BMC Anesthesiology

RESEARCH Open Access

AnesNet: a versatile deep convolutional framework for predicting depth of anesthesia using BIS and multimodal intraoperative signals

Yunjia Qi[1,2†], Xiaoxiao Wang[3†], Yichen Cui[3], Kaixi Liu[1], Lu Hua Chen[4,5,6], Xiangyang Guo[1], Zhengqian Li[1,7*†] and Sichen Liu[7,8*†]

Abstract

Deep anesthesia is associated with delayed recovery, postoperative cognitive dysfunction, and increased long-term mortality. In clinical practice, improper control of anesthesia depth can lead to intraoperative awareness or excessive suppression, thereby elevating the risk of postoperative neurological complications. Therefore, accurate monitoring and prediction of anesthesia depth are of great clinical importance. Currently, the bispectral index (BIS) is the most widely used clinical tool for assessing anesthesia depth by analyzing electroencephalogram (EEG) signals. However, existing BIS monitoring models can only provide real-time assessments and cannot effectively predict the trend of anesthesia depth. To address this limitation, this study proposes AnesNet, a deep learning–based model designed to provide early warnings of anesthesia depth 5, 10, and 15 min in advance. Considering diverse clinical needs, AnesNet supports both regression and classification prediction paradigms, enabling continuous depth forecasting and risk-oriented early warning of clinically unfavorable anesthesia states. A total of 2,803 surgical patients were included in this study. Results show that AnesNet achieved an AUROC of 0.895 in classification tasks and a MAE of 4.90 in regression tasks. Furthermore, feature ablation experiments demonstrated the model's interpretability, enabling quantification of the contribution of individual physiological parameters to the prediction outcomes.

Keywords Artificial Intelligence, Deep Learning, Deep Anesthesia, Bispectral Index, Perioperative Medicine

†Yunjia Qi and Xiaoxiao Wang contributed equally to this work.

†Zhengqian Li and Sichen Liu are designated as co-last authors.

*Correspondence:
Zhengqian Li
zhengqianli@hsc.pku.edu.cn
Sichen Liu
Sichen.Liu@xjtlu.edu.cn
[1]Department of Anesthesiology, Peking University Third Hospital, No. 49, North Garden Street, Haidian District, Beijing 100191, China
[2]Oxford Internet Institute, University of Oxford, Oxford, UK
[3]Research Center of Clinical Epidemiology, Peking University Third Hospital, No. 49, North Garden Street, Haidian District, Beijing 100191, China
[4]Department of Rehabilitation Sciences, Faculty of Health and Social Sciences, The Hong Kong Polytechnic University, Kowloon, Hong Kong SAR, China
[5]Research Institute for Smart Ageing (RISA), The Hong Kong Polytechnic University, Kowloon, Hong Kong SAR, China
[6]Mental Health Research Center (MHRC), The Hong Kong Polytechnic University, Kowloon, Hong Kong SAR, China
[7]Department of Anesthesiology, State Key Laboratory of Vascular Homeostasis and Remodeling, Peking University Third Hospital, Beijing 100191, China
[8]Department of Intelligent Science, School of Advanced Technology, Xi'an Jiaotong-Liverpool University, 111 Ren'ai Road, Suzhou, Jiangsu 215123, China

Introduction

Achieving the appropriate depth of anesthesia is essential for optimal surgical outcomes. A light hypnotic level may cause intraoperative awareness [1], potentially leading to post-traumatic stress disorder or depression [2]. On the other hand, deeper levels of hypnosis have been associated with a range of adverse outcomes, including delayed awakening and recovery [3, 4], potentially postoperative complications such as cognitive dysfunction [5, 6], and possibly long-term mortality [7]. Inadequate intraoperative monitoring of anesthesia depth is considered a significant factor contributing to suboptimal anesthesia management [8], underscoring the need for consistent and accurate monitoring.

The Bispectral Index (BIS) is currently the most widely used method in clinical practice for monitoring intraoperative anesthetic depth. BIS assesses brain activity through electroencephalogram (EEG) signals [9], which are presented as a numerical 'BIS values' ranging from 0 (no brain activity) to 100 (fully awake). BIS values between ≥ 40 and < 60 represent adequate general anesthesia for surgery, and values < 40 represent a deep hypnotic state [10]. Although BIS values are a tool used by anesthesiologists to monitor the depth of anesthesia, by the time an inappropriate depth is indicated by the BIS value, the patient may have already incurred the risks associated with a hypnotic level that is too light or too deep. Therefore, models that can predict the hypnotic level in real time, allowing anesthesiologists to adjust sedation before an inappropriate depth of anesthesia occurs, are critically needed.

A previous study indicated that anesthesiologists prefer to administer deep anesthesia due to the fear of light anesthesia and resulting intraoperative awareness [11]. Other studies have demonstrated a substantial increase in patient exposure to anesthetic drugs with routine care [3, 5, 12]. Collectively, the literature indicates that patients are more likely to be exposed to the risks of deep anesthesia rather than light anesthesia, and clinical tools are needed to ensure that the increased exposure to deep anesthesia does not lead to more adverse outcomes. In particular, accurate predictive models that can identify risk of deep hypnotic events with enough time to adjust anesthetic depth represent an urgent unmet need.

Most previous studies used machine learning approaches to analyze EEG data to approximate BIS in real time [13–18] rather than predicting future BIS values. These studies used features extracted from EEG as the single data source for estimating depth of anesthesia; however, the anesthesia-induced changes in brain activity also affect the central nervous system's regulation of peripheral organ function and vital signs, including heart rate, blood pressure, pulse oxygen saturation, and end-tidal carbon dioxide, indicating that these clinical features may provide important information to predict future depth of anesthesia intraoperatively.

Traditional machine learning methods such as Random Forest (RF) and Support Vector Machine (SVM) perform well on structured, low-dimensional feature data but typically rely on handcrafted feature engineering [19], which limits their ability to directly process high-dimensional, continuous physiological waveform data and to model complex temporal dependencies and nonlinear interactions among multimodal signals. In contrast, convolutional neural networks (CNNs) can automatically learn multi-scale temporal features from raw time-series data and effectively integrate multi-channel physiological signals [13], making them more suitable for modeling BIS and multimodal intraoperative physiological waveforms. Therefore, this study adopts CNN as the primary modeling framework to enable more accurate prediction of future anesthesia depth.

This study aims to develop a predictive model capable of utilizing either single- or multi-channel intraoperative physiological signals to provide real-time early warnings of future trends in anesthesia depth, assessing whether clinically available BIS values, either alone or in combination with other intraoperative physiological signals, can help predict anesthesia depth. We propose a novel convolutional neural network architecture, AnesNet, designed to support the prediction task of anesthesia depth, which can be implemented through two modeling approaches—regression for forecasting future BIS values and classification for identifying imminent deep anesthesia events—while supporting variable input channel configurations to accommodate diverse clinical monitoring scenarios. Compared to traditional models, AnesNet incorporates not only electroencephalographic indicators (BIS), but also multimodal physiological signals including arterial blood pressure (ABP), electrocardiogram (ECG-II), photoplethysmography (PLETH), and end-tidal carbon dioxide ($ETCO_2$). This multimodal integration enhances the model's predictive capability regarding intraoperative anesthesia state fluctuations. We trained and evaluated the model using a large-scale dataset from surgical patients, and conducted ablation experiments to quantify the contribution of each input signal. The results demonstrate that AnesNet offers valuable early-warning capabilities and decision support for anesthesia management, showing strong potential for clinical application.

Methods

To investigate the feasibility of predicting intraoperative deep hypnotic events, we constructed a deep learning pipeline centered around AnesNet. As illustrated in Fig. 1, this framework consists of four major stages: (I) Data Acquisition, in which physiological waveform data were collected from the VitalDB repository containing

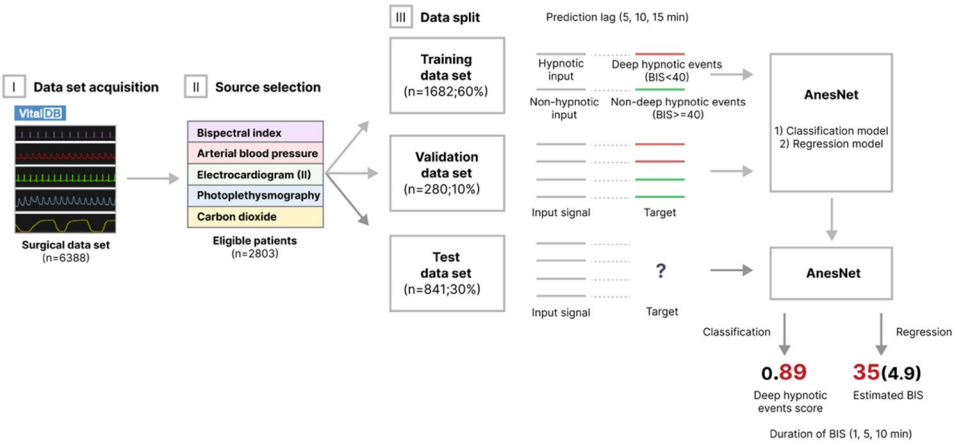

Fig. 1 Overview of the AnesNet framework for intraoperative anesthesia depth prediction

more than 6,000 surgical cases; (II) Signal Selection, where five key physiological indicators (BIS, ABP, ECG-II, PLETH, and $ETCO_2$) were chosen; (III) Data Partitioning, in which 2,803 eligible patients were divided into training, validation, and test sets (60%, 10%, and 30%, respectively); and (IV) Model Training and Evaluation, during which the proposed AnesNet architecture was trained and validated for both classification and regression tasks. This model supports both classification and regression tasks to accommodate different clinical objectives. For each patient, 30-second waveform segments were extracted at 5, 10, and 15 min prior to labeled events (deep hypnotic or non-hypnotic), enabling time-lagged prediction of BIS trends. The final model outputs either a binary probability of deep hypnosis (AnesNet-C) or a continuous BIS score estimate (AnesNet-R), demonstrating both flexibility and high predictive performance across multiple temporal horizons.

Data source

The data used in this study were sourced from the VitalDB database (http://vitaldb.net/data-bank) [20], a publicly available research dataset which provides physiological monitoring data from various devices for 6,388 surgical patients. These patients underwent non-cardiac (general, thoracic, urologic, and gynecologic) surgeries in 10 out of the 31 operating rooms at Seoul National University Hospital, Seoul, Republic of Korea, from June 2016 to August 2017. A waiver of study approval was granted by the Institutional Review Board (IRB) of Peking University Third Hospital because of the use of de-identified data. The original study (VitalDB) was approved by the IRB of Seoul National University Hospital (H-1408-101-605) and written informed consent was waived due to the retrospective nature of the study design. The original study was registered at clinicaltrials.gov (NCT02914444, Principal investigator: Chul-Woo Jung, Date of registration: 2016-09-26).

The database contains clinical information, hemodynamic parameters, and laboratory results recorded during surgery. Five physiological signals were selected as the primary inputs for analysis: ABP, ECG-II, PLETH, $ETCO_2$, and BIS values. ABP and ECG-II represent cardiovascular regulatory function; PLETH captures peripheral perfusion dynamics; and $ETCO_2$ indicates respiratory function, all of which complement BIS by providing additional physiological context for predicting anesthesia depth.

Data preprocessing

We first extracted data containing the aforementioned signals. Arterial blood pressure, electrocardiogram, and pulse oximetry data were obtained using the GE Healthcare Tram-Rac (SNUADC) patient monitor, with a waveform acquisition interval of 1/500 seconds. The CO2 waveform was collected using the Drager Primus anesthesia machine, with a waveform acquisition interval of 1/62.5 s. During the waveform extraction process, all data were resampled to 100 Hz. BIS values were obtained from the Covidien BIS Vista EEG monitor at a sampling rate of 1 Hz.

To ensure data integrity, a rigorous quality-control process was performed. Acceptable signal ranges were defined as follows: BIS, 0–100; ABP, 0–200 mmHg; ECG-II voltage, > -1 mV; PLETH, 0–70; and $ETCO_2$, 0–60 mmHg. Waveforms exhibiting random or irregular fluctuations, sudden spikes, prolonged drifts, harmonic distortion, flat lines, or truncation were excluded from analysis.

After quality control, 2,803 patients were included in the study and randomly divided into training ($n = 1{,}682$), validation ($n = 280$), and test ($n = 841$) sets at a ratio of 6:1:3. A deep hypnotic event was defined as any continuous segment lasting more than 5 min with a BIS value < 40. In addition to BIS values, the other predictive features included waveforms of ABP, ECG-II, PLETH,

and $ETCO_2$. To predict deep hypnotic events, 30-second waveform segments occurring 5, 10, or 15 min before each event were used as input. In contrast, non-hypnotic events were defined as continuous segment of non-hypnotic state lasting more than 20 min, and the input segments were taken 5, 10, or 15 min before and after each non-hypnotic event.

Data analysis

Table 1 summarizes the baseline characteristics of the 2,803 patients included in the training, validation, and

Table 1 Data characteristics

	Training dataset (*n* = 1682, 60%)	Validation dataset (*n* = 280, 10%)	Test dataset (*n* = 841, 30%)	*p*-value
Age, years (SD)	59 (15)	60 (15.5)	59 (14.5)	0.2177
Sex, *n* (%)				0.7063
Male	940 (55.9)	149 (53.2)	466 (55.4)	
Female	742 (44.1)	131 (46.8)	375 (44.6)	
Height, cm (SD)	163 (9.5)	162 (9)	163 (9.5)	0.7015
Weight, kg (SD)	61 (11.5)	61 (12.5)	62 (11.5)	0.6364
Body mass index, kg/m^2 (SD)	23 (4)	23 (4)	23 (3.5)	0.3462
ASA grade, *n* (%)				0.5602
1	352 (21.3)	59 (21.3)	171 (21.0)	
2	1023 (62.0)	183 (66.1)	527 (64.7)	
3	256 (15.5)	33 (11.9)	107 (13.1)	
4	13 (0.8)	1 (0.4)	6 (0.7)	
5	0 (0.0)	0 (0.0)	0 (0.0)	
6	5 (0.3)	1 (0.4)	4 (0.5)	
Preoperative comorbidity, *n* (%)				
Hypertension	569 (33.8)	97 (34.6)	287 (34.1)	0.9610
Diabetes mellitus	201 (12.0)	33 (11.8)	108 (12.8)	0.7921
Pulmonary function test, *n* (%)				0.6136
Normal	1369 (81.4)	235 (84.0)	697 (82.9)	
Obstructive-dominant pattern	218 (13.0)	33 (11.8)	95 (11.3)	
Restrictive-dominant pattern	95 (5.6)	12 (4.3)	49 (5.8)	
Emergency operation, *n* (%)	213 (12.7)	36 (12.9)	103 (12.2)	0.9447
Surgical domain, *n* (%)				0.8693
General surgery	1039 (61.8)	171 (61.1)	526 (62.5)	
Gynaecologic surgery	93 (5.5)	13 (4.6)	46 (5.5)	
Thoracic surgery	528 (31.4)	92 (32.9)	255 (30.3)	
Urogenital surgery	22 (1.3)	4 (1.4)	14 (1.7)	
Approach, *n* (%)				0.9255
Open	683 (40.6)	115 (41.1)	342 (40.7)	
Robotic	64 (3.8)	10 (3.6)	43 (5.1)	
Videoscopic	935 (55.6)	155 (55.4)	456 (54.2)	
Anaesthesia, *n* (%)				0.1466
General	1681 (99.9)	279 (99.6)	841 (100.0)	
Sedation analgesia	1 (0.1)	1 (0.4)	0 (0.0)	
Spinal	0 (0.0)	0 (0.0)	0 (0.0)	
Duration of anaesthesia, min (SD)	241 (106.5)	245 (114.5)	244 (102.5)	0.6797
Postop in-hospital stay, day (SD)	12 (15.5)	13 (21)	11 (11.5)	0.8072
Postop ICU stay, day (SD)	1(5)	1 (2.5)	1 (1.5)	0.9907
In-hospital mortality, *n* (%)	17 (1.0)	1 (0.4)	10 (1.2)	0.4773
Dataset composition				
Deep hypnotic event, *n* (%)	2505 (13.2)	356 (11.5)	1134 (11.7)	
per patient, *n* (SD)	1 (2.5)	1 (3)	1 (2.5)	
Non-deep hypnotic event, *n* (%)	16,520 (86.8)	2732 (88.5)	8583 (88.3)	
per patient, *n* (SD)	10 (7)	10 (7)	10 (8)	

"Per patient, *n* (SD)" represents the mean number of events per patient and the corresponding standard deviation across patients

test datasets. The three datasets were comparable in demographic, anthropometric, and clinical parameters, with no statistically significant differences observed among them (all $p > 0.05$). The mean age of patients was approximately 59 years, and males accounted for about 55% of the population. Most patients were classified as American Society of Anesthesiologists (ASA) grade II and underwent general anesthesia, primarily for general or thoracic surgical procedures. The proportions of emergency operations, surgical approaches, and comorbidities such as hypertension and diabetes mellitus were also balanced across the datasets.

Regarding intraoperative and postoperative variables, the mean anesthesia duration was approximately 240 min, and the average postoperative hospital stay was around 12 days. The incidence of in-hospital mortality was low (< 1.5%) across all groups. Dataset composition analysis showed that deep hypnotic events accounted for roughly 12% of all labeled events, with the remainder representing non-deep hypnotic states. These findings indicate that the dataset partitioning was statistically balanced and suitable for model training and evaluation.

AnesNet

The proposed AnesNet architecture is a multi-channel one-dimensional convolutional neural network (1D-CNN) designed to extract temporal dependencies from physiological waveform data. As illustrated in Fig. 2, the input signals—including BIS, ABP, ECG-II, PLETH, and $ETCO_2$—are concatenated along the channel dimension to form a unified multi-channel input tensor. For each 30-second signal segment sampled at 100 Hz, the temporal length is T = 3000, and the number of input channels is N. The resulting input can therefore be represented as

$$X \in R^{N \times T}$$

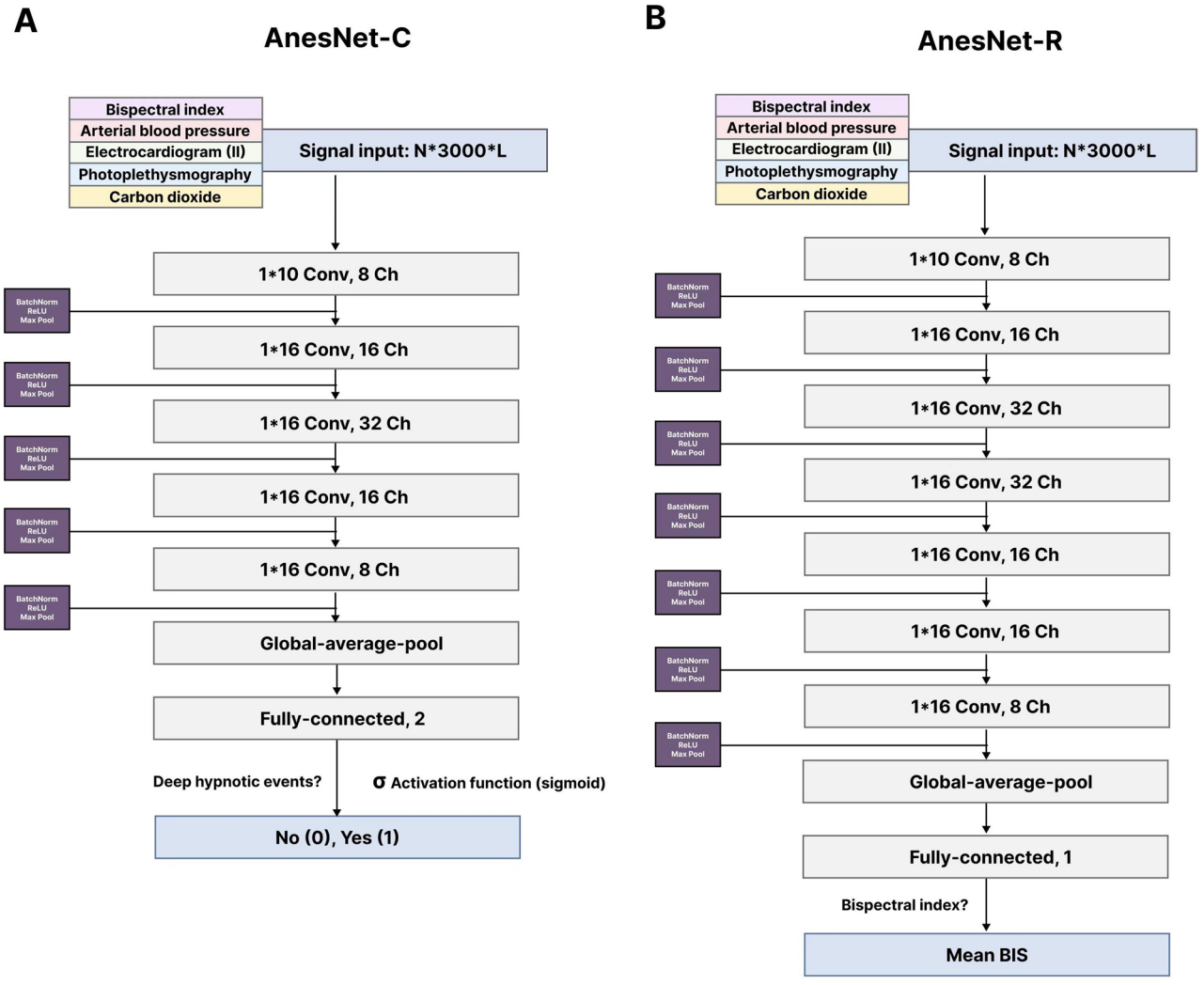

Fig. 2 Architecture of the proposed AnesNet deep learning model for anesthesia depth prediction

Each 1D convolutional layer applies C convolutional kernels of size k to capture local temporal features. The output of the c-th channel at time step t is defined as

$$Y_{c,t} = \sum_{n=1}^{N} \sum_{j=1}^{k} W_{c,n,j} \cdot X_{n,\, t+j-1} + b_c$$

where $W_{c,n,\mathrm{j}}$denotes the j-th weight of the c-th kernel on the n-th input channel; b_c is the bias term. The resulting output feature map can be expressed as

$$Y \in R^{N \times T\prime}$$

Here, $T\prime = T - k + 1$ represents the temporal dimension after convolution. Each convolutional layer is followed by batch normalization, a ReLU activation function, and max-pooling to improve numerical stability and reduce temporal resolution. Stacking multiple convolutional blocks allows the model to progressively capture both short- and long-range temporal dependencies in the physiological signals. A global average pooling layer aggregates the extracted features across time, and a fully connected (dense) layer produces the final task-specific outputs.

Table 2 Performance of the AnesNet-C

Prediction lag	Model	AUROC (95% Confidence Interval)	Sensitivity (95% Confidence Interval)	Specificity (95% Confidence Interval)
5 min	5-parameter multichannel model	0.8953 (0.8858–0.9049)	0.8289 (0.8070–0.8508)	0.8301 (0.8222–0.8381)
	Bispectral-index-only model	0.8920 (0.8826–0.9015)	0.8228 (0.8005–0.8450)	0.8242 (0.8161–0.8322)
	4-parameter multichannel model	0.6207 (0.6041–0.6373)	0.5855 (0.5569–0.6142)	0.5849 (0.5745–0.5953)
10 min	5-parameter multichannel model	0.8648 (0.8509–0.8786)	0.7989 (0.7700-0.8279)	0.8007 (0.7903–0.8110)
	Bispectral-index-only model	0.8519 (0.8370–0.8668)	0.7857 (0.7556–0.8158)	0.7857 (0.7749–0.7965)
	4-parameter multichannel model	0.5988 (0.5781–0.6195)	0.5666 (0.5317–0.6016)	0.5682 (0.5548–0.5815)
15 min	5-parameter multichannel model	0.8421 (0.8262–0.8581)	0.7642 (0.7297–0.7987)	0.7658 (0.7525–0.7791)
	Bispectral-index-only model	0.8233 (0.8050–0.8417)	0.7561 (0.7210–0.7911)	0.7578 (0.7441–0.7715)
	4-parameter multichannel model	0.6232 (0.5982–0.6482)	0.5763 (0.5340–0.6186)	0.5794 (0.5642–0.5947)

Two variants of the model were developed: AnesNet-C (Fig. 2-A) for classification and AnesNet-R (Fig. 2-B) for regression. In AnesNet-C, the output layer employs a sigmoid activation function to map predictions to a probability between 0 and 1, representing the likelihood of an impending deep hypnotic event. In AnesNet-R, the output layer uses a linear activation to estimate the continuous BIS value at the target time point.

In addition, the model's architecture—including the number of layers and channel widths—is adjusted based on the prediction interval (5, 10, or 15 min) to better capturetemporal dynamics over varying time scales. For instance, the 5-minute prediction uses a shallower network with four convolutional layers and progressively increasing channel sizes, whereas the 15-minute prediction employs a deeper network with seven convolutional layers.

Result

The classification performance of AnesNet-C is summarized in Table 2. The results show that the model consistently demonstrates strong predictive ability across 5-, 10-, and 15-minute prediction intervals, achieving AUROC values of 0.8953, 0.8648, and 0.8421, respectively.

To assess the importance of different input features, we conducted ablation studies by building two additional models: a BIS-only single-channel model, and a four-parameter model using only conventional vital signs (ABP, ECG-II, PLETH, and $ETCO_2$). The BIS-only model performed slightly worse than the full hybrid model, suggesting that BIS signals alone are highly informative for predicting deep hypnotic events. In contrast, the four-parameter model showed significantly lower performance than both the full model and the BIS-only model, with AUROC values below 0.63, indicating that conventional vital signs alone are insufficient for accurate prediction of deep anesthesia events in the absence of BIS.

Furthermore, model calibration curves generated via bootstrapping, as shown in Fig. 3, demonstrated good fit and calibration performance of the primary model. Overall, these results suggest that integrating BIS with conventional physiological signals improves predictive performance, while BIS alone provides highly representative information regarding anesthetic depth.

Table 3 summarizes the performance of the regression models across different prediction intervals. The primary hybrid model, which incorporates BIS and the four physiological signals, achieved MAEs of 4.90, 5.44, and 5.65 for the 5-, 10-, and 15-minute intervals, respectively, outperforming all other models in both accuracy and generalization stability. As part of the ablation study, the BIS-only model consistently had slightly higher MAEs than the full model across all time points, reinforcing

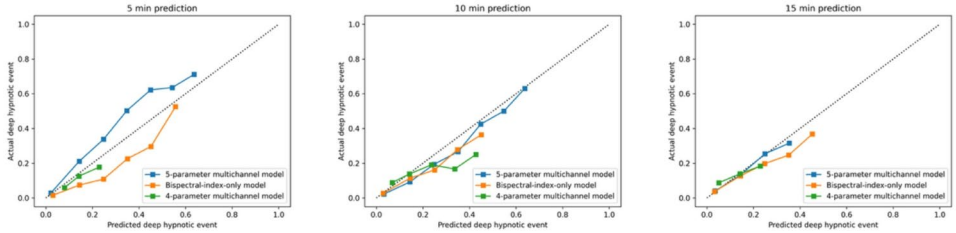

Fig. 3 Bootstrapped calibration curves of the primary model

Table 3 Performance of the AnesNet-R

Prediction lag	Prediction Factors	MAE (95% Confidence Interval)	AUROC (95% Confidence Interval)
5 min	5-parameter multichannel model	4.9024 (4.7654–5.0299)	0.8829 (0.8736–0.8921)
	Bispectral-index-only model	4.9415 (4.8083–5.0587)	0.8828 (0.834–0.8922)
	4-parameter multichannel model	7.1255 (6.9863–7.2663)	0.5772 (0.5605–0.5939)
10 min	5-parameter multichannel model	5.4433 (5.2690–5.6046)	0.8595 (0.8456–0.8733)
	Bispectral-index-only model	5.4435 (5.2726–5.6050)	0.8593 (0.8452–0.8733)
	4-parameter multichannel model	7.1516 (6.9809–7.3322)	0.5642 (0.5421–0.5864)
15 min	5-parameter multichannel model	5.6518 (5.4445–5.8599)	0.8411 (0.8246–0.8576)
	Bispectral-index-only model	5.6633 (5.4466–5.8662)	0.8401 (0.8233–0.8568)
	4-parameter multichannel model	7.6047 (7.3995–7.8109)	0.5425 (0.5179–0.5672)

BIS's dominant role in regression-based prediction. In contrast, the four-parameter model showed markedly higher MAEs (e.g., 7.13 for 5-minute prediction), indicating that conventional vital signs alone are insufficient to accurately reconstruct BIS values. Notably, all models exhibited a modest increase in MAE with longer prediction intervals, reflecting greater uncertainty in forecasting over extended time horizons. Although the regression models' AUROCs were generally slightly lower than those of the classification models, the primary model still maintained AUROCs between 0.84 and 0.88 across all intervals, demonstrating strong discriminative ability.

In summary, the results indicate that BIS is the core information source for predicting future BIS trends; incorporating multiple vital signs in a five-channel hybrid model further improves prediction accuracy; and conventional multiparameter monitoring data, without EEG-based BIS, cannot effectively replace BIS in assessing anesthetic depth.

Discussion

Properly monitoring the depth of anesthesia is crucial during surgical procedures. Deep anesthesia may be associated with increased risks to patients [3, 4], possibly a higher incidence of postoperative complications [3, 5, 6], prolonged hospitalization [21], and potentially elevated mortality rates [7]. Accurately monitoring anesthesia depth and administering appropriate anesthetic agents is challenging due to the variability in patient responses to standard drug doses. This study showed that a deep learning model can predict deep hypnotic events on the basis of BIS values alone or in combination with ABP, ECG-II, PLETH, and ETCO2 5, 10, and 15 min before the deep hypnotic event. The model exhibited satisfactory performance across diverse patients and types of surgery.

Multiple studies have demonstrated frequent, unnecessary, and excessive anesthetic exposure with routine care [3, 5, 12]. Chan et al. reported that routine care resulted in a 21% increase in propofol delivery and a 30% increase in volatile anesthetics compared to BIS value-guided anesthesia [5]. These findings suggest that patients are frequently subjected to an increased risk of deep anesthesia when anesthesiologists rely solely on their experience without the aid of depth-of-anesthesia monitoring.

Traditionally, the depth of anesthesia is approximated through vital signs, including ABP, ECG-II, PLETH, and ETCO2. These monitoring practices are established and considered as standard care protocols. Our findings revealed that a 4-parameter multichannel model, incorporating the aforementioned routinely acquired biosignals, did not achieve satisfactory performance in predicting deep hypnotic events. The model's sensitivity and specificity were between 56% and 58%, indicating that the model failed to detect nearly half of the deep hypnotic events. This could be because patient vital signs may not accurately reflect the depth of anesthesia; rather, they might only reflect the indirect effects of anesthesia on the brain.

During surgery, monitors record and process the brain's spontaneous electrical activity. Several EEG-derived indices, such as the BIS value [18] and Patient State Index [22], have been developed to measure hypnotic levels during anesthesia. Among these, the BIS monitor was the first device approved by the FDA to measure hypnotic

levels and is the most studied method to date. The BIS value is a valuable tool for assessing anesthesia depth, with a BIS < 40 conventionally indicating deep anesthesia [10]. However, by the time anesthesiologists observe a BIS below this threshold, the patient may already be in a deep hypnotic state. This delay in detection puts patients at risk of adverse outcomes if the anesthesiologist is unable to adequately correct the hypnotic state very rapidly.

Predictive models forecasting the risk of deep anesthesia could aid anesthesiologists in proactive intervention, thereby optimizing clinical management and mitigating associated risks. Deep learning algorithms are increasingly used for analyzing biosignal waveforms to predict medical conditions, such as intraoperative hypotension [23]. Efforts have been also made to estimate the depth of anesthesia. However, most previous studies aimed at predicting the depth of anesthesia have focused on approximating BIS values in real time from EEG data, rather than directly predicting them [13, 18]. The model proposed by Alsafy [17], which used hierarchical dispersion entropy coupled with a community graph detection approach, demonstrated a strong correlation between BIS values and the predictive model, achieving a Pearson correlation coefficient of 0.96.

Previously, Hwang et al. [24] proposed a prediction model that achieved a root-mean-square error of 6.614 for the regression task, and an AUROC of 0.937 for the binary classification task of assigning BIS values as either < 60 or ≥ 60. This model predicted BIS values 25 s in advance, which represented the 25 s required to complete the BIS calculation and leaves no time for clinical intervention. However, the 25-second predictive interval did not leave time for the anesthesiologist to adjust drug delivery. In contrast, our model provides anesthesiologists with a reasonable window of time—several minutes—to investigate causes and intervene before a deep hypnotic event occurs.

To our knowledge, this is the first study to propose a deep-learning model for predicting future deep hypnotic events. The BIS-value-only model adequately predicted deep hypnotic events 5 min, 10 min, and 15 min in advance in a diverse patient population that underwent a variety of surgical procedures. Additionally, the hybrid model incorporating biosignals showed greater but not statistically significant AUROCs compared to the BIS-value-only model. BIS monitoring has been specifically developed and validated for assessing the depth of anesthesia, making it a highly specific indicator in this context. The additional biosignals included in the hybrid model may not be directly related to anesthesia depth, which could have limited their effectiveness in improving predictive performance. Further studies of models combining BIS values and other clinical features may help elucidate optimal combinations for predicting deep anesthesia.

One limitation of this study was that it was conducted at a single center, which may limit the generalizability of the findings. A second limitation was the lack of external validation data. Third, a key consideration in interpreting our model's predictive performance is the inherent pharmacodynamic hysteresis—the temporal lag between plasma drug concentration and its clinical effect. It is plausible that the anesthetic states predicted at 5 to 15-minute intervals partially reflect the delayed manifestation of drugs administered at T0. However, our model transcends simple pharmacodynamic lag by capturing individualized physiological responses to both drug titration and surgical stimuli. Unlike population-based PK/PD models, the current approach leverages real-time temporal trends in EEG-based BIS and multiparameter monitoring data to anticipate transitions that are not yet clinically evident. Furthermore, the 15-minute prediction window exceeds the typical time-to-peak effect (tpeak) for common anesthetic agents, suggesting a degree of predictive foresight beyond natural physiological latency. Future research integrating mechanistic PK-PD parameters, such as estimated effect-site concentrations (Ce), will be essential to further disentangle algorithmic prediction from inherent pharmacological delays. Additionally, while drug delivery data, such as end-tidal volatile concentrations or TCI-calculated effect-site concentrations, could offer valuable insights, they were not available for all patients. Including these data in future studies could help refine prediction models.

It is worth noting, the definition of deep anesthesia is based on the BIS value, which is derived from EEG signals recorded from two channels on the forehead. However, recent studies, including the ENGAGES-Canada trials, have challenged the hypothesis that EEG-guided anesthetic depth significantly reduces the incidence of postoperative delirium or mortality [25]. These trials primarily focused on minimizing EEG suppression time, which does not fully address postoperative complications. This highlights the complexity of the relationship between anesthesia depth and postoperative complications. Although our model performed well at predicting deep anesthesia events, this does not imply that minimizing BIS suppression alone will reduce postoperative complications such as delirium. The relationship between anesthesia depth and postoperative outcomes is multifactorial and influenced by patient characteristics and surgical variables. Future research should incorporate additional patient-specific and intraoperative factors to improve predictions of postoperative complications. Furthermore, exploring the potential of multi-channel EEG or full-brain EEG monitoring could offer more precise

assessments of anesthesia depth, potentially improving outcomes, particularly in high-risk populations.

Conclusion

This study proposes AnesNet, a deep neural network–based model for predicting the depth of anesthesia, enabling early warning of intraoperative deep hypnotic events. The model integrates the BIS value with four routine vital signs (ABP, ECG-II, PLETH, and $ETCO_2$), and supports both classification and regression tasks. Validated on real surgical data from 2,803 patients, AnesNet demonstrated strong predictive performance across 5-, 10-, and 15-minute forecasting intervals. Notably, the five-channel hybrid model achieved an AUROC of 0.895 in the classification task and a low MAE of 4.90 in the regression task, significantly outperforming conventional multi-parameter models without BIS input.

Furthermore, ablation experiments revealed that BIS is the core signal for predicting deep anesthesia, while the fusion of multiple parameters further enhances model accuracy and robustness. AnesNet offers a promising tool for refined intraoperative anesthesia management and improved perioperative patient safety. Future studies may incorporate multi-center datasets and full-scalp EEG features to develop more generalizable and interpretable predictive strategies.

Acknowledgements
Not applicable.

Authors' contributions
Yunjia Qi: data analysis or interpretation; Xiaoxiao Wang: writing the paper; Kaixi Liu: review and editing the paper; Lu Hua Chen: review and editing the paper; Xiangyang Guo: review and editing the paper; Zhengqian Li: study concept or design; Sichen Liu: development or design of methodology.

Funding
This work was supported by the Ministry of Science and Technology of the People's Republic of China (grant number STI2030-Major Projects2021ZD0204300), the special fund of the National Clinical Key Specialty Construction Program, P. R. China (2025), National Natural Science Foundation of China (Grant No. 12404538) and Xi'an Jiaotong-Liverpool University, Research Development Fund (Grant No. RDF-22-02-029).

Data availability
The data analysed for the development and validation of the models in this study are available in VitalDB (https://vitaldb.net).

Declarations

Ethics approval and consent to participate
This study used data from a publicly available research dataset, VitalDB. A waiver of study approval was granted by the IRB of Peking University Third Hospital because of the retrospective nature and the use of de-identified data. The original study (VitalDB) was approved by the Institutional Review Board of Seoul National University Hospital (H-1408-101-605).

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 5 January 2026 / Accepted: 16 February 2026
Published online: 21 February 2026

References

1. Ghoneim MM, Block RI, Haffarnan M, et al. Awareness during anesthesia: risk factors, causes and sequelae: a review of reported cases in the literature. Anesth Analg. 2009;108(2):527–35. https://doi.org/10.1213/ane.0b013e318193c634.
2. Bischoff P, Rundshagen I. Awareness under general anesthesia. Dtsch Arztebl Int. 2011;108(1–2):1–7. https://doi.org/10.3238/arztebl.2011.0001.
3. Liu SS. Effects of Bispectral Index monitoring on ambulatory anesthesia: a meta-analysis of randomized controlled trials and a cost analysis. Anesthesiology. 2004;101(2):311–5. https://doi.org/10.1097/00000542-200408000-00010.
4. Long Y, Feng X, Liu H, et al. Effects of anesthetic depth on postoperative pain and delirium: a meta-analysis of randomized controlled trials with trial sequential analysis. Chin Med J (Engl). 2022;135(23):2805–14. https://doi.org/10.1097/CM9.0000000000002449.
5. Chan MT, Cheng BC, Lee TM, et al. BIS-guided anesthesia decreases postoperative delirium and cognitive decline. J Neurosurg Anesthesiol. 2013;25(1):33–42. https://doi.org/10.1097/ANA.0b013e3182712fba.
6. Evered LA, Chan M, Han R, et al. Anaesthetic depth and delirium after major surgery: a randomised clinical trial. Br J Anaesth. 2021;127(5):704–12. https://doi.org/10.1016/j.bja.2021.07.021.
7. Liu YH, Qiu DJ, Jia L, et al. Depth of anesthesia measured by bispectral index and postoperative mortality: A meta-analysis of observational studies. J Clin Anesth. 2019;56:119–25. https://doi.org/10.1016/j.jclinane.2019.01.046.
8. Laferriere-Langlois P, Morisson L, Jeffries S, et al. Depth of Anesthesia and Nociception Monitoring: Current State and Vision For 2050. Anesth Analg. 2024;138(2):295–307. https://doi.org/10.1213/ANE.0000000000006860.
9. Rampil IJ. A primer for EEG signal processing in anesthesia. Anesthesiology. 1998;89(4):980–1002. https://doi.org/10.1097/00000542-199810000-00023.
10. Mathur S, Patel J, Goldstein S et al. Bispectral Index. 2024.
11. Jiang Y, Sleigh J. Consciousness and General Anesthesia: Challenges for Measuring the Depth of Anesthesia. Anesthesiology. 2024;140(2):313–28. https://doi.org/10.1097/ALN.0000000000004830.
12. Wildes TS, Mickle AM, Ben AA, et al. Effect of Electroencephalography-Guided Anesthetic Administration on Postoperative Delirium Among Older Adults Undergoing Major Surgery: The ENGAGES Randomized Clinical Trial. JAMA. 2019;321(5):473–83. https://doi.org/10.1001/jama.2018.22005.
13. Hashimoto DA, Witkowski E, Gao L, et al. Artificial Intelligence in Anesthesiology: Current Techniques, Clinical Applications, and Limitations. Anesthesiology. 2020;132(2):379–94. https://doi.org/10.1097/ALN.0000000000002960.
14. Park Y, Han SH, Byun W, et al. A Real-Time Depth of Anesthesia Monitoring System Based on Deep Neural Network With Large EDO Tolerant EEG Analog Front-End. IEEE Trans Biomed Circuits Syst. 2020;14(4):825–37. https://doi.org/10.1109/TBCAS.2020.2998172.
15. Afshar S, Boostani R, Sanei S. A Combinatorial Deep Learning Structure for Precise Depth of Anesthesia Estimation From EEG Signals. IEEE J Biomed Health Inf. 2021;25(9):3408–15. https://doi.org/10.1109/JBHI.2021.3068481.
16. Connor CW. Open Reimplementation of the BIS Algorithms for Depth of Anesthesia. Anesth Analg. 2022;135(4):855–64. https://doi.org/10.1213/ANE.0000000000006119.
17. Alsafy I, Diykh M. Developing a robust model to predict depth of anesthesia from single channel EEG signal. Phys Eng Sci Med. 2022;45(3):793–808. https://doi.org/10.1007/s13246-022-01145-z.
18. Schmierer T, Li T, Li Y. Harnessing machine learning for EEG signal analysis: Innovations in depth of anaesthesia assessment. Artif Intell Med. 2024;151:102869. https://doi.org/10.1016/j.artmed.2024.102869.
19. Hosseini MP, Hosseini A, Ahi K. A Review on Machine Learning for EEG Signal Processing in Bioengineering. IEEE Rev Biomed Eng. 2021;14:204–18. https://doi.org/10.1109/RBME.2020.2969915.
20. Lee HC, Park Y, Yoon SB, et al. VitalDB, a high-fidelity multi-parameter vital signs database in surgical patients. Sci Data. 2022;9(1):279. https://doi.org/10.1038/s41597-022-01411-5.
21. Sessler DI, Sigl JC, Kelley SD, et al. Hospital stay and mortality are increased in patients having a triple low of low blood pressure, low bispectral index, and low minimum alveolar concentration of volatile anesthesia. Anesthesiology. 2012;116(6):1195–203. https://doi.org/10.1097/ALN.0b013e31825683dc.

22. Carrai R, Martinelli C, Baldanzi F, et al. Is the Patient State Index a reliable parameter as guide to anaesthesiology in cranial neurosurgery? A first intraoperative study and a literature review. Neurophysiol Clin. 2023;53(5):102910. https://doi.org/10.1016/j.neucli.2023.102910.
23. Lee S, Lee HC, Chu YS, et al. Deep learning models for the prediction of intraoperative hypotension. Br J Anaesth. 2021;126(4):808–17. https://doi.org/10.1016/j.bja.2020.12.035.
24. Hwang E, Park HS, Kim HS, et al. Development of a Bispectral index score prediction model based on an interpretable deep learning algorithm. Artif Intell Med. 2023;143:102569. https://doi.org/10.1016/j.artmed.2023.102569.
25. Deschamps A, Ben Abdallah A, Jacobsohn E, et al. Electroencephalography-Guided Anesthesia and Delirium in Older Adults After Cardiac Surgery: The ENGAGES-Canada Randomized Clinical Trial. JAMA. 2024;332(2):112–23. https://doi.org/10.1001/jama.2024.8144.

Publisher's Note