Deep Learning Model for Real-Time Prediction of Intradialytic Hypotension

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

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 ≥20 mm Hg and/or a decrease in mean arterial pressure ≥10 mm Hg on the basis of the initial BP (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≤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: 396–406, 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...
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 $\leq 18$ 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.

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

282,396 hemodialysis sessions (10,053 patients) from September 2004 to December 2018

261,647 sessions (9292 patients) were included in the analysis

20,752 sessions were excluded due to
Age $<18$: 15,858 cases (476 patients)
Less than two BP measurements: 4271 cases (1476 patients)
Laboratory information, not available: 623 cases (260 patients)

Training set (70%)
181,600 sessions (6504 patients)
Validation set (5%)
13,347 sessions (465 patients)
Calibration set (5%)
16,319 sessions (464 patients)
Test set (20%)
50,381 sessions (1859 patients)

**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 $\frac{[2 \times \text{diastolic BP} + \text{systolic BP}]}{3}$. 

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

\[
\text{Precision} = \frac{TP}{TP + FP}
\]

\[
\text{Recall} = \frac{TP}{TP + FN}
\]

\[
\text{F1 score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}
\]

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 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 except at the early time points (Supplemental Figure 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, 0.92 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

<table>
<thead>
<tr>
<th>Table 1. Baseline characteristics of the hemodialysis sessions analyzed in this study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variables</td>
</tr>
<tr>
<td>Age (yr)</td>
</tr>
<tr>
<td>Male, n (%)</td>
</tr>
<tr>
<td>Hemodialysis type, n (%)</td>
</tr>
<tr>
<td>Hemodialysis</td>
</tr>
<tr>
<td>Hemodiafiltration</td>
</tr>
<tr>
<td>Others</td>
</tr>
<tr>
<td>The number of sessions per wk (times)</td>
</tr>
<tr>
<td>The time per session (h)</td>
</tr>
<tr>
<td>Incident hemodialysis, n (%)</td>
</tr>
<tr>
<td>Vascular access, n (%)</td>
</tr>
<tr>
<td>Arteriovenous fistula</td>
</tr>
<tr>
<td>Arteriovenous graft</td>
</tr>
<tr>
<td>Subcutaneously tunneled catheter</td>
</tr>
<tr>
<td>Temporary internal jugular venous catheterization</td>
</tr>
<tr>
<td>Others</td>
</tr>
<tr>
<td>Predialysis weight (kg)</td>
</tr>
<tr>
<td>Setting of ultrafiltration (L)</td>
</tr>
<tr>
<td>Use of anticoagulant, n (%)</td>
</tr>
<tr>
<td>Heparin</td>
</tr>
<tr>
<td>Nafamostat mesilate</td>
</tr>
<tr>
<td>None</td>
</tr>
<tr>
<td>Setting of blood flow rate (ml/min)</td>
</tr>
<tr>
<td>Predialysis systolic BP (mm Hg)</td>
</tr>
<tr>
<td>Predialysis diastolic BP (mm Hg)</td>
</tr>
<tr>
<td>Predialysis mean arterial pressure (mm Hg)</td>
</tr>
<tr>
<td>Predialysis heart rate (beats/min)</td>
</tr>
<tr>
<td>Blood findings</td>
</tr>
<tr>
<td>Hemoglobin (g/dl)</td>
</tr>
<tr>
<td>Albumin (g/dl)</td>
</tr>
<tr>
<td>Calcium (mg/dl)</td>
</tr>
<tr>
<td>Phosphate (mg/dl)</td>
</tr>
<tr>
<td>Sodium (mmol/L)</td>
</tr>
<tr>
<td>Potassium (mmol/L)</td>
</tr>
<tr>
<td>Dialysate sodium (mmol/L)</td>
</tr>
<tr>
<td>Dialysate potassium (mmol/L)</td>
</tr>
<tr>
<td>Dialysate calcium (mmol/L)</td>
</tr>
<tr>
<td>Dialysate bicarbonate (mmol/L)</td>
</tr>
<tr>
<td>Dialysate temperature (°C)</td>
</tr>
</tbody>
</table>

*Total number of patients was 9292.

bThe mean value was 3.7±0.6 h per session.
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.
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 analysis times (Supplemental Figure 3). When the predicted outcome models were evaluated according to the hemodialysis time, although there was a trade-off between precision and recall rates 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 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

---

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

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

IDH-1, intradialytic hypotension defined as nadir systolic BP <90 mm Hg; IDH-2, intradialytic hypotension defined as decrease in systolic BP ≥20 mm Hg and/or decrease in mean arterial pressure ≥10 mm Hg on the basis of BP at initial time point; IDH-3, intradialytic hypotension defined as decrease in systolic BP ≥20 mm Hg and/or decrease in mean arterial pressure ≥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.

*Compared with the receiver operating characteristic curve of RNN model by the Delong test.

---

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

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>RNN (%)</th>
<th>MLP (%)</th>
<th>LightGBM (%)</th>
<th>LR (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IDH-1</td>
<td>0.20</td>
<td>0.40</td>
<td>0.32</td>
<td>0.36</td>
</tr>
<tr>
<td>IDH-2</td>
<td>0.66</td>
<td>1.45</td>
<td>0.85</td>
<td>1.66</td>
</tr>
<tr>
<td>IDH-3</td>
<td>0.43</td>
<td>0.91</td>
<td>0.93</td>
<td>0.51</td>
</tr>
</tbody>
</table>

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 ≥20 mm Hg and/or decrease in mean arterial pressure ≥10 mm Hg on the basis of BP at initial time point; IDH-3, intradialytic hypotension defined as decrease in systolic BP ≥20 mm Hg and/or decrease in mean arterial pressure ≥10 mm Hg on the basis of BP at prediction time point.
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, interventricular 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 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.
Figure 4. | Feature rankings for the model predicting intradialytic hypotension (IDH). (A) IDH-1; (B) IDH-2; (C) IDH-3.
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.DOInc. 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. FI score for predicting intradialytic hypotension in deep and other machine learning and logistic regression models.

Supplemental Table 5. FI 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 (IDH).

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


hemodialysis treatment based on a convolutional neural network.  
*J Biomed Inform* 98: 103271, 2019


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.