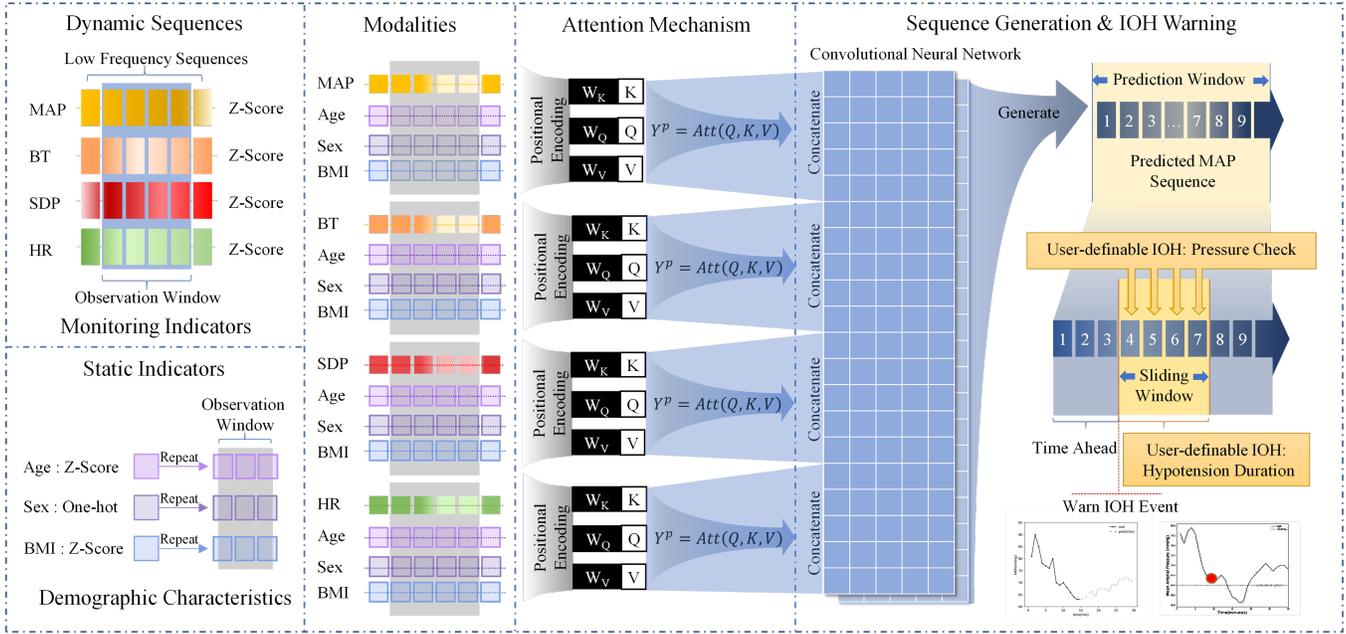
Figure 2: Overview of the CMA composite multi-attention framework. Three static indicators, age, sex, and BMI, are repeatedly concatenated to the four dynamic sequences to form four modalities. The multi-attention mechanism acts on each modality, extracting features and feeding them into the following convolutional neural network to perform MAP sequence prediction. By integrating the short-term blood pressure prediction with the IOH definition, the framework can give early warning for the user-definable IOH events.

Feature Extraction and Fusion The multi-attention mechanism used in our framework has eight heads and eight dimensions. Positional encoded sequences are delivered in $Key - Value$ pairs, thus n groups of inputs can be represented as $(K, V) = [(k_1, v_1), (k_2, v_2), ..., (k_n, v_n)]$, where Key is used to calculate the attention distribution ϵ_i and $Value$ is used to calculate the aggregation. Here both Key and $Value$ for each group are the same, a segment of positional encoded sequences. For a $Query$ representing a future input from a modality to match a Key, we are not to ensure the consistency of Key and $Query$ but the similarity between them. The similarity is the weight of the value. The weighted sum of all values is the attention. The detailed steps are the following:

Step 1: Calculate the similarity of Key and $Query$. The attention value s_i can be calculated using an additive model, dot product model, or cosine similarity. In our framework, we use cosine similarity according to our practice.

$$s_i = F(Q, k_i) \qquad (5)$$

Step 2: Normalize attention values using Softmax function. The weights of significant values can be emphasized.

$$\alpha_i = softmax(s_i) = \frac{exp(s_i)}{\sum_{j=1}^{N} exp(s_j)} \qquad (6)$$

Step 3: Sum up values with weights:

$$Attention((K, V), Q) = \sum_{i=1}^{N} \alpha_i v_i \qquad (7)$$

Finally, the attention value can be summarized as:

$$
\begin{aligned}
att((K, V), q) &= \sum_{i=1}^{N} \alpha_i v_i \\
&= \sum_{i=1}^{N} \frac{exp(s(K_i, q))}{\sum_j exp(s(K_j, q))}
\end{aligned} \qquad (8)
$$

IOH Warning by Short-term MAP Prediciton

Network Configuration for MAP Prediction The attention-weighted features are fed into a lightweight convolutional neural network. The network contains two convolutional layers, a max-pooling layer using the ELU activation function, and three dense layers. The first convolutional layer has a kernel of $(n_{indicators} + n_{P.E.})/2$ to reduce the dimension by half, where $n_{indicators}$ is the total amount of dynamic and static indicators and $n_{P.E.}$ is to match the size of the positional encoding for all sequences. The second convolutional layer reduces the kernel size by half with a kernel size of $(n_{indicators} + n_{P.E.})/4$. The output of the convolutional layers and the max pooling layer is concatenated, reshaped, and fed into the following dense layers. Layer normalization is applied to stabilize the learning process before the last dense layer generates the MAP sequence.

User-definable IOH-event Prediction With the short-term arterial pressure prediction sequence, we can set up an IOH definition to search IOH events on the sequence and issue the warning once found. Since the IOH definition in our

design is only used for searching the hypotension events, the anesthesiologists can make any setting based on their experience before the surgery. We also simplify the IOH event detection under different IOH definitions by two key variables, IOH risk threshold, and duration, for a better user experience. As shown in Algorithm 1, the IOH event detection can be easily customized by adjusting the duration t and the IOH risk threshold l as needed.

Experiments

We evaluated the performance of the CMA framework in a set of experiments on two real-world datasets to answer the following four questions:

- Q1: How is the MAP short-term prediction performance of CMA compared to various baselines?
- Q2: How is the IOH events warning performance of CMA compared to state-of-the-arts solution?
- Q3: How do the composite modalities contribute to the model performance?
- Q4: How is the performance of CMA is affected by the vital signs sampling rate?

DataSets In the experiments, we employed two real-world datasets, Tongji dataset from Tongji Hospital and an open-access real-world dataset. The Tongji dataset consists of 387,291 patient records collected between October 2015 and April 2021. The Institutional Review Board (IRB) of Tongji Hospital approved this study and waived the requirement for patient informed consent as part of the study approval (IRB number: 20210304). The other dataset, VitalDB (Lee and Jung 2018), consists of 6,388 patient records. Our experimental datasets were built on exacting the mentioned dynamic and static indicators from these two datasets. The dynamic sequences in Tongji dataset are collected every 30 seconds. The sampling rate of the dynamic sequences in VitalDB is between 500Hz and 100Hz. We used the sampling rate at 30 seconds per value on both datasets in our experiments to allow perfect data alignment. In this case, our sampling rate is 3000 times lower than those works (Lee et al. 2021; Davies et al. 2020; Lee et al. 2022) using 100Hz data sampling rate. After removing those records with a missing rate above 20%, we have 21,011 patient records from Tongji dataset and 1,560 patient records from VitalDB. We divided these records into training, validation, and test instances in a proportion of 70%, 20%, and 10%, respectively.

Experimental Environment Setting The network used in our CMA framework is implemented by Keras using the Tensorflow 2 backend with the default setting except for the following ones. In the training phase, we use an Adam optimizer with a batch size of 256, and the early stopping is enabled with the patience of 256. The embedding dimension size d and the depth convolutional neural layers L are set as 64 and 2, respectively. The size of the observation window was 15 minutes. The predicted sequence lengths were set to 3, 5, 10, or 15 minutes unless otherwise stated. We used the IOH definition of MAP staying below 65mm Hg for at least 1 minute unless otherwise stated. We used the same model and parameter settings for both Tongji and VitalDB
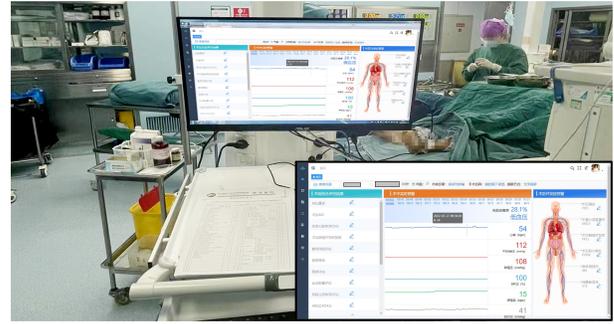
Figure 3: As a clinical trial, our CMA framework has been deployed in the intelligent operating room at Tongji Hospital. The display screen is divided into three sections: preoperative assessment, MAP short-term prediction, and IOH real-time early warning. All related information has been blurred to protect patient privacy.

Algorithm 1: Intraoperative Hypotensive Events Detection

Input: y_{pred}: a prediction output sequence of MAP values for d_{IOH} minutes in form of a list, t: the IOH minimum lasting duration, MAP_{max}: the maximum MAP value from the dataset, and MAP_{min}: the minimum MAP value from the dataset, l: the IOH risk threshold.

Output:h: a boolean representing the ocurrence of an IOH event.

1: $y = MAP_{min}$
2: **for** i in $range(0, d_{IOH} - t + 1)$ **do**
3:      $y_1 = max(y_{pred}[i : i + t])$
4:      $y = min(y, y_1)$
5: **end for**
6: $p = 1 - (y - MAP_{min})/(MAP_{max} - MAP_{min})$
7: $h = p > l$
8: **return** h

---

datasets. All experiments were conducted on a machine with Intel Xeon Gold CPU x2, 512 GB DDR4 memory, and AMD Radeon Instinct MI50 with 32GB GPU memory. Some testing tasks of Tongji dataset were run as a clinical trial, as shown in Figure 3, in Tongji Hospital with ethical approval.

Performance Benchmarks To better evaluate the performance of CMA on IOH events early warning, we chose the below algorithms as performance benchmarks in our experiments. Note that most performance benchmarks do not have enough technical details to reproduce. We thus adopted their best results for a fair comparison. We also indicated those missing results as N/A.

- HPI (Hatib et al. 2018; Davies et al. 2020): A commercial system with proprietary monitors and machine learning algorithms was developed and used HPI (*Hypotension Prediction Index*) to predict IOH events. We used both the most well-known 2018 results (HPI-2018) and the latest 2020 results (HPI-2020).
- BRNN (Jeong et al. 2019): It is a bidirectional recurrent neural network followed by dense layers. It uses vital

Method	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
CMA/Tongji	1.82	2.22	2.56	2.85	3.09	3.31	3.49	3.66	3.81	3.95	4.07	4.19	4.30	4.41	4.50
CMA-Dynamic/Tongji	1.90	2.28	2.61	2.89	3.13	3.34	3.53	3.69	3.85	3.98	4.11	4.23	4.34	4.44	4.54
GRU/Tongji	1.94	2.32	2.64	2.92	3.16	3.37	3.55	3.72	3.86	4.00	4.12	4.24	4.35	4.45	4.55
Transformer/Tongji	1.90	2.29	2.68	3.25	3.76	4.20	4.63	5.06	5.37	5.71	6.07	6.42	6.78	7.14	7.50
CMA/External	3.05	3.63	4.11	4.52	4.85	5.14	5.39	5.61	5.80	5.97	6.12	6.25	6.38	6.49	6.59
CMA-Dynamic/External	3.25	3.83	4.31	4.69	5.02	5.30	5.55	5.76	5.95	6.12	6.26	6.40	6.52	6.64	6.74
GRU/External	3.31	3.85	4.34	4.75	5.08	5.37	5.63	5.86	6.06	6.23	6.39	6.53	6.66	6.78	6.89
Transformer/External	3.26	3.82	4.41	4.99	5.58	6.16	6.75	7.34	7.92	8.51	9.09	9.67	10.25	10.83	11.42

Time (minutes), Error (mm Hg)

Figure 4: Mean absolute error of the predicted mean arterial pressure measured by mm Hg

Method	Dataset	Time	AUC	Accuracy	Precision	Recall	Specificity
CMA	The Tongji low frequency dataset	5min	0.963	94.10%	80.00%	82.80%	96.10%
		10min	0.93	91.30%	72.00%	73.60%	94.60%
		15min	**0.912**	90.50%	70.50%	68.40%	94.60%
	The external low frequency dataset (VitalDB)	5min	0.93	90.60%	65.50%	75.00%	93.30%
		10min	0.858	86.10%	50.90%	60.50%	90.30%
		15min	**0.834**	86.00%	50.00%	52.50%	91.40%
HPI-2018	The high frequency dataset (Arterial waveform)	5min	0.97	N/A	N/A	91.80%	92.20%
		10min	0.95	N/A	N/A	89.30%	89.50%
		15min	**0.95**	N/A	N/A	87.50%	87.30%
HPI-2020	The high frequency dataset (Arterial waveform)	5min	0.926	N/A	N/A	85.8%	85.8%
		10min	0.895	N/A	N/A	81.7%	81.7%
		15min	**0.879**	N/A	N/A	80.6%	80.6%
MC-CNN	The high frequency dataset (VitalDB)	5min	0.932	N/A	N/A	85.8%	85.8%
		10min	0.912	N/A	N/A	835.%	83.4%
		15min	**0.897**	N/A	N/A	81.4%	81.4%
RFM-IBP	The high frequency dataset (Arterial waveform)	5min	N/A	98.90%	65.20%	44.10%	N/A

Table 1: IOH event classification performance

signs and the anesthesia-related data from the induction phase of anesthesia to predict IOH events.

- MAP-NE (Tang et al. 2019): A least-squares and reduced-order model uses combined MAP and norepinephrine (NE) to predict the mean arterial blood pressure of the sepsis patients who received NE.
- RFM-IBP (Lee et al. 2022): A random forest model predicts hypotension 5 minutes ahead of its occurrence throughout the entire period of anesthesia.
- MC-CNN (Lee et al. 2021): A 1-D convolutional neural network predicts MAP and detects IOH events.
- GRU (Xiuyun et al. 2018): A gated recurrent unit network is commonly used machine learning model for time series data processing.
- Transformer (Yang et al. 2022): A popular deep learning architecture is good at processing text sequence data.

Evaluation Metrics We used a commonly-used metric *Mean Absolute Error* (MAE) for prediction performance evaluation, which is defined below:

$$MAE = \frac{1}{T} \sum_{t=1}^{n} |y_t - \hat{y}_t| \tag{9}$$

MAE measures the deviation between the estimates and ground truth of MAP. The lower the value, the more minor deviation between estimates and ground truth generated by the model. The IOH event warning as a classification task was evaluated by accuracy, precision, recall, specificity and *area under ROC* (AUC) in our experiments.

MAP Prediction Performance Comparison (Q1)

The results of MAP short-term prediction tests are shown in Table 2. We can observe that CMA consistently has the best performance in all tests demonstrating its effectiveness in short-term MAP prediction. The transformer algorithm could barely keep up with CMA when the time horizon was small. When the time horizon increased to 10, the performance of the transformer algorithm fell far behind CMA. Figure 4 visualizes the MAP prediction of CMA and the top-

3 performance benchmarks on Tongji and external datasets by MAE, where the brighter yellow means better performance on the test. It is not surprising that the prediction performance of CMA outperformed all benchmarks on every time horizon.

IOH Warning Comparison (Q2)

IOH Early Warning Performance   The test results of issuing IOH early warning are shown in Table 1. It is interesting to observe that the performance of CMA ranked either 1 or 2 in every aspect of all tests. Note that our solution is the only one that uses low sampling rate data, while the other benchmarks rely on high sampling rates to achieve the best performance. Besides, CMA had better accuracy and AUC on the Tongji dataset than the external one. We believe this result is likely caused by CMA can better capture the procedural and implicit nature of IOH events early warning from more patient records.

To evaluate how CMA adopted the change of IOH definitions, we used two IOH definitions to investigate its performance variation. The IOH definition was set to the MAP values staying below 65 mm Hg lasting for 1 minute or below 65 mm Hg lasting for 2 minutes. The test results are shown in Figure 5 (a). The results show that CMA can achieve an AUC of 0.912 and 0.941 for issuing IOH event early warning when using different IOH definitions on the Tongji dataset. The results of CMA on the VitalDB dataset were not as good as the Tongji dataset but still kept acceptable on IOH early warning. Overall, CMA demonstrated a competent adaptation to different IOH definitions while maintaining reliable performance. Please note that these results were achieved without involving any model reconstruction or retraining. This feature allows CMA has a low dependency on computing resources during the run time.

Modality Effects (Q3)

We performed ablation experiments to analyze the effects of multiple modalities in CMA. We constructed three tasks to predict MAP with all seven indicators (CMA), four indicators (MAP, age, sex, and BMI, named CMA-4f), and four dynamic indicators (Dynamic). These algorithms were tested on Tongji and VitalDB datasets to examine how the performance could be affected. As shown in Figure 5(b), the MAE of CMA-4f was 5.8 mm Hg, slightly better than the performance of the transformer algorithm listed in Table 2.

Impact of Sampling Rate (Q4)

We studied the impact of using different sampling rates on our framework. The tested sampling rates were set to 30 seconds, 1 minute, and 2 minutes. The test results are shown in Figure 5 (b), (c), and (d). CMA at 30s interval achieved the best prediction performance. When the data frequency is reduced from 30 seconds to 1 minute, the MAE of CMA and GRU does not increase significantly. At this time, the MAE of CMA using only dynamic indicators (Dynamic) has risen by 6%. When the frequency is further reduced to 2 minutes, the MAE values of all methods increase significantly. Even the MAE of CMA using all indicators (CMA) increased by

Time	3min	10min	15min
Our Tongji validation dataset			
CMA	2.56	3.95	**4.50**
CMA-Dynamic Only	2.61	3.982	4.54
GRU	2.64	4.00	4.55
Transformer	2.68	5.715	7.50
Our external validation dataset (VitalDB)			
CMA	4.11	5.97	6.59
CMA-Dynamic Only	4.31	6.12	6.74
GRU	4.34	6.23	6.89
Transformer	4.41	8.51	11.42
Other validation datasets			
BRNN	8.7	N/A	N/A
MC-CNN	N/A	7.5	7.88
MAP-NE	3	5.5	N/A

Table 2: Mean absolute error of MAP short-term prediction methods on different time horizons
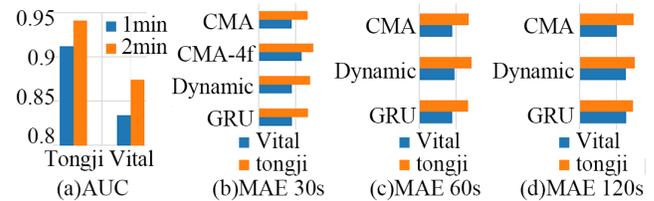
Figure 5: Simple tests for user-defined IOH definition, modalities, and different sampling rates

12.9% compared to the 30-second frequency. CMA at 2 minutes interval still shows acceptable performance on arterial pressure prediction, but the long gap between two consecutive readings severely limits its performance and has no practical value. CMA-30 and CMA-1 were both acceptable solutions, but CMA-30 has the most desirable performance.

User-definable IOH Event Prediction

Conclusion and Outlook

We proposed a composite multi-attention framework to issue an early warning for intraoperative hypotension events. The proposal is the first attempt to use only four vital signals with a low sampling rate and three common demographic characteristics as model input. We use this framework to predict the short-term mean arterial pressure instead of the intraoperative hypotension event. This strategy enables us not directly to involve the intraoperative hypotension definition in the model training, thus saving computational cost when adopting a new definition in the model for different patients. The experimental results showed that CMA outperformed state-of-the-art models on short-term arterial pressure prediction and intraoperative hypotension event early warning.

In the future, we aim to complete our clinical trial in the Tongji Hospital. According to the trial results, we will further adjust the algorithm to improve its performance and make it applicable to more medical devices. We will also plan to share the deidentified and approved Tongji dataset with the public.

Acknowledgments

References

Asai, N.; Doi, C.; Iwai, K.; Ideno, S.; Seki, H.; Kato, J.; Yamada, T.; Morisaki, H.; and Shigeno, H. 2019. Proposal of anesthetic dose prediction model to avoid post-induction hypotension using electronic anesthesia records. In *Proceedings of 2019 Twelfth International Conference on Mobile Computing and Ubiquitous Network (ICMU)*, 1–4. IEEE.

Davies, S. J.; Vistisen, S. T.; Jian, Z.; Hatib, F.; and Scheeren, T. W. 2020. Ability of an arterial waveform analysis–derived hypotension prediction index to predict future hypotensive events in surgical patients. *Anesthesia & Analgesia*, 130(2): 352–359.

Friedman, L.; and Komogortsev, O. V. 2019. Assessment of the effectiveness of seven biometric feature normalization techniques. *IEEE Transactions on Information Forensics and Security*, 14(10): 2528–2536.

Hatib, F.; Jian, Z.; Buddi, S.; Lee, C.; Settels, J.; Sibert, K.; Rinehart, J.; and Cannesson, M. 2018. Machine-learning algorithm to predict hypotension based on high-fidelity arterial pressure waveform analysis. *Anesthesiology*, 129(4): 663–674.

Inada, R.; Doi, C.; Yamasaki, Y.; Shigeno, H.; and Seki, H. 2021. Intraoperative Hypotension Prediction System by Considering Personal Lifestyle and Medical History. In *Proceedings of 2021 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)*, 489–494. IEEE.

Jeong, Y.-S.; Kang, A. R.; Jung, W.; Lee, S. J.; Lee, S.; Lee, M.; Chung, Y. H.; Koo, B. S.; and Kim, S. H. 2019. Prediction of blood pressure after induction of anesthesia using deep learning: A feasibility study. *Applied Sciences*, 9(23): 5135.

Kendale, S.; Kulkarni, P.; Rosenberg, A. D.; and Wang, J. 2018. Supervised machine-learning predictive analytics for prediction of postinduction hypotension. *Anesthesiology*, 129(4): 675–688.

Lee, H.-C.; and Jung, C.-W. 2018. Vital Recorder—a free research tool for automatic recording of high-resolution time-synchronised physiological data from multiple anaesthesia devices. *Scientific Reports*, 8(1): 1–8.

Lee, S.; Lee, H.-C.; Chu, Y. S.; Song, S. W.; Ahn, G. J.; Lee, H.; Yang, S.; and Koh, S. B. 2021. Deep learning models for the prediction of intraoperative hypotension. *British Journal of Anaesthesia*, 126(4): 808–817.

Lee, S.; Lee, M.; Kim, S.-H.; and Woo, J. 2022. Intraoperative Hypotension Prediction Model Based on Systematic Feature Engineering and Machine Learning. *Sensors*, 22(9): 3108.

Salmasi, V.; Maheshwari, K.; Yang, D.; Mascha, E. J.; Singh, A.; Sessler, D. I.; and Kurz, A. 2017. Relationship between intraoperative hypotension, defined by either reduction from baseline or absolute thresholds, and acute kidney and myocardial injury after noncardiac surgery: a retrospective cohort analysis. *Anesthesiology*, 126(1): 47–65.

Shimasaki, C. 2020. Chapter 32 - Biotechnology Products and Their Customers: Developing a Successful Market Strategy. In Shimasaki, C., ed., *Biotechnology Entrepreneurship (Second Edition)*, 481–495. Academic Press. ISBN 978-0-12-815585-1.

Song, Q.; Li, J.; Li, C.; Guo, H.; and Huang, R. 2022. Fully attentional network for semantic segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 2280–2288.

Tang, Y.; Brown, S.; Sorensen, J.; and Harley, J. B. 2019. Reduced Rank Least Squares for Real-Time Short Term Estimation of Mean Arterial Blood Pressure in Septic Patients Receiving Norepinephrine. *IEEE Journal of Translational Engineering in Health and Medicine*, 7: 1–9.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. *Advances in Neural Information Processing Systems*, 30.

Wang, Z.; Liu, X.; Lin, J.; Yang, C.; and Li, H. 2020. Multi-attention based cross-domain beauty product image retrieval. *Science China Information Sciences*, 63(2): 1–3.

Wesselink, E.; Kappen, T.; Torn, H.; Slooter, A.; and Van Klei, W. 2018. Intraoperative hypotension and the risk of postoperative adverse outcomes: a systematic review. *British Journal of Anaesthesia*, 121(4): 706–721.

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.; and Veelo, D. P. 2020. 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*, 323(11): 1052–1060.

Xiuyun, G.; Ying, W.; Yang, G.; Chengzhi, S.; Wen, X.; and Yimiao, Y. 2018. Short-term load forecasting model of GRU network based on deep learning framework. In *Proceedings of 2018 2nd IEEE Conference on Energy Internet and Energy System Integration (EI2)*, 1–4. IEEE.

Yang, L.; Li, J.; Dong, R.; Zhang, Y.; and Smyth, B. 2022. NumHTML: Numeric-Oriented Hierarchical Transformer Model for Multi-task Financial Forecasting. *arXiv preprint arXiv:2201.01770*.

Zhang, L.; Xiang, W.; Zhao, C.; Zhao, H.; Li, R.; and Wu, R. 2022. Co-promotion Predictions of Financing Market and Sales Market: A Cooperative-Competitive Attention Approach. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 9040–9047.