








# Deep Learning Model for Real-Time Prediction of Intradialytic Hypotension

Hojun Lee <sup>1</sup>, Donghwan Yun <sup>2,3</sup>, Jayeon Yoo,<sup>1</sup> Kiyoon Yoo,<sup>1</sup> Yong Chul Kim <sup>3</sup>, Dong Ki Kim <sup>3</sup>, Kook-Hwan Oh <sup>3</sup>, Kwon Wook Joo <sup>3</sup>, Yon Su Kim,<sup>2,3</sup> Nojun Kwak,<sup>1</sup> and Seung Seok Han <sup>2,3</sup>

## Abstract

**Background and objectives** Intradialytic hypotension has high clinical significance. However, predicting it using conventional statistical models may be difficult because several factors have interactive and complex effects on the risk. Herein, we applied a deep learning model (recurrent neural network) to predict the risk of intradialytic hypotension using a timestamp-bearing dataset.

**Design, setting, participants, & measurements** We obtained 261,647 hemodialysis sessions with 1,600,531 independent timestamps (*i.e.*, time-varying vital signs) and randomly divided them into training (70%), validation (5%), calibration (5%), and testing (20%) sets. Intradialytic hypotension was defined when nadir systolic BP was <90 mm Hg (termed intradialytic hypotension 1) or when a decrease in systolic BP  $\geq 20$  mm Hg and/or a decrease in mean arterial pressure  $\geq 10$  mm Hg on the basis of the initial BPs (termed intradialytic hypotension 2) or prediction time BPs (termed intradialytic hypotension 3) occurred within 1 hour. The area under the receiver operating characteristic curves, the area under the precision-recall curves, and F1 scores obtained using the recurrent neural network model were compared with those obtained using multilayer perceptron, Light Gradient Boosting Machine, and logistic regression models.

**Results** The recurrent neural network model for predicting intradialytic hypotension 1 achieved an area under the receiver operating characteristic curve of 0.94 (95% confidence intervals, 0.94 to 0.94), which was higher than those obtained using the other models ( $P < 0.001$ ). The recurrent neural network model for predicting intradialytic hypotension 2 and intradialytic hypotension 3 achieved area under the receiver operating characteristic curves of 0.87 (interquartile range, 0.87–0.87) and 0.79 (interquartile range, 0.79–0.79), respectively, which were also higher than those obtained using the other models ( $P \leq 0.001$ ). The area under the precision-recall curve and F1 score were higher using the recurrent neural network model than they were using the other models. The recurrent neural network models for intradialytic hypotension were highly calibrated.

**Conclusions** Our deep learning model can be used to predict the real-time risk of intradialytic hypotension.

CJASN 16: ●●●–●●●, 2021. doi: <https://doi.org/10.2215/CJN.09280620>

## Introduction

Intradialytic hypotension (IDH) is an important complication during hemodialysis because it is associated with subsequent cardiovascular morbidity and mortality (1,2). Despite the heterogeneity in defining IDH, it occurs in more than 10% of hemodialysis sessions (3) and up to 50% of elderly patients (4). Several risk factors have been described, including female sex, high weight gain, high ultrafiltration rate, and comorbidities (5–8), but these factors are dependent on the characteristics of each study, and there may be unknown factors related with IDH that are still to be discovered. Cardiovascular and uremic mechanisms underlie IDH (9,10), but other mechanisms may also be associated with it. This casual diversity makes predicting IDH difficult, although a lot of resources and studies on IDH are available.

Artificial intelligence models have changed the paradigm of clinical decision making from diagnosis to treatment (11). Among these models, deep learning (12), which is a subfield of machine learning, stands out

because it can improve the overall performance of health care, particularly in diagnostic imaging and pathologic processes, and in the synthetic judgment of big data flow (13). Deep learning can learn and characterize flow from a variety of data types, and can thus develop a model from time-varying sequential inputs. This complex function may allow the model to predict the risk of IDH in real time. Herein, we addressed whether a deep learning model could be used to predict the risk of IDH within 1 hour at any time point using the data from 261,647 hemodialysis sessions with over 1,600,000 independent timestamps (*i.e.*, time-varying vital signs), and compared its performance with the performances of other machine learning and logistic regression models.

## Materials and Methods

### Data Source and Study Population

The institutional review board of Seoul National University Hospital approved the study design

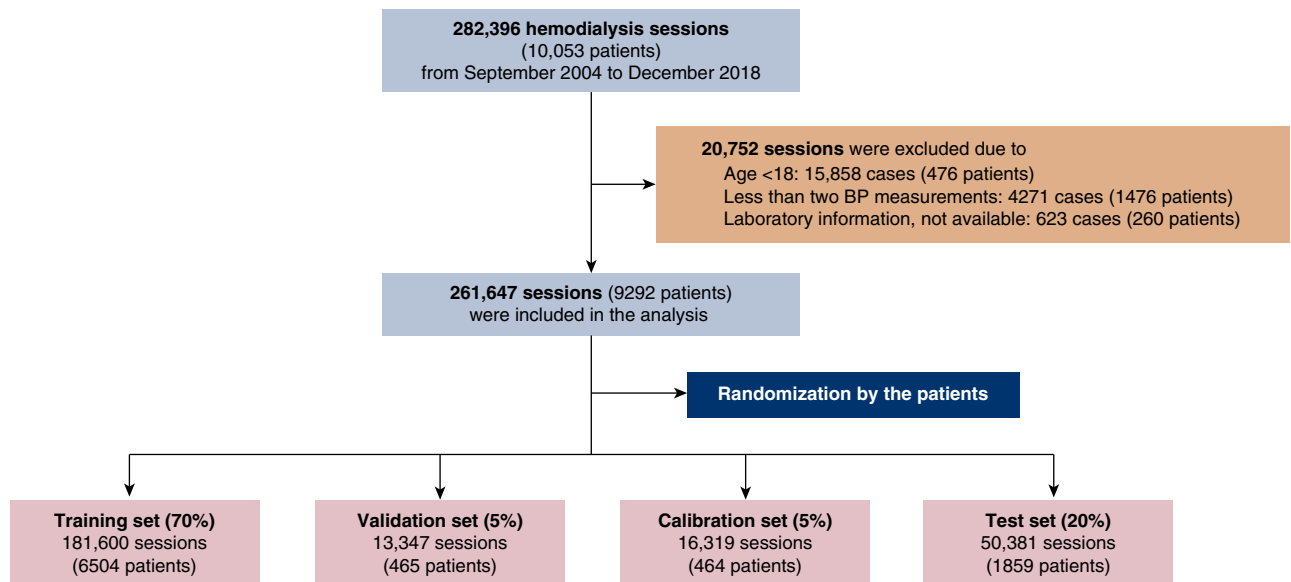
<sup>1</sup>Department of Intelligence and Information, Seoul National University, Seoul, Korea

<sup>2</sup>Department of Biomedical Sciences, Seoul National University College of Medicine, Seoul, Korea

<sup>3</sup>Department of Internal Medicine, Seoul National University College of Medicine, Seoul, Korea

### Correspondence:

Dr. Seung Seok Han, Department of Internal Medicine, Seoul National University College of Medicine, 103 Daehak-ro, Jongno-gu, Seoul 03080, Korea, or Prof. Nojun Kwak, Department of Transdisciplinary Studies, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Korea. Email: hansway80@gmail.com or nojunk@snu.ac.kr



**Figure 1.** | The flow chart of study data retrieval and splitting.

(no. H-1904–143–1029), which was conducted in accordance with the principles of the Declaration of Helsinki. Electronic health records have been available since October 2004. Accordingly, 282,396 sessions from 10,053 patients who underwent hemodialysis at Seoul National University Hospital were retrospectively reviewed from October 2004 to December 2018. Hemodialysis sessions from patients aged <18 years ( $n=15,859$ ) and those with insufficient clinical information ( $n=4267$ ) and no laboratory findings ( $n=623$ ) were excluded. A total of 261,647 sessions (9292 patients) remained for analysis, and the data comparing the characteristics of the hemodialysis sessions that were included and excluded from the analyses are presented in Supplemental Table 1.

We randomized the patients into a training set (70%) to develop the model, a validation set (5%) to validate the model, a calibration set (5%) to calibrate the model, and a testing set (20%) to test the performance of the model, wherein the patient-session ratio and IDH incidence were similarly distributed between sets. The flow chart of study data retrieval and splitting is shown in Figure 1.

#### Timestamps of the Hemodialysis Session

Timestamps were on the basis of the BP measurements. When there were no specific episodes or complications during hemodialysis, the vital signs were monitored every 1 hour (left image in Figure 2A). However, if vital signs were unstable, they were monitored more frequently and the hemodialysis settings were adjusted (right image in Figure 2A). The total number of independent timestamps was 1,600,531 out of 261,647 hemodialysis sessions.

#### Study Outcomes

Three definitions were used for IDH. IDH-1 was defined when intradialytic nadir systolic BP was <90 mm Hg within 1 hour (2). When IDH was defined as a decrease in systolic BP of  $\geq 20$  mm Hg and/or a decrease in mean

arterial pressure of  $\geq 10$  mm Hg (14,15) within 1 hour, the reference BPs were determined at initial (IDH-2) or prediction (IDH-3) time point. Each IDH criterion was treated as a separate binary outcome (*i.e.*, 0 or 1). Mean arterial pressure was calculated as  $([2 \times \text{diastolic BP}] + \text{systolic BP})/3$ .

#### Study Variables

The dataset contained information at initial and any time points during hemodialysis, such as age, sex, vital signs, comorbidities, medications, and laboratory findings. A list of features is available in Supplemental Material.

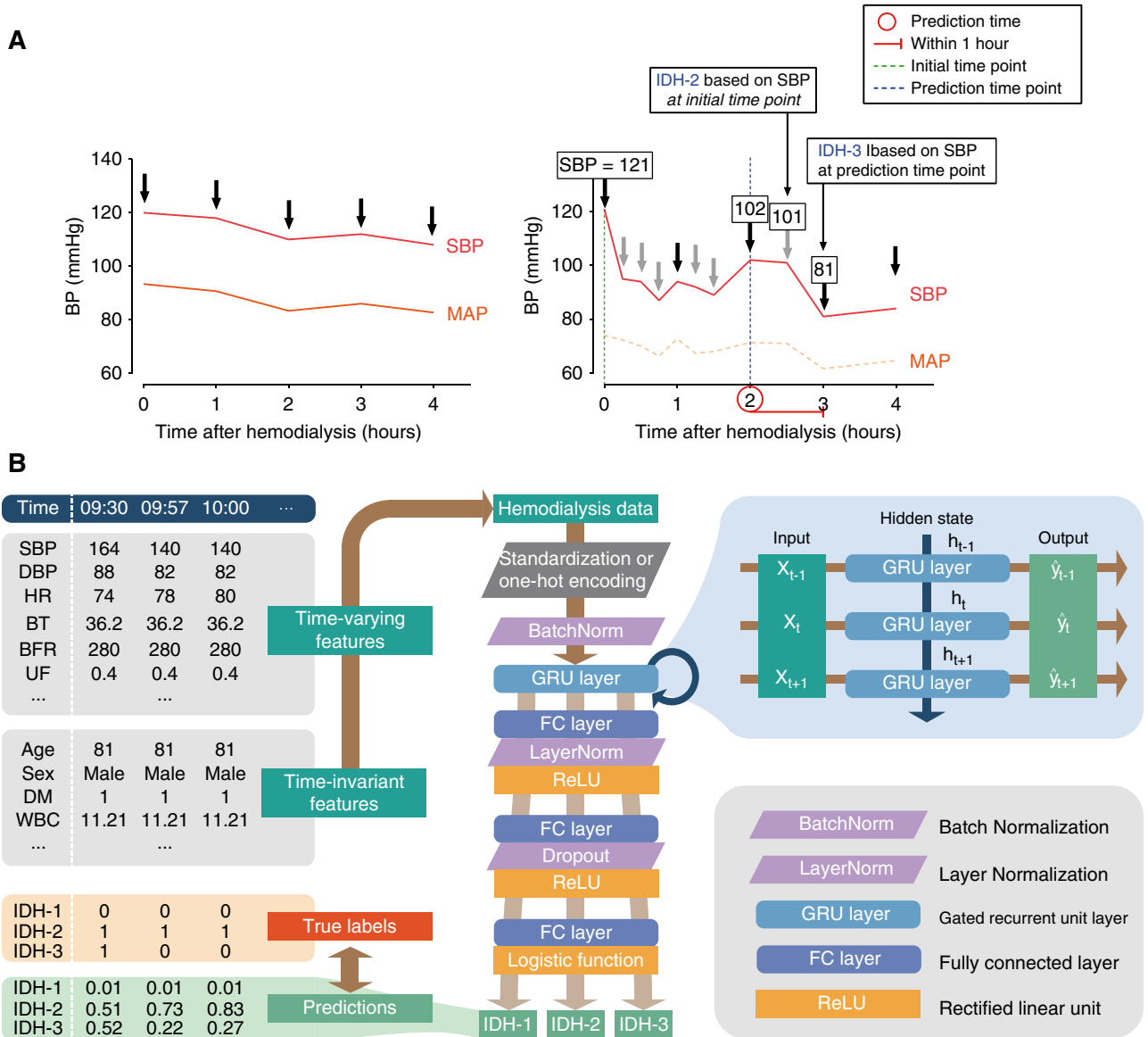
#### Model Development

The recurrent neural network model is suitable for time series datasets (16). We used the gated recurrent unit in the recurrent neural network. The gated recurrent unit combines information on current time ( $t$ ) with previous time ( $t-1$ ). The baseline cell of the gated recurrent unit has update and reset gates, and the results of these gates reflect previous information. A single recurrent neural network model was developed to predict three definitions of IDH. The recurrent neural network architecture of the study is given in Figure 2B.

We also developed multilayer perceptron model, Light Gradient Boosting Machine model, and the logistic regression model to compare their performances with the recurrent neural network model. These models could not handle all of the temporal information, and thus, the data at only two timestamps (prediction time point and its previous one) were used. The detailed methods were presented in Supplemental Material and we summarized the characteristics of the models in Supplemental Table 2.

#### Model Evaluation

The performances of the models were evaluated using three criteria: the area under the receiver operating



**Figure 2. | Development of recurrent neural network model.** (A) Illustrative example of sessions with stable vital signs (left) and intradialytic hypotension (IDH) (right). The risk of IDH within 1 hour at a certain time point (red circle) was calculated. When IDH was defined as a decrease in systolic BP  $\geq 20$  mm Hg and/or a decrease in mean arterial pressure  $\geq 10$  mm Hg, the reference BPs were determined at initial (IDH-2) or prediction (IDH-3) time point. Black and gray arrows indicate routine and additional monitoring of BPs, respectively. (B) Architecture of the proposed recurrent neural network model. Briefly, time-varying and time-invariant features were embedded in the cells with multilayer perceptron. The deepening effect was obtained by inserting fully connected layers between cells, and the learning was stabilized using the layer normalization. IDH-1, intradialytic hypotension defined as nadir systolic BP  $< 90$  mm Hg; IDH-2, intradialytic hypotension defined as decrease in systolic BP  $\geq 20$  mm Hg and/or decrease in mean arterial pressure  $\geq 10$  mm Hg on the basis of BP at initial time point; IDH-3, intradialytic hypotension defined as decrease in systolic BP  $\geq 20$  mm Hg and/or decrease in mean arterial pressure  $\geq 10$  mm Hg on the basis of BP at prediction time point. SBP, systolic BP; DBP, diastolic BP; HR, heart rate; BT, body temperature; BFR, blood flow rate; UF, ultrafiltration; DM, diabetes mellitus; WBC, white blood cell count; BatchNorm, batch normalization; LayerNorm, layer normalization; GRU, gated recurrent unit; FC, fully connected; ReLU, rectified linear unit.

characteristic curve, the area under the precision-recall curve, and F1 score, as follows.

$$\text{Recall} : \frac{TP}{TP + FN}$$

$$\text{Precision} : \frac{TP}{TP + FP}$$

$$\text{F1 score} : \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

TP, true positive; FP, false positive; and FN, false negative.

The curves were plotted by varying the thresholds, and the area under the curves were compared using the DeLong test. Interpretation of the area under the receiver operating characteristic curves and the area under the precision-recall curves was achieved by comparing the values among the

**Table 1. Baseline characteristics of the hemodialysis sessions analyzed in this study**

Variables	Total (n=261,647) <sup>a</sup>
Age (yr)	62±15
Male, n (%)	151,531 (58)
<b>Hemodialysis type, n (%)</b>	
Hemodialysis	243,155 (92)
Hemodiafiltration	17,123 (7)
Others	1369 (0.5)
The number of sessions per wk (times)	3±0.5
The time per session (h)	4.0 (3.8–4.0) <sup>b</sup>
Incident hemodialysis, n (%)	27,230 (10)
<b>Vascular access, n (%)</b>	
Arteriovenous fistula	183,330 (70)
Arteriovenous graft	14,110 (5)
Subcutaneously tunneled catheter	46,994 (18)
Temporary internal jugular venous catheterization	12,848 (5)
Others	4345 (2)
Predialysis weight (kg)	58±12
Setting of ultrafiltration (L)	2.0 (1.2–2.7)
<b>Use of anticoagulant, n (%)</b>	
Heparin	172,023 (66)
Nafamostat mesilate	43,198 (16)
None	46,694 (18)
Setting of blood flow rate (ml/min)	250 (230–280)
Predialysis systolic BP (mm Hg)	139 (121–155)
Predialysis diastolic BP (mm Hg)	73 (65–82)
Predialysis mean arterial pressure (mm Hg)	95 (85–106)
Predialysis heart rate (/min)	75 (66–86)
<b>Blood findings</b>	
Hemoglobin (g/dl)	10.4 (9.5–11.2)
Albumin (g/dl)	3.7 (3.1–4.0)
Calcium (mg/dl)	8.9 (8.3–9.4)
Phosphate (mg/dl)	4.4 (3.4–5.4)
Sodium (mmol/L)	137 (135–139)
Potassium (mmol/L)	4.6 (4.1–5.2)
<b>Dialysate findings</b>	
Dialysate sodium (mmol/L)	137±1.3
Dialysate potassium (mmol/L)	2.5±0.3
Dialysate calcium (mmol/L)	1.5±0.1
Dialysate bicarbonate (mmol/L)	33.8±1.0
Dialysate temperature (°C)	36.7±0.5

<sup>a</sup>Total number of patients was 9292.  
<sup>b</sup>The mean value was 3.7±0.6 h per session.

models. All *P* values were two sided, and values <0.05 were considered significant. In the decision curve analysis, models were converted to a logistic regression using probability theory. Calibrations were performed using the Platt scaling method (17). All of the performance indices were measured using the testing set.

### Feature Set-Ablation Analysis

To examine the effect of each feature set on the model performance, we performed a feature set-ablation analysis, which measured changes in performance by eliminating one set from the model inputs. The features were categorized into six sets: A (age and sex), B (hemodialysis-related features), C (vital signs), D (comorbidities), E (laboratory findings), and F (medications). Variables included in the sets are summarized in Supplemental Table 3.

### Feature Ranking Analysis

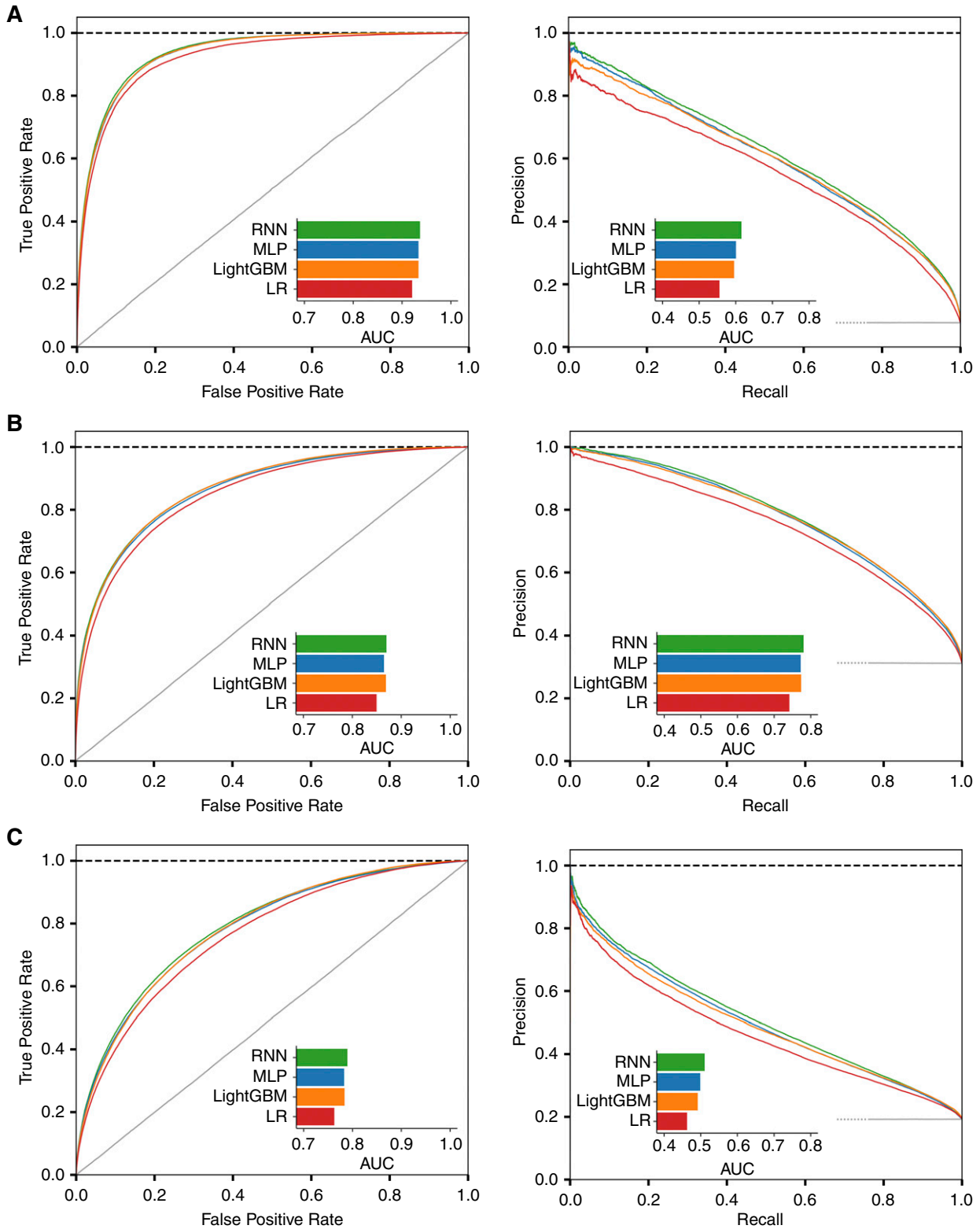
We also performed a feature ranking analysis according to the method used in a previous study (18). Briefly, each

feature was dropped one by one from the testing dataset when the model was developed, and the prediction results were compared with the reference prediction results with all of the features. The detailed explanations and mathematical equations are shown in the Supplemental Material.

## Results

### Study Cases

The mean age of the hemodialysis sessions in this study was 62±15 years, and over 90% of the sessions were for conventional hemodialysis. The median value of the number of BP measurements was 5 (interquartile range [IQR], 5–9), and the median time interval was 30 (IQR, 20–59) minutes. Nadir systolic BP <90 mm Hg (IDH-1) occurred in 11% (*n*=27,971) of the sessions (3827 patients), whereas decreases in systolic BP of ≥20 mm Hg or decreases in mean arterial pressure of ≥10 mm Hg (IDH-2 and IDH-3) occurred in 49% (*n*=128,404 in 8216 patients) and 51% (*n*=134,434 in 8415 patients) during hemodialysis, respectively. We found that the rate of IDH did not oscillate



**Figure 3. | Receiver operating characteristic (left) and precision-recall (right) curves for prediction of intradialytic hypotension (IDH). (A)** Prediction of IDH-1. (B) Prediction of IDH-2. (C) Prediction of IDH-3. RNN, recurrent neural network; MLP, multilayer perceptron; LightGBM, Light Gradient Boosting Machine; LR, logistic regression; AUC, area under the curve.

**Table 2. Area under the curves for predicting intradialytic hypotension for the four models used in this study**

Outcome	Models	Area Under the Receiver Operating Characteristic Curve (95% Confidence Interval)	P value <sup>a</sup>	Area Under the Precision-Recall Curve (95% Confidence Interval)
IDH-1	RNN	0.94 (0.94 to 0.94)		0.62 (0.61 to 0.62)
	MLP	0.93 (0.93 to 0.94)	<0.001	0.60 (0.60 to 0.60)
	LightGBM	0.93 (0.93 to 0.94)	<0.001	0.50 (0.59 to 0.60)
	LR	0.92 (0.92 to 0.92)	<0.001	0.56 (0.55 to 0.55)
IDH-2	RNN	0.87 (0.87 to 0.87)		0.78 (0.78 to 0.78)
	MLP	0.87 (0.86 to 0.87)	<0.001	0.77 (0.77 to 0.78)
	LightGBM	0.87 (0.87 to 0.87)	0.001	0.77 (0.77 to 0.78)
	LR	0.85 (0.85 to 0.85)	<0.001	0.74 (0.74 to 0.74)
IDH-3	RNN	0.79 (0.79 to 0.79)		0.51 (0.51 to 0.51)
	MLP	0.78 (0.78 to 0.79)	<0.001	0.50 (0.50 to 0.50)
	LightGBM	0.78 (0.78 to 0.79)	<0.001	0.49 (0.49 to 0.49)
	LR	0.76 (0.76 to 0.77)	<0.001	0.46 (0.46 to 0.47)

IDH-1, intradialytic hypotension defined as nadir systolic BP <90 mm Hg; IDH-2, intradialytic hypotension defined as decrease in systolic BP  $\geq$ 20 mm Hg and/or decrease in mean arterial pressure  $\geq$ 10 mm Hg on the basis of BP at initial time point; IDH-3, intradialytic hypotension defined as decrease in systolic BP  $\geq$ 20 mm Hg and/or decrease in mean arterial pressure  $\geq$ 10 mm Hg on the basis of BP at prediction time point. RNN, recurrent neural network; MLP, multilayer perceptron; LightGBM, Light Gradient Boosting Machine; LR, logistic regression.  
<sup>a</sup>Compared with the receiver operating characteristic curve of RNN model by the Delong test.

except at the early time points (Supplemental Figure 1). Other baseline characteristics of the hemodialysis sessions are given in Table 1.

### Testing Results

Among 304,720 timestamps of the test set, IDH-1, IDH-2, and IDH-3 were found in 8%, 31%, and 19%, respectively. The predicted results for different threshold values (0.1, 0.3, 0.5, 0.7, and 0.9) are presented in confusion matrices (Supplemental Figure 2). Although the distributions of normal (*i.e.*, no IDH) and abnormal (*i.e.*, IDH) labels in IDH-1 and IDH-3 seemed to be imbalanced because of the low occurrences of IDH-1 (8%) and IDH-3 (19%), the confusion matrices displayed well-calibrated distributions according to the thresholds. These results indicated that no critical classification errors existed. Using the testing dataset, the area under the receiver operating characteristic curve for predicting IDH-1 was highest in the recurrent neural network model (0.94, 95% confidence interval [95% CI], 0.94 to 0.94), followed by Light Gradient Boosting machine (0.93; 95% CI, 0.93 to 0.94), multilayer perceptron (0.93; 95% CI, 0.93 to 0.94), and logistic regression (0.92; 95% CI, 0.92 to 0.92) (all *P* values were <0.001) (Figure 3A). The area under the precision-recall curves for predicting IDH-1 were 0.62 (IQR, 0.61–0.62) in recurrent neural network, 0.60 (IQR, 0.60–0.60) in multilayer perceptron, 0.60 (IQR, 0.60–0.60) in Light Gradient Boosting Machine, and 0.56 (IQR, 0.55–0.56) in logistic regression. The area under the receiver operating characteristic curves and the area under the precision-recall curves for predicting IDH-2 and IDH-3 were also highest in the recurrent neural network model compared with their values in the other machine learning and logistic regression models (Figure 3, B and C). The area under the curve results for all four models are summarized in Table 2. Similarly, the F1 scores in the recurrent neural network model were higher than those in the other models, irrespective of threshold levels (Supplemental Table 4).

The recall and precision rates of the recurrent neural network model were evaluated according to the hemodialysis times (Supplemental Figure 3). When the predicted probability exceeded a specified operating threshold, the prediction was considered positive. The predictive recall and precision rates of IDH were similar throughout the hemodialysis time, although there was a trade-off between precisions and recalls after 4 hours.

The decision curve analyses showed that the ratio between cost and benefit was different depending on the definition of IDH (Supplemental Figure 4). The net benefit of the recurrent neural network model was higher than that of other models, and the logistic regression model had the poorest benefit compared with the other models.

### Calibration of Models

After calibrating models using the calibration dataset, the results were plotted against the percentage of positive labels in the buckets using the testing dataset (Supplemental

**Table 3. Expected calibration errors after calibrating the four models used in this study**

Outcomes	Models			
	RNN (%)	MLP (%)	LightGBM (%)	LR (%)
IDH-1	0.20	0.40	0.32	0.36
IDH-2	0.66	1.45	0.85	1.66
IDH-3	0.43	0.91	0.93	0.51

RNN, recurrent neural network; MLP, multilayer perceptron; LightGBM, Light Gradient Boosting Machine; LR, logistic regression; IDH-1, intradialytic hypotension defined as nadir systolic BP <90 mm Hg; IDH-2, intradialytic hypotension defined as decrease in systolic BP  $\geq$ 20 mm Hg and/or decrease in mean arterial pressure  $\geq$ 10 mm Hg on the basis of BP at initial time point; IDH-3, intradialytic hypotension defined as decrease in systolic BP  $\geq$ 20 mm Hg and/or decrease in mean arterial pressure  $\geq$ 10 mm Hg on the basis of BP at prediction time point.

**Table 4. Performance of the recurrent neural network model after ablation of the feature set**

Outcome	Models	Area Under the Receiver Operating Characteristic Curve (95% Confidence Interval)	$\Delta$ Area Under the Receiver Operating Characteristic Curve	Area Under the Precision-Recall Curve (95% Confidence Interval)	$\Delta$ Area Under the Precision-Recall Curve
IDH-1	Remove set A	0.94 (0.94 to 0.94)	0	0.62 (0.61 to 0.62)	0
	Remove set B	0.93 (0.93 to 0.93)	0.01	0.60 (0.59 to 0.60)	0.02
	Remove set C	0.89 (0.89 to 0.89)	0.05	0.44 (0.44 to 0.44)	0.18
	Remove set D	0.93 (0.93 to 0.93)	0.01	0.60 (0.60 to 0.61)	0.01
	Remove set E	0.94 (0.94 to 0.94)	0.001	0.61 (0.61 to 0.62)	0.002
	Remove set F	0.94 (0.94 to 0.94)	0	0.62 (0.61 to 0.62)	0
IDH-2	Remove set A	0.87 (0.87 to 0.87)	0.001	0.78 (0.78 to 0.78)	0.002
	Remove set B	0.86 (0.86 to 0.86)	0.01	0.76 (0.76 to 0.76)	0.02
	Remove set C	0.73 (0.73 to 0.74)	0.14	0.57 (0.56 to 0.57)	0.22
	Remove set D	0.86 (0.86 to 0.86)	0.01	0.77 (0.77 to 0.78)	0.01
	Remove set E	0.87 (0.87 to 0.87)	0.001	0.78 (0.78 to 0.78)	0.001
	Remove set F	0.87 (0.87 to 0.87)	-0.001	0.78 (0.78 to 0.78)	0
IDH-3	Remove set A	0.79 (0.79 to 0.79)	0.001	0.51 (0.51 to 0.51)	0.003
	Remove set B	0.77 (0.77 to 0.78)	0.02	0.48 (0.48 to 0.48)	0.03
	Remove set C	0.70 (0.70 to 0.71)	0.09	0.38 (0.37 to 0.38)	0.14
	Remove set D	0.78 (0.77 to 0.78)	0.01	0.49 (0.49 to 0.49)	0.02
	Remove set E	0.79 (0.79 to 0.79)	0.002	0.51 (0.51 to 0.51)	0.003
	Remove set F	0.79 (0.79 to 0.79)	0	0.51 (0.51 to 0.51)	0.002

IDH-1, intradialytic hypotension defined as nadir systolic BP <90 mm Hg; IDH-2, intradialytic hypotension defined as decrease in systolic BP  $\geq$ 20 mm Hg and/or decrease in mean arterial pressure  $\geq$ 10 mm Hg on the basis of BP at initial time point; IDH-3, intradialytic hypotension defined as decrease in systolic BP  $\geq$ 20 mm Hg and/or decrease in mean arterial pressure  $\geq$ 10 mm Hg on the basis of BP at prediction time point. Each set contains features as follows: A, age and sex; B, hemodialysis-related features; C, vital signs; D, comorbidities; E, laboratory findings; and F, medications.

Figure 5). The corresponding expected calibration errors of the recurrent neural network model were lower than those of the other models, except for the IDH-1 prediction (Table 3).

### Feature-Set Ablation Analysis

Subsequently, the model performance without each feature set was measured to ascertain the contribution of features in developing the recurrent neural network model (Table 4). When set C (vital signs) was ablated in the model, the area under the receiver operating characteristic curve and the area under the precision-recall curve decreased the most from the baseline model with all features. The decrease in F1 scores was the largest when set C was ablated in the model (Supplemental Table 5). The degree of performance degradation was the next worst when set B (hemodialysis-related features) was ablated. These data suggest that time-varying vital signs contributed the most to the construction of the model, followed by hemodialysis settings.

### Feature Ranking Analysis

To estimate the contribution degree of each feature in predicting the IDH risk, we also performed a feature ranking analysis. Accordingly, the features that contributed most to the model performance were time-varying vital signs and hemodialysis settings (Figure 4).

### Subgroup Analysis

We performed the subgroup analysis for the incident and prevalent hemodialysis, and the model performed well for both the incident and prevalent sessions

(Supplemental Table 6). We further performed a subgroup analysis according to the tertiles of the ultrafiltration volumes (first tertile,  $\leq$ 1.1 kg per session; second tertile, 1.2–2.1 kg per session; and third tertile,  $\geq$ 2.2 kg per session), and we found that the model performance was similar between high and low ultrafiltration settings (Supplemental Table 7).

We also evaluated the model performances in the sessions with or without the history of IDH. The history of IDHs within 7 days were 16%, 65%, and 67% using the IDH-1, IDH-2, and IDH-3 criteria, respectively. Among the sessions with no previous IDH history within 7 days, the prevalence of IDH-1, IDH-2, and IDH-3 was 8%, 40%, and 48%, respectively. Irrespective of the IDH histories, our model successfully predicted the risk of IDH (Supplemental Table 8).

### Sensitivity Analysis

For sensitivity analysis, the recurrent neural network model was constructed after randomizing sessions, instead of patients, into training (70%), validation (5%), calibration (5%), and testing (20%) sets. This was performed to balance the characteristics of the patients and number of hemodialysis sessions per patient. Despite using this randomization method, the recurrent neural network model predicted IDH in a similar manner as before (Supplemental Table 9).

When the echocardiographic findings including left ventricular ejection fraction, left ventricular end-diastolic dimension, left ventricular end-systolic dimension, inter-ventricular septum thickness, and left ventricular mass were used as an additional input feature of the recurrent neural network model ( $n=227,640$ ; 87% sessions), the IDH predictability did not improve significantly (Supplemental

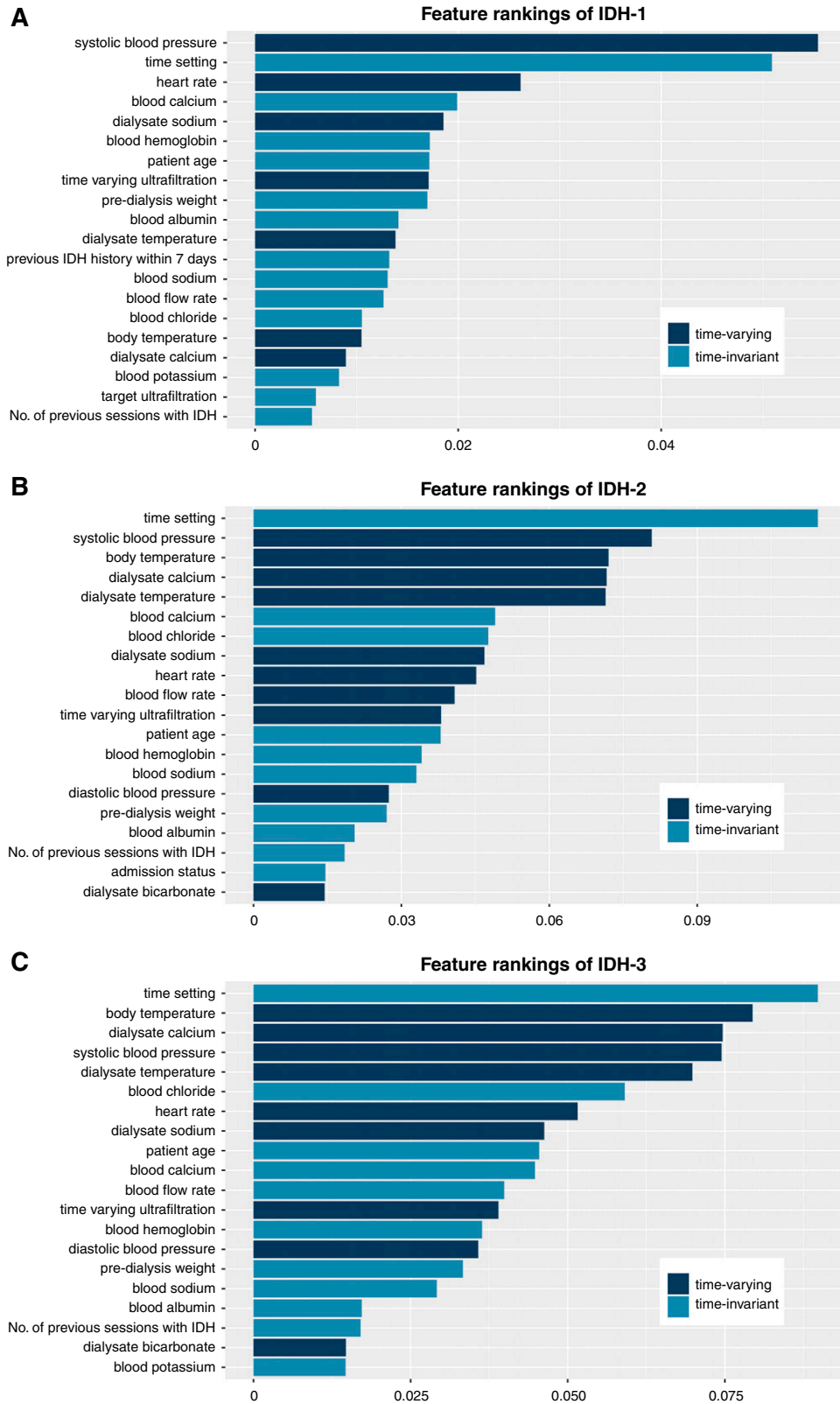


Figure 4. | Feature rankings for the model predicting intradialytic hypotension (IDH). (A) IDH-1; (B) IDH-2; (C) IDH-3.



Table 10). This suggests the current routinely monitored features are sufficient to predict real-time IDH.

We performed an additional sensitivity analysis with other definitions of IDH: systolic BP <100 mm Hg within 1 hour (termed IDH-4), and a decrease in systolic BP of  $\geq 20$  mm Hg or a decrease in mean arterial pressure of  $\geq 10$  mm Hg from initial BP within 1 hour (termed IDH-5) (4). The performances of models for the IDH-4 and IDH-5 events are summarized in Supplemental Table 11, and the recurrent neural network model was still greater than other models in predicting IDH-4 and IDH-5.

## Discussion

Despite the clinical implications of IDH for cardiovascular morbidity and mortality, its prediction is not easy because several interactive factors, time-varying vital signs, and hemodialysis settings need be considered. We addressed this issue using the recurrent neural network model to consider time-varying features that other machine learning and conventional logistic regression models cannot use. The recurrent neural network model was better than the other tested models at predicting the risk of IDH, indicating the continuous and real-time prediction of IDH during hemodialysis may be achievable.

Artificial intelligence, including deep learning, has the potential to enhance the efficacy and easiness of clinical pathways. Particularly, unprecedented advances in artificial intelligence have enabled digital radiologic and pathologic imaging diagnosis beyond human limits and boundaries (19). The application of artificial intelligence has also grown in the nephrology field, particularly for predicting AKI. Indeed, the recurrent neural network model successfully predicted AKI (20), and the results were more accurate than those obtained using other machine learning (*e.g.*, gradient boosting machine) (21) and discrete-time logistic regression (22) models, probably because of the time-varying features in the recurrent neural network model.

Although machine learning models have been used to predict IDH (23–25), these models did not consider time-varying features such as BPs. We used the recurrent neural network model to predict IDH because time-varying vital signs and hemodialysis settings or their trends are known to strongly affect predictions (26,27). As expected, the recurrent neural network model performed better than the other machine learning or logistic regression models tested. The results of the feature set-ablation analysis and feature ranking analyses also support the contribution of time-varying vital signs and hemodialysis settings in constructing the IDH-prediction model.

Some methods to control and reduce the risk of IDH have been reported (4,28,29), but despite this, IDH still occurs in approximately 10% of hemodialysis sessions (3). It is important to reduce this rate to reduce the risk of cardiovascular diseases that can develop as a result. Predicting IDH in advance may help establish an intervention plan. To the best of our knowledge, there are no artificial intelligence-based alarm systems for predicting the risk of IDH in real time. The present models will be implemented in clinical practice as an alarm service, although external validation is warranted. Several issues need to be solved before the present model can

be implemented in clinical practice, including digital recording of information, incorporation of monitoring records into the clinical records, and determining the management steps if there is risk of IDH. However, the performance results, feature ranking and feature set-ablation assays, and subgroup and sensitivity analyses will be a basis of future studies to overcome these issues.

In this study, the recurrent neural network model performed better than the other models for differently defined IDHs and maintained its predictive power in various subgroups. It suggests the recurrent neural network model could be well applied to various clinical situations and demands. Because there is a trade-off between precision and recall, the optimal threshold may differ depending on clinical demands (*e.g.*, active surveillance with high recall), and changes in threshold values will alter the predictive capacity of the model.

Despite the promising results, certain limitations need to be considered. Machine learning models, including deep learning, are black boxes, which makes it difficult to understand how predictions are arrived at, and how much each variable or timestamp contributes to the results. Nevertheless, our feature set-ablation analysis may be able to provide underlying clues and an intervention strategy for IDH. The study design was retrospective, so the results produced by the recurrent neural network model may change depending on the characteristics of the dataset used (30). Unidentified factors such as cardiac monitoring (*e.g.*, cardiac arrhythmia), dialysis vintage, and medical records (*e.g.*, written notes) may provide additional information on the risk of IDH and further improve the model's performance. Dialysis settings may differ between countries and populations, which will limit the application of the present models to other patient subsets. The difference in characteristics between patients included and excluded in analyses may also limit the model applicability. Regarding this issue, we provide the framework of the models in the Supplemental Methods. This will be helpful in adjusting the models according to the population characteristics, although this needs to be validated in future and independent cohorts.

Until now, the risk of IDH has been difficult to estimate. Our results suggest real-time prediction of IDH may be possible using a deep learning model, and the application of such a model could enable clinicians to cope with IDH earlier than before. This approach will have a positive effect in all hemodialysis situations, and will be especially useful in small hospitals that have small doctor numbers and during home hemodialysis. These results may motivate researchers to apply deep learning models to predict other outcomes of hemodialysis.

## Disclosures

H. Lee reports receiving research funding from Institute of Information & Communications Technology Planning & Evaluation (IITP) and having other interests/relationships with V.DO Inc. J. Yoo reports receiving research funding from IITP and having other interests/relationships with Recobell. K. Joo reports receiving research funding from Baxter. K.-H. Oh reports receiving research funding from Fresenius Medical Care, Korea. K. Yoo reports receiving research funding from IITP. All remaining authors have nothing to disclose.

## Funding

This work was supported by Institute of Information & Communications Technology Planning & Evaluation, which is funded by the Ministry of Science and ICT, Republic of Korea, grant 2019-0-01367 (to N. Kwak). The funders had no role in study design, data collection or analysis, decision to publish, or preparation of the manuscript.

## Supplemental Material

This article contains the following supplemental material online at <http://cjasn.asnjournals.org/lookup/suppl/doi:10.2215/CJN.09280620/-/DCSupplemental>.

Supplemental Methods.

Supplemental Table 1. Characteristics of hemodialysis sessions that were included and excluded from the analyses.

Supplemental Table 2. Description of the four models used in this study.

Supplemental Table 3. Summarized table of feature sets of the model input.

Supplemental Table 4. F1 score for predicting intradialytic hypotension in deep and other machine learning and logistic regression models.

Supplemental Table 5. F1 scores of the recurrent neural network after ablation of the feature set.

Supplemental Table 6. Area under the curves for predicting intradialytic hypotension in incidence and prevalent hemodialysis sessions.

Supplemental Table 7. Area under the curves for predicting intradialytic hypotension according to the tertiles of ultrafiltration.

Supplemental Table 8. Area under the curves for predicting intradialytic hypotension on the basis of a history of intradialytic hypotension.

Supplemental Table 9. Performance of the recurrent neural network model after using sessions-stratified randomization.

Supplemental Table 10. Performance of the recurrent neural network model in sessions with and without echocardiographic information.

Supplemental Table 11. Area under the curves for predicting the differently defined intradialytic hypotension.

Supplemental Figure 1. Exploratory data analysis of the rate of intradialytic hypotension (IDH).

Supplemental Figure 2. Confusion matrix plot for (A) intradialytic hypotension 1 (IDH-1), (B) IDH-2, and (C) IDH-3. The case numbers are given in each cell.

Supplemental Figure 3. Precision-recall graph according to the hemodialysis time. The data output thresholds were set as 0.1, 0.3, 0.5, 0.7, and 0.9. (A) intradialytic hypotension 1 (IDH-1); (B) IDH-2; and (C) IDH-3.

Supplemental Figure 4. Decision curve analysis of recurrent neural network (RNN) and three other models. (A) Intradialytic hypotension 1 (IDH-1); (B) IDH-2; and (C) IDH-3.

Supplemental Figure 5. Platt scaling plot used to calibrate the models. (A) Intradialytic hypotension 1 (IDH-1); (B) IDH-2; (C) IDH-3. Bin size = 0.05.

## References

- Stefánsson BV, Brunelli SM, Cabrera C, Rosenbaum D, Anum E, Ramakrishnan K, Jensen DE, Ståhlhammar NO: Intradialytic hypotension and risk of cardiovascular disease. *Clin J Am Soc Nephrol* 9: 2124–2132, 2014
- Flythe JE, Xue H, Lynch KE, Curhan GC, Brunelli SM: Association of mortality risk with various definitions of intradialytic hypotension. *J Am Soc Nephrol* 26: 724–734, 2015
- Kuipers J, Verboom LM, Ipema KJR, Paans W, Krijnen WP, Gaillard CAJM, Westerhuis R, Franssen CFM: The prevalence of intradialytic hypotension in patients on conventional hemodialysis: A systematic review with meta-analysis. *Am J Nephrol* 49: 497–506, 2019
- Santoro A, Mancini E, Basile C, Amoroso L, Di Giulio S, Usberti M, Colasanti G, Verzetti G, Rocco A, Imbasciati E, Panzetta G, Bolzani R, Grandi F, Polacchini M: Blood volume controlled hemodialysis in hypotension-prone patients: A randomized, multicenter controlled trial. *Kidney Int* 62: 1034–1045, 2002
- Sands JJ, Usvyat LA, Sullivan T, Segal JH, Zabetakis P, Kotanko P, Maddux FW, Diaz-Buxo JA: Intradialytic hypotension: Frequency, sources of variation and correlation with clinical outcome. *Hemodial Int* 18: 415–422, 2014
- Davenport A, Cox C, Thuraisingham R: Blood pressure control and symptomatic intradialytic hypotension in diabetic haemodialysis patients: A cross-sectional survey. *Nephron Clin Pract* 109: c65–c71, 2008
- Koomans HA, Braam B, Geers AB, Roos JC, Dorhout Mees EJ: The importance of plasma protein for blood volume and blood pressure homeostasis. *Kidney Int* 30: 730–735, 1986
- Rocha A, Sousa C, Teles P, Coelho A, Xavier E: Frequency of intradialytic hypotensive episodes: Old problem, new insights. *J Am Soc Hypertens* 9: 763–768, 2015
- Reeves PB, Mc Causland FR: Mechanisms, clinical implications, and treatment of intradialytic hypotension. *Clin J Am Soc Nephrol* 13: 1297–1303, 2018
- National Kidney Foundation: KDOQI clinical practice guideline for hemodialysis adequacy: 2015 update [published correction appears in *Am J Kidney Dis* 67: 534, 2016]. *Am J Kidney Dis* 66: 884–930, 2015
- Burlacu A, Iftene A, Busoiu E, Cogeana D, Covic A: Challenging the supremacy of evidence-based medicine through artificial intelligence: The time has come for a change of paradigms. *Nephrol Dial Transplant* 35: 191–194, 2020
- LeCun Y, Bengio Y, Hinton G: Deep learning. *Nature* 521: 436–444, 2015
- Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, Cui C, Corrado G, Thrun S, Dean J: A guide to deep learning in healthcare. *Nat Med* 25: 24–29, 2019
- Kooman J, Basci A, Pizzarello F, Canaud B, Haage P, Fouque D, Konner K, Martin-Malo A, Pedrini L, Tattersall J, Tordoir J, Vennegoor M, Wanner C, ter Wee P, Vanholder R: EBPG guideline on haemodynamic instability. *Nephrol Dial Transplant* 22[Suppl 2]: ii22–ii44, 2007
- K/DOQI Workgroup: K/DOQI clinical practice guidelines for cardiovascular disease in dialysis patients. *Am J Kidney Dis* 45 [Suppl 3]: S1–S153, 2005
- Hochreiter S, Schmidhuber J: Long short-term memory. *Neural Comput* 9: 1735–1780, 1997
- Platt JC: Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In: *Advances in Large Margin Classifiers*, edited by Smola A, Bartlett P, Scholkopf B, Schuurmans D, Cambridge, MIT Press, 1999, pp 61–74
- Ginley B, Lutnick B, Jen KY, Fogo AB, Jain S, Rosenberg A, Walavalkar V, Wilding G, Tomaszewski JE, Yacoub R, Rossi GM, Sarder P: Computational segmentation and classification of diabetic glomerulosclerosis. *J Am Soc Nephrol* 30: 1953–1967, 2019
- Niazi MKK, Parwani AV, Gurcan MN: Digital pathology and artificial intelligence. *Lancet Oncol* 20: e253–e261, 2019
- Tomašev N, Glorot X, Rae JW, Zielinski M, Askham H, Saraiva A, Mottram A, Meyer C, Ravuri S, Protsyuk I, Connell A, Hughes CO, Karthikesalingam A, Cornebise J, Montgomery H, Rees G, Laing C, Baker CR, Peterson K, Reeves R, Hassabis D, King D, Suleyman M, Back T, Nielson C, Ledsam JR, Mohamed S: A clinically applicable approach to continuous prediction of future acute kidney injury. *Nature* 572: 116–119, 2019
- Koyner JL, Carey KA, Edelson DP, Churpek MM: The development of a machine learning inpatient acute kidney injury prediction model. *Crit Care Med* 46: 1070–1077, 2018
- Simonov M, Ugwuowo U, Moreira E, Yamamoto Y, Biswas A, Martin M, Testani J, Wilson FP: A simple real-time model for predicting acute kidney injury in hospitalized patients in the US: A descriptive modeling study. *PLoS Med* 16: e1002861, 2019

23. Thakur SS, Abdul SS, Chiu HS, Roy RB, Huang PY, Malwade S, Nursetyo AA, Li YJ: Artificial-intelligence-based prediction of clinical events among hemodialysis patients using non-contact sensor data. *Sensors (Basel)* 18: 2833, 2018
24. Barbieri C, Cattinelli I, Neri L, Mari F, Ramos R, Brancaccio D, Canaud B, Stuard S: Development of an artificial intelligence model to guide the management of blood pressure, fluid volume, and dialysis dose in end-stage kidney disease patients: Proof of concept and first clinical assessment. *Kidney Dis* 5: 28–33, 2019
25. Lin CJ, Chen CY, Wu PC, Pan CF, Shih HM, Huang MY, Chou LH, Tang JS, Wu CJ: Intelligent system to predict intradialytic hypotension in chronic hemodialysis. *J Formos Med Assoc* 117: 888–893, 2018
26. Wang F, Wang Y, Tian Y, Zhang P, Chen J, Li J: Pattern recognition and prognostic analysis of longitudinal blood pressure records in hemodialysis treatment based on a convolutional neural network. *J Biomed Inform* 98: 103271, 2019
27. Park J, Rhee CM, Sim JJ, Kim YL, Ricks J, Streja E, Vashistha T, Tolouian R, Kovesdy CP, Kalantar-Zadeh K: A comparative effectiveness research study of the change in blood pressure during hemodialysis treatment and survival. *Kidney Int* 84: 795–802, 2013
28. Selby NM, McIntyre CW: A systematic review of the clinical effects of reducing dialysate fluid temperature. *Nephrol Dial Transplant* 21: 1883–1898, 2006
29. Sadowski RH, Allred EN, Jabs K: Sodium modeling ameliorates intradialytic and interdialytic symptoms in young hemodialysis patients. *J Am Soc Nephrol* 4: 1192–1198, 1993
30. Keane PA, Topol EJ: With an eye to AI and autonomous diagnosis. *NPJ Digit Med* 1: 40, 2018

**Received:** June 10, 2020 **Accepted:** December 8, 2020

H.L. and D.Y. contributed equally to this work.

Published online ahead of print. Publication date available at [www.cjasn.org](http://www.cjasn.org).

See related editorial, “Leveraging Deep Learning to Improve Safety of Outpatient Hemodialysis,” on pages xxx–xxx.

## **Supplemental Material**

### **Deep Learning Model for Real-Time Prediction of Intradialytic Hypotension**

Hojun Lee, Donghwan Yun, Jayeon Yoo, Kiyoon Yoo, Yong Chul Kim, Dong Ki Kim, Kook-Hwan Oh, Kwon Wook Joo, Yon Su Kim, Nojun Kwak and Seung Seok Han

#### Supplemental Methods

Supplemental Table 1. Characteristics of hemodialysis sessions that were included and excluded from the analyses

Supplemental Table 2. Description of the four models used in the present study

Supplemental Table 3. Summarized table of feature sets of the model input

Supplemental Table 4. F1 score for predicting intradialytic hypotension in deep and other machine learning, and logistic regression models

Supplemental Table 5. F1 scores of the recurrent neural network after ablation of the feature set

Supplemental Table 6. Area under the curves for predicting intradialytic hypotension in incidence and prevalent hemodialysis sessions

Supplemental Table 7. Area under the curves for predicting intradialytic hypotension according to the tertiles of ultrafiltration

Supplemental Table 8. Area under the curves for predicting intradialytic hypotension based on a history of intradialytic hypotension

Supplemental Table 9. Performance of the recurrent neural network model after using sessions-stratified randomization

Supplemental Table 10. Performance of the recurrent neural network model in sessions with and without echocardiographic information

Supplemental Table 11. Area under the curves for predicting the differently defined intradialytic hypotension

Supplemental Figure 1. Exploratory data analysis of the rate of intradialytic hypotension (IDH).

Supplemental Figure 2. Confusion matrix plot for (A) IDH-1, (B) IDH-2, and (C) IDH-3. The case numbers are given in each cell.

Supplemental Figure 3. Precision-recall graph according to the hemodialysis time. The data output thresholds were set as 0.1, 0.3, 0.5, 0.7, and 0.9. (A) IDH-1. (B) IDH-2. (C) IDH-3.

Supplemental Figure 4. Decision curve analysis of recurrent neural network (RNN) and three other models. (A) Intradialytic hypotension (IDH)-1. (B) IDH-2. (C) IDH-3.

Supplemental Figure 5. Platt scaling plot used to calibrate the models. (A) Intradialytic hypotension (IDH)-1. (B) IDH-2. (C) IDH-3. Bin size = 0.05.

## Supplemental Methods

### Study variables

The baseline clinical information was age, sex, vital signs (systolic blood pressure, diastolic blood pressure, heart rate, and body temperature), hemodialysis settings (type of hemodialysis [hemodialysis, hemodiafiltration, hemoperfusion, hemofiltration, hemofiltration reinfusion, and supplementary ultrafiltration], blood flow rate, dialysate flow rate, target and time-varying amounts of ultrafiltration, time setting, access route [arteriovenous fistula and graft, temporary catheterization via internal jugular and femoral vein, subcutaneously-tunneled catheter, sheath for coronary angiography, and Hickman catheter], pre-dialysis weight, use of anti-coagulant [heparin and nafamostat mesilate], priming fluid [normal saline, half saline, and red blood cells], dialysate [e.g., Hemo B Dex 0.1% and 0.15%, and Hemotrate-B1] and dialyzer [e.g., APS-15U, APS-21U, Rexeed-13LX, Rexeed-18LX, BLS 812G, BLS 812SD, BLS 814SD, BLS 816SD, BLS 819SD, NC 1485, PHF0714, SG30, Adsorba, polyflux 14, polyflux 14H, polyflux 14L, polyflux 14S, polyflux 170H, polyflux 17L, polyflux 17S, polyflux 6H, polyflux 8L, polyflux S, Theranova 400, F4 HPS, F5 HPS, F6 HPS, FX, FX paed, FX5, FX8, FX40, FX50, FX80, FB 130T, and Sureflux 130E-GA]), the dialysate temperature and concentrations of sodium, potassium, calcium, and bicarbonate, incident or prevalent sessions, admission status, the presence of comorbidities (e.g., diabetes mellitus, hypertension, cardiovascular disease, and kidney transplantation), the number of session per week, the history of IDH within one week, total number of sessions with IDH within one week, and medications used before initiating hemodialysis. Laboratory blood findings were measured at the beginning of the hemodialysis sessions, including white blood cells, hemoglobin, platelet, cholesterol, albumin, glucose, calcium, phosphate, uric acid, blood urea nitrogen, creatinine, sodium,

potassium, chloride, and total carbon dioxide. There were no missing variables.

Echocardiographic information before hemodialysis sessions including left ventricular ejection fraction, left ventricular end-diastolic dimension, left ventricular end-systolic dimension, interventricular septum thickness, and left ventricular mass was available for 227,640 (87%) sessions and 7256 (78%) patients. This information was used in a sensitivity analysis.

## **Model development**

Statistical analyses were performed using R software (version 3.5.1; The Comprehensive R Archive Network: <http://cran.r-project.org>) and Python (version 3.6.8; Python Software Foundation: <http://www.python.org>). The PyTorch 1.3 was used as a deep learning framework throughout this process (1).

The categorical and continuous variables of the baseline characteristics are presented as proportions and means  $\pm$  standard deviation, respectively. The dataset was treated as follows:  $S = [(x_{1,1}, y_{1,1}), \dots, (x_{1,L_1}, y_{1,L_1}), (x_{2,1}, y_{2,1}), \dots, (x_d, L_d, y_d, L_d)]$ , where  $d$  and  $L_d$  indicate the number of dialysis cases and the frame number of the  $d$ th dialysis, respectively. The ground truth labels were denoted as  $y_{p,q} = [y_{p,q, \text{IDH-1}}, y_{p,q, \text{IDH-2(initial)}}, y_{p,q, \text{IDH-2(present)}}]$ , where 0 was normal and 1 was abnormal (i.e., IDH). When converting the data in the training dataset into vectors, the continuous features were standardized with a mean of 0 and a variance of 1, and the categorical features were transformed into binary variables (i.e. 0 or 1) by one-hot encoding. The dataset  $S$  was used as the training data to train the recurrent neural network, multilayer perceptron, Light Gradient Boosting Machine, and logistic regression models.

Binary cross-entropy loss was used as the loss function for the recurrent neural network to

calculate the difference between actual labels and predictions. We used the Adam optimization method as the optimizer (2). The pseudocode for the recurrent neural network is given below.

Pseudocode 1: Predicting IDH Using recurrent neural network Evaluation	
1	<b>Function</b> IDH-prediction (Trained Network, $X_{p,1:L_p}$ , $\sigma$ )
	<b>Input:</b> Trained Network: BatchNorm, RNN, MLP-1, MLP-2, MLP-3;
	$\sigma$ : logistic function;
	$X_{p,1:L_p} = (x_1, x_2, \dots, x_{L_p})$ : entries of p-th dialysis of test data;
	<b>Output :</b> $\hat{y}_{p,1:L_p} = [(\hat{y}_{IDH-1}, \hat{y}_{IDH-2}(init), \hat{y}_{IDH-3}(present)), \dots, (\hat{y}_{p,IDH-1}, \hat{y}_{p,IDH-2}(init), \hat{y}_{p,IDH-3}(present))]$ : predictions for p-th dialysis
2	Timestamp index $t \leftarrow 1$
3	<b>for</b> $t=1$ to $L_p$ :
4	$x_{t,batchnorm} \leftarrow \text{BatchNorm}(x_t)$
5	$h_t \leftarrow \text{RNN}(x_{t,batchnorm}, h_{t-1})$
6	$\hat{y}_{IDH-1} \leftarrow \sigma(\text{MLP-1}(h_t))$
7	$\hat{y}_{IDH-2}(init) \leftarrow \sigma(\text{MLP-2}(h_t))$
8	$\hat{y}_{IDH-3}(present) \leftarrow \sigma(\text{MLP-3}(h_t))$
9	$t \leftarrow t+1$
10	
11	$y_{p,1:L_p} = [(y_{1,IDH-1}, y_{1,IDH-2}(init), y_{1,IDH-3}(present)), \dots, (y_{L_p,IDH-1}, y_{L_p,IDH-2}(init), y_{L_p,IDH-3}(present))]$ : ground truth labels of p-th dialysis



RNN, recurrent neural network; MLP, multilayer perceptron

The multilayer perceptron algorithm consists of a series of non-linear functions and fully-connected layers that are affine transforms as follows:  $\hat{y}_{p,q} = \sigma_n \circ f_n \circ \sigma_{n-1} \circ f_{n-1} \circ \dots \circ \sigma_1 \circ f_1(x_{p,q})$  (3). The  $\sigma_j$  is the  $j$ th non-linear function (e.g.,  $\sigma_j(x) = \max(0, x)$ ) and the  $f_j$  is the  $j$ th fully-connected layer (i.e., affine transform). Throughout this calculation, the multilayer perceptron can extract meaningful information on higher dimensions of the input vector. For a probability model, the last  $\sigma_n$  is a logistic function. The binary cross-entropy loss and the Adam optimization methods were used (2). The architecture of the multilayer perceptron is shown below.

Layer	Shape
BatchNorm	260 (input feature size)
Fully connected layer + ReLU	$260 \times 256$
Fully connected layer + BatchNorm + ReLU	$256 \times 256$
Fully connected layer + ReLU	$256 \times 256$
Fully connected layer + ReLU	$256 \times 256$
Fully connected layer	$256 \times 3$
Logistic function	3

BatchNorm, batch normalization; ReLU, rectified linear unit

Light Gradient Boosting Machine combines learned ‘Learners’ after learning several weak ‘Learners’ (4). Throughout the learning and predicting process for weak ‘Learners’, the error is lowered by gradient boosting on residuals of incorrectly predicted results. The Light Gradient Boosting Machine method is faster than the other Gradient Boosting Machine methods such as extreme gradient boosting (4). Logistic regression calculates the weighted sum of the feature vector that is derived from regression coefficients and feature values.

### Feature ranking analysis

To estimate how much features contribute to the prediction of IDH, we use the feature ranking method proposed in the previous paper (5). This method drops each feature one by one from the test dataset when the model inference and compares the prediction results to the reference prediction result which is gained without losing any features. Large prediction differences between dropped data and full-featured data represent that the dropped features have contributed much more when the model makes predictions.

$$score_{f_{drop}} = \frac{1}{N} \sum_{i=1}^N |p_{i,f_{drop}} - p_i| \dots (A)$$

To apply this feature ranking method to our approach, we modify it to suit our settings as shown in equation (A). In eq (A),  $f_{drop}$  is the feature we focus on and drop from the input data.  $p_i$  means the reference prediction result of the i-th data which the model infers using all features and  $p_{i,f_{drop}}$  means that the prediction result of the i-th data when the feature  $f_{drop}$  is dropped. These absolute values are averaged over all dataset of size N. The scores were calculated for IDH-1, IDH-2, and IDH-3 respectively.

The batch norm layer serves as a standardization. To drop each feature, we have set each output of the batch norm layer as 0. The prediction result may be higher or lower than reference

prediction, however, to measure the degree of difference between dropped feature data and full featured data we average the absolute value of differences.

## References

1. Paszke A, Gross S, Chintala S, Chanan G, Yang E, DeVito Z, Lin Z, Desmaison A, Antiga L, Lerer A: Automatic differentiation in PyTorch. *In NIPS-W2017*.
2. Kingma DP, Ba J: Adam: A method for stochastic optimization. *arXiv preprint arXiv:14126980*, 2014
3. Pal SK, Mitra S: Multilayer perceptron, fuzzy sets, and classification. *IEEE Trans Neural Netw* 3: 683-697, 1992.
4. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu T-Y: LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *31st Conference on Neural Information Processing Systems (NIPS 2017)* 3148-3156, 2017.
5. Ginley B, Lutnick B, Jen KY, Fogo AB, Jain S, Rosenberg A, et al.: Computational Segmentation and Classification of Diabetic Glomerulosclerosis. *J Am Soc Nephrol*, 30: 1953-1967, 2019.

Supplemental Table 1. Characteristics of hemodialysis sessions that were included and excluded from the analyses

Variables	Excluded (n = 20,752)	Included (n = 261,647)
Age (years)	19 ± 20	62 ± 15
Male (%)	58	58
Diabetes mellitus (%)	15	48
Hypertension (%)	57	68

Supplemental Table 2. Description of the four models used in the present study

Models	Descriptions	Comparison with other models
Logistic regression	Output the probability of a certain event happening by using a linear classifier followed by a logistic function. It is able to model the linear relationship between the input features and the target.	No inherent way of modeling temporal information
Light Gradient Boosting Machine	Composed of ensemble of tree-based models. It is able to make accurate predictions by iteratively boosting the errors of individual models.	Able to compute feature importance of the input data Able to model non-linear relationships No inherent way of modeling temporal information
Multilayer perceptron	A class of deep neural network that can model complex relationships by using multiple hidden layers. Similar to above methods, only fixed-size input is taken.	No explicit way of computing feature importance (black-box model). Feature selection method should be applied. No inherent way of modeling temporal information
Recurrent neural network	A class of deep neural network that assumes sequential data as inputs with temporal relationship. It does this by continuously utilizing previous sequences for predicting the target of current sequence.	Inherently models input sequentially No explicit way of computing feature importance (black-box model). Feature selection method should be applied.

Supplemental Table 3. Summarized table of feature sets of the model input

Feature sets	Variables
A (age and sex)	Age and sex
B (hemodialysis-related features)	Type of hemodialysis, dialysate flow rate, blood flow rate, target and time-varying amounts of ultrafiltration, target ultrafiltration, time setting, vascular access route, pre-dialysis weight, anti-coagulants, priming fluids, dialyzer type, dialysate sodium concentration, dialysate potassium concentration, dialysate calcium concentration, dialysate bicarbonate concentration, dialysate temperature, and incident or prevalent hemodialysis.
C (vital signs)	Systolic blood pressure, diastolic blood pressure, heart rate, and body temperature.
D (clinical information)	Diabetes mellitus, hypertension, cardiovascular disease, kidney transplant donor, kidney transplant recipient, number of dialysis session per week, previous history of IDH within 7 days, and total number of sessions with IDH within 7 days.
E (laboratory findings)	White blood cell count, hemoglobin, platelet, cholesterol, albumin, glucose, calcium, phosphate, uric acid, blood urea nitrogen, creatinine, sodium, potassium, chloride, and bicarbonate.
F (medications)	Beta-blockers, calcium channel blockers, angiotensin-converting enzyme inhibitors, aldosterone receptor blockers, diuretics, lipid-lowering agents, minoxidil, aspirin, adenosine diphosphate receptor inhibitors, warfarin, oral hypoglycemic agents, insulin, allopurinol, febuxostat, erythropoietin-stimulating agents, calcium-based phosphate binder, and non-calcium-based phosphate binders.

Supplemental Table 4. F1 score for predicting intradialytic hypotension in deep and other machine learning, and logistic regression models

Threshold	Models	Outcome		
		IDH-1	IDH-2	IDH-3
0.1	RNN	0.5028	0.6116	0.4391
	MLP	0.5096	0.6114	0.4391
	LightGBM	0.5025	0.6127	0.4301
	LR	0.4788	0.5754	0.4102
0.3	RNN	0.5844	0.6950	0.4981
	MLP	0.5717	0.6916	0.4846
	LightGBM	0.5759	0.6963	0.4800
	LR	0.5498	0.6763	0.4484
0.5	RNN	0.5197	0.6732	0.3682
	MLP	0.4765	0.6637	0.3522
	LightGBM	0.4912	0.6702	0.3186
	LR	0.4523	0.6271	0.2612
0.7	RNN	0.3542	0.5631	0.1714
	MLP	0.2829	0.5479	0.1685
	LightGBM	0.3002	0.5589	0.0961
	LR	0.2546	0.4873	0.0829
0.9	RNN	0.1275	0.3357	0.0163
	MLP	0.0706	0.3120	0.0102
	LightGBM	0.0367	0.2689	0.0014
	LR	0.0454	0.2547	0.0071

IDH-1, intradialytic hypotension defined as nadir systolic blood pressure <90 mmHg; IDH-2, intradialytic hypotension defined as decrease in systolic blood pressure  $\geq 20$  mmHg and/or decrease in mean arterial pressure  $\geq 10$  mmHg based on blood pressure at initial time point; IDH-3, intradialytic hypotension defined as decrease in systolic blood pressure  $\geq 20$  mmHg and/or decrease in mean arterial pressure  $\geq 10$  mmHg based on blood pressure at prediction

time point.

BP, blood pressure; RNN, recurrent neural network, MLP, multilayer perceptron; LightGBM,

Light Gradient Boosting Machine; LR, logistic regression.



Supplemental Table 5. F1 scores of the recurrent neural network after ablation of the feature set

Threshold	Models	Outcome		
		IDH-1	IDH-2	IDH-3
0.1	Remove set A	0.5051	0.6066	0.4366
	Remove set B	0.5002	0.5950	0.4174
	Remove set C	0.4140	0.5086	0.3683
	Remove set D	0.4965	0.6133	0.4321
	Remove set E	0.5043	0.6072	0.4317
	Remove set F	0.5041	0.6103	0.4354
0.3	Remove set A	0.5818	0.6918	0.4983
	Remove set B	0.5700	0.6836	0.4766
	Remove set C	0.4413	0.5592	0.3716
	Remove set D	0.5733	0.6896	0.4813
	Remove set E	0.5809	0.6943	0.4972
	Remove set F	0.5834	0.6948	0.4965
0.5	Remove set A	0.5129	0.6762	0.3705
	Remove set B	0.4760	0.6594	0.3266
	Remove set C	0.2851	0.4156	0.1161
	Remove set D	0.5120	0.6560	0.3315
	Remove set E	0.5028	0.6697	0.3569

0.7	Remove set F	0.5154	0.6739	0.3630
	Remove set A	0.3438	0.5697	0.1614
	Remove set B	0.2991	0.5270	0.1257
	Remove set C	0.1018	0.1605	0.0122
	Remove set D	0.3652	0.5340	0.1286
	Remove set E	0.3328	0.5538	0.1454
0.9	Remove set F	0.3471	0.5647	0.1657
	Remove set A	0.1190	0.3358	0.0105
	Remove set B	0.0880	0.2398	0.0051
	Remove set C	0.0064	0.0046	0.0000
	Remove set D	0.1381	0.3119	0.0068
	Remove set E	0.1119	0.3264	0.0095
	Remove set F	0.1263	0.3295	0.0154

---

IDH-1, intradialytic hypotension defined as nadir systolic blood pressure <90 mmHg; IDH-2, intradialytic hypotension defined as decrease in systolic blood pressure  $\geq 20$  mmHg and/or decrease in mean arterial pressure  $\geq 10$  mmHg based on blood pressure at initial time point; IDH-3, intradialytic hypotension defined as decrease in systolic blood pressure  $\geq 20$  mmHg and/or decrease in mean arterial pressure  $\geq 10$  mmHg based on blood pressure at prediction time point.

Each set contains features as follows: A, age and sex; B, hemodialysis-related features; C, vital signs; D, comorbidities; E, laboratory findings;

and F, medications.

Supplemental Table 6. Area under the curves for predicting intradialytic hypotension in incidence and prevalent hemodialysis sessions

Outcomes	Groups	AUROC (95% CI)	AUPRC (95% CI)
IDH-1	Incident	0.944 (0.941–0.947)	0.652 (0.650–0.653)
	Prevalent	0.936 (0.935–0.938)	0.610 (0.608–0.612)
IDH-2	Incident	0.872 (0.868–0.876)	0.783 (0.781–0.784)
	Prevalent	0.869 (0.868–0.871)	0.780 (0.779–0.782)
IDH-3	Incident	0.772 (0.766–0.777)	0.473 (0.471–0.475)
	Prevalent	0.792 (0.790–0.795)	0.516 (0.515–0.518)

AUROC, area under the receiver operating characteristic curve; CI, confidence interval;

AUPRC, area under the precision-recall curve; IDH, intradialytic hypotension.

Supplemental Table 7. Area under the curves for predicting intradialytic hypotension according to the tertiles of ultrafiltration

Outcomes	Ultrafiltration	AUROC (95% CI)	AUPRC (95% CI)
IDH-1	1 <sup>st</sup> tertile	0.951 (0.948–0.953)	0.667 (0.665–0.668)
	2 <sup>nd</sup> tertile	0.935 (0.932–0.937)	0.593 (0.591–0.594)
	3 <sup>rd</sup> tertile	0.926 (0.924–0.929)	0.590 (0.588–0.592)
IDH-2	1 <sup>st</sup> tertile	0.870 (0.868–0.872)	0.745 (0.744–0.747)
	2 <sup>nd</sup> tertile	0.874 (0.872–0.876)	0.789 (0.787–0.790)
	3 <sup>rd</sup> tertile	0.862 (0.860–0.864)	0.799 (0.797–0.800)
IDH-3	1 <sup>st</sup> tertile	0.788 (0.784–0.792)	0.477 (0.475–0.479)
	2 <sup>nd</sup> tertile	0.799 (0.796–0.803)	0.525 (0.523–0.527)
	3 <sup>rd</sup> tertile	0.780 (0.777–0.784)	0.525 (0.523–0.526)

AUROC, area under the receiver operating characteristic curve; CI, confidence interval;

AUPRC, area under the precision-recall curve; IDH, intradialytic hypotension.

Supplemental Table 8. Area under the curves for predicting intradialytic hypotension based on a history of intradialytic hypotension

Outcomes	History of IDH	AUROC (95% CI)	AUPRC (95% CI)
IDH-1	Absent	0.938 (0.936–0.941)	0.502 (0.500–0.504)
	Present	0.860 (0.857–0.863)	0.664 (0.662–0.665)
IDH-2	Absent	0.884 (0.881–0.886)	0.748 (0.746–0.749)
	Present	0.858 (0.856–0.860)	0.791 (0.789–0.792)
IDH-3	Absent	0.792 (0.788–0.796)	0.446 (0.445–0.448)
	Present	0.784 (0.781–0.786)	0.528 (0.526–0.530)

IDH, intradialytic hypotension; AUROC, area under the receiver operating characteristic curve; CI, confidence interval; AUPRC, area under the precision-recall curve.

Supplemental Table 9. Performance of the recurrent neural network model after using sessions-stratified randomization

Outcome	AUROC (95% CI)	AUPRC (95% CI)
IDH-1	0.943 (0.941–0.944)	0.659 (0.657–0.661)
IDH-2	0.877 (0.876–0.878)	0.788 (0.787–0.789)
IDH-3	0.793 (0.791–0.794)	0.508 (0.506–0.510)

AUROC, area under the receiver operating characteristic curve; CI, confidence interval;

AUPRC, area under the precision-recall curve; IDH, intradialytic hypotension.

Supplemental Table 10. Performance of the recurrent neural network model in sessions with and without echocardiographic information

Outcomes	Model without echocardiographic information		Model with echocardiographic information	
	AUROC (95% CI)	AUPRC (95% CI)	AUROC (95% CI)	AUPRC (95% CI)
IDH-1	0.938 (0.936–0.939)	0.643 (0.642–0.645)	0.938 (0.936–0.939)	0.643 (0.642–0.645)
IDH-2	0.875 (0.874–0.876)	0.783 (0.781–0.784)	0.875 (0.873–0.876)	0.782 (0.781–0.784)
IDH-3	0.800 (0.797–0.802)	0.522 (0.521–0.524)	0.799 (0.797–0.801)	0.522 (0.520–0.524)

AUROC, area under the receiver operating characteristic curve; CI, confidence interval; AUPRC, area under the precision-recall curve; IDH, intradialytic hypotension.

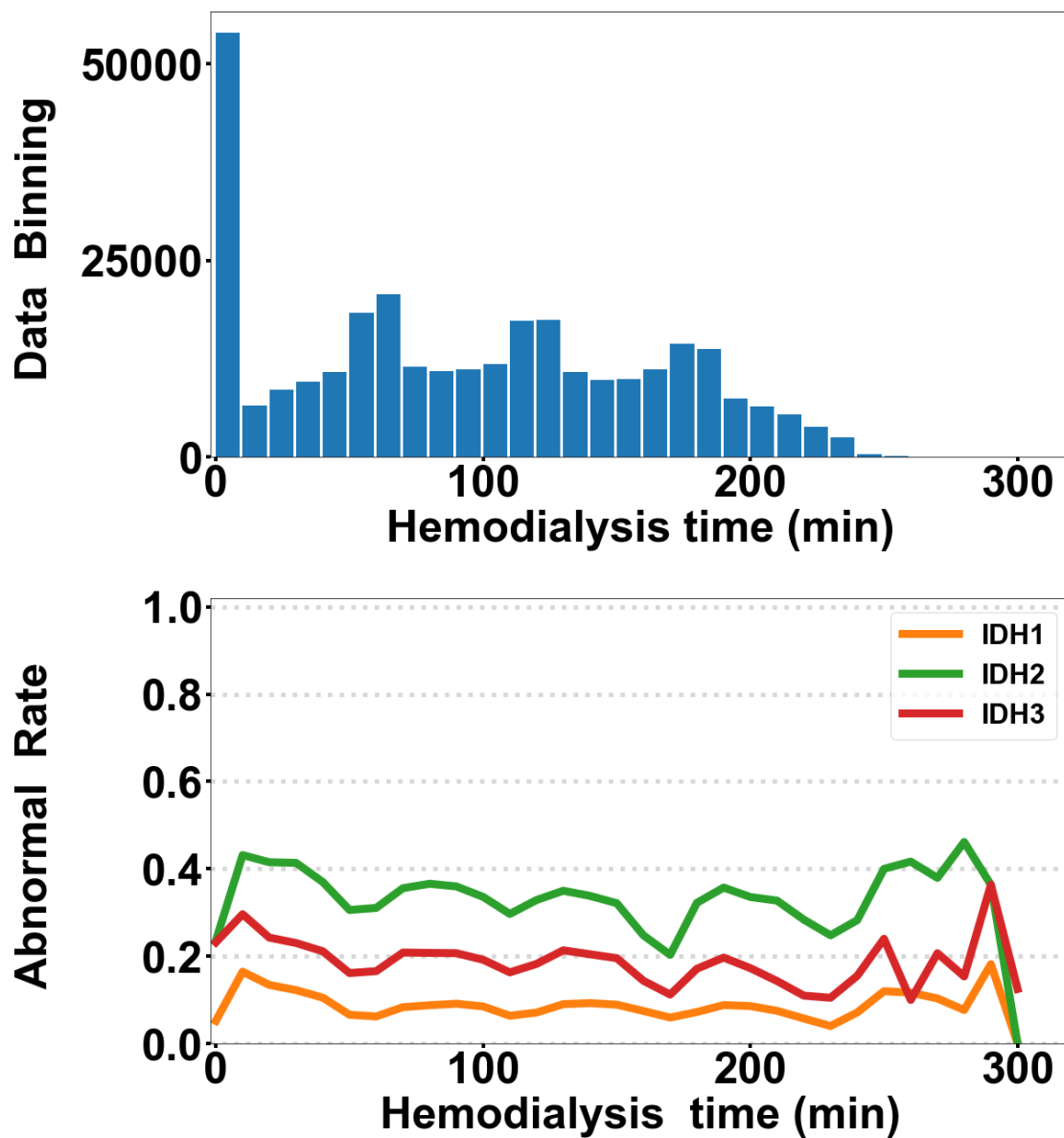


Supplemental Table 11. Area under the curves for predicting the differently defined intradialytic hypotension

Outcomes	Models	AUROC (95% CI)	<i>P</i> value	AUPRC (95% CI)
IDH-4	RNN	0.930 (0.929–0.932)		0.742 (0.740–0.744)
	MLP	0.928 (0.926–0.929)	<0.001	0.731 (0.729–0.732)
	LightGBM	0.928 (0.927–0.929)	<0.001	0.731 (0.730–0.733)
	Logistic regression	0.916 (0.914–0.917)	<0.001	0.694 (0.692–0.696)
IDH-5	RNN	0.888 (0.887–0.890)		0.724 (0.722–0.726)
	MLP	0.884 (0.882–0.885)	<0.001	0.715 (0.714–0.717)
	LightGBM	0.887 (0.885–0.888)	<0.001	0.715 (0.714–0.717)
	Logistic regression	0.872 (0.871–0.874)	<0.001	0.687 (0.685–0.688)

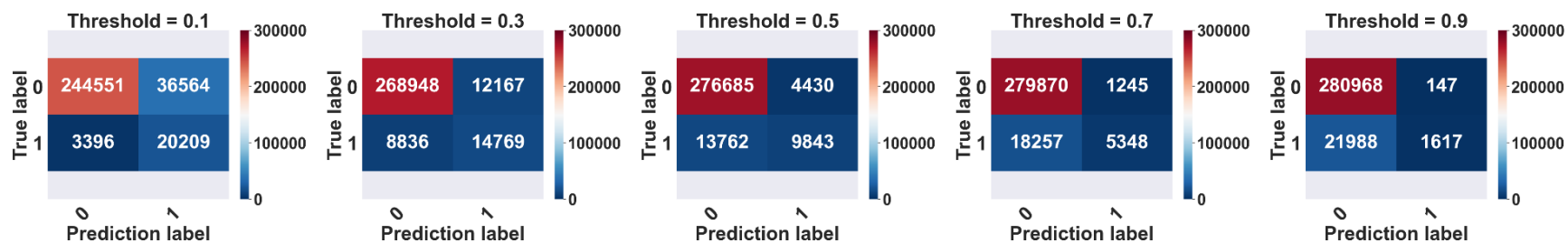
AUROC, area under the receiver operating characteristic curve; CI, confidence interval; AUPRC, area under the precision-recall curve; IDH, intradialytic hypotension; RNN, recurrent neural network, MLP, multilayer perceptron; LightGBM, Light Gradient Boosting Machine; LR, logistic regression.

Supplemental Figure 1. Exploratory data analysis of the rate of intradialytic hypotension (IDH).

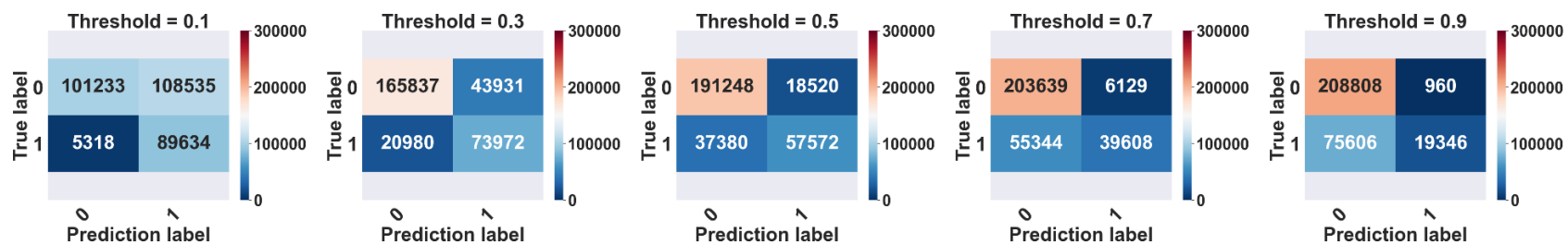


Supplemental Figure 2. Confusion matrix plot for (A) IDH-1, (B) IDH-2, and (C) IDH-3. The case numbers are given in each cell.

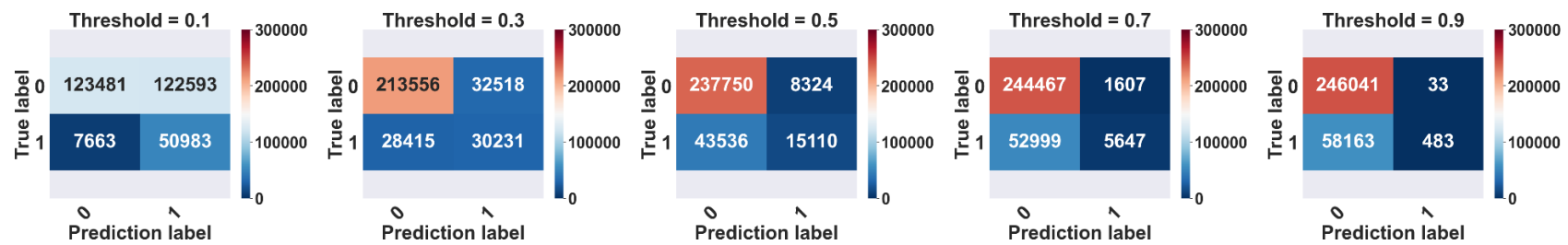
(A)



(B)

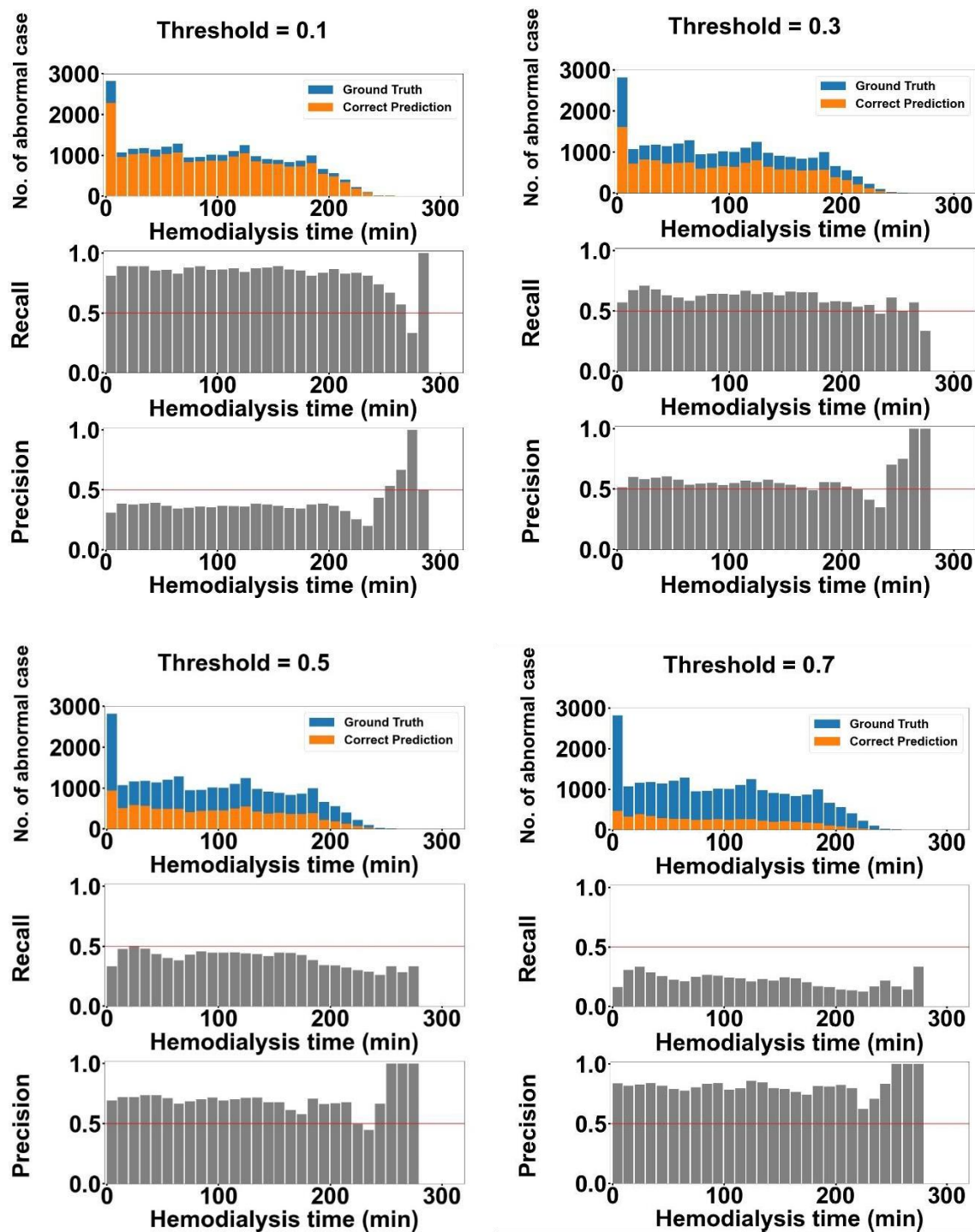


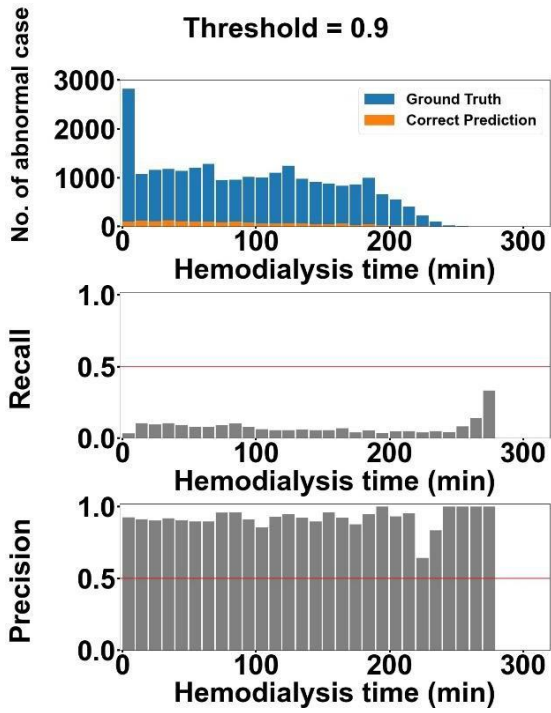
(C)



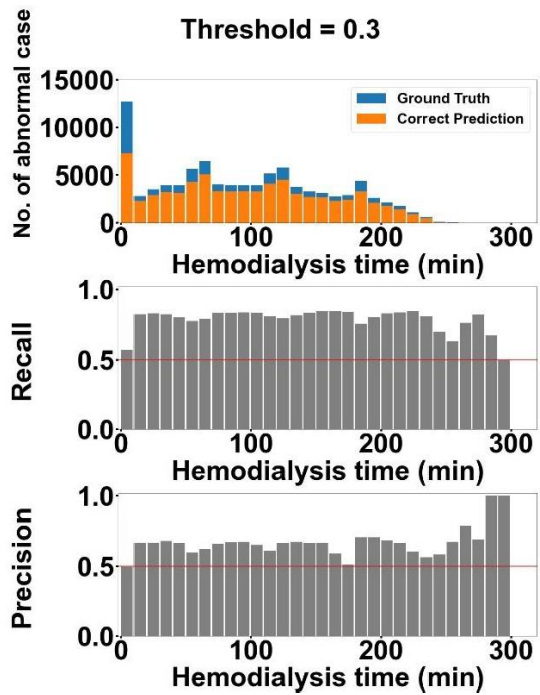
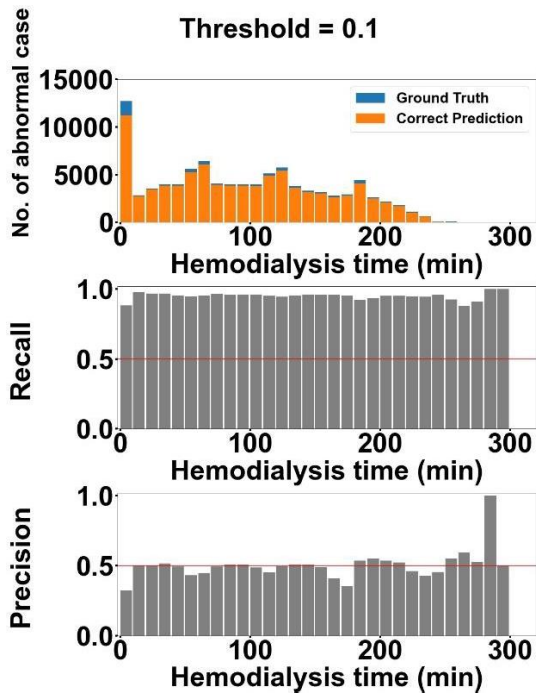
Supplemental Figure 3. Precision-recall graph according to the hemodialysis time. The data output thresholds were set as 0.1, 0.3, 0.5, 0.7, and 0.9. (A) IDH-1. (B) IDH-2. (C) IDH-3.

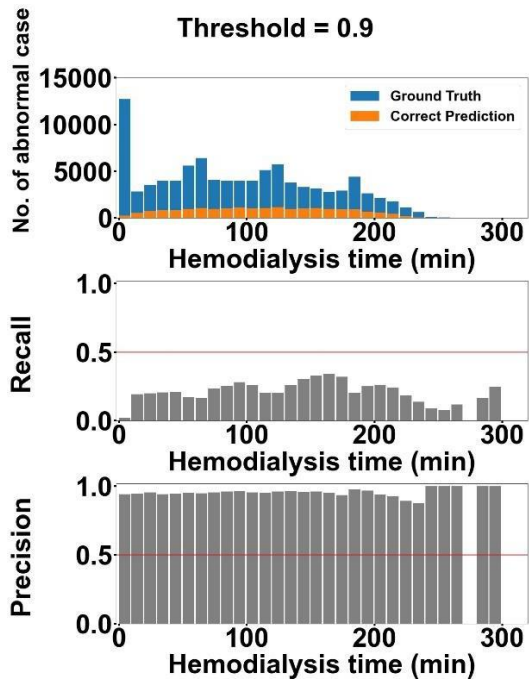
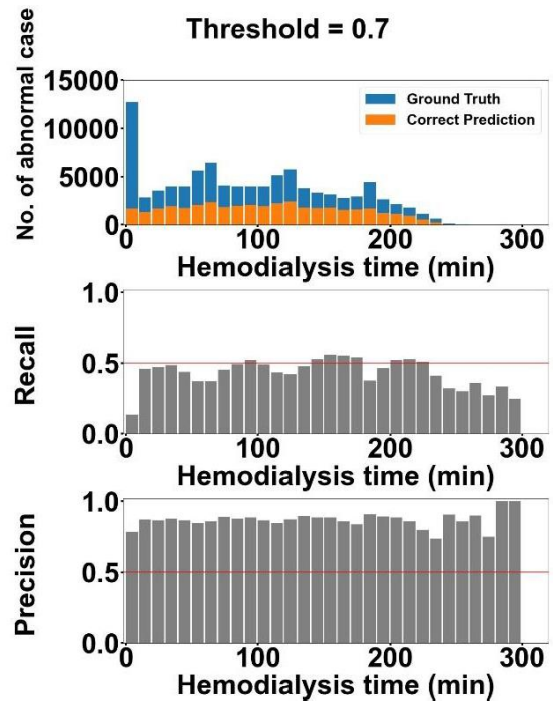
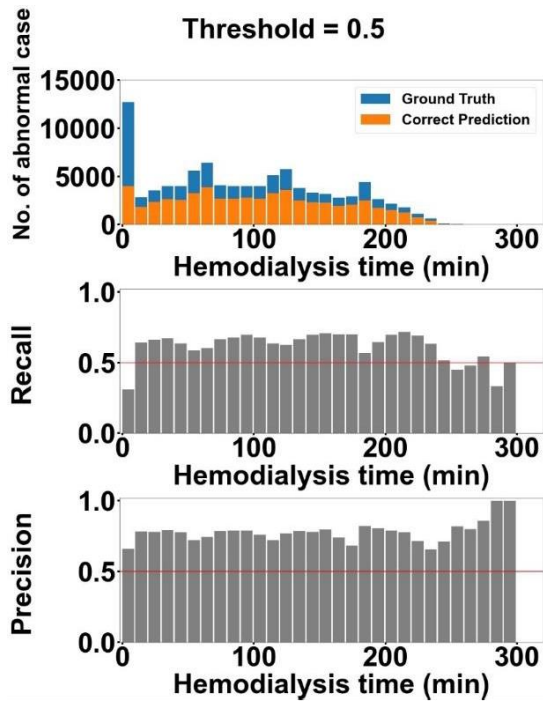
(A)



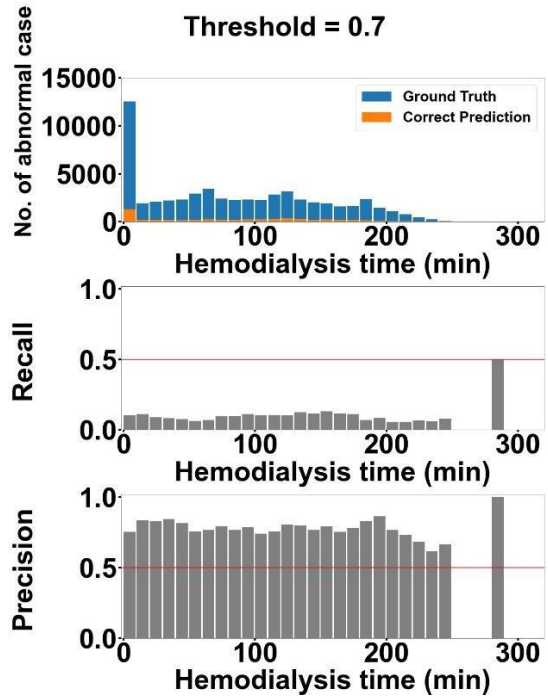
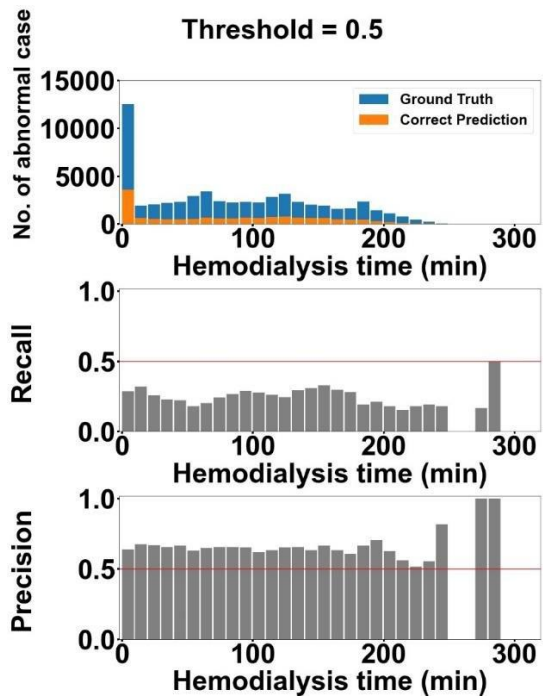
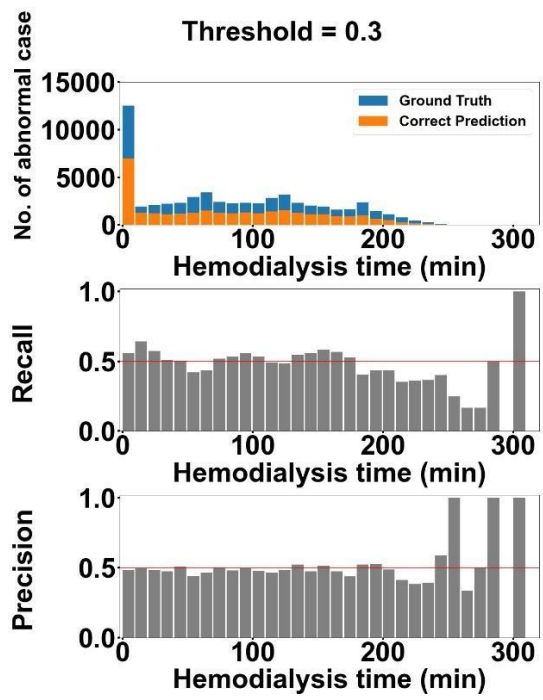
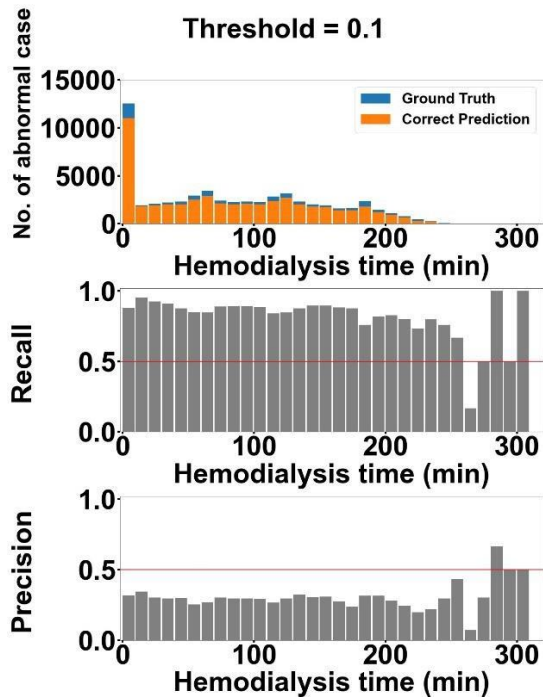


(B)

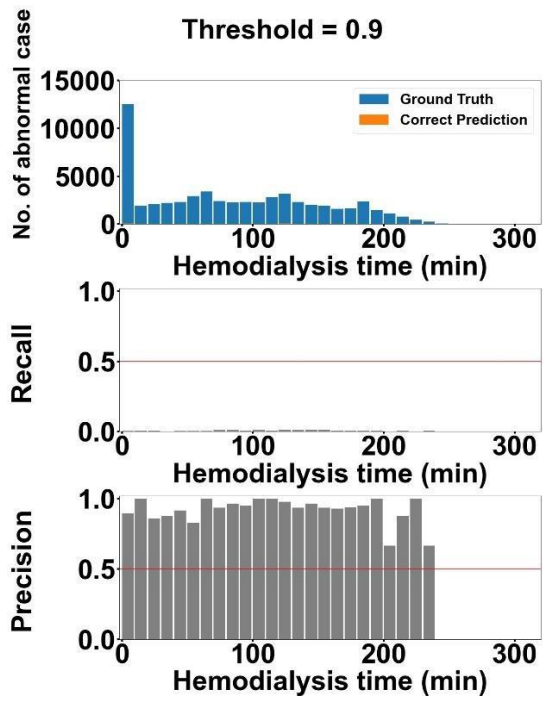




(C)

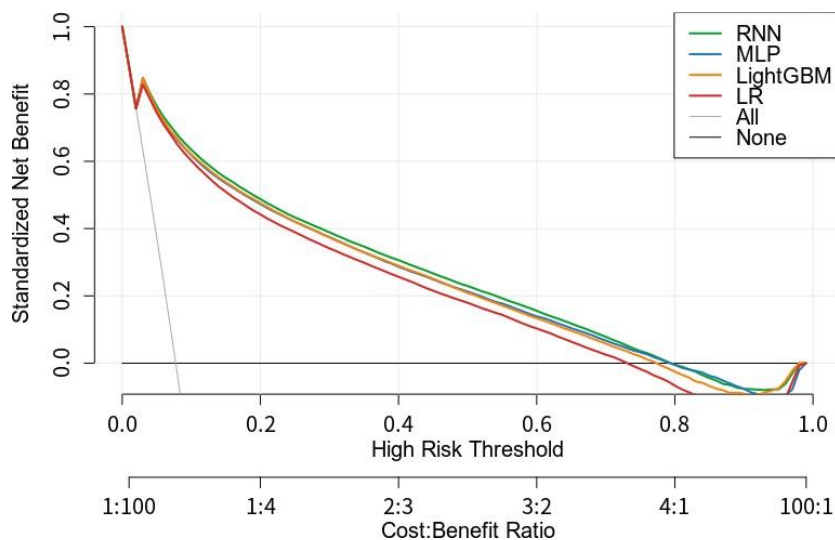




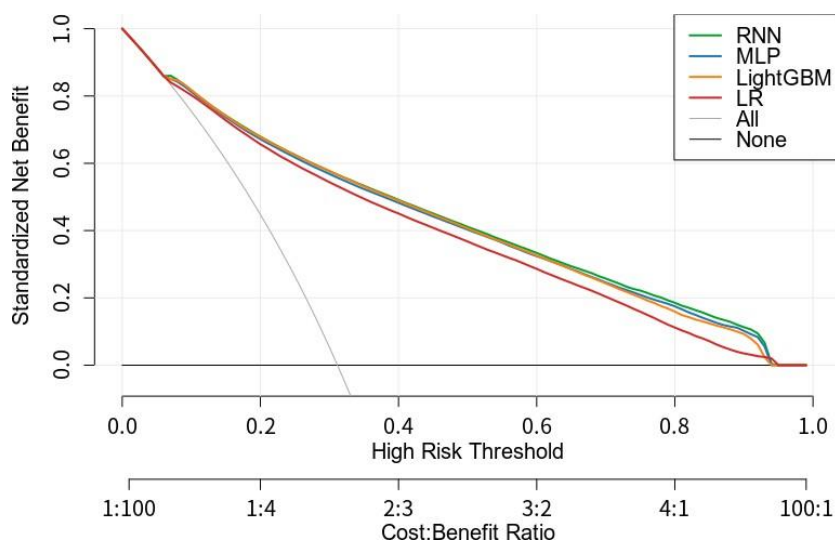


Supplemental Figure 4. Decision curve analysis of recurrent neural network (RNN) and three other models. (A) Intradialytic hypotension (IDH)-1. (B) IDH-2. (C) IDH-3. MLP, multilayer perceptron; LightGBM, Light Gradient Boosting Machine; LR, logistic regression.

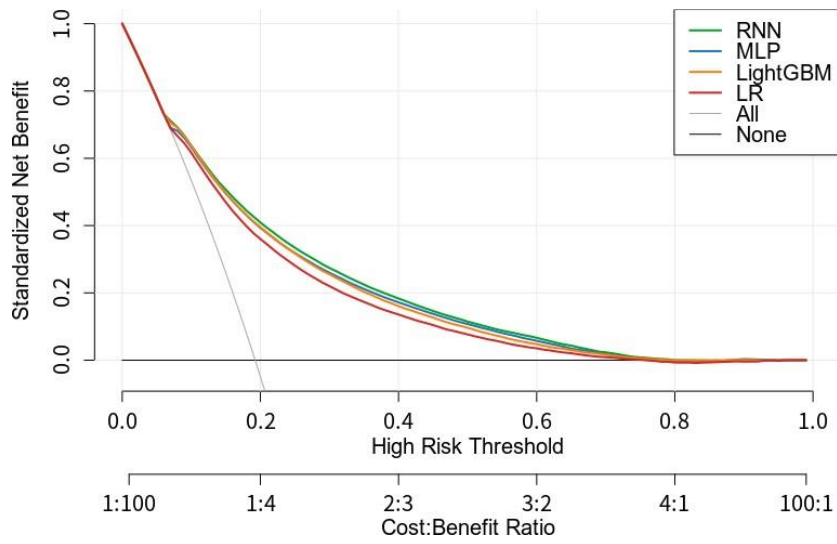
(A)



(B)

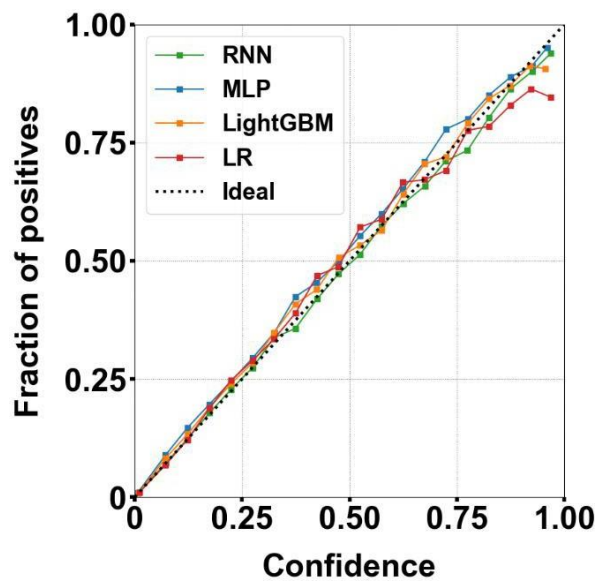


(C)

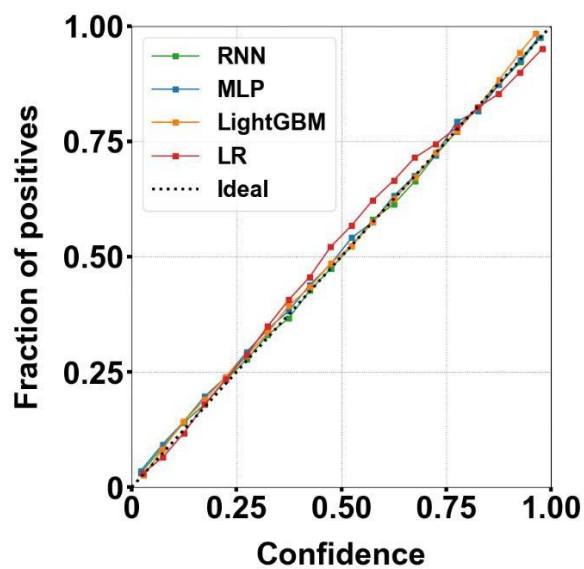


Supplemental Figure 5. Platt scaling plot used to calibrate the models. (A) Intradialytic hypotension (IDH)-1. (B) IDH-2. (C) IDH-3. Bin size = 0.05. RNN, recurrent neural network; MLP, multilayer perceptron; LightGBM, Light Gradient Boosting Machine; LR, logistic regression.

(A)



(B)



(C)

