BMC Nephrology

RESEARCH

Open Access

Research on the development of an intelligent prediction model for blood pressure variability during hemodialysis

Zhijian Ren[1,2], Minqiao Zhang[3], Pingping Wang[4], Kanan Chen[1], Jing Wang[1], Lingping Wu[1], Yue Hong[1], Yihui Qu[1], Qun Luo[1] and Kedan Cai[1*]

Abstract

Objective Blood pressure fluctuations during dialysis, including intradialytic hypotension (IDH) and intradialytic hypertension (IDHTN), are common complications among patients undergoing maintenance hemodialysis. Early prediction of IDH and IDHTN can help reduce the occurrence of these fluctuations. With the development of artificial intelligence, machine learning and deep learning models have become increasingly sophisticated in the field of hemodialysis. Utilizing machine learning to predict blood pressure fluctuations during dialysis has become a viable predictive method.

Methods Our study included data from 67,524 hemodialysis sessions conducted at Ningbo No.2 Hospital and Xiangshan First People's Hospital from August 1, 2019, to September 30, 2023. 47,053 sessions were used for model training and testing, while 20,471 sessions were used for external validation. We collected 45 features, including general information, vital signs, blood routine, blood biochemistry, and other relevant data. Data not meeting the inclusion criteria were excluded, and feature engineering was performed. The definitions of IDH and IDHTN were clarified, and 10 machine learning algorithms were used to build the models. For model development, the dialysis data were randomly split into a training set (80%) and a testing set (20%). To evaluate model performance, six metrics were used: accuracy, precision, recall, F1 score, ROC-AUC, and PR-AUC. Shapley Additive Explanation (SHAP) method was employed to identify eight key features, which were used to develop a clinical application utilizing the Streamlit framework.

Results Statistical analysis showed that IDH occurred in 56.63% of hemodialysis sessions, while the incidence of IDHTN was 23.53%. Multiple machine learning models (e.g., CatBoost, RF) were developed to predict IDH and IDHTN events. XGBoost performed the best, achieving ROC-AUC scores of 0.89 for both IDH and IDHTN in internal validation, with PR-AUC scores of 0.95 and 0.78, and high accuracy, precision, recall, and F1 scores. The SHAP method identified pre-dialysis systolic blood pressure, BMI, and pre-dialysis mean arterial pressure as the top three important features. It has been translated into a convenient application for use in clinical settings.

*Correspondence:
Kedan Cai
caikedan@ucas.edu.cn

Full list of author information is available at the end of the article

Conclusion Using machine learning models to predict IDH and IDHTN during hemodialysis is feasible and provides clinically reliable predictive performance. This can help timely implement interventions during hemodialysis to prevent problems, reduce blood pressure fluctuations during dialysis, and improve patient outcomes.

Introduction

End-stage renal disease (ESRD) poses a substantial public health challenge due to its increasing prevalence worldwide. Currently, over 3 million individuals receive treatment for ESRD globally, with this number continuing to grow, driven by an aging population and the rising incidence of diabetes and cardiovascular diseases [1, 2]. In mainland China alone, the China National Renal Data System (CNRDS) reported 916,647 hemodialysis (HD) patients as of December 2023. HD remains the most widely utilized renal replacement therapy; however, it is frequently complicated by hemodynamic fluctuations during treatment, including intradialytic hypertension (IDHTN) and intradialytic hypotension (IDH) [3]. These complications significantly impact the safety and adequacy of HD, as frequent changes in organ perfusion during dialysis can affect cardiac, central nervous system, gastrointestinal, and vascular access function, ultimately increasing the risk of mortality [4]. Effective management of these complications is, therefore, crucial for optimizing HD outcomes [5].

IDH and IDHTN have been most strongly associated with cardiovascular events and mortality. A retrospective analysis of 39,497 HD patient records revealed that IDH is significantly linked to myocardial infarction, hospitalization for heart failure or volume overload, and both cardiovascular and all-cause mortality [6]. Furthermore, IDH has been shown to cause cerebral ischemia, an additional form of end-organ damage. It is also associated with the accelerated loss of residual renal function and vascular access thrombosis. In addition to these long-term adverse outcomes, IDH and IDHTN frequently result in considerable patient discomfort during HD sessions, manifesting as dizziness, weakness, headache, nausea, and vomiting. These symptoms substantially reduce the quality of life for HD patients [8].

At present, nephrologists typically address IDH and IDHTN reactively, relying on clinical experience to intervene after these complications occur. Accurate risk prediction of hemodynamic fluctuations could enable dialysis staff to implement preventative measures, reducing their occurrence and impact. Consequently, there is a pressing need for improved prevention and management strategies for IDH and IDHTN to mitigate long-term adverse outcomes and enhance the quality of life for HD patients [6]. However, despite extensive research efforts, accurate prediction remains challenging due to the complex and multifactorial nature of these conditions across diverse clinical settings [7].

At present, medical artificial intelligence (AI) is playing an increasingly important role in various fields of nephrology, from diagnosis to treatment [8]. Machine learning (ML), a major branch of AI, offers promising tools for analyzing large-scale HD data to predict and manage blood pressure fluctuations [9].

Unlike other studies that use a single machine learning model to predict IDH, we collected a dataset from HD patients at Ningbo No. 2 Hospital and Xiangshan First People's Hospital to develop machine learning models for predicting both IDH and IDHTN. Totally 45 features were gathered from each HD session, extracted from a predefined time window prior to the start of dialysis, to analyze the factors influencing blood pressure fluctuations. The data from Ningbo No.2 Hospital was used as the training and internal validation set, while the data from Xiangshan First People's Hospital served as the external validation set. We developed and tested 10 machine learning models to predict blood pressure fluctuations in HD patients. Unlike previous studies that focused solely on predicting IDH and IDHTN, we compared the performance of these 10 models and selected the best-performing one. The 8 important features were selected, and a clinical application was developed.

Methods

Data collection and processing

Totally 276 patients were included who underwent hemodialysis between August 1, 2019, and September 30, 2023. Data on the hemodialysis process were stored using the Hua Mai Healthcare system at Ningbo No.2 Hospital and Xiangshan First People's Hospital. The exclusion criteria were as follows: (1) age under 18 years, (2) missing data ≥ 70%, (3) dialysis due to acute kidney injury (AKI), and (4) missing pre-dialysis systolic or diastolic blood pressure data. From Ningbo No.2 Hospital, 232 hemodialysis patients were included. After excluding 3 patients due to being under 18 years old, 4 patients due to AKI, 3 patients with ≥ 70% missing data, and 4 patients with missing pre-dialysis blood pressure data, finally, 218 patients encompassing 47,053 hemodialysis sessions were remained. This portion of the data was used for model training and internal validation. From Xiangshan First People's Hospital, 44 hemodialysis patients were included; this portion of the data was used for external validation. After excluding 1 patient due to AKI, 1 patient with ≥ 70% missing data, and 2 patients with missing pre-dialysis blood pressure data, finally 40 patients encompassing 20,471 hemodialysis sessions were remained.

After initial data processing, outliers were also excluded. Outliers included patients with pre-dialysis and intradialysis blood pressure < 30 mmHg or > 300 mmHg, as measured by the Yuwell YE655D sphygmomanometer. After excluding outliers, we compared various imputation methods, including Mean, KNN, Multiple, Random Forest, Median, and Decision Tree imputation. The performance of these methods was evaluated using four metrics: R^2, Mean Squared Error (MSE), Mean Absolute Error (MAE), and Relative Error (RE). Ultimately, we selected the random forest imputation method to fill in the missing values.

Additionally, data integrity, consistency, accuracy, and distribution checks were conducted, and the results met the required standards. After processing, data from 258 patients encompassing 67,524 hemodialysis sessions were included in the study. Figure 1 shows the entire workflow of this study [10, 11].

Definition of IDH and IDHTN

According to the Kidney Disease Outcome Quality Initiative (K/DOQI) guidelines, IDH was defined as a decrease in systolic blood pressure ≥ 20 mmHg or mean arterial pressure ≥ 10 mmHg during dialysis accompanied by related symptoms [12]. IDHTN was defined as an increase in systolic blood pressure ≥ 10 mmHg during or after dialysis over the last 3 or 4 sessions, accompanied by cardiovascular events [13]. Due to missing data on symptoms during dialysis and cardiovascular events, in this study, the authors defined IDH as a decrease in systolic blood pressure ≥ 20 mmHg or mean arterial pressure ≥ 10 mmHg during dialysis. IDHTN was defined as an increase in systolic blood pressure ≥ 10 mmHg during or after dialysis.

Feature selection

We selected 45 patient characteristics for analysis, including primary diseases (chronic glomerulonephritis, hypertension, diabetes mellitus, gout, other primary diseases), blood pressure measurements (pre-dialysis SBP, pre-dialysis DBP, 1-hour SBP, 2-hour SBP, 3-hour SBP, 4-hour SBP, 1-hour DBP, 2-hour DBP, 3-hour DBP, 4-hour DBP, pre-dialysis MAP, 1-hour MAP, 2-hour MAP, 3-hour MAP, 4-hour MAP), demographic characteristics (male, female, age, pre-dialysis weight, BMI, height), laboratory tests (WBC, hemoglobin, hematocrit, platelet, albumin, calcium, potassium, phosphorus, sodium), dialysis adequacy indexes (KT/V, URR), dialysis settings (ultrafiltration volume, ultrafiltration rate, dry weight, dialysis duration, dialysis frequency), and ultrasound and imaging examinations (cardiothoracic ratio (CTR), left ventricular mass index (LVMI)), as detailed in Table 1 [14]. Eight primary disease characteristics were processed as nominal features and converted to one-hot encoding.

For data processing, apart from converting primary diseases to one-hot encoding, we deleted abnormal data values, setting them as missing values, and used random forest imputation for missing values. According to definitions, we labeled normal blood pressure during dialysis as 0, hypotension during dialysis as 1, and hypertension during dialysis as 2 [15, 16].

Model development and comparison

We obtained a cohort dataset of hemodialysis patient sessions, with data from Ningbo No.2 Hospital used for model training and testing (80% for training, 20% for internal validation) to avoid overfitting. The remaining data from Xiangshan First People's Hospital were used for external validation. Ten machine learning models were used: Support Vector Machines (SVM), k-Nearest Neighbors (KNN), Decision Trees (DT), Random Forest (RF), Logistic Regression (LR), Naive Bayes (NB), Extreme Gradient Boosting (XGBoost), LightGBM, CatBoost, and Adaptive Boosting (AdaBoost). The final model parameters were optimized using grid search and manual tuning. Model performance was evaluated using six metrics: receiver operating characteristic area under the curve (ROC-AUC), precision-recall area under the curve (PR-AUC), accuracy, precision, recall, and F1-score [17]. Based on this, calibration and decision curves were further employed to comprehensively compare and evaluate models with similar performance.

Feature selection and model explanation

Proper interpretation of machine learning models is also an important and challenging task. The SHAP method was employed to rank the importance of input features and interpret the predictive model results. SHAP is based on cooperative game theory and provides an unified measure of feature importance, ensuring consistent and accurate explanations. SHAP offers both global and local explanations, elucidating the association between input features and IDH/IDHTN. Global explanations give an overall view of feature importance across the entire dataset, while local explanations provide insights into how features impact individual predictions. In this study, various SHAP visualizations were utilized for feature analysis, including the SHAP Bar and Dot Plot, which provides a high-level overview of feature impacts; the SHAP Force Plot, which illustrates individual predictions and the contributions of each feature; and the SHAP Heatmap, which highlights interactions between features across multiple observations. These tools collectively enhance the interpretability of the predictive models and ensure that the results are both transparent and understandable [18, 19].

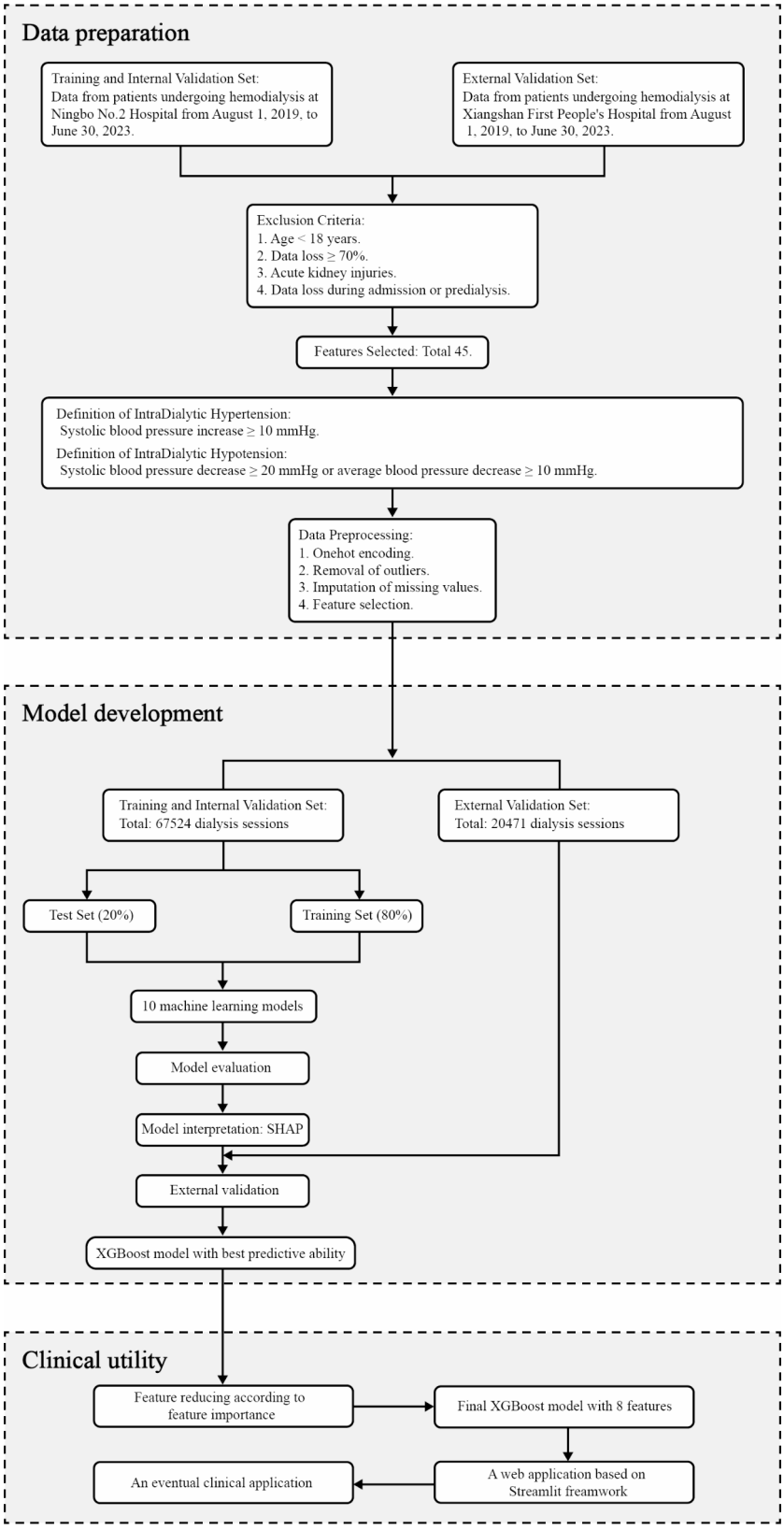

Fig. 1 Flow chart of the study design

Table 1 Baseline characteristics

	Training and internal validation cohort (*n* = 218)	External validation cohort (*n* = 40)	All cohort (*n* = 258)
Male(n;%)	141; 64.68%	21; 52.50%	162; 62.79%
Female(n;%)	77; 35.32%	19; 47.50%	96; 37.21%
Age(year)	61.8 ± 14.21	61.7 ± 11.21	61.77 ± 13.36
Height(cm)	165.57 ± 8.64	163.38 ± 9.03	164.9 ± 8.82
Pre-dialysis weight(kg)	60.53 ± 12.78	58.98 ± 11.07	60.06 ± 12.3
BMI(kg/m2)	21.98 ± 3.66	21.97 ± 2.92	21.98 ± 3.45
CGN(chronic glomerulonephritis)(n;%)	97; 44.50%	23; 57.50%	120; 46.51%
HTN(hypertension)(n;%)	162; 74.31%	40; 100.00%	202; 78.29%
DM(diabetes mellitus)(n;%)	33; 15.14%	8; 20.00%	41; 15.89%
Gout(n;%)	14; 6.42%	3; 7.50%	17; 6.59%
Others(n;%)	113; 51.83%	17; 42.50%	130; 50.39%
Pre-dialysis heart rate	77.69 ± 13.58	73.43 ± 12.29	76.39 ± 13.34
Pre-dialysis SBP(mmHg)	145.01 ± 23.59	150.25 ± 21.53	146.61 ± 23.11
Pre-dialysis DBP(mmHg)	75.09 ± 13.53	81.25 ± 13.05	76.97 ± 13.68
1-hour SBP(mmHg)	137.89 ± 22.52	140.22 ± 20.23	138.61 ± 21.87
2-hour SBP(mmHg)	135.69 ± 22.49	139.53 ± 19.73	136.87 ± 21.76
3-hour SBP(mmHg)	135.5 ± 23.11	139.40 ± 20.05	136.69 ± 22.29
4-hour SBP(mmHg)	132.26 ± 23.76	136.65 ± 20.30	133.60 ± 22.85
1-hour DBP(mmHg)	74.13 ± 13.05	78.57 ± 12.94	75.49 ± 13.18
2-hour DBP(mmHg)	74.78 ± 13.36	79.10 ± 12.59	76.10 ± 13.28
3-hour DBP(mmHg)	75.42 ± 13.62	79.72 ± 12.75	76.73 ± 13.51
4-hour DBP(mmHg)	74.65 ± 14.01	78.84 ± 12.67	75.93 ± 13.75
Pre-dialysis MAP(mmHg)	98.39 ± 14.22	104.25 ± 14.06	100.18 ± 14.42
1-hour MAP(mmHg)	95.38 ± 13.90	99.12 ± 14.15	96.53 ± 14.08
2-hour MAP(mmHg)	95.08 ± 14.37	99.24 ± 13.75	96.36 ± 14.31
3-hour MAP(mmHg)	95.44 ± 14.73	99.61 ± 13.95	96.72 ± 14.62
4-hour MAP(mmHg)	93.85 ± 15.30	98.11 ± 14.00	95.16 ± 15.05
WBC(white blood cell)(*10^9/L)	6.18 ± 2.08	5.31 ± 1.33	5.91 ± 1.92
Hb(hemoglobin)(g/L)	110.31 ± 17.75	108.40 ± 13.95	109.72 ± 16.70
Hct(hematocrit)(%)	33.7 ± 5.28	33.88 ± 4.32	33.76 ± 5.01
Plt(platelet)(*10^9/L)	185.72 ± 56.27	167.05 ± 51.29	180.01 ± 55.46
Alb(albumin)(g/L)	38.3 ± 4.40	38.48 ± 3.33	38.35 ± 4.10
Ca(calcium)(mmol/L)	2.27 ± 0.23	2.33 ± 0.24	2.29 ± 0.23
K(potassium)(mmol/L)	4.63 ± 0.80	5.01 ± 0.95	4.74 ± 0.87
P(phosphorus)(mmol/L)	1.67 ± 0.58	1.81 ± 0.58	1.71 ± 0.58
Na(sodium)(mmol/L)	138.89 ± 3.23	139.59 ± 11.79	139.11 ± 7.06
KT/V(urea clearance index)	1.57 ± 0.56	1.51 ± 0.26	1.55 ± 0.49
URR(urea reduction ratio)	0.71 ± 0.08	71.28 ± 6.72	22.30 ± 32.73
CTR(cardiothoracic ratio)	0.49 ± 0.05	0.55 ± 0.06	0.51 ± 0.06
LVMI(left ventricular mass index)(g/m2)	101.52 ± 42.20	49.39 ± 30.36	85.57 ± 45.77
DW(dry weight)(kg)	57.9 ± 12.48	57.60 ± 11.35	57.81 ± 12.14
UFV(ultrafiltration volume)(L)	2.63 ± 1.01	2.63 ± 0.85	2.63 ± 0.96
UFR(ultrafiltration rate)(%)	0.66 ± 0.25	0.66 ± 0.22	0.66 ± 0.24
DF(dialysis frequency)(per week)	2.8 ± 0.52	3.97 ± 0.13	3.16 ± 0.70
DD(dialysis duration)(h)	3.98 ± 0.14	3.97 ± 0.13	3.98 ± 0.13

SBP: Systolic blood pressure; DBP: Diastolic blood pressure; MAP: Mean arterial pressure

Webpage deployment tool

Webpage deployment tool based on the Streamlit framework: A web application was developed on the Python-based Streamlit framework. When the values of the 9 corresponding features from the final model are provided, the application can generate predictions of blood pressure changes during dialysis, including normal blood pressure, IDH, and IDHTN [20].

Statistical analysis

The data analysis, model development, evaluation, and creation of the web-based application were conducted using several software tools. Python version 3.10.5 (https://www.python.org) was used for coding and model building, while SPSS Statistics version 27.0 (https://www.ibm.com/spss) was employed for statistical analysis. Additionally, GraphPad Prism version 10.3.0 (https://www.graphpad.com) was utilized for creating high-quality graphs and visualizations.

Results

Patient characteristics

The final cohort included 67,524 hemodialysis sessions. The data from 47,053 sessions at Ningbo No.2 Hospital were used for model training and testing, while the data from 20,471 sessions at Xiangshan First People's Hospital were reserved for external validation. Comprehensive data collection encompassed general information, vital signs, complete blood counts, and blood biochemistry, totaling 35 features. Detailed baseline statistics of these features are provided in Table 1.

In data preprocessing, outliers were identified and removed or marked as missing values, followed by data imputation. 6 imputation methods—Mean, KNN, Multiple, RF, Median, and Decision Tree—were evaluated. The performance of these methods was assessed using four metrics: R^2, MSE, MAE, and RE. Detailed comparative results are provided in Table 2. Based on the comprehensive analysis of the results, the Random Forest imputation method demonstrated the best performance and was therefore selected for imputing missing values.

A thorough statistical analysis of blood pressure changes in hemodialysis patients was conducted. The main findings are illustrated in Fig. 2.The proportion of patients experiencing various blood pressure conditions during dialysis was evaluated, with 20.84% maintaining normal blood pressure, 56.63% developing IDH, and 22.53% experiencing IDHTN (Fig. 2A). The distribution of systolic blood pressure across different time points during dialysis was examined, revealing a stable trend with minimal fluctuations (Fig. 2B). The analysis of IDH incidents over time highlighted an upward trend, with the occurrence of IDH increasing as dialysis progressed (Fig. 2C). A similar examination of IDHTN showed an initial rise, peaking at the third hour, followed by a decline by the fourth hour (Fig. 2D).Statistical data indicate that IDH remains the most prevalent type of blood pressure fluctuation during dialysis. The progressively increasing trend of IDH suggests a rising risk of hypotension as dialysis duration extends, underscoring the importance of vigilant blood pressure management [4].

Table 2 Performance comparison of imputation methods

Method	R^2	MSE	MAE	RE
Mean	0.6326	0.0831	0.2209	0.2545
KNN	0.7395	0.0626	0.1827	0.2104
Multiple	0.8484	0.0414	0.1462	0.1685
Random Forest	0.8821	0.0341	0.1281	0.1477
Median	0.6543	0.0798	0.2150	0.2476
Decision Tree	0.8600	0.0380	0.1350	0.1556

R^2: Coefficient of Determination; MSE: Mean Squared Error; MAE: Mean Absolute Error; RE: Relative Error

Model development and performance comparison

Through the analysis of results from 10 machine learning models (Tables 3 and 4), the XGBoost, RF, and CatBoost models consistently demonstrated superior performance across most evaluation metrics for both IDH and IDHTN predictions. Internal validation results, including ROC and PR curves for each model, are shown in Fig. 3. Among all models, XGBoost, RF, and CatBoost achieved the highest areas under the ROC and PR curves. To further assess the performance of these three models, calibration and decision curves were employed. Figure 4A and B indicate that XGBoost outperformed RF and CatBoost, providing the most reliable predictions. The XGBoost algorithm is an ensemble method based on decision trees, where individual trees are built sequentially. During the construction of each decision tree, weights are assigned to independent variables to optimize predictions for the target outcome. XGBoost uses a gradient boosting framework, iteratively refining the model by adding decision trees that correct errors from previous iterations. This iterative refinement makes XGBoost a highly robust and accurate predictive model [21].

Identification of the final model

Upon comparing the performance metrics, XGBoost was identified as the final model due to its superior results. For IDH prediction, XGBoost achieved a high ROC-AUC of 0.89 and a PR-AUC of 0.95. Additionally, it maintained a balanced performance with an accuracy of 0.84, precision of 0.87, recall of 0.93, and an F1-score of 0.90. Similarly, for IDHTN prediction, XGBoost performed admirably, with a ROC-AUC of 0.89, PR-AUC of 0.78, accuracy of 0.85, precision of 0.77, recall of 0.64, and an F1-score of 0.70. These results underscore the robustness

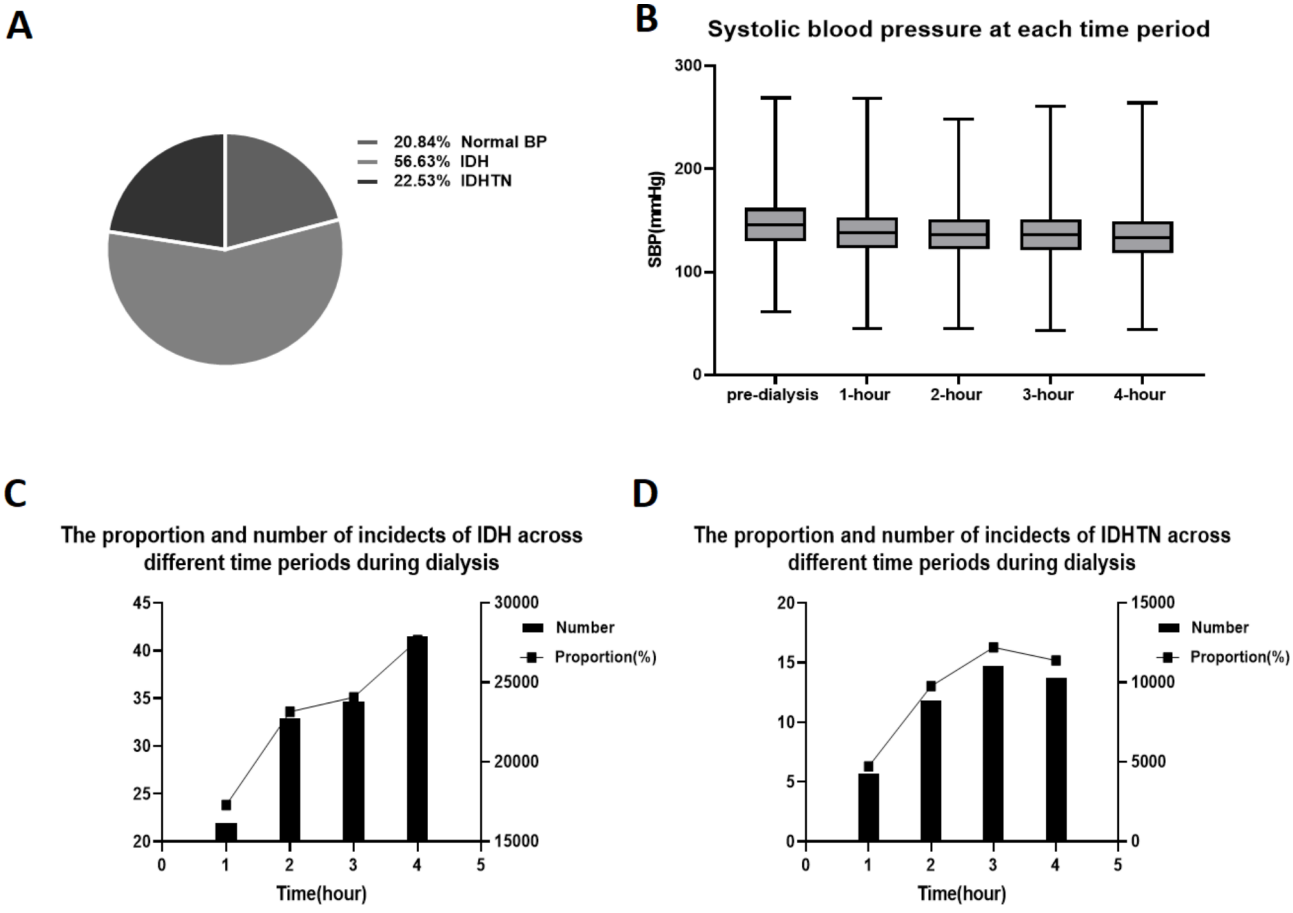

Fig. 2 Data analysis of blood pressure changes during dialysis. (**A**) the proportion of patients with different blood pressure conditions during dialysis. (**B**) the distribution of systolic blood pressure (SBP) at various time points during dialysis. (**C**) the proportion and number of IDH incidents at different times during dialysis. (**D**) the proportion and nuer of IDHTN incidents at different times during dialysis. IDH: Intradialytic Hypotension; IDHTN: Intradialytic Hypertension

Table 3 IDH machine learning results

	XGBoost	SVM	KNN	DT	RF	LR	NB	AdaBoost	LightGBM	CatBoost
ROC-AUC	0.89	0.87	0.84	0.72	0.89	0.79	0.74	0.81	0.88	0.89
PR-AUC	0.95	0.94	0.92	0.89	0.95	0.90	0.85	0.91	0.95	0.95
accuracy	0.84	0.82	0.81	0.76	0.83	0.76	0.71	0.77	0.83	0.84
precision	0.87	0.83	0.84	0.84	0.85	0.79	0.82	0.80	0.85	0.86
recall	0.93	0.93	0.90	0.83	0.92	0.90	0.76	0.90	0.93	0.92
F1-score	0.90	0.88	0.87	0.83	0.88	0.84	0.79	0.85	0.88	0.89

Table 4 IDHTN machine learning results

	XGBoost	SVM	KNN	DT	RF	LR	NB	AdaBoost	LightGBM	CatBoost
ROC-AUC	0.89	0.87	0.84	0.72	0.89	0.79	0.74	0.81	0.88	0.89
PR-AUC	0.78	0.76	0.72	0.66	0.78	0.60	0.55	0.62	0.77	0.79
accuracy	0.85	0.82	0.81	0.76	0.83	0.76	0.71	0.77	0.83	0.84
precision	0.77	0.76	0.72	0.59	0.75	0.63	0.50	0.64	0.76	0.77
recall	0.64	0.54	0.59	0.61	0.62	0.42	0.59	0.46	0.59	0.62
F1-score	0.70	0.63	0.65	0.60	0.68	0.50	0.54	0.54	0.67	0.69

XGBoost: Extreme Gradient Boosting; SVM: Support Vector Machine; KNN: K-Nearest Neighbors; DT: Decision Tree; RF: Random Forest; LR: Logistic Regression; NB: Naive Bayes; AdaBoost: Adaptive Boosting; LightGBM: Light Gradient Boosting Machine; CatBoost: Categorical Boosting; ROC-AUC: Receiver Operating Characteristic - Area Under the Curve; PR-AUC: Precision-Recall - Area Under the Curve

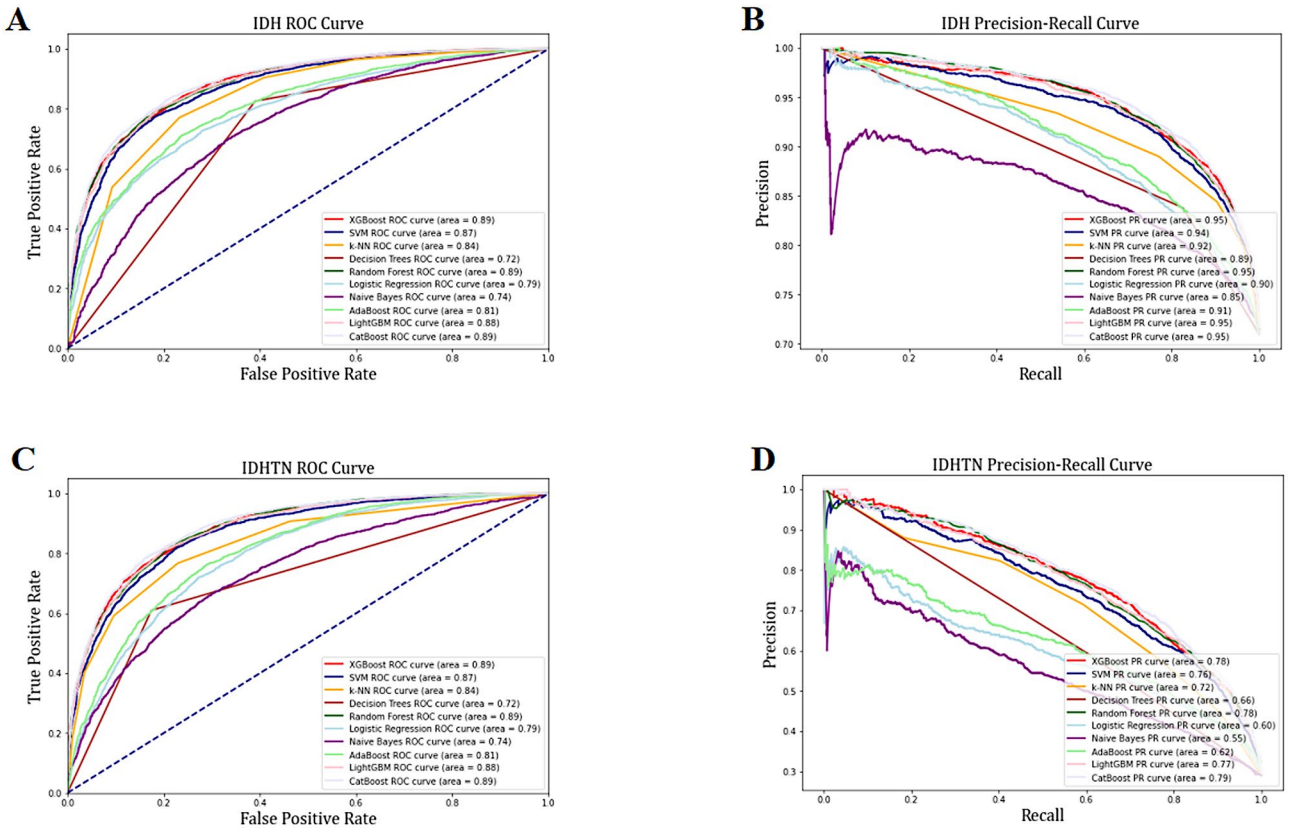

Fig. 3 The internal validation results of 10 machine learning models for predicting intradialytic hypotension and intradialytic hypertension, including ROC and PR curves. (**A**) the ROC curves of the 10 models for predicting intradialytic hypotension, (**B**) the PR curves of the 10 models for predicting intradialytic hypotension, (**C**) the ROC curves of the 10 models for predicting intradialytic hypertension, (**D**) the PR curves of the 10 models for predicting intradialytic hypertension

and reliability of the XGBoost model for predicting both IDH and IDHTN.

External validation of the final model

The robustness of the XGBoost model was further tested through external validation using an independent dataset of 20,471 hemodialysis sessions. The model's performance remained consistently high, with ROC-AUC and PR-AUC scores comparable to those obtained during internal validation. Among the three models—XGBoost, RF, and CatBoost—all achieved ROC-AUC values of at least 0.93 for predicting IDH and IDHTN. The PR-AUC values for predicting IDH were at least 0.97, while the PR-AUC values for predicting IDHTN were slightly lower but still at least 0.85. Figure 5 illustrates the ROC and PR curves for both IDH and IDHTN predictions during external validation, confirming the model's generalizability and effectiveness in a real-world setting.

Model explanation

To enhance the interpretability of the XGBoost model, SHAP was employed. SHAP assigns each feature an importance value for a particular prediction, leveraging Shapley values from cooperative game theory to provide consistent and interpretable explanations of model outputs. It decomposes a prediction into contributions from each feature, ensuring fairness and accuracy in feature impact analysis. The SHAP summary bar plot and dot plot for all 35 features used in the prediction are shown in Supplemental Fig. 1 and Supplemental Fig. 2.The impact of the top 8 features on the model's output is highlighted in a SHAP summary bar plot, with pre-dialysis SBP and BMI emerging as significant predictors, as presented in Fig. 6A.The impact of the top 8 features on the model's output is highlighted in a SHAP summary dot plot, as showed in Fig. 6B.The color gradient ranging from blue to red represents the magnitude of the feature values for each data point. Red corresponds to higher values of the respective feature, while blue indicates lower values. This color-coding helps to illustrate how the variation in feature values affects the SHAP values and the model's prediction. In Fig. 6C, the X-axis represents standardized pre-dialysis SBP values, while the Y-axis indicates the contribution of pre-dialysis SBP to the model's prediction. The color gradient, ranging from blue to red, corresponds to pre-dialysis DBP

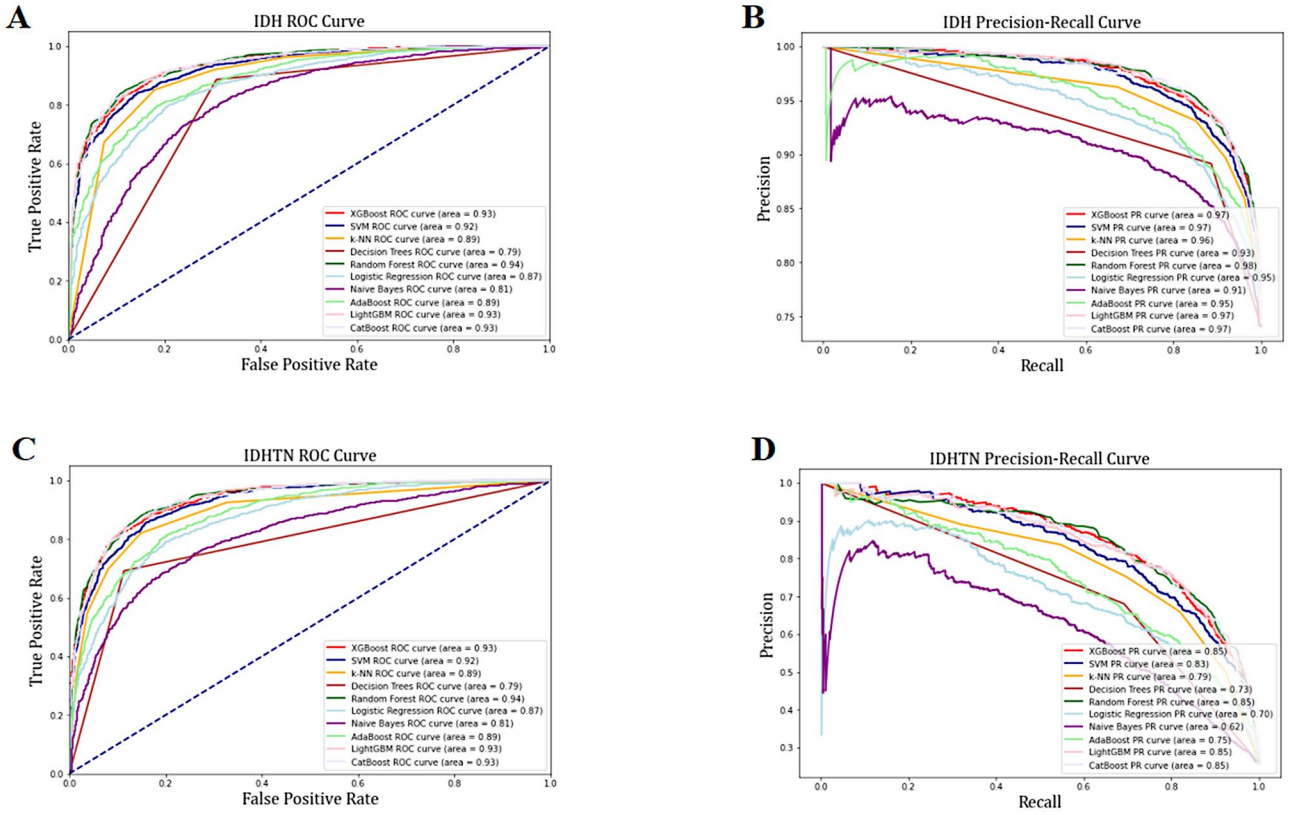

Fig. 4 (**A**) Calibration curves for XGBoost, RandomForest, and CatBoost models. (**B**) Decision curves for XGBoost, RandomForest, and CatBoost models

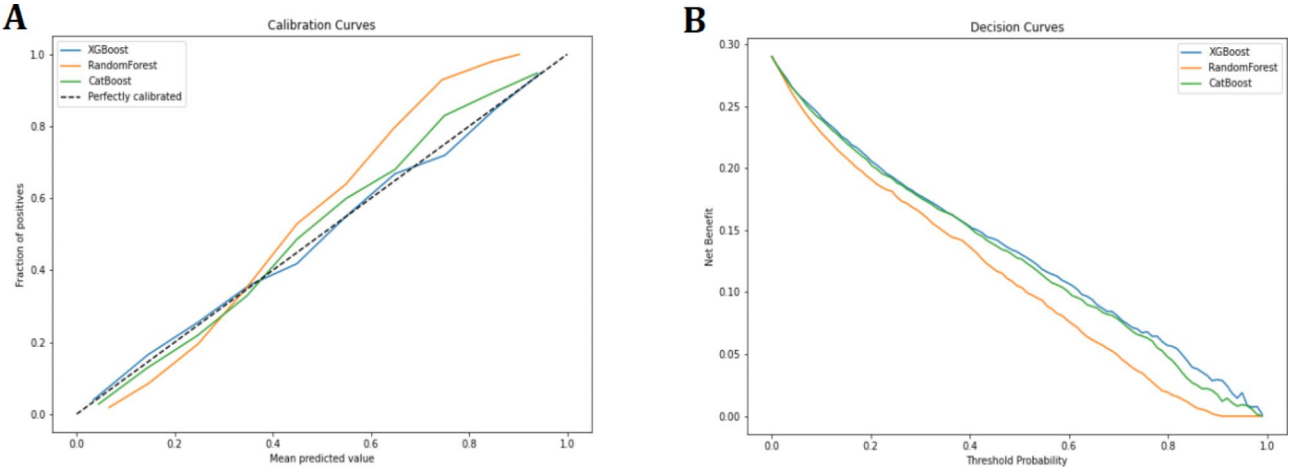

Fig. 5 The external validation results of 10 machine learning models for predicting intradialytic hypotension and intradialytic hypertension, including ROC and PR curves. (**A**) the ROC curves of the 10 models for predicting intradialytic hypotension, (**B**) the Precision-Recall curves of the 10 models for predicting intradialytic hypotension, (**C**) the ROC curves of the 10 models for predicting intradialytic hypertension, (**D**) the Precision-Recall curves of the 10 models for predicting intradialytic hypertension

values. There is a clear negative correlation between pre-dialysis SBP and its SHAP value, suggesting that higher pre-dialysis SBP reduces the predicted outcome. Additionally, lower DBP values (blue) correspond to higher SHAP values for SBP, whereas higher DBP values (red) correspond to lower SHAP values, indicating an interaction effect where DBP modulates the influence of SBP on the prediction. The SHAP force plot provides a detailed explanation for a specific sample, visually illustrating how individual features contribute to the model's prediction. Red arrows represent features that increase the prediction, while blue arrows represent features that decrease it. For example, "pre-dialysis MAP" and "pre-dialysis SBP" push the prediction higher, while "hematocrit,"

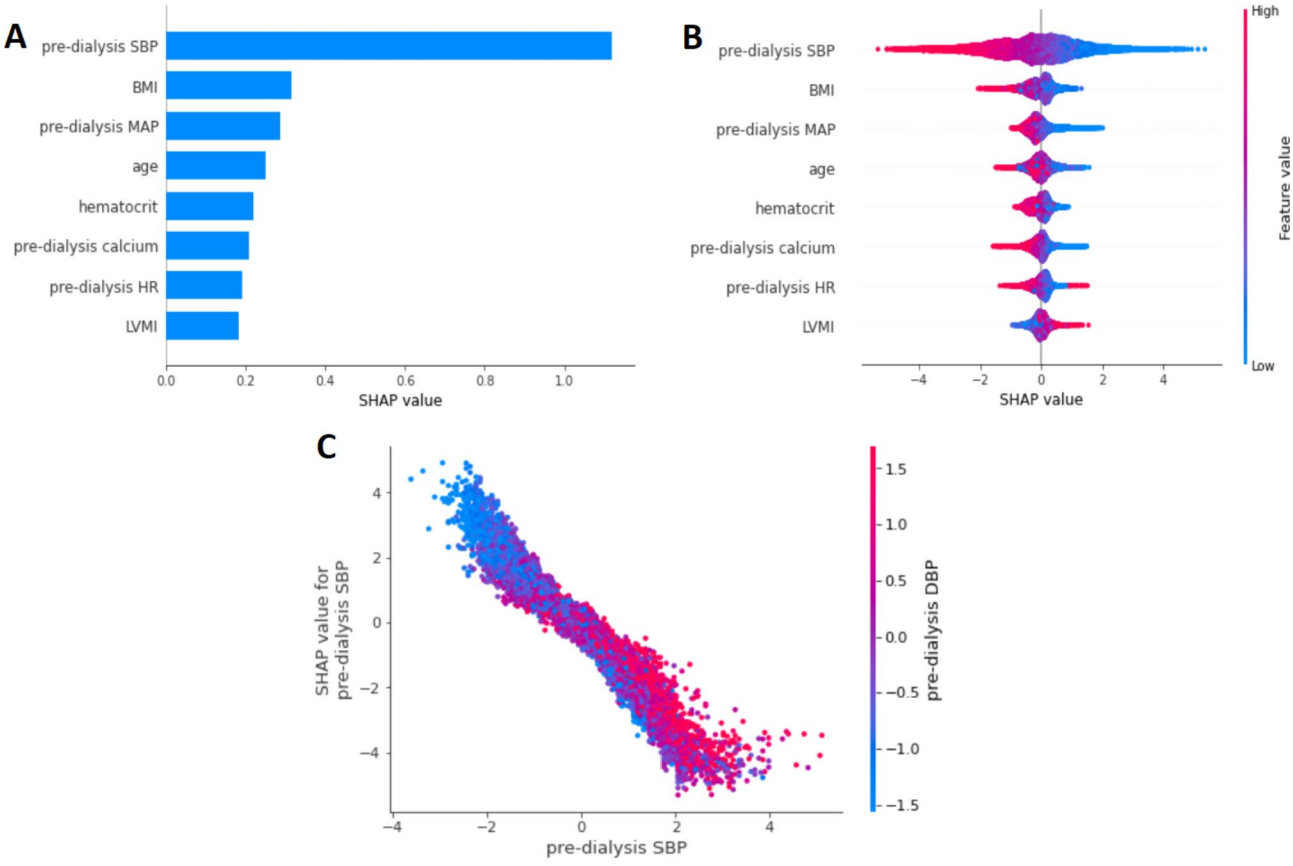

Fig. 6 (**A**) SHAP summary bar plot. (**B**) SHAP summary dot plot. (**C**) Scatter plot of pre-dialysis SBP vs. its SHAP values, colored by pre-dialysis DBP, indicating a negative relationship. Red indicates that a higher feature value has the corresponding impact, as indicated by the x-axis, on model output. Blue indicates the impact of lower feature values on model output

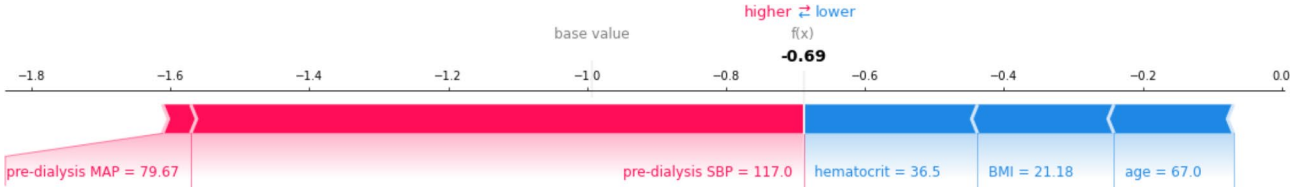

Fig. 7 SHAP force plot. Feature contributions: pre-dialysis MAP (79.67) and pre-dialysis SBP (117.0): Positive influence. Hematocrit (36.5), BMI (21.18), and age (67.0): Negative influence. Red indicates higher prediction, blue indicates lower prediction

"BMI," and "age" have a negative influence, pulling the prediction lower, as presented in Fig. 7. Additionally, a heatmap represents the correlation matrix of SHAP values for the eight most important features is displayed in Fig. 8. The color gradient spans from red to teal, where red indicates a strong positive correlation (close to 1) and teal indicates a strong negative correlation (close to -1). The intensity of the color reflects the strength of the correlation; deeper shades represent stronger correlations, while lighter shades indicate weaker correlations. Each cell corresponds to the correlation between two features, with feature names labeled along the y-axis and x-axis. This heatmap offers a visual understanding of the relationships and interdependencies among the selected feature [22, 23].

Clinical utility

Based on the SHAP summary bar plot and SHAP summary dot plot, we identified the 8 most important features, including (1. Pre-dialysis MAP, 2. LVMI, 3. Age, 4. Hematocrit, 5. Pre-dialysis Calcium, 6. Pre-dialysis Heart Rate, 7. BMI, and 8. Pre-dialysis SBP). The simplified model's ROC-AUC was 0.88, PR-AUC was 0.77, Accuracy was 0.83, Precision was 0.74, Recall was 0.62, and F1-Score was 0.68. Its ROC-AUC curve, PR-AUC curve, calibration curve, and decision curve are shown in Supplemental Fig. 3. Using the 8 features and adding

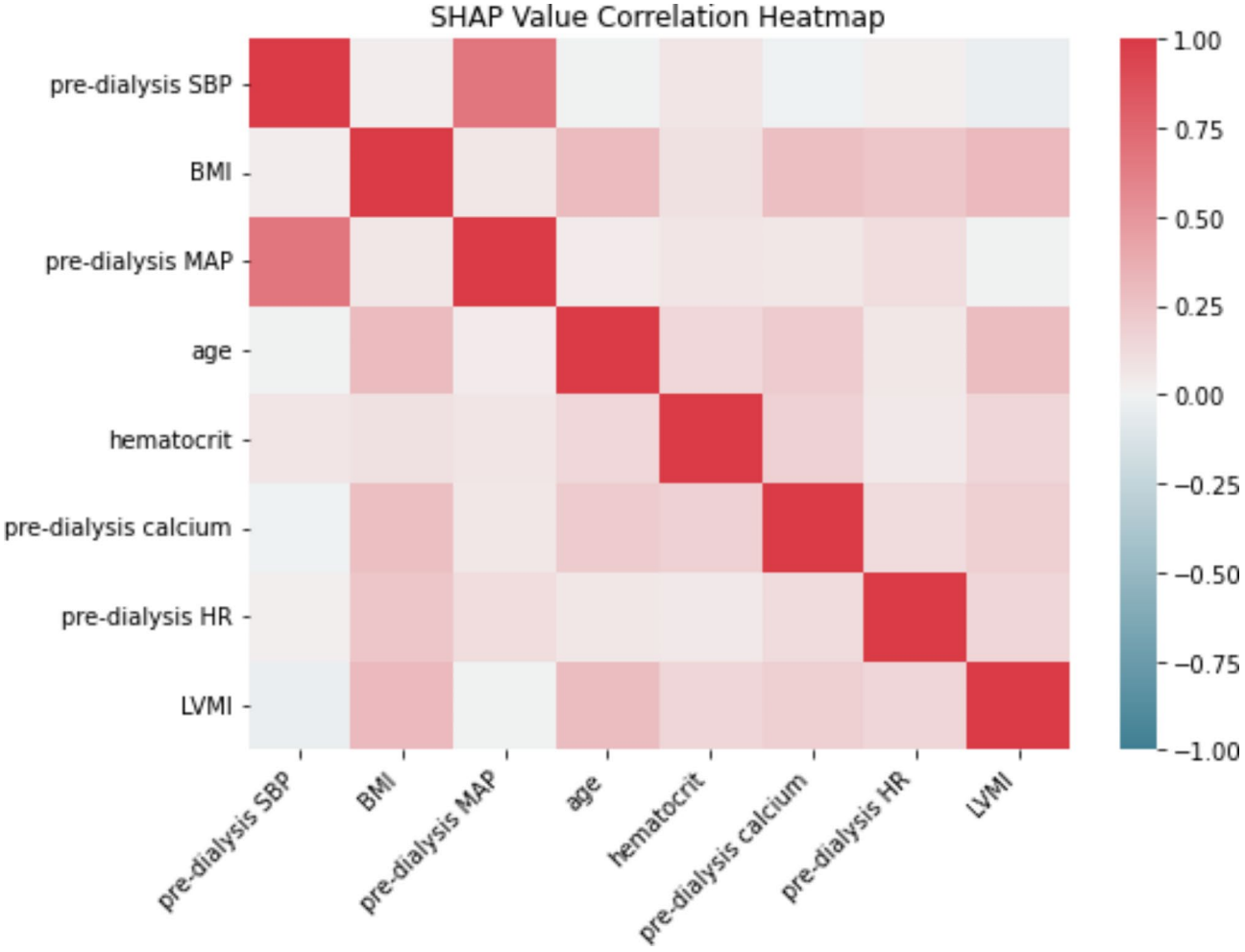

Fig. 8 The heatmap of SHAP values for the most important 8 features. The color range spans from red to teal, with red indicating positive correlation (correlation coefficient close to 1) and teal indicating negative correlation (correlation coefficient close to -1). The intensity of the color represents the strength of the correlation. Each cell in the heatmap represents the correlation between two features, with the labels on the left and bottom indicating the specific feature names. This heatmap provides a visual representation of the relationships between the selected features

ultrafiltration volume, we streamlined our XGBoost model and developed a web-based prediction application using the Streamlit framework. The software interface is shown in Fig. 9. This application allows users to input relevant clinical data and generate predictive outcomes, such as the likelihood of IDH [24].

Discussion

Our study presented the development and validation of machine learning models for predicting IDH and IDHTN in hemodialysis patients, utilizing extensive data from two hospitals. We demonstrated the feasibility and potential of these models to improve patient outcomes through accurate, real-time prediction of blood pressure fluctuations during dialysis. A comparative analysis of ten machine learning models revealed that XGBoost consistently outperformed other algorithms, such as Random Forest (RF) and CatBoost, across multiple metrics, including ROC-AUC, PR-AUC, accuracy, precision, recall, and F1-score. XGBoost's superior performance was attributed to its ability to handle complex data structures and capture intricate relationships between features, with high ROC-AUC and PR-AUC values in both internal and external validations underscoring its reliability and generalizability. This study was the first to showcase multiple machine learning models for predicting IDH and IDHTN, particularly addressing the unclear and complex mechanisms underlying these conditions, which involved factors such as pre-dialysis SBP and MAP. Our models, based on data from Hua Mai Healthcare System, offered a promising tool for identifying patients who could benefit from targeted interventions [25].

A key challenge in deploying machine learning models in clinical practice was their interpretability. We addressed this issue using SHAP methods, which provided both global and local explanations for model

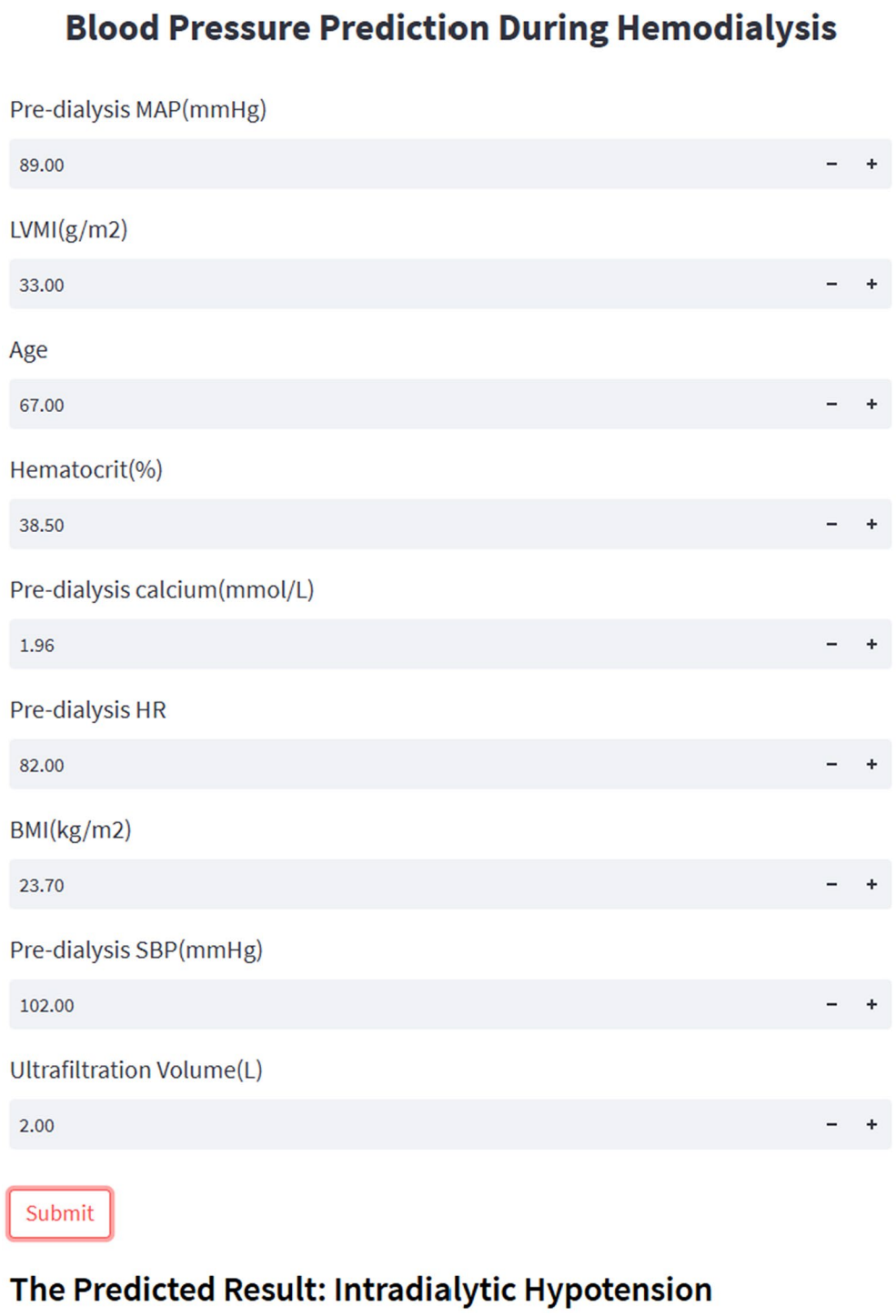

Fig. 9 Convenient application for clinical utility. The final prediction model with nine clinical features is deployed for use in predicting intradialytic blood pressure outcomes. Upon entering the actual values for each of the nine features (pre-dialysis MAP, LVMI, age, hematocrit, pre-dialysis calcium, pre-dialysis HR, BMI, pre-dialysis SBP and ultrafiltration volume), the application automatically classifies the prediction into categories such as "normal blood pressure", "intradialytic hypotension" or "intradialytic hypertension" In this instance, based on the entered data, the result was classified as "intradialytic hypotension"

predictions. SHAP analysis highlighted pre-dialysis SBP and pre-dialysis MAP as significant predictors of IDH and IDHTN. This insight aligned with clinical understanding, as blood pressure dynamics were critical indicators of dialysis outcomes. SHAP summary bar plots, dot plots, force plots, and heatmaps clearly illustrated the impact of various features on the model's predictions [26].

Implementing our web-based prediction application in clinical settings could transform the management of hemodialysis patients through proactive interventions. By accurately predicting IDH and IDHTN, the application enabled healthcare providers to adjust treatment regimens, such as fluid removal rates and antihypertensive medication dosages, to mitigate adverse events. This real-time predictive capability could significantly enhance patient safety and the overall effectiveness of dialysis treatment. Moreover, integrating the application into hemodialysis systems for real-time prediction, validated through clinical trials, held promise for reducing complications and improving patient outcomes. This advancement contributed to medical technology and aligned with broader socioeconomic goals of reducing healthcare costs associated with managing ESRD complications [27].

Our study highlighted the importance of controlling pre-dialysis blood pressure changes as a significant factor in promoting IDH and IDHTN. Although potassium and phosphate were generally considered the most impactful electrolytes for hemodialysis patients, our study found that pre-dialysis calcium had the greatest influence on blood pressure changes during dialysis, followed by pre-dialysis phosphate. This finding diverged from current understanding and warranted further investigation [28].

Despite the encouraging results, our study had several limitations. The dataset was sourced from two hospitals in China, which may have limited the generalizability of the findings to other populations. Additionally, while the model performed well in predicting IDH and IDHTN, it did not account for other dialysis complications. Data collection and labeling were potentially influenced by biases from individual healthcare providers and institutional protocols, and the dataset, spanning nearly two years, resulted in a limited sample size. Furthermore, missing and anomalous data could have impacted the accuracy of the analysis. The model also did not consider potential interventions, such as the use of antihypertensive drugs or changes in ultrafiltration settings after treatment initiation, which may have introduced inaccuracies, as only historical patient data were available for training [29]. Furthermore, pre-dialysis SBP, BMI, and pre-dialysis MAP were identified as the three most influential factors affecting intradialytic blood pressure changes. The clinical relevance of these variables in predicting IDH and IDHTN warrants further investigation.

In future studies, we plan to conduct in-depth analyses of key feature variables, enhance the functionality and user interface of the web-based application, and expand the dataset to include a broader and more diverse patient population. Expanding the study area to incorporate multicenter datasets will strengthen the generalizability of the findings. Additionally, integrating more clinical parameters will improve the model's comprehensiveness [30]. We also aim to standardize monitoring equipment for hemodialysis patients and control environmental variability to minimize research bias. Based on the application model, timely interventions during hemodialysis—such as adjusting ultrafiltration volume or administering blood pressure-regulating medications—could be implemented to stabilize patients' blood pressure. Expanding and refining the training and testing datasets, along with optimizing the model code, will be essential for developing a robust clinical application capable of real-time blood pressure prediction. Usability testing of the software and collecting feedback from hemodialysis patients will further guide improvements, ensuring the software meets clinical needs effectively [31].

Overall, this paper presented multiple machine learning models for predicting IDH and IDHTN, and identified key factors influencing blood pressure changes during dialysis using SHAP methods. We demonstrated the practical utility of machine learning in predicting these fluctuations and highlighted critical features that required attention. Moving forward, we would focus on refining and enhancing the clinical application software to ensure its readiness for mature use in clinical practice. This included further developing the predictive model to increase its maturity, improving its interpretability for clinicians, and ensuring its broad acceptance across different institutions.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12882-025-03959-x.

Supplementary Material 1

Acknowledgements

We thanked Ningbo No.2 Hospital and Xiangshan First People's Hospital for their data collection support.

Author contributions

Zhijian Ren conducted model development, software creation, and manuscript writing. Minqiao Zhang and Pingping Wang organized the data and created the charts. Kanan Chen, Jing Wang, Lingping Wu, Yue Hong, and Yihui Qu collected the data. Qun Luo and Kedan Cai provided manuscript guidance.

Funding

This work was supported in part by funds from the Zhejiang Provincial Medical and Health Project (2024KY336), the Provincial and Municipal Joint Medical Key Discipline (2022-S03), the Ningbo Health Science and Technology Plan Project (2023Y89), the Ningbo Major Research and Development Plan Project(2024Z213)and the Ningbo City Public Welfare Research Program (2024S162).

Data availability

Approval for this study was obtained from the ethics committee of Ningbo No.2 Hospital. As the datasets were considered the property of these institutions, data release to third parties requires permission from the institutional Health Data Oversight Committee (HDOC). Requests for data can be made to the HDOC Administrator at Ningbo No.2 Hospital and will be

evaluated based on principles such as public benefit, fairness, transparency, and responsible stewardship. The HDOC ensures no sale or barter of data, restricts redisclosure, limits data to the minimum necessary, and reviews for conflicts of interest. Data releases require a data use agreement, and requests are typically reviewed within 2–4 weeks. Once approved, the Technology Development Group will facilitate the data provision.

Declarations

Ethics approval and consent to participate

In China, all experimental protocols were approved by the Ethics Committee of Ningbo Second Hospital (Ethics Approval Number: KY202301501) and the Ethics Committee of Xiangshan County First People's Hospital (Ethics Approval Number: XYYJ-2024-660). This study utilized de-identified data from patients undergoing hemodialysis at Ningbo Second Hospital and Xiangshan County First People's Hospital between August 1, 2019, and September 30, 2023. The data no longer contained personally identifiable information, and the analysis posed minimal risk to participants. In accordance with Article 5(1) of Sect. 5, Chap. 2, from the 2023 Guidelines for the Establishment of Ethics Committees for Clinical Research Involving Humans, as well as the Declaration of Helsinki, the Ethics Committees granted a waiver of informed consent due to the retrospective nature of the study.

Competing interests

The authors declare no competing interests.

Clinical trial number

Not applicable.

Author details

[1]Department of Nephrology, Ningbo No.2 Hospital, Ningbo, PR China
[2]Department of Nephrology, Ninghai County Hospital of Traditional Chinese Medicine, Ningbo, PR China
[3]Department of Nephrology, the First People's Hospital of Xiangshan, Ningbo 315700, PR China
[4]Department of Rehabilitation, Ninghai First Hospital, Ningbo, PR China

Received: 29 August 2024 / Accepted: 10 January 2025
Published online: 17 February 2025

References

1. Viecelli AK, Lok CE. Hemodialysis vascular access in the elderly-getting it right. Kidney Int. 2019;95(1):38–49. https://doi.org/10.1016/j.kint.2018.09.016.
2. Yang X, Zhao D, Yu F, et al. An optimized machine learning framework for predicting intradialytic hypotension using indexes of chronic kidney disease-mineral and bone disorders. Comput Biol Med. 2022;145:105510. https://doi.org/10.1016/j.compbiomed.2022.105510.
3. Hirawa N. Blood pressure management in hemodialysis patients. Hypertens Res. 2023;46(7):1807–9. https://doi.org/10.1038/s41440-023-01279-x.
4. Gómez-Pulido JA, Gómez-Pulido JM, Rodríguez-Puyol D, et al. Predicting the Appearance of Hypotension during Hemodialysis Sessions using machine learning classifiers. Int J Environ Res Public Health. 2021;18(5):2364. https://doi.org/10.3390/ijerph18052364.
5. Keane DF, Raimann JG, Zhang H, et al. The time of onset of intradialytic hypotension during a hemodialysis session associates with clinical parameters and mortality. Kidney Int. 2021;99(6):1408–17. https://doi.org/10.1016/j.kint.2021.01.018.
6. Kanbay M, Ertuglu LA, Afsar B, et al. An update review of intradialytic hypotension: concept, risk factors, clinical implications and management. Clin Kidney J. 2020;13(6):981–93. https://doi.org/10.1093/ckj/sfaa078.
7. Dashtban A, Mizani MA, Pasea L, et al. Identifying subtypes of chronic kidney disease with machine learning: development, internal validation and prognostic validation using linked electronic health records in 350,067 individuals. EBioMedicine. 2023;89:104489. https://doi.org/10.1016/j.ebiom.2023.104489.
8. Hu J, Xu J, Li M, et al. Identification and validation of an explainable prediction model of acute kidney injury with prognostic implications in critically ill children: a prospective multicenter cohort study. EClinicalMedicine. 2024;68:102409. https://doi.org/10.1016/j.eclinm.2023.102409.
9. Greener JG, Kandathil SM, Moffat L, Jones DT. A guide to machine learning for biologists. Nat Rev Mol Cell Biol. 2022;23(1):40–55. https://doi.org/10.1038/s41580-021-00407-0.
10. Graterol Torres F, Molina M, Soler-Majoral J, et al. Evolving concepts on inflaatory biomarkers and malnutrition in chronic kidney disease. Nutrients. 2022;14(20):4297. https://doi.org/10.3390/nu14204297.
11. Yang X, Zhao D, Yu F, et al. Boosted machine learning model for predicting intradialytic hypotension using serum biomarkers of nutrition. Comput Biol Med. 2022;147:105752. https://doi.org/10.1016/j.compbiomed.2022.105752.
12. Flythe JE, Xue H, Lynch KE, et al. Association of mortality risk with various definitions of intradialytic hypotension. J Am Soc Nephrol. 2015;26(3):724–34. https://doi.org/10.1681/ASN.2014020222.
13. Bansal N, Artinian NT, Bakris G, et al. American Heart Association Council on the kidney in Cardiovascular Disease; Council on Cardiovascular and Stroke nursing; and Council on Epidemiology and Prevention. Hypertension in patients treated with In-Center maintenance hemodialysis: current evidence and Future opportunities: A Scientific Statement from the American Heart Association. Hypertension. 2023;80(6):e112–22. https://doi.org/10.1161/HYP.0000000000000230.
14. Clift AK, Mahon H, Khan G, et al. Identifying patients with undiagnosed small intestinal neuroendocrine tumours in primary care using statistical and machine learning: model development and validation study. Br J Cancer. 2024;131(2):305–11. https://doi.org/10.1038/s41416-024-02736-1.
15. Hecking M, Madero M, Port FK, et al. Fluid volume management in hemodialysis: never give up! Kidney Int. 2023;103(1):2–5. https://doi.org/10.1016/j.kint.2022.09.021.
16. Yoo D, Divard G, Raynaud M, et al. A machine learning-driven virtual Biopsy System for kidney transplant patients. Nat Coun. 2024;15(1):554. https://doi.org/10.1038/s41467-023-44595-z.
17. Kishi S, Kadoya H, Kashihara N. Treatment of chronic kidney disease in older populations. Nat Rev Nephrol. 2024;20(9):586–602. https://doi.org/10.1038/s41581-024-00854-w.
18. Kotanko P, Zhang H, Wang Y. Artificial Intelligence and Machine Learning in Dialysis: Ready for Prime Time? Clin J Am Soc Nephrol. 2023;18(6):803–5. https://doi.org/10.2215/CJN.0000000000000089.
19. Correa S, Mc Causland FR. Leveraging Deep Learning to Improve Safety of Outpatient Hemodialysis. Clin J Am Soc Nephrol. 2021;16(3):343–4. https://doi.org/10.2215/CJN.00450121.
20. Huang JC, Tsai YC, Wu PY, et al. Predictive modeling of blood pressure during hemodialysis: a comparison of linear model, random forest, support vector regression, XGBoost, LASSO regression and ensele method. Comput Methods Programs Biomed. 2020;195:105536. https://doi.org/10.1016/j.cmpb.2020.105536.
21. Lee H, Yun D, Yoo J, et al. Deep learning model for real-time prediction of Intradialytic Hypotension. Clin J Am Soc Nephrol. 2021;16(3):396–406. https://doi.org/10.2215/CJN.09280620.
22. Sars B, van der Sande FM, Kooman JP. Intradialytic hypotension: mechanisms and outcome. Blood Purif. 2020;49(1–2):158–67. https://doi.org/10.1159/000503776.
23. Wang F, Wang Y, Tian Y, et al. Pattern recognition and prognostic analysis of longitudinal blood pressure records in hemodialysis treatment based on a convolutional neural network. J Biomed Inf. 2019;98:103271. https://doi.org/10.1016/j.jbi.2019.103271.
24. Correa S, Scovner KM, Tumlin JA, et al. Electrolyte Changes in Contemporary hemodialysis: a secondary analysis of the monitoring in Dialysis (MiD) Study. Kidney360. 2021;2(4):695–707. https://doi.org/10.34067/KID.0007452020.
25. GBD 2021 Causes of Death Collaborators. Global burden of 288 causes of death and life expectancy decomposition in 204 countries and territories and 811 subnational locations, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021. Lancet. 2024;403(10440):2100–2132. doi: 10.1016/S0140-6736(24)00367-2. Epub 2024 Apr 3. Erratum in: Lancet. 2024 May 18;403(10440):1988.
26. .Zamanzadeh D, Feng J, Petousis P, et al. Data-driven prediction of continuous renal replacement therapy survival. Nat Coun. 2024;15(1):5440. https://doi.org/10.1038/s41467-024-49763-3.
27. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD + AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;385:e078378. https://doi.org/10.1136/bmj-2023-078378.
28. Liu P, Sawhney S, Heide-Jørgensen U, et al. Predicting the risks of kidney failure and death in adults with moderate to severe chronic kidney

disease: multinational, longitudinal, population based, cohort study. BMJ. 2024;385:e078063. https://doi.org/10.1136/bmj-2023-078063.
29. Charkviani M, Gregoire JR. Enhancing Nephrology Education: a 10-Year experience on Fellow-Led Quality Improvement projects in a Hemodialysis Unit. Mayo Clin Proc. 2024;99(8):1342–5. https://doi.org/10.1016/j.mayocp.2024.04.003.
30. Yang IN, Liu CF, Chien CC, et al. Personalized prediction of intradialytic hypotension in clinical practice: development and evaluation of a novel AI dashboard incorporating risk factors from previous and current dialysis sessions. Int J Med Inf. 2024;190:105538. https://doi.org/10.1016/j.ijmedinf.2024.105538.
31. Mehta MC, Katz IT, Jha AK. Transforming Global Health with AI. N Engl J Med. 2020;382(9):791–3. https://doi.org/10.1056/NEJMp1912079.

Publisher's note