Prediction of Bispectral Index during Target-controlled Infusion of Propofol and Remifentanil

A Deep Learning Approach

Hyung-Chul Lee, M.D., Ho-Geol Ryu, M.D., Ph.D., Eun-Jin Chung, M.D., Chul-Woo Jung, M.D., Ph.D.

ABSTRACT

Background: The discrepancy between predicted effect-site concentration and measured bispectral index is problematic during intravenous anesthesia with target-controlled infusion of propofol and remifentanil. We hypothesized that bispectral index during total intravenous anesthesia would be more accurately predicted by a deep learning approach.

Methods: Long short-term memory and the feed-forward neural network were sequenced to simulate the pharmacokinetic and pharmacodynamic parts of an empirical model, respectively, to predict intraoperative bispectral index during combined use of propofol and remifentanil. Inputs of long short-term memory were infusion histories of propofol and remifentanil, which were retrieved from target-controlled infusion pumps for 1,800 s at 10-s intervals. Inputs of the feed-forward network were the outputs of long short-term memory and demographic data such as age, sex, weight, and height. The final output of the feed-forward network was the bispectral index. The performance of bispectral index prediction was compared between the deep learning model and previously reported response surface model.

Results: The model hyperparameters comprised 8 memory cells in the long short-term memory layer and 16 nodes in the hidden layer of the feed-forward network. The model training and testing were performed with separate data sets of 131 and 100 cases. The concordance correlation coefficient (95% CI) were 0.561 (0.560 to 0.562) in the deep learning model, which was significantly larger than that in the response surface model (0.265 [0.263 to 0.266], *P* < 0.001).

Conclusions: The deep learning model–predicted bispectral index during target-controlled infusion of propofol and remifentanil more accurately compared to the traditional model. The deep learning approach in anesthetic pharmacology seems promising because of its excellent performance and extensibility. (ANESTHESIOLOGY 2017; XXX:00-00)

TOTAL intravenous anesthesia (TIVA) using target-controlled infusion of propofol is widely used for anesthesia and sedation. Target-controlled infusion is performed using a computer-assisted infusion pump that calculates the amount of propofol needed to rapidly achieve the target plasma concentration or the effect-site concentration (Ce) of propofol, which is predicted by three-compartment pharmacokinetic (PK) and pharmacodynamic (PD) models.[1,2] During target-controlled infusion, the propofol Ce is assumed to correlate with the level of hypnosis. However, the bispectral index (BIS), the most widely used measure of hypnosis, does not always agree with the model-driven Ce of propofol, especially during anesthesia induction and recovery.[3]

The discrepancy between the predicted Ce of propofol and the measured hypnotic effect is believed to be related to PK–PD modeling. A PK model with a limited number of compartments and covariates may be insufficient to accurately account for propofol kinetics.[4] The effect measures adopted in PD models are different from those used in clinical practice such as BIS.[5] In addition, the synergistic effect

What We Already Know about This Topic

- The combined effects of propofol and remifentanil on the bispectral index have been characterized using isobole and response surface models
- Deep learning is a kind of machine learning based on a set of algorithms to model high-level abstractions in data using multiple linear and nonlinear transformations

What This Article Tells Us That Is New

- An empirical model was developed from propofol and remifentanil dosing histories and demographic data to predict bispectral index during total intravenous anesthesia target-controlled infusions using a deep learning approach
- The deep learning model had less error in predicting bispectral index during anesthesia induction, maintenance, and recovery periods than the response surface model
- The generalizability of the deep learning model is very dependent on the training data set

of remifentanil on hypnosis is not considered when using the propofol and remifentanil target-controlled infusions. Finally, the traditional PK–PD models built using a small

<zdoi;10.1097/ALN.0000000000001892>

number of study participants in a limited experimental setting lack the capacity for a variety of surgical situations.

The discrepancy between predicted BIS and measured BIS during TIVA could be minimized if the actual clinical data are adequately interpreted.[6,7] However, clinical big data with many noise signals and hidden confounders are not easily analyzed using traditional PK–PD modeling tools. We hypothesized that deep learning, a type of machine learning based on a set of algorithms to model high level abstractions in represented data using multiple linear and nonlinear transformations, could better interpret the dose–response relationship of anesthetic drugs represented in clinical data. Deep learning has recently gained attention as a tool for solving complex issues that were too complicated to analyze using conventional statistical/mathematical methods without overfitting.[8] The main goal of this study is to build an empirical model that predicts changes in BIS during the target-controlled infusions of propofol and remifentanil better than the traditional mechanistic PK–PD model through a deep learning approach.

Materials and Methods

Case Selection

The data were retrieved from our registry that stores basic information and vital signs data of surgical patients in our institution. The registry construction was approved by the institutional review board of Seoul National University Hospital (Seoul, Korea; approval No. H-1408-101-605) and registered at publicly accessible clinical trial registration site (ClinicalTrial.gov, NCT02914444). The retrospective use of registry data for the current study was additionally approved (H-1610-057-798) by the institutional review board. The data from patients who underwent general surgery between June 2016 and September 2016 were assessed. TIVA cases without any anesthetic adjuncts were enrolled after review of registry data.

Practice of TIVA

During the study period, the choice of TIVA and volatile anesthesia was at the discretion of the on-duty anesthesiologist. Patients did not receive premedication. Routine monitoring such as electrocardiogram, noninvasive blood pressure, pulse oximetry, and BIS monitor (BIS VISTA; Medtronic, Ireland) was applied to patients before anesthesia induction. Effect-site target-controlled infusion of propofol and remifentanil was used for induction and maintenance of anesthesia. Use of the Schnider et al.[2] or modified Marsh et al.[1] model for propofol target-controlled infusion was at the discretion of the attending anesthesiologist. The Minto et al.[9] model was used for remifentanil target-controlled infusion. Target Ce values of propofol and remifentanil were set at 3 to 5 µg/ml and 4 to 6 ng/ml, respectively, during anesthesia induction. After loss of response to verbal command, rocuronium (0.6 to 1.2 mg/kg) was administered to facilitate

tracheal intubation. Mechanical ventilation was maintained with a tidal volume of 6 to 8 ml/kg and respiratory rate of 10 to 20/min. Propofol target Ce was titrated to keep the BIS between 40 and 60, and the remifentanil target was adjusted in response to systemic blood pressure during anesthesia maintenance. Rocuronium (0.15 mg/kg) was intermittently administered to maintain neuromuscular block during surgery. At the end of surgery, reversal of neuromuscular block was performed with neostigmine (0.04 mg/kg) and glycopyrrolate (0.01 mg/kg). Propofol and remifentanil administrations were stopped, and the patient was awakened by prodding and calling in a loud tone when BIS increased beyond 60. Extubation was done immediately after notice of sufficient spontaneous ventilation indicated by negative inspiration pressure less than –20 mmHg. Patients recovered from anesthesia were transferred to the postanesthesia care unit or the intensive care unit.

Data Collection

Vital signs data in the registry were recorded with the Vital Recorder program, which was developed by the authors for recording of time-synced data from multiple anesthesia devices including patient monitor, anesthesia machine, BIS monitor, cardiac output monitor, and target-controlled infusion pumps (the program is freely downloadable from the website, https://vitaldb.net; accessed September 1, 2017). Among the collected variables in the registry, demographic data (age, sex, weight, and height), BIS data, and propofol and remifentanil target-controlled infusion data were used. The BIS data were BIS values and signal quality index collected from BIS VISTA at 1-s intervals. Propofol and remifentanil data included cumulative infusion volumes and Ce values of two drugs retrieved from target-controlled infusion pump (Orchestra Base Primea with module DPS; Fresenius Kabi AG, Germany) at an interval of 1 s.

Visual inspection of target-controlled infusion and BIS data was conducted for the period from the start of propofol or remifentanil infusion to the end of BIS measurement. The following cases were excluded: (1) volatile anesthetic was additionally used during TIVA, (2) BIS less than 80 was observed at the start of drug infusion, (3) more than 300 s of data loss was observed, (4) the cumulated infusion volume was not 0 when the first BIS value was recorded, and (5) target-controlled infusion pump was incidentally reset during anesthesia.

Data Preparation

The data of selected cases were randomly allocated to three sets of data: training, validation, and testing data sets. Both the training and validation data sets were used for modeling. The testing data set is the remaining cases not used for modeling but was used to test the performance of the final model.[10]

Input variables of the deep learning model included the PK–PD covariates (age, sex, weight, and height) and

2

propofol and remifentanil infusion histories. The output was BIS value. Data preprocessing such as normalization of input data and smoothing of BIS values was performed before modeling. Age, sex, weight, and height were normalized with mean and SD of the training data set.

Preprocessing of propofol and remifentanil infusion histories data was performed with training and validation data in terms of dose normalization and time range setting. The original infusion history data retrieved from the target-controlled infusion pump were the accumulated infusion volumes that are updated every 10 s according to the target-controlled infusion algorithm by Shafer and Gregg.[11] The 10-s doses of two drugs were calculated from the infused volumes and drug concentrations. As the maximum infusion rate of target-controlled infusion pump was set at 500 ml/h, the maximum 10-s doses of propofol and remifentanil were 27.8 mg and 27.8 µg. Normalization was performed by dividing the 10-s dose by 12 to meet the 10-s dose between 0 and 2.5, which is the maximum input value without saturation in hard-sigmoid function in machine learning. The time range of data was determined considering the context-sensitive decrement time of propofol. If we assume that hypnosis is maintained at propofol Ce of 3 µg/ml during surgery and the patient recovers at 1 µg/ml, the calculated 66% decrement time will be 25 min after 3 h of propofol infusion according to the modified Marsh model.[12] We assumed that at least 1,800 s of data (180 data points) were required to track the cumulative effect of propofol on BIS values during the recovery period. The infusion history of remifentanil was also calculated for 1,800 s to account for the hypnotic synergy of remifentanil.

Selection and smoothing were performed for BIS values of the training data set. BIS values with signal quality index greater than 50 were selected, and then locally weighted scatterplot smoothing (LOWESS) with smoothing parameter 0.03 was applied to the original BIS values to reduce calculation error during training.[13] The unprocessed BIS value was used for validation and testing data sets. Because the BIS is an indexed value between 0 and 100, the value was normalized to 0 to 1 by dividing the BIS value by 100.

Model Building

The empirical model consisted of two neural networks such as long short-term memory and feed-forward neural network, which were sequentially connected to simulate a linked PK–PD model of propofol (fig. 1). A grid search method was used to optimize the hyperparameter, the numbers of memory cells and nodes in the neural networks. A total of 18 combinations (the number of memory cells in long short-term memory = 2, 4, 6, 8, 10, and 12; the number of nodes in the hidden layer = 8, 16, and 32) were tested by applying a fivefold cross-validation technique. Thereafter, the combination of 8 memory cells and 16 nodes representing the smallest validation error was selected as the fixed model architecture to train the deep learning model (fig. 2).

Time series data of propofol were input to 8 memory cells of the long short-term memory. Each memory cell in

the long short-term memory served as a compartment of the PK model and calculated the amount of propofol to vary in the compartments using the propofol infusion history. Remifentanil had the same long short-term memory structure as propofol. Outputs of propofol and remifentanil from long short-term memory layer, as well as four covariates, were the inputs of the feed-forward neural network. The feed-forward neural network was constructed with 1 input layer (8 input nodes for propofol, 8 input nodes for remifentanil, and 4 input nodes for covariates), 1 hidden layer (16 nodes), and 1 output layer (1 output node of BIS value) to simulate two interacting PD models of propofol and remifentanil. The activation functions used were hard sigmoid, hyperbolic tangent, rectified linear unit, and sigmoid functions for gate units of long short-term memory, memory cell of long short-term memory, hidden layer of feed-forward neural network and output layer of feed-forward neural network, respectively.

During training, weights of nodes were calculated with *ADAM* optimizer, a gradient descent optimization algorithm (https://arxiv.org/abs/1412.6980; accessed July 23, 2015). Selection of an optimal model without overfitting was performed using the training and validation data sets. In general, the model is trained to find the optimal weights of the nodes using the training data set during one training epoch. The trained model is then applied to the validation data set to calculate the validation error, which is the absolute difference between the model predicted BIS and the measured BIS. The validation errors gradually decrease as the training epoch repeats; however, the errors paradoxically increase when overfitting occurs. Therefore, model selection is made when the validation error is at its nadir. The final model was applied to the testing data set for external validation. The learning was performed with the authors' own program written in the Python language using the Keras library (https://github.com/fchollet/keras; accessed October 21, 2016). A workstation with Xeon CPU (E3-1230V5; Intel Corporation, USA) and Nvidia GTX 1080 GPU (GeForce GTX 1080 G1; GIGA-BYTE Technology Co., Ltd., Taiwan) was used for the deep learning.

BIS Prediction by Response Surface Model

The model performance was compared between the deep learning model and previously reported response surface model. The response surface model predicted BIS was calculated by the equation of Short *et al.*,[14] which is based on the modified hierarchy model of Bouillon *et al.*[15] that measures combined effect of propofol and remifentanil on hypnosis. The input for the response surface model were propofol Ce and remifentanil Ce, which were calculated by either the Schnider or the modified Marsh model and the Minto model, respectively, and directly recorded from the target-controlled infusion pump.

Statistical Analysis

Descriptive statistics were used to describe demographics, BIS, and drug use in the training, validation, and testing groups. The performance of BIS prediction was compared
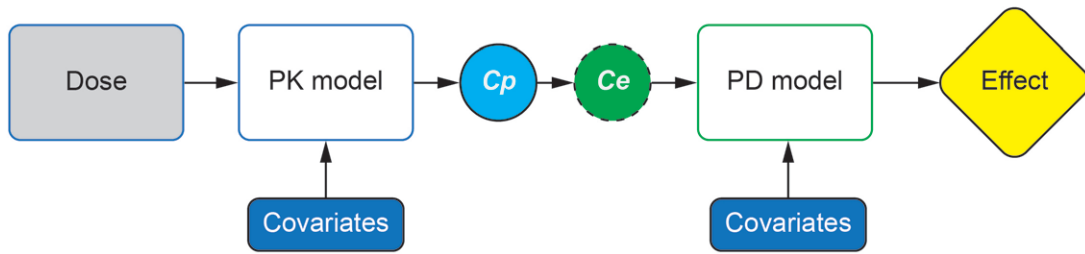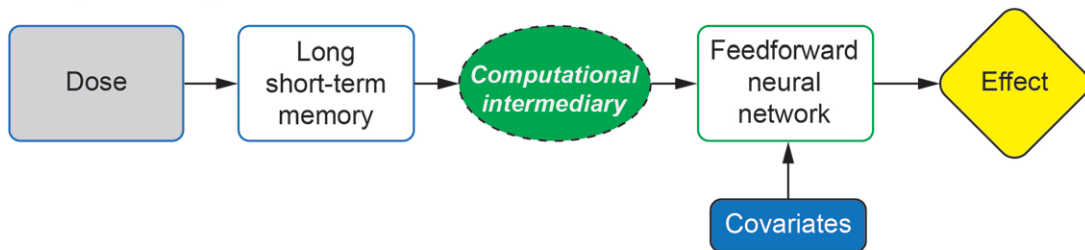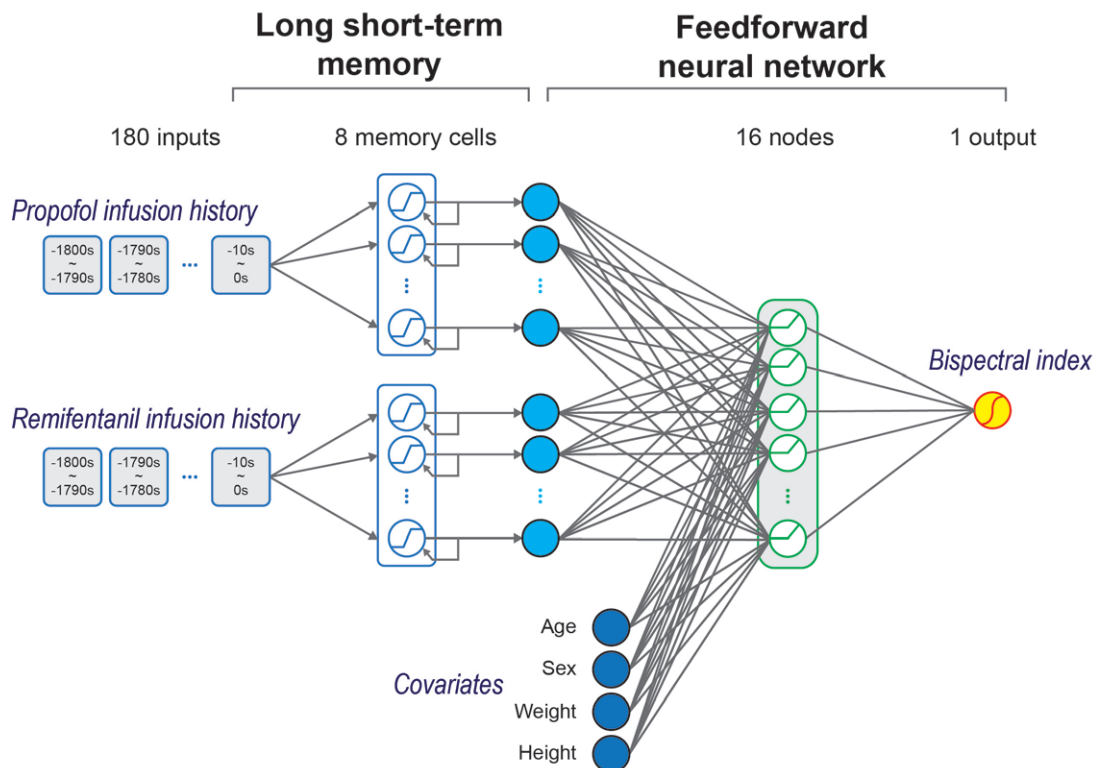
Fig. 1. Schematic representation of the deep learning model. (*A*) The traditional PK (pharmacokinetic)–pharmacodynamic (PD) model requires a PK–PD intermediary like plasma concentration (Cp) or effect-site concentration (Ce). The model has covariates in the PK part to enhance Cp to Ce relationship and in the PD part illustrating Ce to effect relationship. However, the deep learning model is designed to calculate computational intermediaries and has covariates input in the feed-forward neural network (virtual PD) part only. (*B*) Structure of deep learning model. Deep learning model of propofol and remifentanil is built to predict bispectral index during anesthesia. Long short-term memory and feed-forward neural network are sequenced to model the linked PK–PD and interaction of propofol and remifentanil. The inputs are infusion histories of propofol and remifentanil and covariates like age, sex, weight, and height, and the output is the bispectral index.
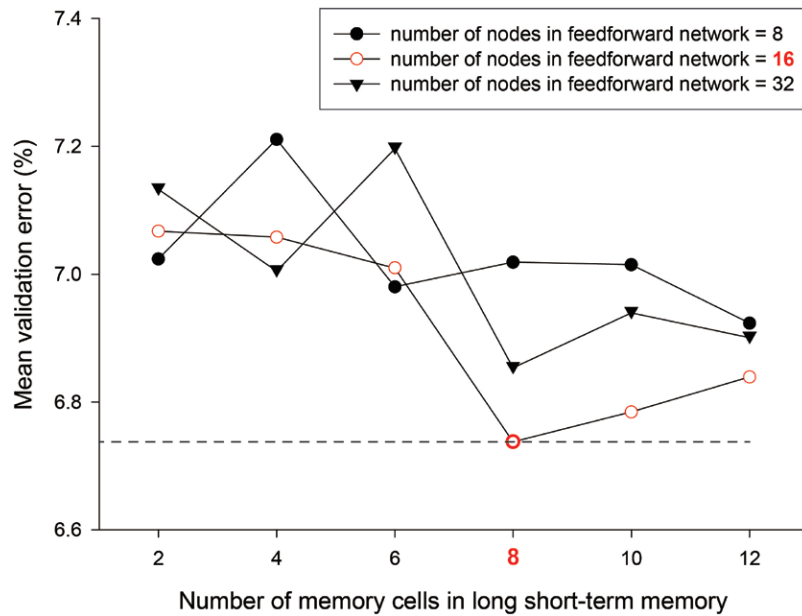
Fig. 2. Results of hyperparameter optimization with fivefold cross-validation. A grid search method was used for hyperparameter optimization. A total of 18 combinations have been tested for the number of memory cells in long short-term memory and the number of nodes in the hidden layer of feed-forward neural network. The combination representing the smallest validation error was selected as the hyperparameter of the model (8 memory cells and 16 nodes).

between the deep learning model and the response surface model using the testing data set. The model fit was summarized with Lin's[16] concordance correlation coefficient, which evaluates the precision and accuracy between two measurements at the same time. In addition, comparisons were performed separately for the induction (from the start of propofol infusion to 10 min later), maintenance, and recovery (from the stop of propofol infusion to the end of anesthesia) periods of anesthesia. The method of performance measurement was used for comparison.[17] Performance error (PE) was calculated as ([measured BIS – predicted BIS]/predicted BIS). Median performance error (MDPE) and median absolute performance error (MDAPE) are the median of PE and median of absolute PE during anesthesia periods, respectively. Root mean square error (RMSE) is the square root of the mean square error and represents the sample SD of the differences between predicted and measured BIS values. MDPE, MDAPE, and RMSE were compared between two models using paired *t* test.

The data are expressed as means ± SD (range) or absolute numbers. Statistical analysis was performed with SPSS 21 (IBM, USA), and *P* < 0.05 was considered significant.

Results

The number of cases in the registry was 1,223 during the study period, and 417 (34.1%) cases were performed with TIVA. After exclusion, 231 cases were selected, and 101, 30, and 100 cases were assigned to the training, validation, and testing groups, respectively. The general characteristics of the three groups are described in table 1.

The total number of data points were 2,038,389 (927,104 for training data set, 219,723 for validation data set, and 89,562 for testing data set). The time required for building a deep learning model was on average about 1.8 h. The final model had 6.23 training error and 6.46 validation error as a BIS value.

The concordance correlation coefficient (95% CI) were 0.561 (0.560 to 0.562) in the deep learning model, which was significantly larger than that in the response surface model (0.265 [0.263 to 0.266]; *P* < 0.001). The Pearson correlation coefficient (measure of precision) and bias correction factor (measure of accuracy) were 0.622 and 0.902 for the deep learning model and 0.378 and 0.701 for the response surface model.

MDPE, MDAPE, and RMSE of deep learning model were significantly smaller than those of response surface model during all periods of anesthesia (*P* < 0.001; table 2). Figure 3 shows the PEs of both the deep learning and the response surface models during the three anesthesia periods in all cases. The PE of the response surface model shows more negative deviation during induction period and more positive deviation with larger range during maintenance and recovery periods compared to those of the deep learning model. Supplemental Digital Content 1 (http://links.lww.com/ALN/B542) shows the individual plots of the measured and model-predicted BIS values in 100 testing group cases. The cases with the smallest and largest prediction errors are selected and shown in figure 4.

The weight matrix of the final model are provided in Excel format in Supplemental Digital Content 2 (http://links.lww.com/ALN/B543). A machine-readable weight matrix in

Table 1. Data Set Characteristics

	Training Data Set	Validation Data Set	Testing Data Set
N	101	30	100
Age (year)	58 ± 14 (18–87)	53 ± 16 (20–85)	55 ± 14 (23–85)
Sex (M/F)	44/57	15/15	39/61
Weight (kg)	61 ± 9 (41–88)	62 ± 10 (47–88)	62 ± 11 (38–92)
Height (cm)	161 ± 8 (140–180)	163 ± 10 (145–180)	162 ± 8 (144–186)
Anesthesia duration (h)	2.6 ± 1.7 (0.6–9.7)	2.6 ± 1.7 (0.5–6.6)	2.5 ± 1.6 (0.7–7.9)
Median BIS	44 ± 6 (31–64)	44 ± 6 (37–60)	44 ± 5 (34–57)
Propofol			
Total dose (g)	1.0 ± 0.6 (0.3–2.9)	1.1 ± 0.8 (0.4–2.9)	1.1 ± 0.7 (0.3–3.2)
Median Ce (µg/ml)	3.0 ± 0.5 (2.0–4.0)	3.1 ± 0.4 (2.0–4.0)	3.1 ± 0.5 (2.0–4.1)
Remifentanil			
Total dose (mg)	1.3 ± 1.3 (0.2–9.5)	1.5 ± 1.4 (0.2–5.1)	1.2 ± 1.1 (0.1–5.2)
Median Ce (ng/ml)	3.8 ± 1.0 (0.9–7.0)	4.0 ± 0.9 (2.6–6.5)	3.8 ± 1.0 (2.0–6.0)

The Data are numbers and means ± SD (range).
BIS = bispectral index; Ce = effect-site concentration.

Table 2. Comparison of Errors between the Deep Learning Model and the Response Surface Model during Three Anesthesia Periods

Anesthesia Period	MDPE (%)		MDAPE (%)		RMSE	
	Deep Learning	Response Surface	Deep Learning	Response Surface	Deep Learning	Response Surface
All	−0.1 ± 11.9	25.0 ± 17.6	13.9 ± 5.3	28.5 ± 14.3	9 ± 2	15 ± 4
Induction	−3.1 ± 14.6	−15.7 ± 15.3	15.2 ± 7.3	22.6 ± 8.9	11 ± 4	19 ± 6
Maintenance	1.1 ± 13.1	28.2 ± 19.1	14.1 ± 6.5	30.2 ± 16.8	9 ± 2	14 ± 5
Recovery	−5.3 ± 14.9	19.9 ± 21.0	14.8 ± 7.8	23.7 ± 16.8	11 ± 6	15 ± 6

The data are means ± SD. All values are significantly smaller in the deep learning model compared with the response surface model ($P < 0.001$).
MDAPE = median absolute performance error; MDPE = median performance error; RMSE = root mean square error.

hierarchical data format (HDF), the program code written in the Python language, and the raw data file in comma-separated values (CSV) format are accessible from the publicly open data repository (https://osf.io/d8gs2; accessed September 1, 2017).

Discussion

In the current study, we successfully developed an empirical model from dosing histories of propofol and remifentanil and demographic data to predict BIS during TIVA through a deep learning approach. The deep learning model had less error in predicting BIS during the induction, maintenance, and recovery periods of anesthesia compared to the traditional mechanistic model.

The use of the artificial neural network to build prediction models is not uncommon in medical research. Steady-state plasma drug concentration was predicted with multilayer feed-forward neural network, which showed less prediction error than the nonlinear mixed effects modeling method.[18] An artificial neural network using 10 common clinical parameters predicted a BIS value of less than 60 after bolus injection of propofol more accurately than clinicians.[19] A simple feed-forward neural network predicted residual neuromuscular block using the degree of spontaneous neuromuscular recovery before reversal and the time elapsed after

reversal.[20] The feed-forward neural network showed better performance in predicting the occurrence of postoperative nausea and vomiting[21] or hypotension after induction of general anesthesia[22] than traditional and statistical diagnostic models. The artificial neural network has also been extensively applied to interpret complicated data such as electroencephalographic (EEG) signals. A feed-forward neural network was trained to build a novel index of anesthesia depth from raw EEG signal, which was comparable to a BIS value with correlation coefficient of 0.94.[23] A recurrent neural network was capable of differentiating three anesthetic states using preprocessed EEG with an accuracy as high as 99.6%.[24] A feed-forward neural network model that combined preprocessed EEG with multiple vital signs to build a new depth of anesthesia index was tested for prediction of anesthesia level, and the index showed less error and higher prediction accuracy than BIS.[25]

The pharmacodynamic drug interaction between propofol and remifentanil has been traditionally explained by isobole and response surface models.[14,15] Recently, Short *et al.*[14] estimated BIS value from the propofol Ce and remifentanil Ce using a response surface model. The predicted BIS showed good accordance with measured BIS with a MDPE of 8 ± 24% and a MDAPE of 25 ± 13%. However, the artificial neural network approach showed better performance
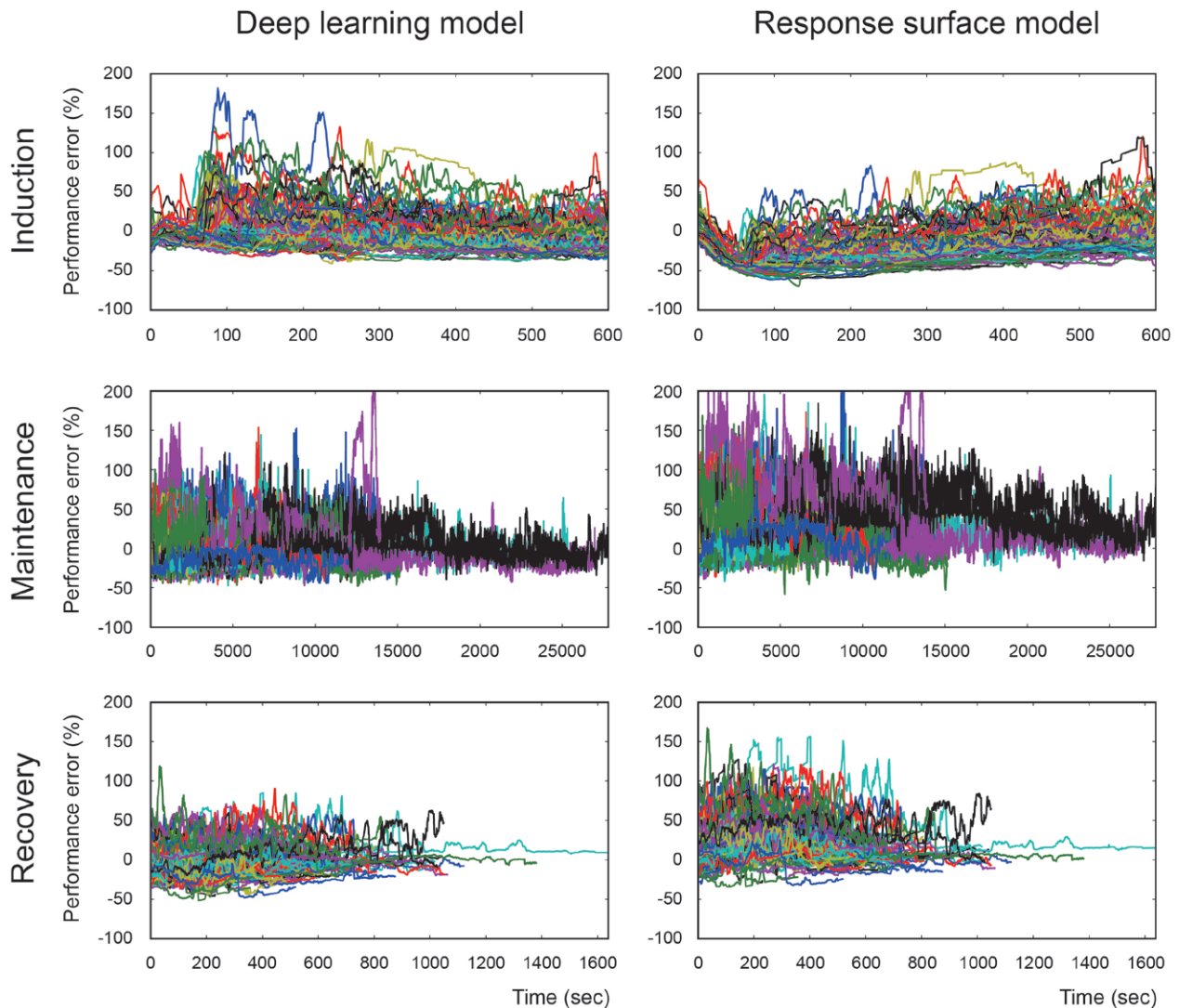
Fig. 3. Performance errors of deep learning model and response surface model during anesthesia induction, maintenance, and recovery periods. Performance errors were significantly smaller in the deep learning model compared to the response surface model during three periods of anesthesia ($P < 0.001$).

than the traditional response surface model in predicting BIS during propofol and remifentanil target-controlled infusions. Gambús *et al.*[26] adopted a fuzzy logic-based artificial neural network (Adaptive Neuro Fuzzy Inference System, ANFIS) to predict BIS from the combination of propofol Ce and remifentanil Ce during sedation-analgesia for endoscopic procedure. MDPE, MDAPE, and RMSE were –5.83, 15.85, and 13.25%, respectively, in the validation group when procedural stimulus was present, which were far less than the errors in the study of Short *et al.*[14] In the current study, the performance of our model far surpasses the response surface model and ANFIS. Table 2 shows that the errors of our model were significantly smaller than those of response surface model during whole anesthesia periods. The errors of our model seem only slightly smaller than those of ANFIS. However, the ANFIS model was built using calculated Ce, which tends to be inaccurate in the dynamic phase,

and was tested only in the steady state. The ANFIS model may be less applicable to induction and recovery periods of anesthesia. The superior predictive power of our model in the dynamic phase may come from the successful use of long short-term memory to process time-series data and appropriate interpretation of pharmacodynamic drug interaction by the use of the feed-forward neural network.

Generally, the empirical model optimized for describing the data well has the disadvantage that there is no biologic basis, and parameter interpretation is difficult. Furthermore, the empirical model may have less predictive power than the mechanistic PK–PD model, because overfitting is likely to occur during complicated modeling using a large number of parameters. To address the weaknesses of empirical modeling, we designed a model architecture similar to the traditional mechanistic PK–PD model and used state-of-the-art computational methods such as deep learning. The
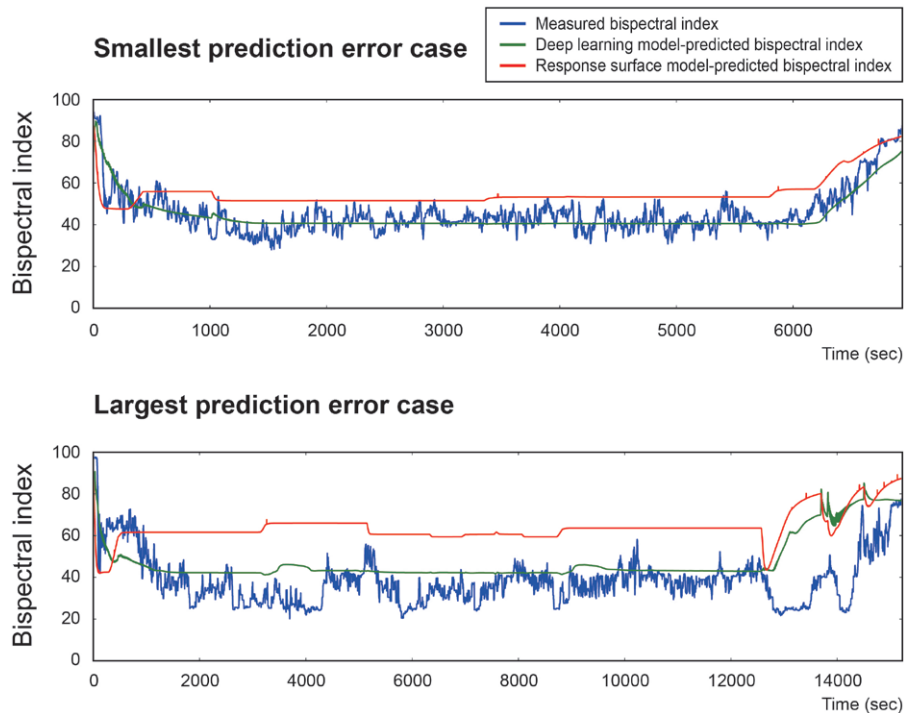
Fig. 4. Cases with the smallest and largest prediction errors predicted by deep learning. Measured and model-predicted bispectral index values are plotted. The root mean square errors of the cases with the smallest and largest prediction errors fitted by the deep learning model were 5.5 and 15.4, respectively.

long short-term memory in this study has a substantial difference as well as theoretical similarity from the traditional PK model. In the 180 consecutive input nodes of the long short-term memory, the previous time node affects the next time node, and the change in the amount of drug over time in the final node is perfectly linear, as in the traditional PK model. However, our long short-term memory model calculates the computational intermediary in the virtual drug compartments without the assumption of pharmacokinetic intermediary like the plasma concentration or Ce, a source of error in traditional PK–PD models. The feed-forward neural network calculates the nonlinear dose-response relationship between the computational intermediary of propofol in the compartments and measured BIS. Contrary to a simple feed-forward neural network that performs a task similar to multiple linear regression analysis, a feed-forward neural network with a hidden layer can approximate any nonlinear functions by increasing the number of nodes in the layer.[27] In addition, the effect of covariates and the combined effect of propofol and remifentanil were also estimated in the hidden layer of the feed-forward neural network. The covariates were entered in the PD part, which is more prone to error than the PK part, because it partially improved overall model performance in our preliminary test.

The main advantage of our deep learning model architecture is its extensibility in various aspects. First, the need for frequent blood sampling and drug concentration analysis, which is a major limitation of traditional PK–PD studies because of cost or ethical concerns, is not required. Because our deep learning model requires only dosing history and measured effect, PK–PD studies in vulnerable subjects may be more readily performed. Second, the effects of various covariates can be easily tested in our deep learning model. The high-dimensionality problem in traditional covariate modeling can be eliminated because our model directly relates covariates with effect rather than PK–PD parameters.[28] Various covariates affecting propofol PK–PD, such as cardiac output and hemorrhage, can be readily added to input nodes of the feed-forward neural network and tested in the deep learning model.[29,30] Third, the combined effect of more than two drugs can be modeled by adding another long short-term memory inputs. Finally, it is a great extensibility option to use rapidly developing hardware, software, and algorithms in the field of machine learning. Furthermore, the clinical application of our study results may be considered. The BIS prediction curve can be drawn on the display of target-controlled infusion pump to guide optimal dosing of two synergistic drugs. The results of deep learning can be applied immediately to current target-controlled infusion devices, because, unlike learning process, low computer performance is enough to calculate the BIS from the inputs and node weights.[31]

However, our deep learning approach has some limitations. First, interpreting the deep learning results is difficult. Useful information from traditional PK–PD studies such as volume and clearance of each compartment, maximum effect, Hill coefficient, and fixed and random effects of covariates cannot be identified in the deep learning model. The weight value of each node, the result of the deep learning, is neither

intuitive nor informative. We set up drug compartments similar to the assumptions of the PK model, but the interaction between the eight drug compartments, which were designed to improve performance, is still difficult to understand. Conflicts of interest in performance enhancement and ease of understanding are difficult to solve in the empirical modeling. Second, the generalizability of the deep learning model is very dependent on the training data set. Nonlinear mixed effects modeling may predict values beyond the range of data used in the experiment through extrapolation, but the deep learning is less capable of estimating the values beyond the learned range. Adding more cases from different patient populations (*i.e.*, obesity), various infusion schemes, and repeated learning from the accumulated data will lead to a more robust model. Third, we cannot claim that the problem of serial correlation between the predicted BIS values is completely resolved by our model. Although the previous BIS value is not the input for the next BIS prediction, and the weight assignment for the time series input is optimized using the long short-term memory, learning could have been biased by unique output patterns during general anesthesia such as rapid decrease and gradual increase in BIS, which are typical during the induction and recovery phases, respectively. This problem may be improved by learning a variety of drug administration schemes and the resulting BIS patterns. Finally, unexpected situations in clinical practice like a change in the infusion rate of carrier fluid or the disconnection of fluid line might have reduced the performance of the deep learning model. Problems related to retrospective design may be solved by prospective studies or by increasing the number of cases.

In conclusion, our study demonstrated that the deep learning model is superior to traditional PK–PD model in predicting BIS during propofol and remifentanil target-controlled infusions in surgical patients. The major advantage of the deep learning approach is its performance and extensibility. We expect that the accumulation of clinical big data will make the deep learning model more powerful and extend its application to a variety of clinical situations in the future.

Research Support

Competing Interests

The authors declare no competing interests.

Correspondence

Address correspondence to Dr. Jung: Seoul National University College of Medicine, Seoul National University Hospital, 101 Daehak-ro, Jongno-gu, Seoul 03080, Korea. jungcwoo@gmail.com. Information on purchasing reprints may be found at www.anesthesiology.org or on the masthead page at the beginning of this issue. ANESTHESIOLOGY's articles are made freely accessible to all readers, for personal use only, 6 months from the cover date of the issue.

References

1. Marsh B, White M, Morton N, Kenny GN: Pharmacokinetic model driven infusion of propofol in children. Br J Anaesth 1991; 67:41–8
2. Schnider TW, Minto CF, Gambus PL, Andresen C, Goodale DB, Shafer SL, Youngs EJ: The influence of method of administration and covariates on the pharmacokinetics of propofol in adult volunteers. ANESTHESIOLOGY 1998; 88:1170–82
3. Coppens M, Van Limmen JG, Schnider T, Wyler B, Bonte S, Dewaele F, Struys MM, Vereecke HE: Study of the time course of the clinical effect of propofol compared with the time course of the predicted effect-site concentration: Performance of three pharmacokinetic-dynamic models. Br J Anaesth 2010; 104:452–8
4. Björnsson MA, Norberg A, Kalman S, Karlsson MO, Simonsson US: A two-compartment effect site model describes the bispectral index after different rates of propofol infusion. J Pharmacokinet Pharmacodyn 2010; 37:243–55
5. Soehle M, Kuech M, Grube M, Wirz S, Kreuer S, Hoeft A, Bruhn J, Ellerkmann RK: Patient state index vs. bispectral index as measures of the electroencephalographic effects of propofol. Br J Anaesth 2010; 105:172–8
6. Jeleazcov C, Ihmsen H, Schmidt J, Ammon C, Schwilden H, Schüttler J, Fechner J: Pharmacodynamic modelling of the bispectral index response to propofol-based anaesthesia during general surgery in children. Br J Anaesth 2008; 100:509–16
7. Wiczling P, Bienert A, Sobczyński P, Hartmann-Sobczyńska R, Bieda K, Marcinkowska A, Malatyńska M, Kaliszan R, Grześkowiak E: Pharmacokinetics and pharmacodynamics of propofol in patients undergoing abdominal aortic surgery. Pharmacol Rep 2012; 64:113–22
8. LeCun Y, Bengio Y, Hinton G: Deep learning. Nature 2015; 521:436–44
9. Minto CF, Schnider TW, Egan TD, Youngs E, Lemmens HJ, Gambus PL, Billard V, Hoke JF, Moore KH, Hermann DJ, Muir KT, Mandema JW, Shafer SL: Influence of age and gender on the pharmacokinetics and pharmacodynamics of remifentanil: I. Model development. ANESTHESIOLOGY 1997; 86:10–23
10. Rumelhart DE, Widrow B, Lehr MA: The basic ideas in neural networks. Commun ACM 1994; 37:87–92
11. Shafer SL, Gregg KM: Algorithms to rapidly achieve and maintain stable drug concentrations at the site of drug effect with a computer-controlled infusion pump. J Pharmacokinet Biopharm 1992; 20:147–69
12. Struys MM, De Smet T, Depoorter B, Versichelen LF, Mortier EP, Dumortier FJ, Shafer SL, Rolly G: Comparison of plasma compartment *versus* two methods for effect compartment–controlled target-controlled infusion for propofol. ANESTHESIOLOGY 2000; 92:399–406
13. Cleveland WS: Robust locally weighted regression and smoothing scatterplots. J Am Stat Assoc 1979; 74:829–36
14. Short TG, Hannam JA, Laurent S, Campbell D, Misur M, Merry AF, Tam YH: Refining target-controlled infusion: An assessment of pharmacodynamic target-controlled infusion of propofol and remifentanil using a response surface model of their combined effects on bispectral index. Anesth Analg 2016; 122:90–7
15. Bouillon TW, Bruhn J, Radulescu L, Andresen C, Shafer TJ, Cohane C, Shafer SL: Pharmacodynamic interaction between propofol and remifentanil regarding hypnosis, tolerance of laryngoscopy, bispectral index, and electroencephalographic approximate entropy. ANESTHESIOLOGY 2004; 100:1353–72
16. Lin LI: A concordance correlation coefficient to evaluate reproducibility. Biometrics 1989; 45:255–68

17. Varvel JR, Donoho DL, Shafer SL: Measuring the predictive performance of computer-controlled infusion pumps. J Pharmacokinet Biopharm 1992; 20:63–94

18. Brier ME, Zurada JM, Aronoff GR: Neural network predicted peak and trough gentamicin concentrations. Pharm Res 1995; 12:406–12

19. Lin C-S, Li Y-C, Mok M s., Wu C-C, Chiu H-W, Lin Y-H: Neural network modeling to predict the hypnotic effect of propofol bolus induction. Amia 2002 2002:450–3

20. Laffey JG, Tobin E, Boylan JF, McShane AJ: Assessment of a simple artificial neural network for predicting residual neuromuscular block. Br J Anaesth 2003; 90:48–52

21. Peng SY, Wu KC, Wang JJ, Chuang JH, Peng SK, Lai YH: Predicting postoperative nausea and vomiting with the application of an artificial neural network. Br J Anaesth 2007; 98:60–5

22. Lin CS, Chang CC, Chiu JS, Lee YW, Lin JA, Mok MS, Chiu HW, Li YC: Application of an artificial neural network to predict postinduction hypotension during general anesthesia. Med Decis Making 2011; 31:308–14

23. Ortolani O, Conti A, Di Filippo A, Adembri C, Moraldi E, Evangelisti A, Maggini M, Roberts SJ: EEG signal processing in anaesthesia: Use of a neural network technique for monitoring depth of anaesthesia. Br J Anaesth 2002; 88:644–8

24. Srinivasan V, Eswaran C, Sriraam N: EEG based automated detection of anesthetic levels using a recurrent artificial neural network. Int J Bioelectromagn 2005; 7:267–70

25. Sadrawi M, Fan SZ, Abbod MF, Jen KK, Shieh JS: Computational depth of anesthesia via multiple vital signs based on artificial neural networks. Biomed Res Int 2015; 2015:536863

26. Gambús PL, Jensen EW, Jospin M, Borrat X, Martínez Pallí G, Fernández-Candil J, Valencia JF, Barba X, Caminal P, Trocóniz IF: Modeling the effect of propofol and remifentanil combinations for sedation-analgesia in endoscopic procedures using an Adaptive Neuro Fuzzy Inference System (ANFIS). Anesth Analg 2011; 112:331–9

27. Hornik K: Approximation capabilities of multilayer feedforward networks. Neural Networks 1991; 4:251–7

28. Verotta D. Covariate modeling in population PK/PD models: An open problem. Adv Pharmacoepidem Drug Safety 2012; S1:006.

29. Upton RN, Ludbrook GL, Grant C, Martinez AM: Cardiac output is a determinant of the initial concentrations of propofol after short-infusion administration. Anesth Analg 1999; 89:545–52

30. Johnson KB, Egan TD, Kern SE, McJames SW, Cluff ML, Pace NL: Influence of hemorrhagic shock followed by crystalloid resuscitation on propofol: A pharmacokinetic and pharmacodynamic analysis. ANESTHESIOLOGY 2004; 101:647–59

31. Beam AL, Kohane IS: Translating artificial intelligence into clinical care. JAMA 2016; 316:2368–9