Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Predicting intraoperative hypotension using scalogram images and dual deep learning models

Geon Lee
Soonchunhyang University
misoon lee
Soonchunhyang University Hospital
sanghyun kim
Soonchunhyang University Hospital
Jiyoung Woo
jywoo@sch.ac.kr

Soonchunhyang University

Article

Keywords: Hypotension prediction, Invasive blood pressure, Vision transformer, Deep learning, Signal processing, Intraoperative hypotension

Posted Date: March 20th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8666700/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Predicting intraoperative hypotension using scalogram images and dual deep learning models

Geon Lee[1], Misoon Lee[2], Sanghyun Kim[2], Jiyoung Woo[1,*]

1 Dept. ICT Convergence, Soonchunhyang University, Asan, South Korea
2 Dept. Anesthesiology and Pain Medicine, Soonchunhyang University Bucheon Hospital, Bucheon, South Korea

¤ 22-15, Soonchunhyang-ro, Sinchang-myeon, Asan-si, Chungcheongnam-do, Republic of Korea

* jywoo@sch.ac.kr

Abstract

Intraoperative hypotension (IOH) is associated with an increased risk of postoperative complications, including myocardial injury, renal damage, and mortality. While AI-based models such as the Hypotension Prediction Index (HPI) have been commercialized, concerns remain regarding overestimated performance due to selection bias and limited interpretability. This study aims to develop a robust, multi-modal deep learning model that improves early prediction of IOH using Invasive Blood Pressure (IBP) waveforms, with an emphasis on class-specific performance and fair evaluation protocols. We collected a large dataset of intraoperative arterial waveform recordings and annotated hypotension events. We designed a transformer-based architecture—integrating Vision Transformer (ViT) for scalogram images and TabNet for feature-engineered ABP data—to autonomously learn dynamic arterial patterns while maintaining interpretability. The proposed multi-modal model achieved competitive performance, particularly outperforming baselines in minority-class recall, which is crucial for imbalanced medical tasks. At a 5-minute prediction horizon, the model demonstrated an AUROC of 0.94 and an AUPRC of 0.78. To address the issue of selection bias, we redefined classes based on clinical occurrence; while prior studies often required the entire observation window to remain strictly below or above certain thresholds, we defined a hypotensive event as any window containing at least one mean arterial pressure (MAP) measurement of $\leq$ 65 mmHg. Furthermore, to ensure rigorous performance evaluation and prevent data leakage, we implemented patient-level cross-validation. This approach advances the field by ensuring fairness in model evaluation and providing actionable insights in the intraoperative environment

Keywords: Hypotension prediction, Invasive blood pressure, Vision transformer, Deep learning, Signal processing, Intraoperative hypotension

February 19, 2026

1 Introduction

Arterial hypotension that occurs during anesthesia significantly increases the risk of postoperative complications such as myocardial infarction and acute kidney injury [1]. Hence, careful monitoring of the patient's hemodynamic changes is essential during anesthesia. Immediate interventions are required upon detecting hypotension to maintain hemodynamic stability. There have also been attempts to observe the decision-making process of anesthesiologists specializing in hemodynamic management of patients after heart surgery and to create a decision-making system based on this [2]. If hemodynamic fluctuations can be predicted in advance, it would facilitate safer anesthetic management. To monitor hypotension, invasive blood pressure (IBP) is routinely monitored in the operating room. IBP measurements, obtained by catheter insertion, are especially important for the management of high-risk surgical patients.

There are high expectations for the application of artificial intelligence models to respond to hypotension, and many attempts have been made. Currently, Edwards Lifesciences is the only commercially available model to predict intraoperative hypotension. Their model extracts 3,022 individual features from arterial blood pressure and generates combinatorial features to predict hypotension 5-15 min ahead using machine learning. They introduced a hypotension prediction index (HPI) to indicate the risk of the occurrence of hypotension [3]. Subsequent studies have validated the effectiveness of HPI in predicting hypotension, demonstrating its reliability and precision [4], [5]. However, recent studies on hypotension prediction have raised concerns that selection bias could inflate the reported performance of HPI [6], [7].

According to their observations, HPI defines a hypotensive event as a period during which the mean arterial pressure (MAP) remained below 65 mmHg for at least 1 minute. Positive samples were selected as data from 5, 10, and 15 minutes before the onset of IOH. In contrast, non-hypotensive events were defined as intervals in which hypotension did not occur within the preceding 20 minutes and the MAP remained above 75 mmHg for 30 minutes, with data points at the midpoint of this interval designated as negative samples. According to this criterion, the MAP at the time of negative samples is always above 75 mmHg, and all cases where the current MAP is below 75 mmHg within the data are treated as hint of hypotension occurring, while only cases where the current MAP is above 75 mmHg are treated as normal. As a result, the model can easily predict future hypotension based solely on the current MAP without the need for complex feature extraction. Criticism has been raised that "this selection bias" has led to an overestimation of performance, sparking debate over fair performance evaluation methods. Simulation studies have shown that significant performance gains can be achieved by adding bias to the data [8, 9]. In addition, some studies have argued that AI models for predicting IOH do not outperform traditional regression-based methods [10, 11].

On the other hand, the researchers who supported Edwards Lifesciences conducted further retrospective and prospective analyses in 2024, once again demonstrating a high area under the receiver operating characteristic curve score for HPI in both

approaches [12]. Futhermore, they emphasizes that early intervention through predictive alerts can reduce the duration of exposure to hypotension. Still, many researchers argue that even if predictions can be made using simple indicators, AI can provide earlier signals by comprehensively considering multiple variables and that it offers practical advantages such as enabling the establishment of automated alert systems in clinical setting. Although there is an issue such as selection bias in some AI research, they still need to prove their clinical value beyond existing monitoring methods [13]. Rather than saying that all AI research is unnecessary for IOH prediction, it is necessary to solve the problems of selection bias and performance overestimation.

In this study, we will design problems and construct training samples to avoid selection bias and performance overestimation. Also, we will propose a multi-modal model designed to capture a sufficient range of features, aiming to improve the early prediction accuracy of hypotension. Specifically, our model converts blood pressure into scalogram images using the wavelet transform and feeds this image into a Vision Transformer model. In parallel, we process the anesthesia record through TabNet, a deep learning architecture that uses features from tabular data. By combining two models, we construct a framework for predicting hypotension. To enhance the prediction performance of our model, we utilize different mother wavelets to generate different scalogram images. This study aims to predict hypotension 1–5 minutes before it occurs, helping anesthesiologists monitor patients during surgery. To the best of our knowledge, this is the first study to implement a multi-modal deep learning framework that integrates a rectangular-patch Vision Transformer with TabNet for intraoperative hypotension prediction. Unlike conventional models that rely on square partitioning or static feature extraction, our approach preserves the temporal continuity of arterial waveforms while providing feature-level interpretability. By concurrently addressing selection bias through a refined evaluation protocol, this research establishes a more robust and fair standard for AI-based hemodynamic monitoring

2 Related Works

The earliest research primarily focused on identifying factors influencing hypotension prediction. Reich et al. aimed to identify the relationship between hypotension and preoperative variables, such as sex, age, height, preoperative medications, and anesthetic induction methods using Fisher's exact test [14]. Their findings emphasized that the incidence of post-induction hypotension is closely related to specific key factors, including age 50 years or older, the presence of hypotension prior to induction, and the use of propofol during induction. Ko et al. identified lactate level, blood in the nasogastric tube, and systolic blood pressure as risk factors for hypotension in normotensive patients with non-variceal upper gastrointestinal bleeding [15]. Lee et al. assessed the shock index, which is defined as heart rate divided by systolic blood pressure, the modified shock index, and the age shock index as predictors of hypotension [16].

Furthermore, Ghosh et al.'s work utilized symbolic aggregate approximation (SAX), a method that transforms time-series data into symbolic sequences by dividing the data into segments, computing average values for each segment, and mapping them to symbols,for predicting hypotension [17]. Park et al. used multivariate negative binomial analysis to predict hypotension during dialysis by generating heart rate variability characteristics and incorporating them with patient clinical characteristics [18].

The key works are listed in Table 1. Machine learning-based hypotension prediction has been widely utilized in research because of its ability to generate variables from blood pressure data to make predictions, enabling an understanding of the contributions of variables to hypotension prediction. SVM does not have the ability to calculate

variable importance either, but it has been used for model comparison because of its strong performance.

Deep learning-based hypotension prediction methods require less preprocessing compared to machine learning-based models, making them advantageous for easier model creation and comparable or superior performance. However, these methods may lack the interpretability to explain the factors influencing hypotension prediction. Several studies are underway to address these limitations, such as eXplainable AI (XAI) [23]. However, studies using XAI have shown lower performance compared to those of existing studies.

Meanwhile, VitalDB [20] provides the open dataset for researchers. VitalDB collects and provides vital sign from patients undergoing surgery. MIMIC dataset collects the dataset from ICU patients. Moon et al. [24] used the VitalDB and developed CNN-based deep learning models. Their model has demonstrated high performance, but they also apply undersampling to the test data to mitigate data imbalance, which may lead to a lack of reality. Jo et al. [28] developed a multi-modal model to predict intraoperative hypotension using arterial blood pressure, electroencephalogram and electrocardiogram. Using 14,140 surgery cases from VitalDB, they aimed to forecast hypotensive events 3, 5, 10, and 15 minutes before occurrence. The best performance was achieved when combining IBP and EEG waveforms with a high AUROC of 0.935 at 5 minutes before event.

In real-world situations, severe class imbalance occurs due to the rarity of hypotension, resulting in lower performance when applied in practice compared to the reported results. Choe et al. [21] proposed a model combining CNN and RNN. When building a prediction model for hypotension N minutes in advance, significant issues may arise in the process of sampling time-series data. If existing studies have overlooked these sampling issues and proceeded directly to model training, the performance of the resulting models could be overestimated. If the goal is to predict N minutes in advance, the prediction should be made using only the data exactly N minutes before the first occurrence of hypotension. However, existing studies may have generated samples by including hypotension-related data that occur continuously after the N-minute prior time point. Since vital data changes most significantly just before the onset of hypotension, these samples can excessively inflate the model's performance. If continuous hypotension samples are used without deletion, the model may overfit to specific patterns, reducing its generalization ability.

In the operating room scenario, prolonged hypotension lasting 30 minutes is rare, as immediate interventions are typically performed during surgery. Previous studies that assume such extended hypotensive periods may not be directly applicable in the operating room scenario.

In addition, many studies in the conventional medical field rely on the area under receiver operating characteristics (AUROC) as a performance metric. However, due to the highly imbalanced nature of intraoperative hypotension data, AUROC may not accurately predict the class of interest. In an imbalanced dataset where the positive class is very scarce, a model can achieve a high AUROC value even if it simply predicts all data as negative. This is because the overwhelmingly large number of negative instances makes it easy to obtain high specificity, and the ROC curve tends to be positioned close to the upper left corner. Consequently, a high AUROC value can be misleading, suggesting excellent model performance when, in reality, the model's predictive power for the positive class is very low. In cases of class imbalance, the evaluation of performance using metrics such as the F1 score and the area under the precision recall curve (AUPRC) proves to be more effective than AUROC [29]. AUPRC utilizes recall and precision. Because true positive rate (TPR) is the same as recall, the difference between AUROC and AUPRC lies in the difference between false positive

rate (FPR) and precision. FPR is calculated as the ratio between the number of negative events wrongly classified as positive and the total number of actual negative events. When dealing with datasets containing a minimal number of actual positive events, adjusting the prediction threshold may not significantly impact to FPR due to the larger number of actual negative events in the denominator. However, because precision is calculated as the ratio between the number of actual positive events classified as positive and the total number of events classified as positive by the model, it becomes relatively sensitive to changes in the threshold compared to the FPR. Therefore, AUPRC is more informative for the detection of rare diseases.

In particular, predicting hypotension n min before it occurs allows adequate lead time for appropriate action to be taken. To ensure this lead time, the first hypotension prediction must be made at least n min before the initial occurrence of hypotension. Subsequent hypotensive events should be excluded from the prediction because they do not provide the required n min of advance notification. Assuming hypotension predictions are made 1 min in advance at intervals of 1 min and t is the time at which the first hypotension is observed. The prediction must be made at least $t - n$ min in advance to be effective. When hypotension occurs again at $t + 1$ min, a prediction made at $t + 1 - n$ min is meaningless because it does not allow the required n min of lead time for the hypotension observed at t. Many previous studies have overlooked this issue, focusing solely on predicting consecutive hypotensive events. However, we must ensure that predictions are made based on the first time point in the sequence of hypotensive events to guarantee the targeted lead time. This approach ensures that medical personnel have ample time to intervene effectively. Furthermore, utilizing a multimodal deep learning model to convert features and raw signal data into images from a single IBP data to improve the predictive power has not been explored in existing studies.

To address these issues, we propose a model that considers the data imbalance issue and the prediction of the first hypotensive event.

3 Materials and Methods

3.1 Data Preprocessing

IBP data was recorded at a frequency of 100 Hz. The samples were generated from a 140 s window with a 10 s overlap. The first 20 s of the window were observed to extract features indicative of the hypotension event, while the last 60 s were observed to determine whether hypotension occurs. A 60 s lead time window is implemented between the feature extraction and class observation windows to enable a 1 min advance prediction. This process is repeated by sliding the windows with the overlapping interval, as shown in Fig. 1.

To solve the selection bias and the overestimation of performance, we set the following rules in generating samples. First, we define the class and sample carefully to avoid the selection bias. In the original HPI study, the hypotension sample was defined as data before the onset of hypotension, whereas normal data was defined as a 30-minute period of MAP > 75 with no hypotension within 20 minutes before or after. This is a very strict definition of normal data and improves the shortcomings of class prediction based on MAP alone, which is based on the current MAP in addition to future events. In this study, we defined classes by looking only at future data, not at the current map. We eliminated selection bias by using the same criteria to define hypotensive and normal samples based on MAP at a future point in time. A MAP of at most 65 mmHg defines a hypotension event, whereas a MAP of at least 75 mmHg defines a normal event [1]. The gray zone represents the interval between hypotension and normal events. Including this gray zone, data that are considered outliers, such as those with minimal

or excessive variation in values, are excluded from the analysis, along with instances in which hypotension first occurred in the training and delay windows.

Secondly, to reduce the model performance overestimation, as mentioned in the Introduction, we focus on the first hypotension. To ensure this lead time, the first hypotension prediction must be made at least n min before the initial occurrence of hypotension. Subsequent hypotensive events should be excluded from the prediction because they do not provide the required n min of advance notification. As shown in Fig. 1, consecutive hypotension samples are deleted (cross-out line).

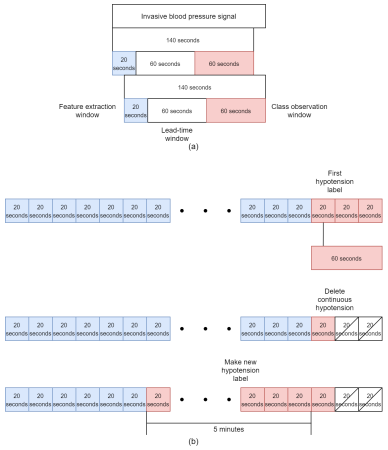


Fig 1. Method of data windowing (a) and re-labeling (b)

Third, to overcome the performance decrease due to the data imbalance issue in predicting initial hypotension, we apply augmentation and preprocessing techniques on initial hypotension events, similar to the process shown in Fig. 1-b. When the first hypertension event was detected, the lead time was decreased to 1-5 min by changing the labels from normal to hypotension events for the 5 min preceding that point. Sequential hypotension events occurring after this point are excluded from the analysis. In summary, we refined the IBP data to predict hypotension events that occur 1 to 5 min before the first hypotension event. Following data preprocessing and outlier removal, some outlier data, exhibit distinct behavior due to external pressure or errors in electronic devices. By applying a Fourier transform to input data and decomposing them into frequency components, we can detect highly captured frequencies. However, abnormal blood pressure data result in anomalies during Fourier transformation, such as showing high values at low frequencies or displaying maximum values when it has only zero frequency. To remove these anomalies, we apply the Fourier transform to IBP data and exclude cases where the smallest frequency of the peak value is less than 2 Hz and the index of the peak value is zero as outliers.

Following this processing, IBP data are finally partitioned into 441,386 normal and 48,110 hypotension windows, totaling 489,496 windows for 912 patients, which are then used as input to the model in two steps.

3.2 Wavelet Transform

The wavelet transform allows for frequency domain analysis over time. Although spectral analysis, a conventional method of time-series analysis, is based on the Fourier transform, it lacks time domain information. Time-frequency analysis has been proposed to address this limitation [30]; however, interpreting components outside a fixed frequency range can be challenging. The wavelet transform has emerged as a solution to the limitations of these two previous time-series analysis methods [31].

The wavelet transform utilizes specific basis functions, known as wavelet functions, to convert time-series data into frequency-domain data. These basis functions are crucial in extracting various features of the signal [32]. The formula for the wavelet transform involves scaling and shifting a wavelet function to analyze a signal.

$$W(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t)\, \psi^* \left(\frac{t-b}{a} \right) dt \tag{1}$$

where a represents scaling and b represents shifting. ψ^* denotes the wavelet function, which in the wavelet transform, also known as the mother wavelet, and $*$ signifies a complex number operation. Based on the equation, the wavelet transform shifts the mother wavelet along the time axis and calculates the correlation coefficient with the original signal $x(t)$. The mother wavelet is subsequently scaled and the process is repeated.

Commonly used mother wavelets include Morlet, Mexican Hat, Paul, and Bump. The Morlet wavelet, widely used in studies, excels in localizing time and frequency [33] and is calculated using Eq. (2). The Morlet wavelet captures the oscillatory behavior of signals. The Mexican hat wavelet detects non-oscillatory and extremely fluctuating points. as expressed in Eq. (3) [34]. [35]. The Paul and Bump wavelets are expressed using Eq. (4) [36] and Eq. (5), respectively [35]. The Paul wavelet detects an adjustable order for varying degrees of oscillation. The Bump wavelet detects smooth, compact support with non-oscillatory characteristics. The Bump wavelet, with its banded shape, is better suited for localizing frequencies than the Morlet wavelet.

$$\psi_0(\eta) = \pi^{-\frac{1}{4}} e^{im\eta} e^{-\frac{\eta^2}{2}} \tag{2}$$

$$\psi(\eta) = \frac{(-1)^{m+1}}{\sqrt{\Gamma\left(\frac{m+1}{2}\right)}} \frac{d^m}{d\eta^m} e^{\frac{\eta^2}{2}} \tag{3}$$

$$\psi(\eta) = \frac{2^m i^m m!}{\sqrt{\pi\,(2m)!}} (1 - i\eta)^{-(m+1)} \tag{4}$$

$$\psi(\eta) = e^{1 - \frac{1}{1 - \left(\frac{\eta-\mu}{\sigma^2}\right)^2}}, \qquad X[\mu - \sigma, \mu + \sigma] \tag{5}$$

The Morlet wavelet utilizes a non-dimensional parameter η and a wavenumber m which is typically set to 6. In the Mexican Hat wavelet, Γ represents the gamma function, with m set to 2. m is typically set to 4 in the Paul wavelet. In the Bump wavelet, μ and σ represent the mean and standard deviation, respectively. In this experiment, we used a mean and standard deviation of 5 and 0.6, respectively. The dimension of the wavelet-transformed scalogram images is 100 x 2000, which is too large for model training, so we reduced the image size to 50 x 200.

3.3 Feature engineering

To enhance the model performance by deriving various features and implementing the multimodal model, the multimodal model would allow us to leverage the advantages of deep learning while incorporating the advantages of existing machine learning-based studies.

Initially, we generated statistical features such as max, min, interquartile range(IQR), skewness (a measure of symmetry), kurtosis (a measure of whether the data is heavy-tailed or light-tailed relative to a normal distribution), and root mean square(RMS) which is the average of the signal strength. Furthermore, we incorporated peak features. Because our model aims to predict hypotension, we also identified peaks

corresponding to the lower part of the IBP as valleys. The features of valleys were derived similarly to peaks, except for the max and min features. In our study, we adopted the method of previous research in establishing the threshold for valley calculation [37]. However, setting the absolute threshold for valleys at a MAP of at most 65 mmHg was appropriate due to that the IBP is aggregated into 20-s windows and this results in variations in mean and variance across windows. To address this, we first normalized the 20s window and utilized valley height and distance between valleys to set the threshold for analysis. In addition to features, such as count and mean, we also generated features related to the peak interval and the crest factor. The crest factor represents the ratio of the peak value to other values and serves as an indicator of the peak anomaly. The change analysis is to segment data into distinct parts and compare them to identify change points in the mean or variance. The following features can be derived from the number of change points, their intervals and the amount of change before and after the change point.

For the frequency domain analysis, the Fourier transform that extracts features from a signal is used. When applied to IBP data, a spectrum image is generated with frequency and power on the x- and y-axes, respectively. We selected the five highest values of power and their corresponding frequencies as features of the Fourier transform.

Furthermore, we extracted variables that have clinical significance, such as empirical MAP, PPmax, PPmin, and PPV [3]. MAP empirical is a modified version of the formula for MAP calculated as follows.

$$MAP_{empirical} = \frac{SYS + (2 * DIA)}{3} \tag{6}$$

In this equation, MAP empirical was calculated using the peak value as systolic blood pressure and the valley value as diastolic blood pressure. During feature engineering, there are cases where the number of peaks and valleys differed a lot. While most of these cases were excluded during preprocessing, some cases were not fully processed owing to the relatively small difference in the number of peaks and valleys, resulting in unavailable MAP empirical values. Therefore, we took the additional measure of setting the MAP empirical value to NA to eliminate cases where the number of peaks and valleys varied during feature engineering. Pulse pressure variation (PPV) refers to a change in pulse pressure, which is determined using Eq. (7) based on the max and min values of the change in pulse pressure (PPmax and PPmin, respectively).

$$PPV\,(\%) = \frac{PP_{max} - PP_{min}}{\frac{PP_{max}+PP_{min}}{2}} * 100 \tag{7}$$

Finally, we incorporated the approximate entropy [38] and sample entropy [39] features to quantify the unpredictability of time-series fluctuations. Approximate entropy converts n raw time-series data into a vector of size m and establishes a filtering level r. By calculating the distance of each m dimensional vector from the others m, counting the number of vectors with a distance of at most r, and dividing by the total number of vectors, we obtain the ratio. A logarithm can be applied to this ratio and subsequently averaged using Eq. (8), where C represents the number of vectors with a distance of r or less.

$$\Phi^m(r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \ln C_i^m(r) \tag{8}$$

The calculation of approximate entropy involves determining the difference between m and $m+1$ from Eq. (8) as shown in Eq. (9).

$$\text{ApEn} = \Phi^m(r) - \Phi^{m+1}(r) \tag{9}$$

Sample entropy was developed as a solution to the limitations of approximate entropy. It offers the advantage of being less sensitive to data length and is computed by excluding self-similarity. Initially, we calculated sample entropy for $\Phi_i'^m(r)$ without applying the logarithm, as follows.

$$\Phi'^m(r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} C_i^m(r) \tag{10}$$

Subsequently, when calculating sample entropy, the logarithm was applied as shown in Eq. (11).

$$\mathrm{SaEn} = \log \Phi'^m(r) - \log \Phi'^{m+1}(r) = \log \frac{\Phi'^m(r)}{\Phi'^{m+1}(r)} \tag{11}$$

Finally, 41 features were extracted from IBP data to generate tabular data, as shown in Table 2. By applying four types of wavelet transforms to the IBP data, scalogram images are generated. In addition, numerous features are generated through feature engineering to produce tabular data. The final model, shown in Figure 2, is trained as a multimodal model using these diverse data types.

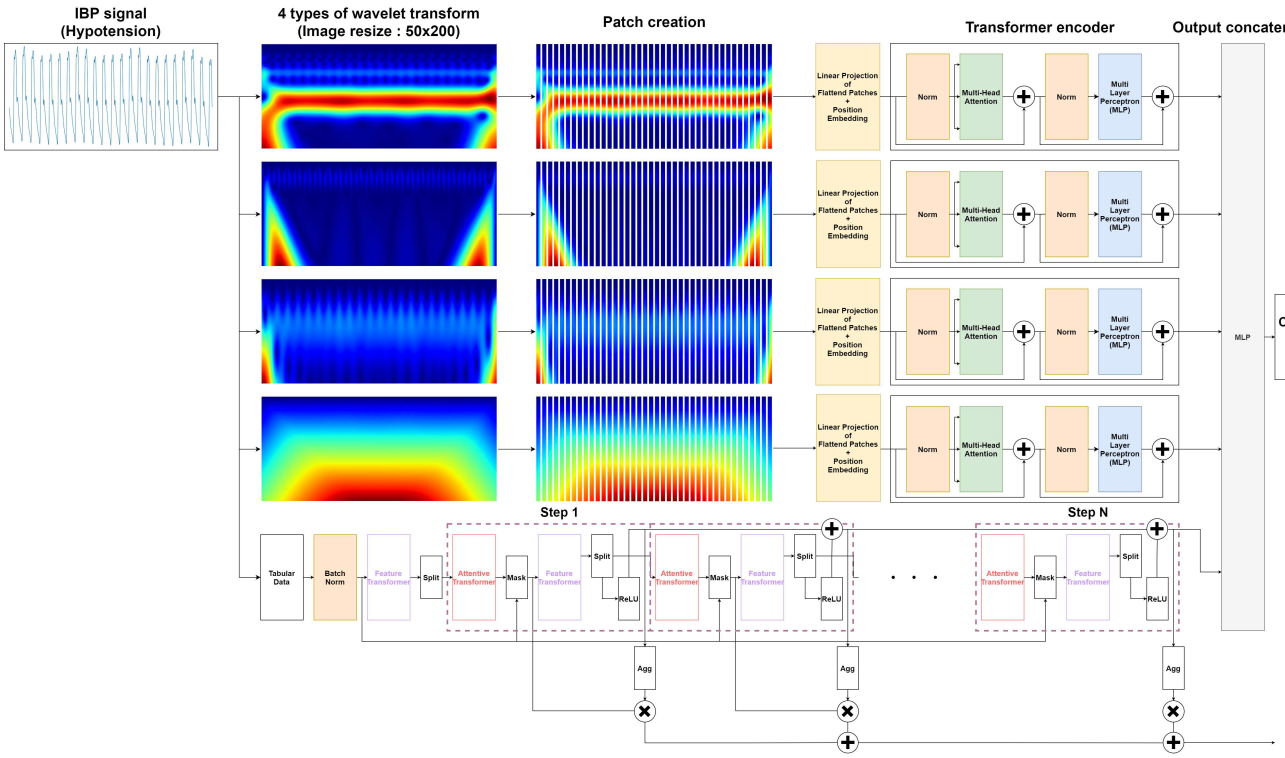


Fig 2. Multimodal framework with vision transformer and TabNet

To integrate images and tabular data within a single model, we use the vision transformer [40] and TabNet [41], known for their effectiveness in their respective domains. The scalogram images are segmented into rectangular patches and used as input to the transformer encoder to generate the result. Simultaneously, the tabular data of features is processed in a three-step process to determine feature importance and generate results.

Table 1. Summary of related works

Year	Author	Method	Algorithm	Data	Result
Machine learning-based hypotension prediction					
2018	Hatib et al. [3]	Extract 3,022 individual features and 2,603,125 combinatorial features from IBP	Logistic regression	MIMIC-II database, Edwards lifesciences database, UCI database	External validation AUROC 0.95(before 5min) 0.92(before 10min) 0.91(before 15min)
2020	Moghadam et al. [19]	scalar features(SBP, DBP etc)	Linear regression, SVM, KNN	MIMIC-III database	Accuracy 0.95 Sensitivity 0.85 Specificity 0.96
22022	Lee et al. [20]	Features extracted from IBP	Random Forest	Soonchunhyang Bucheon Hospital	Precision 0.88 Recall 0.52 F1 score 0.66 AUROC 0.96
2021	Choe et al. [21]	Make 20 features from IBP	CNN and RNN-based ensemble model	VitalDB database	Precision 0.64 Recall 0.70 F1 score 0.67 AUROC 0.96 AUPRC 0.72
2021	Lee et al. [22]	Use IBP, ECG, PPG, CO2	CNN-based deep learning model	VitalDB database	AUROC 0.93 Sensitivity 0.85 Specificity 0.85
2023	Hwang et al. [23]	Use SHapley Additive exPlanation(SHAP) model	eXplainable AI(XAI)	Asan Medical Center Database, VitalDB database	AUROC 0.91
2024	Moon et al. [24]	IBP and multi-channel(PLE, ECG, CO2) signal and frequency domain features	CNN-based deep learning model	VitalDB database	AUROC 0.95 Specificity 0.86 Sensitivity 0.90
Hypotension prediction using both deep learning and machine learning					
2008	Lin et al. [25]	Use 14 features from SBP and DBP	Linear regression, Artificial neural networks(ANN), Simplified ANN	Private data	Accuracy 0.78 Sensitivity 0.76 Specificity 0.76
2020	Kang et al. [26]	Scalar vitals(SBP, HR etc)	Random forest, Naïve bayes, logistic regression, ANN	Soonchunhyang Bucheon Hospital	AUROC 0.84
2020	Lee et al. [27]	27 features from scalar vitals	CNN, DNN, Random forest, Xgboost	Soonchunhyang Bucheon Hospital	Accuracy 0.74 Precision 0.76 Recall 0.76
2022	Jo et al. [28]	IBP, ECG, EEG	CNN	VitalDB	AUROC 0.97

Table 2. Summary of features

Types	Features	Counts
Statistical features	Max, Min, Mean, Sd(Standard deviation), Skewness, Kurtosis, RMS, IQR, Number of data under the mean	9
Peak features	Number of peak, Mean, Max, Sd, Crest factor, Mean of peaks interval, Sd of peaks interval	7
Valley features	Number of valley, Mean, Min, Sd, Crest factor, Mean of valleys interval, Sd of valleys interval	7
Change point features	The existence of changes in mean and variance	2
Fourier features	Top 5 power, Frequency of top 5 power	10
Blood pressure features	MAP_empirical, PPmax, PPmin, PPV	4
Entropy features	Approximate entropy, Sample entropy	2

3.4 Proposed Rectangular-patch Vision Transformer on Image-converted IBP

Many studies have applied convolutional neural network (CNN) models [42] to images generated by wavelet or short-time Fourier transform methods in signal processing. Wavelet and short-time Fourier transform methods convert time-series signal data into time-frequency representations. This transformation allows for the extraction of features from both the time and frequency domains, which can then be visualized as images. Therefore, CNN-based models can effectively capture spatial patterns within these images. The vision transformer is a model that applies the transformer architecture, commonly used in natural language processing, to image classification [40]. By dividing images into patches, flattening them, and incorporating positional embedding, the vision transformer enriches the input data with positional information. The vision transformer, recognized as a state-of-the-art algorithm, has demonstrated excellent performance in conventional image classification and is now being extended to applications in the medical field.

Since the x-axis of our scalogram images represents the time domain, square partitioning can lead to a misinterpretation of temporal continuity. To address this issue, we modified the positional embedding by resizing the patches from square to rectangular shapes. This modification ensures that values on the same x-axis can be accommodated within a single patch, as shown in the patch creation section of Fig. 2. By resizing the patches to a rectangular shape, our model ensures that broader time-domain information is integrated within a single embedding, allowing the transformer encoder to better learn the sequential progression leading to hypotensive events. To the best of our knowledge, this is the first study to implement a rectangular-patch Vision Transformer for time-series data. Our approach preserves the temporal continuity of arterial waveforms.

3.5 TabNet on Features

In previous research that focused on tabular data, deep learning-based methods have often underperformed compared to machine learning-based methods. This difference can be attributed to the effectiveness of tree-based ensemble models, such as Xgboost and LightGBM, which offer optimal performance with respect to training time and ease of interpretation. TabNet [41], on the other hand, is a deep learning model designed to mirror the feature selection capabilities of these tree-based models in the network structure.

The TabNet encoder operates through sequential multistep processing involving N decision steps. Each step comprises a feature transformer, an attentive transformer, and a feature masking step. The feature transformer processes features and comprises four common blocks, each following the connection structure of a fully connected layer - batch normalization - gated linear unit (GLU) [43]. Notably, TabNet employs ghost batch normalization (GBN) during the batch normalization process. GBN divides a large batch into smaller virtual batches (ghost batches) and performs batch normalization for each virtual batch independently, thus mitigating the statistical instability that may occur when using large batches [44]. The GLU regulates the flow of information from the preceding layers. The process involves multiplying the elements of the input with a layer that has been transformed into a probability value by applying a sigmoid function.

In TabNet, the attentive transformer is responsible for determining feature selection in machine learning. It adjusts the strength of features during training by using masks to enable certain features while blocking others. These two transformers generate masks for each step and feature attribute, providing TabNet interpretability that conventional deep learning models may not afford.

3.6 Multimodal Model

In multimodal learning, various types of inputs must be processed to produce a single result, requiring a fusion process during learning. Previous studies [45] have categorized multimodal learning into early, late, and joint fusion. Early fusion involves combining two different data types into a single dataset before training the model, sometimes using preprocessing to merge data through various transformations. This approach allows for the early utilization of correlations between features to enhance performance, requiring only one training step [46]. Late fusion involves training two different data types on separate models and then combining the results. While this approach allows for individual modification of each model [47], it does not leverage feature-level correlations. A more effective fusion method that utilizes the advantages of these fusion methods is joint fusion. This technique combines features learned in the middle layer of neural networks with features from other data, using them as input for the final model [48]. Unlike late fusion, the entire model can be trained by backpropagating the loss function through the process. In this study, the fusion model was utilized where the features generated by the two models were combined into multilayer perceptron to learn the two models by backpropagation to predict intraoperative hypotension.

4 Experiment and Results

4.1 Data Collection

In this study, IBP from 7,321 patients in 15 operating rooms at Soonchunhyang University Bucheon Hospital was collected between December 30, 2019 and October 30, 2020. IBP was recorded using a vital recorder [49], with data collection approved by the hospital's review board (approval No. SCHDB(Soonchunhyang Dongban)_IRB_2021-11-015). We accessed this dataset at 2022-03-15. All data were fully anonymized before we accessed them. Informed consent was obtained from all participants or their legal guardians, and all procedures were conducted in compliance with relevant guidelines and regulations. Owing to the sporadic nature of blood pressure data collection during surgery, datasets without blood pressure data were excluded. Among the collected data, 912 patients had available invasive blood pressure data and initially interpolation was applied to the dataset. Current polynomial-based interpolation methods, such as CUBIC, have shown limitations when applied to data utilized in this study, as shown in Fig. 3 (b). Therefore, a new interpolation method was deemed necessary. Given the constantly repeated characteristics of the IBP data, we determined that using data preceding the missing values would suffice for interpolation. Consequently, for the missing values, interpolation was performed by identifying the first point where the missing value occurred and the point with a slope (difference from the previous value) most similar to it, within a window of the same size as the missing value. This approach yielded superior interpolation results compared with conventional interpolation methods, as shown in Figure 3 (c). The interpolation was applied to data containing 8,933 missing values.

4.2 Model Performance

Data partitioning was conducted using a patient-stratified 5-fold cross-validation. By grouping data points by individual patient identifiers before distribution into folds, we eliminated the risk of 'within-patient' information leakage, where temporally adjacent samples from the same individual could bias the validation performance. To assess whether combining the two models enhances performance, the results of the proposed model were compared with predictions made using the vision transformer alone and

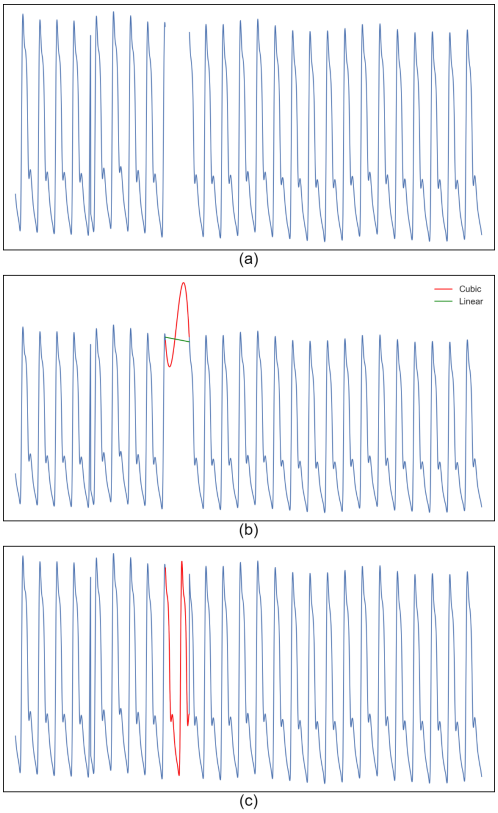


Fig 3. Sample of data imputation, (a) original data with NA. (b) data with cubic and linear imputation. (c) data with original imputation

TabNet alone. Additionally, the final model was compared with existing models that performed well in predicting intraoperative hypotension. In the vision transformer component, the scalogram image was divided into patches sized 50 x 2, resulting in 100 patches. The multi-head attention layer within the transformer utilized four heads, and the transformer block was constructed by sequentially passing through layers of sizes 128 and 64, repeated eight times.

The TabNet component utilized 16 decision layers, 16 attention layers, 3 steps, and 2 blocks each shared between decision steps and decision step dependent by default. The relaxation factor for computing the prior in the attentive transformer layer was set to 1.3, with epsilon set to 1e-15 to address issues arising from the inclusion of zeros. The batch size utilized for GBN was eight, the momentum for the exponential moving average was set to 0.98, the mask for the attentive transformer was set to softmax, and the lambda sparse, the mask sparsity loss coefficient, was set to 1e-3. Finally, we utilized a model with 348 layers: 244 layers of vision transformers, 94 layers of TabNet encoders, and eight layers to aggregate results and make predictions, with 2 input layers.

The training was repeated for 1000 epochs with a learning rate of 0.001, weight decay of 0.0001, and class weights in a normal:hypotension ratio of 8:2. During training, we set the training to stop early if the F1 score in the validation data set did not decrease for more than 70 iterations. Subsequently, we selected the weight that exhibited superior performance in the test dataset by comparing the best weight with the weight at the current iteration and utilized it as the weight for that fold.

Table 3. Model performances comparison of related studies and ablation studies

Method	Precision	Recall	F1 score	AUROC	AUPRC
Random Forest with feature engineering [37]	0.88	0.52	0.66	0.96	N/A
XAI [23]	N/A	N/A	N/A	0.91	N/A
STEP-OP [50]	0.65	0.70	0.67	0.96	0.72
HPI [3]	N/A	N/A	N/A	0.95	N/A
Vision Transformer(Only)	0.44	0.85	0.58	0.94	0.75
TabNet(Only)	0.48	0.84	0.52	0.94	0.76
Our model (combine)	**0.68**	0.78	**0.73**	**0.94**	**0.78**

A comparison between our proposed model and related studies, rounded to three decimal places, is listed in Table 3.While the proposed model showed slightly lower AUROC compared to some baselines, it achieved the highest F1-score (0.73) and AUPRC (0.78), indicating a more robust capability in detecting rare hypotensive events in imbalanced clinical data. Although precision decreased compared to the model with random forest after feature engineering, this 0.20 decrease was offset by a significant increase in recall (0.26). This improvement is noteworthy, as models with higher recall are highly valued in the medical field [51]. Furthermore, our proposed model outperformed previous studies in terms of the F1 score and AUPRC, which are metrics that better account for data imbalance, although AUROC was 0.02 lower than STEP-OP.

The findings of an experiment conducted to assess the efficacy of multimodal training are listed in Table 3. When comparing the performance of intraoperative hypotension between the vision transformer and TabNet models individually, and then comparing those results to the combined results of multimodal training using both models, a notable improvement in precision of over 0.20 was observed. However, the recall experienced a slight decrease of approximately 0.08. This shows that the multimodal model, which combines both models, achieves better performance for hypotension prediction than when each model is used independently.

In terms of AUROC, the difference between the three models was negligible,

attributed to the lower prevalence of hypotension compared to normal cases. In contrast, AUPRC demonstrated superior results with multimodal training.

4.3 Exploratory Analysis

By analyzing the mask in TabNet, we determined that all features were used to predict intraoperative hypotension. To validate this, we constructed a comparison table to discern any discrepancies between hypotension and normal data.

Table 4 presents a comparison of normal and hypotension data per data set, rather than per patient based on hypotension occurrence. Significant differences were evident in all features. In particular, statistical features indicate that hypotension data consistently exhibited lower values compared to normal data for all features, except for skewness and the number of data below the mean. The latter metric is particularly meaningful as it signifies a prevalence of values below the average. Skewness and kurtosis values, which characterize the distribution of data, reveal distinct differences. The hypotension data showed positive skewness and negative kurtosis, indicating the presence of unusually high values and a flatter peak compared to the normal distribution, respectively. In general, the hypotension data showed lower values and a wider range compared to normal data. The analysis of peak and valley features revealed that the mean, minimum and maximum values were lower for hypotension data compared with those for normal data. Additionally, the Fourier features indicated that the power of hypotension data was consistently lower overall. Notably, the occurrence of hypotension data lagged behind normal data initially but gradually shifted to occur earlier after the third time point. The first dominant frequency signified the most common oscillation period within the data, with the power of this frequency indicating its strength and prevalence. The hypotensive data exhibited a lower frequency in the first dominant frequency.

5 Conclusion and Discussion

In this study, we proposed a multimodal model that employs dozens of features and scalar images - rather than thousands of features extracted from IBP - to minimize complex preprocessing and predict intraoperative hypotension 1–5 minutes in advance without requiring additional patient information. In algorithm perspective, we proposed a novel rectangular-patch ViT for waverform data. To mitigate selection bias and performance overestimation in intraoperative hypotension (IOH) prediction, we propose a refined sample generation framework with three key strategies: (1) class redefinition for selection bias reduction: We redefined hypotension and normal classes based solely on future MAP values ($\leq$65 mmHg for hypotension, $\geq$75 mmHg for normal) and minimized ambiguous gray zone data and outliers to ensure unbiased class separation. (2) Event-Based analysis to solve data leakage problem and preformance inflation: We mitigated data leakage and performance inflation by focusing exclusively on the first hypotensive event per case. We focused only on the first hypotensive event per case and deleted consecutive hypotensive events, ensuring predictions are made at least n minutes before onset, while excluding all subsequent hypotensive phases. (3) Imbalance Mitigation: We applied data augmentation by relabeling the 1–5 minute periods preceding the first hypotensive event to enhance minority class representation, while discarding later hypotensive segments to prevent data skew.

Our work has the following limitations. We highlighted the limitations of evaluating models using AUROC in existing medical studies, as it may not adequately address data imbalance. Our model outperformed existing studies by using metrics that better account for the imbalance. However, external validation is required to ascertain the model's efficacy beyond the Soonchunhyang University Bucheon Hospital dataset. In

Table 4. IBP feature characteristics accoring to the class

Features	Hypotension		Normal		p-Value
	Mean	Standard Deviation	Mean	Standard Deviation	
max	110.531	16.436	133.26	16.543	¡0.001
min	49.09	8.475	66.127	9.014	¡0.001
mean	70.579	9.59	91.802	10.418	¡0.001
sd	16.446	4.43	18.93	4.915	¡0.001
skewness	0.862	0.278	0.662	0.217	¡0.001
kurtosis	-0.411	0.991	-0.806	0.931	¡0.001
rms	72.573	9.826	93.83	10.707	¡0.001
IQR	23.936	8.743	31.716	11.07	¡0.001
the number of data under the mean	1270.784	81.658	1226.01	74.063	¡0.001
the number of peaks	24.869	5.653	23.611	5.253	¡0.001
mean of peaks interval	83.562	21.0	87.898	25.789	¡0.001
sd of peaks interval	10.912	23.938	14.213	32.916	¡0.001
min of peaks	100.574	15.31	105.187	15.632	¡0.001
mean of peaks	105.187	15.632	128.463	16.141	¡0.001
sd of peaks	2.734	1.614	2.413	1.25	¡0.001
crest factor (peak)	1.526	0.137	1.421	0.097	¡0.001
the number of valleys	19.273	7.754	22.951	5.469	¡0.001
mean of valleys interval	110.765	82.535	90.393	34.912	¡0.001
sd of valleys interval	45.927	68.921	17.7	38.339	¡0.001
max of valleys	52.443	11.643	71.289	9.759	¡0.001
mean of valleys	50.583	11.257	68.865	9.425	¡0.001
sd of valleys	1.245	0.955	1.44	0.953	¡0.001
crest factor (valley)	0.661	0.126	0.704	0.073	¡0.001
existence of mean change	0.246	0.431	0.183	0.387	¡0.001
existence of variance change	0.094	0.291	0.061	0.239	¡0.001
first strongest frequency	26.811	8.924	24.84	6.684	¡0.001
second strongest frequency	49.132	14.997	48.596	11.73	¡0.001
third strongest frequency	65.41	31.113	69.387	30.991	¡0.001
fourth strongest frequency	66.346	46.122	75.689	43.815	¡0.001
fifth strongest frequency	70.716	51.728	81.16	49.076	¡0.001
power of first strongest frequency	7.681	2.685	9.387	3.023	¡0.001
power of second strongest frequency	4.39	1.442	4.846	1.458	¡0.001
power of third strongest frequency	2.227	0.931	2.285	0.846	¡0.001
power of fourth strongest frequency	1.433	0.685	1.543	0.609	¡0.001
power of fifth strongest frequency	1.097	0.582	1.197	0.506	¡0.001
MAP empirical	70.278	9.425	88.71	9.428	¡0.001
PPmax	55.694	13.594	63.513	13.766	¡0.001
PPmin	49.575	12.508	57.881	12.892	¡0.001
PPV	11.841	9.358	9.438	7.083	¡0.001
Approximate entropy	0.199	0.064	0.189	0.054	¡0.001
Sample entropy	0.127	0.053	0.125	0.045	<¡0.001

addition, rather than relying solely on backward findings, it is essential to include forward analyses for a more complete validation [12]. To confirm the practical efficacy of our model, it is necessary to develop a software interface and conduct prospective clinical trials. This will allow us to investigate whether the IOH alerts effectively support anesthesiologists in decision-making. Key performance indicators in a clinical setting will include the reduction of alarm fatigue and the optimization of alert timing. Future studies should rigorously monitor whether frequent false alarms disrupt the surgical workflow or if any critical hypotensive phases remain undetected, thereby ensuring both the safety and reliability of the AI-driven assistant.

Third, the interpretability of the model remains a challenge. Despite our efforts to enhance the interpretability for features using TabNet, it exhibited the same lack of interpretability for scalogram images as many other deep learning models. In the future, We aim to integrate advanced explainable AI techniques, such as Grad-CAM or attention-map visualizations to provide universal interpretability across both tabular and image-based inputs.

Author Contributions

Conceptualization, G.L. and J.W.; methodology, G.L. and J.W.; validation, M.L. and S.K.; software, G.L.; resources, M.L. and S.K.; data curation, M.L. and S.K.; writing—original draft, G.L; writing—review and editing, J.W.; supervision, J.W.; All authors have read and agreed to the published version of the manuscript.

Funding declaration

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. RS-2023-00218176), National Research Foundation of Korea funded by the Ministry of Education (NRF-2020R1I1A3056858) and Soonchunhyang University Research Fund.

Data Availability Statement

The IBP data and the source code are available at the dropbox link (https://www.dropbox.com/scl/fo/mml7mry9il4it4abmb6va/AKsgACpDMEfAFt31bgFk8IY?rlke and will be public at the github after acceptance.

References

1. Salmasi V, Maheshwari K, Yang D, Mascha EJ, Singh A, Sessler DI, et al. 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. 2017;126(1):47–65.

2. Ross E, Lehan J, Sia T. Hybrid intelligent system for hemodynamic management in cardiac surgical patients. Artificial Intelligence in Medicine. 2009;46(2):131–140.

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

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

5. Dong S, Wang Q, Wang S, Zhou C, Wang H. Hypotension prediction index for the prevention of hypotension during surgery and critical care: A narrative review. Computers in Biology and Medicine. 2024; p. 107995.

6. Vistisen ST, Enevoldsen J. CON: The hypotension prediction index is not a validated predictor of hypotension. European Journal of Anaesthesiology— EJA. 2024;41(2):118–121.

7. Jacquet-Lagrèze M, Larue A, Guilherme E, Schweizer R, Portran P, Ruste M, et al. Prediction of intraoperative hypotension from the linear extrapolation of mean arterial pressure. European Journal of Anaesthesiology— EJA. 2022;39(7):574–581.

8. Vistisen ST, Enevoldsen J. Artificial intelligence for predicting intra-operative hypotension: You get what you ask for. European Journal of Anaesthesiology. 2024;41(1):63–65. doi:10.1097/EJA.0000000000001939.

9. Enevoldsen J, Vistisen ST. Adding bias to artificial intelligence for predicting intra-operative hypotension enhances apparent predictive performance. European Journal of Anaesthesiology. 2023;40(4):290–292. doi:10.1097/EJA.0000000000001693.

10. Mulder M, Kamphuis W, Poterman M, Veelo D, de Vries J, Geerts BF. Comparison of hypotension prediction index alarm performance with arterial pressure threshold alarms. Anesthesiology. 2024;140(4):417–430. doi:10.1097/ALN.0000000000004990.

11. Saugel B, Maheshwari K, Sessler DI. Artificial intelligence for predicting peri-operative hypotension: from hype through disillusionment to reality. British Journal of Anaesthesia. 2023;130(4):418–422. doi:10.1016/j.bja.2023.02.029.

12. Davies SJ, Sessler DI, Jian Z, Fleming NW, Mythen M, Maheshwari K, et al. Comparison of differences in cohort (forward) and case control (backward) methodologic approaches for validation of the Hypotension Prediction Index. Anesthesiology. 2024;141(3):443.

13. Michard F, Mulder MP, Gonzalez F, Sanfilippo F. AI for the hemodynamic assessment of critically ill and surgical patients: focus on clinical applications. Annals of Intensive Care. 2025;15(1):26.

14. Reich DL, Hossain S, Krol M, Baez B, Patel P, Bernstein A, et al. Predictors of hypotension after induction of general anesthesia. Anesthesia & Analgesia. 2005;101(3):622–628.

15. Ko BS, Kim YJ, Jung DH, Sohn CH, Seo DW, Lee YS, et al. Early risk score for predicting hypotension in normotensive patients with non-variceal upper gastrointestinal bleeding. Journal of Clinical Medicine. 2019;8(1):37.

16. Lee K, Jang JS, Kim J, Suh YJ. Age shock index, shock index, and modified shock index for predicting postintubation hypotension in the emergency department. The American journal of emergency medicine. 2020;38(5):911–915.

17. Ghosh S, Feng M, Nguyen H, Li J. Hypotension risk prediction via sequential contrast patterns of ICU blood pressure. IEEE journal of biomedical and health informatics. 2015;20(5):1416–1426.

18. Park S, Kim WJ, Cho NJ, Choi CY, Heo NH, Gil HW, et al. Predicting intradialytic hypotension using heart rate variability. Scientific reports. 2019;9(1):2574.

19. Moghadam MC, Abad EMK, Bagherzadeh N, Ramsingh D, Li GP, Kain ZN. A machine-learning approach to predicting hypotensive events in ICU settings. Computers in biology and medicine. 2020;118:103626.

20. Lee HC, Park Y, Yoon SB, Yang SM, Park D, Jung CW. VitalDB, a high-fidelity multi-parameter vital signs database in surgical patients. Scientific Data. 2022;9(1):279.

21. Choe S, Park E, Shin W, Koo B, Shin D, Jung C, et al. Short-term event prediction in the operating room (STEP-OP) of five-minute intraoperative hypotension using hybrid deep learning: retrospective observational study and model development. JMIR Medical Informatics. 2021;9(9):e31311.

22. Lee S, Lee HC, Chu YS, Song SW, Ahn GJ, Lee H, et al. Deep learning models for the prediction of intraoperative hypotension. British journal of anaesthesia. 2021;126(4):808–817.

23. Hwang E, Park YS, Kim JY, Park SH, Kim J, Kim SH. Intraoperative Hypotension Prediction Based on Features Automatically Generated Within an Interpretable Deep Learning Model. IEEE Transactions on Neural Networks and Learning Systems. 2023;.

24. Moon JH, Lee G, Lee SM, Ryu J, Kim D, Sohn KA. Frequency Domain Deep Learning With Non-Invasive Features for Intraoperative Hypotension Prediction. IEEE Journal of Biomedical and Health Informatics. 2024;.

25. Lin CS, Chiu JS, Hsieh MH, Mok MS, Li YC, Chiu HW. Predicting hypotensive episodes during spinal anesthesia with the application of artificial neural networks. Computer methods and programs in biomedicine. 2008;92(2):193–197.

26. Kang AR, Lee J, Jung W, Lee M, Park SY, Woo J, et al. Development of a prediction model for hypotension after induction of anesthesia using machine learning. PloS one. 2020;15(4):e0231172.

27. Lee J, Woo J, Kang AR, Jeong YS, Jung W, Lee M, et al. Comparative analysis on machine learning and deep learning to predict post-induction hypotension. Sensors. 2020;20(16):4575.

28. Jo YY, Jang JH, Kwon Jm, Lee HC, Jung CW, Byun S, et al. Predicting intraoperative hypotension using deep learning with waveforms of arterial blood pressure, electroencephalogram, and electrocardiogram: Retrospective study. Plos one. 2022;17(8):e0272055.

29. Jeni LA, Cohn JF, De La Torre F. Facing imbalanced data–recommendations for the use of performance metrics. In: 2013 Humaine association conference on affective computing and intelligent interaction. IEEE; 2013. p. 245–251.

30. Smith LC, Turcotte DL, Isacks BL. Stream flow characterization and feature detection using a discrete wavelet transform. Hydrological processes. 1998;12(2):233–249.

31. Misiti M, Misiti Y, Oppenheim G, Poggi JM. Wavelets and their Applications. John Wiley & Sons; 2013.

32. Lee H, Lee J, Yoo C. Selecting a mother wavelet for univariate wavelet analysis of time series data. Journal of Korea Water Resources Association. 2019;52(8):575–587.

33. Jevrejeva S, Moore J, Grinsted A. Influence of the Arctic Oscillation and El Niño-Southern Oscillation (ENSO) on ice conditions in the Baltic Sea: The wavelet approach. Journal of Geophysical Research: Atmospheres. 2003;108(D21).

34. Baliunas S, Frick P, Sokoloff D, Soon W. Time scales and trends in the central England temperature data (1659–1990): A wavelet analysis. Geophysical Research Letters. 1997;24(11):1351–1354.

35. Jiang Q, Suter BW. Instantaneous frequency estimation based on synchrosqueezing wavelet transform. Signal Processing. 2017;138:167–181.

36. De Moortel I, Munday S, Hood AW. Wavelet analysis: the effect of varying basic wavelet parameters. Solar Physics. 2004;222(2):203–228.

37. Lee S, Lee M, Kim SH, Woo J. Intraoperative hypotension prediction model based on systematic feature engineering and machine learning. Sensors. 2022;22(9):3108.

38. Pincus SM; IEEE. Approximate entropy: a complexity measure for biological time series data. 1991; p. 35–36.

39. Richman JS, Lake DE, Moorman JR. Sample entropy. 2004;384:172–184.

40. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:201011929. 2020;.

41. Arik SÖ, Pfister T. Tabnet: Attentive interpretable tabular learning. 2021;35(8):6679–6687.

42. Keerthi Krishnan K, Soman K. CNN based classification of motor imaginary using variational mode decomposed EEG-spectrum image. Biomedical engineering letters. 2021;11(3):235–247.

43. Chen K, Wang Q, He Z, Chen K, Hu J, He J. Convolutional sequence to sequence non-intrusive load monitoring. the Journal of Engineering. 2018;2018(17):1860–1864.

44. Hoffer E, Hubara I, Soudry D. Train longer, generalize better: closing the generalization gap in large batch training of neural networks. Advances in neural information processing systems. 2017;30.

45. Akkus C, Chu L, Djakovic V, Jauch-Walser S, Koch P, Loss G, et al.. Multimodal Deep Learning; 2023.

46. Snoek CG, Worring M, Smeulders AW. Early versus late fusion in semantic video analysis. In: Proceedings of the 13th annual ACM international conference on Multimedia; 2005. p. 399–402.

47. Atrey PK, Kankanhalli MS, Oommen JB. Goal-oriented optimal subset selection of correlated multimedia streams. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM). 2007;3(1):2–es.

48. Huang SC, Pareek A, Seyyedi S, Banerjee I, Lungren MP. Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. NPJ digital medicine. 2020;3(1):136.

49. Lee HC, Jung CW. Vital Recorder—a free research tool for automatic recording of high-resolution time-synchronised physiological data from multiple anaesthesia devices. Scientific reports. 2018;8(1):1527.

50. Choe S, Park E, Shin W, Koo B, Shin D, Jung C, et al. Short-term event prediction in the operating room (STEP-OP) of five-minute intraoperative hypotension using hybrid deep learning: retrospective observational study and model development. JMIR Medical Informatics. 2021;9(9):e31311.

51. Powers DM. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. arXiv preprint arXiv:201016061. 2020;.