RESEARCH

Predicting dry weight change in Hemodialysis patients using machine learning

Hiroko Inoue[1†], Megumi Oya[2,3†], Masashi Aizawa[1], Kyogo Wagatsuma[2,3], Masatomo Kamimae[2,3], Yusuke Kashiwagi[1], Masayoshi Ishii[1,4], Hanae Wakabayashi[1,4], Takayuki Fujii[5], Satoshi Suzuki[5], Noriyuki Hattori[4,6], Narihito Tatsumoto[1], Eiryo Kawakami[2,3*] and Katsuhiko Asanuma[1,4*]

Abstract

Background Machine Learning has been increasingly used in the medical field, including managing patients undergoing hemodialysis. The random forest classifier is a Machine Learning method that can generate high accuracy and interpretability in the data analysis of various diseases. We attempted to apply Machine Learning to adjust dry weight, the appropriate volume status of patients undergoing hemodialysis, which requires a complex decision-making process considering multiple indicators and the patient's physical conditions.

Methods All medical data and 69,375 dialysis records of 314 Asian patients undergoing hemodialysis at a single dialysis center in Japan between July 2018 and April 2020 were collected from the electronic medical record system. Using the random forest classifier, we developed models to predict the probabilities of adjusting the dry weight at each dialysis session.

Results The areas under the receiver-operating-characteristic curves of the models for adjusting the dry weight upward and downward were 0.70 and 0.74, respectively. The average probability of upward adjustment of the dry weight had sharp a peak around the actual change over time, while the average probability of downward adjustment of the dry weight formed a gradual peak. Feature importance analysis revealed that median blood pressure decline was a strong predictor for adjusting the dry weight upward. In contrast, elevated serum levels of C-reactive protein and hypoalbuminemia were important indicators for adjusting the dry weight downward.

Conclusions The random forest classifier should provide a helpful guide to predict the optimal changes to the dry weight with relative accuracy and may be useful in clinical practice.

Keywords Machine learning, Random Forest classifier, Importance analysis, Dry Weight, Hemodialysis

†Hiroko Inoue and Megumi Oya contributed equally to this work.

*Correspondence:
Eiryo Kawakami
eiryo.kawakami@chiba-u.jp
Katsuhiko Asanuma
kasanuma@chiba-u.jp
[1]Department of Nephrology, Graduate School of Medicine, Chiba University, Chuo-ku, Chiba, Japan
[2]Department of Artificial Intelligence Medicine, Graduate School of Medicine, Chiba University, Chuo- ku, Chiba, Japan
[3]Advanced Data Science Project (ADSP), RIKEN Information R&D and Strategy Headquarters, RIKEN, Yokohama, Kanagawa, Japan
[4]Department of Artificial Kidney, Chiba University Hospital, Chuo-ku, Chiba, Japan
[5]Department of Nephrology, Seirei Sakura Citizen hospital, Sakura, Chiba, Japan
[6]Department of Emergency and Critical Care Medicine, Graduate School of Medicine, Chiba University, Chiba, Japan

Inoue *et al. BMC Nephrology*   (2023) 24:196

Page 2 of 10

Background

Machine learning (ML) has been increasingly used in the medical field for diagnosing and predicting illnesses based on multiple parameters and background characteristics of patients [1]. ML has demonstrated a pivotal role in estimating the onset of acute kidney injury [2–5] and the therapeutic responses of diabetic nephropathy [6] and IgA nephropathy [7]. In dialysis treatment, ML has also shown to be useful in the adjustment of the erythropoiesis-stimulating agent dosage for renal anemia [8–11], prediction of the occurrence of hypotension [12–14], and evaluation of fluid volume for patients undergoing dialysis [15]. In healthcare systems, ML has the potential to improve the selection of appropriate investigation and therapeutic processes, possibly resulting in improved prognosis in hemodialysis patients.

Heart failure is a major cause of mortality among patients undergoing hemodialysis [16]. Many laboratory and clinical parameters are associated with the incidence and progression of heart failure, including hypertension, anemia, serum calcium levels, and excess extracellular volume [17]. Among these, dry weight (DW) is pivotal in controlling the volume status in hemodialysis patients [17]. DW is defined as "body weight with adequate fluid volume without excessive hypotension during dialysis and with minimal cardiovascular burden in the long term [18]." Therefore, maintaining an appropriate DW is crucial for the prevention of heart failure and results in the reduction of mortality in hemodialysis patients [19].

Clinically, physicians determine DW by considering multiple indicators, such as blood pressure, increased body weight between dialysis sessions, cardiothoracic ratio, pleural effusion, edema, brain natriuretic peptide (BNP), and other blood tests [20]. Revising and changing a patient's DW according to their condition, including infectious diseases, diet, and physical activity at the time, is necessary. That is, DW cannot be determined by simple calculation, without specific indicators, and the weight of each parameter varies per patient. Trends in fluid volume changes may exist with changes in appetite, physical activity, and acute illness but are challenging to predict accurately. Therefore, automating information collection and processing, supporting diagnosis, and treatment using ML is urgently needed.

The random forest (RF) classifier is a ML method that generates high accuracy in the data analysis of various diseases, such as cardiovascular disease [21], stroke [22], cataracts [23], and ovarian cancer [24], because it can consider interactions between variables and is not affected by possible outliers. The RF classifier is one of the class identification methods in which data and explanatory variables are randomly divided to create multiple decision trees, and the final classification is achieved by the majority vote [25]. Because the correlations between each decision tree are weakened, overfitting is suppressed, which improves prediction performance. In addition, the RF classifier has high interpretability, such as the ability to calculate the relative importance of input variables, and is considered helpful for cases with many explanatory variables.

Previous studies using ML have shown that there are significant errors in the predictions of DW [26, 27]. Therefore, in the present study, we aimed to apply an RF classifier to predict the adjustments of DW made by dialysis specialists using many explanatory variables.

Methods

Patients and variables

We conducted a retrospective observational study at the Seirei Sakura Citizen Hospital in Chiba, Japan. Figure 1 shows the subject selection process. The study included patients who underwent hemodialysis at the facility twice or thrice a week for at least three months and were aged 20 years or older. All dialysis records were extracted from the dialysis system (Nikkiso Co., Ltd., Tokyo, Japan) from July 2018 to April 2020. A unique format was created to extract dialysis records from the dialysis system. The records were aggregated for each patient, and patients with more than 50 dialysis records were selected. The dialysis records within the first three months from the initiation of dialysis were excluded because of the unstable volume status in that period. CHDF and apheresis sessions are excluded to ensure consistency of conditions.

The clinical parameters, backgrounds, and medications of the eligible patients were collected from the electronic medical record system and anonymized. Laboratory tests were routinely performed twice a month at the beginning of the week. Table 1 lists the variables used in the extracted data.

Dialysis and dry weight

Generally, patients undergo hemodialysis twice or thrice a week for 240 min per session. In every session, nephrologists are attending near the hemodialysis patients and checking the hemodialysis procedure being correctly done. Hemodialysis patients usually have regular laboratory tests twice a month, meaning they are tested every two or three weeks. The DW of each patient was assessed and adjusted by several nephrologists periodically and when clinically indicated, by referring to multiple indicators such as blood pressure, clinical findings, chest x-ray findings, laboratory test results, sonographic measurement of the inferior vena cava diameter, and blood volume monitor, as recommended by the Japanese Society for Dialysis Therapy guidelines [18] and K/DOQI clinical practice guidelines [28].

The default dialysate sodium concentration was 140 mmol/L, and the dialysate calcium concentration was

Inoue *et al. BMC Nephrology*   (2023) 24:196

Page 3 of 10
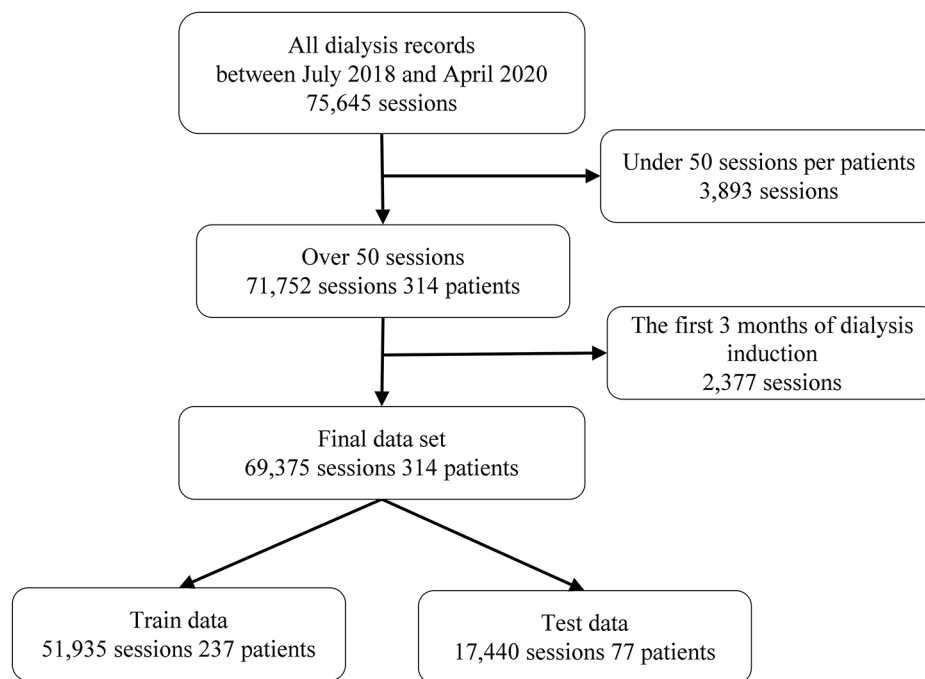
Fig. 1 Process of selecting the subjects

Table 1 Input variables used in analyses

category	items
Demographic and anthropometric data	Age, gender, race, height, date of introduction of dialysis, name of primary disease
Laboratory test results	Pre- and post-dialysis WBC, RBC, Hb, Ht, Plt, Ret, TP, Alb, ALP, ChE, BUN, Cre, Na, K, Cl, Ca, IP, intact PTH, Fe, UIBC, FER, BNP, CRP
Prescription data	Drug name, dosage, number of days prescribed, and start date
Medical record entries that include the following keywords	Pulmonary congestion, pleural effusion, oxygen administration, edema, hypotension
Dialysis record	Membrane area of dialyzer, blood flow rate, injected drugs during dialysis, cardiothoracic ratio, median values of vital signs (blood pressure, pulse rate, body temperature), intradialytic hypotension

WBC, white blood cell; RBC, red blood cell; Hb, hemoglobin; Ht, hematocrit; Plt, platelet; Ret, reticulocyte; TP, total protein; Alb, albumin; ALP, alkaline phosphatase; ChE, cholinesterase; BUN, blood urea nitrogen; Cre, creatinine; Na, sodium; K, potassium; Cl, chlorine; Ca, calcium; IP, inorganic phosphorus; PTH, parathyroid hormone; Fe, iron; UIBC, unsaturated iron binding capacity; FER, ferritin; BNP, brain natriuretic peptide; CRP, C-reactive protein

2.75 mmol/L. The dialysate flow rate was 500 ml/min, and the dialysate temperature was 36.0 °C for almost all patients.

Data preprocessing

Table 1 shows the list of input variables used in this analysis. The pre-dialysis blood test results are used from the day of the test until the day before the next blood test. The post-dialysis laboratory test results are used as the data of the following dialysis date and are valid until the next test is performed. Each drug in the prescription data was divided into categories based on its efficacy, and the top 50 categories with the highest number of prescriptions were assigned an identification code. We manually confirmed the presence or absence of clinical findings from medical record entries containing the following keywords: pleural effusion, edema, oxygen demand, and hypotension. Dialyzers were expressed as membrane areas. Median values were obtained from all vital signs (including blood pressure and heart rate) measurements during each dialysis session and is only used on the same day. Intradialytic hypotension was defined as a fall in systolic blood pressure of 20 or more during a dialysis session. Among the above, dummy variables were used, except for continuous variables. The missing values were filled with the previous or average values before splitting the train and test set (Supplementary Table S1). 75% of the data (237 patients) were randomly selected for the training dataset, and the remaining 25% (77 patients) were used as the test dataset to evaluate the performance of the prediction. In the dataset, physicians adjusted the DW upward (downward) from the previous session in approximately 2.5% (3.0%) of all dialysis sessions. Whether these changes were made was used as the objective variable in machine learning.

Supervised machine learning and importance analysis

There are many machine learning algorithms that can be applied to this task. In our initial examination, we have found that RF produces better results than other algorithms such as XGBoost and ANN, since the number of explanatory variables in our data is not very large and the task is a simple binary prediction of Dry Weight change. Our previous study also showed that RF performs reproducibly better than other algorithms in a binary classification task based on the similar clinical parameters as in this study [24]. For these reasons, we have applied RF classifier to train the models to predict the change in DW. The RF is an ensemble learning method that improves performance by training multiple models. In the RF models, training is performed using multiple decision trees. In the training process, the dataset was randomly selected for each decision tree using bootstrap sampling. The RF prediction is achieved by taking the majority vote of each decision tree. At each dialysis session, two RF classifier models were trained separately to determine whether "the DW should be adjusted upward or not" and "the DW should be adjusted downward or not," because we hypothesized that the factors involved in Dry Weight increase were different from those involved in Dry Weight decrease, and it would be important to build separate models to account for the different contributions of the explanatory variables. The label data were set as the actual DW change by the nephrologists of the hospital. The RandomForestClassifier in the Python package scikit-learn was used in the analysis. GridSearchCV was applied to automatically optimize the hyperparameters of the RF classifier. For hyperparameter optimization, we use the following hyperparameters candidates for two models: max_depth: [None, 2, 3, 4, 5, 6], max_features: ["auto", 6, 12, 24, 48], criterion: ["gini", "entropy"]. The labels of the dataset in this study were imbalanced (the number of sessions "without changing DW" was relatively high compared to the number of sessions "adjusting DW upward" or "downward"). To regulate this imbalance, the Synthetic Minority Oversampling Technique (SMOTE) algorithm was applied [29]. SMOTE is a method of oversampling that increases the minority population of imbalanced data. SMOTE in the Python package imbalanced learning was used in this study. For each dialysis session for each patient, the two models predicted the probabilities of "the DW should be adjusted upward" and "the DW should be adjusted downward," denoted as "P_{up}" and "P_{down}" scores, respectively, with both scores ranging between 0 and 1. The probabilities of the classified predictions were rescaled using the probability calibration [30–32]. The trained RF classifier models were calibrated using CalibratedClassifierCV in the scikit-learn package. The overall performance of each model was evaluated using an area under the curve

(AUC) of receiver operating characteristic (ROC) curve for the test data set. We also calculated accuracy, precision, recall, and F1 at the optimal probability threshold with the highest Youden Index. To monitor changes in a patient, we examined the probability changes for each patient. In addition, the predicted probability changes were evaluated in the 30 hemodialysis sessions before and after the DW change. The feature importance was obtained from the trained models, which show the contribution of each feature to the prediction.

Results

Patient characteristics

We retrospectively collected 69,375 dialysis records from 314 Asian patients during the observation period. The average observation period was 17.5 months. 73% of the patients were men. The mean age of the patients was 66.4 years old, and the median dialysis vintage was 4.0 years. The most common primary disease for dialysis induction was diabetic nephropathy (42.4%), followed by chronic glomerulonephritis (19.4%), nephrosclerosis (16.6%), and others (17.8%) (Table 2).

Prediction of dry weight change with RF classifiers

To learn the physician's decision on how to set the DW for the next dialysis session, two RF classifier models were trained separately with 51,935 dialysis records from 237 patients and validated for accuracy using 17,440 records from 77 patients.

As the result of GridsearchCV, we got (max_depth=3, max_features=12, criterion = "gini") for DW up model, and (max_depth=5, max_features=12, criterion = "entropy") for DW down model. After the calibration of the models, the optimal probability thresholds with the highest Youden were 0.204 for the DW up model, and 0.114 for the DW down model. The AUC, accuracy, precision, recall, and F1 score of the two models for the test data set were (AUC=0.70, accuracy=0.656, precision=0.045, recall=0.676, F1 score=0.084) for DW up model, and (AUC=0.74, accuracy=0.618, precision=0.055, recall=0.751, F1 score=0.102) for DW down model, respectively (Fig. 2). We also examined the relationship between the actual DW changes and the predicted probabilities of the ML models, P_{up} and P_{down}, in two representative cases. In the first case, shown in Fig. 3a and b, and 3c, the machine learning prediction, P_{up}, was greater than 0.5 in the majority of instances when the DW actually changed upward. P_{down} was often higher before and after the DW was changed downward, but the values were smaller than those of P_{up} and often did not show a clear peak. In the second case, shown in Fig. 3d and e, and 3f, the actual DW remained fixed at a constant value despite the many P_{up} peaks in the first half of the case. The P_{down} was low in the first half, but

Table 2 Characteristics of patients and dialysis parameters in the training and test data

	All (n = 314)			Train data (n = 237)		Test data (n = 77)	
Demographics							
Age, years	66.4		± 12.4	66.4	± 12.5	66.3	± 12.3
Gender (Male/Female)	224	/90		170	/67	54	/23
Race, n (%)							
Asian (Japanese)	312	(99.4)		235	(99.2)	77	(100.0)
Asian (Southeast Asian)	2	(0.6)		2	(0.8)	0	(0.0)
Dialysis vintage, years	4.0	(2.0–8.0)		4.0	(2.0-8.8)	4.0	(2.0–6.0)
Body mass index, kg/m2	22.2		± 4.3	22.1	± 3.8	22.6	± 5.4
Primary disease, n (%)							
Diabetic nephropathy	133		(42.4)	95	(40.1)	38	(49.4)
Chronic glomerulonephritis	61		(19.4)	51	(21.5)	10	(13.0)
Nephrosclerosis	52		(16.6)	40	(16.9)	12	(15.6)
Polycystic kidney disease	12		(3.8)	11	(4.6)	1	(1.3)
Others	56		(17.8)	40	(16.9)	16	(20.8)
Dialysis parameter							
UF (L/session)	2.4		± 0.97	2.4	± 0.95	2.6	± 1.03
UFR (L/hr)	0.61		± 0.24	0.60	± 0.24	0.64	± 0.25
Dry Weight (kg)	59.7		± 14.2	59.3	± 12.8	60.8	± 17.6
Dialysis mode, n (%)							
HD	5047		(7.3)	3238	(6.2)	1809	(10.4)
HD + ECUM	306		(0.4)	146	(0.3)	160	(0.9)
OHDF	63,700		(91.8)	48,320	(93.0)	15,380	(88.2)
OHDF + ECUM	254		(0.4)	185	(0.4)	69	(0.4)
ECUM	68		(0.1)	46	(0.1)	22	(0.1)

Values are expressed as mean ± standard deviation, median (interquartile range), or percent frequency

UF, ultrafiltration; UFR, ultrafiltration rate; HD, hemodialysis; ECUM, extracorporeal ultrafiltration method; OHDF, online hemodiafiltration

persistently quite high, close to 1.0, around the end of the recording. Interestingly, P_{up} was also quite high during this period, indicating a conflict between the two machine learning models.

Thereafter, we examined the changes in the prediction probabilities of ML in the 30 hemodialysis sessions before and after the actual DW change for all cases included in the test dataset. The average P_{up} had a sharp peak at the time of the actual upward change in DW and then declined immediately (Fig. 4a). However, the change in P_{down} was slow and remained high for a while after the actual downward change in DW occurred (Fig. 4b).

Identification of key factors in the prediction of DW changes

To clarify the parameters on which the determination of DW changes in the ML process is based, we calculated the variable importance of the two ML models. Figure 5 shows the relative importance of the input variables in each model, listed in order of importance values.

For the model to predict upward DW changes, variables related to blood pressure are of great importance (Fig. 5a). The top three most important variables were the median value of systolic blood pressure, mean blood pressure, and diastolic blood pressure measured several times during a single dialysis session.

Systolic blood pressure was also an essential factor in predicting downward DW changes (Fig. 5b). However, the two most important variables were C-reactive protein (CRP) and pre-dialysis albumin levels, while blood pressure was relatively less important.

Next, the transition of the most important features was examined in 30 hemodialysis sessions before and after the actual DW change. Blood pressure, an important predictor of the upward DW change, showed a marked decrease at the time of the actual upward DW change, followed by rapid recovery (Fig. 6a and b, and 6c). However, even the systolic blood pressure, which declined the most, only dropped from 141 mm Hg to 133 mm Hg on average over 30 dialysis sessions.

CRP, an important indicator of the downward DW change, showed a sharp increase just before the actual DW change, but its peak occurred a little later than the actual DW change (Fig. 6d). The change in albumin was slower and began to increase immediately after the actual DW change (Fig. 6e). Systolic blood pressure, which showed a sharp peak around the upward DW change, showed a slow change before and after the downward DW change (Fig. 6f). CRP levels rose from 1.2 mg/dL to 2.1 mg/dL, albumin levels declined from 3.18 g/dL to 3.12 g/dL, and the median systolic blood pressure rose
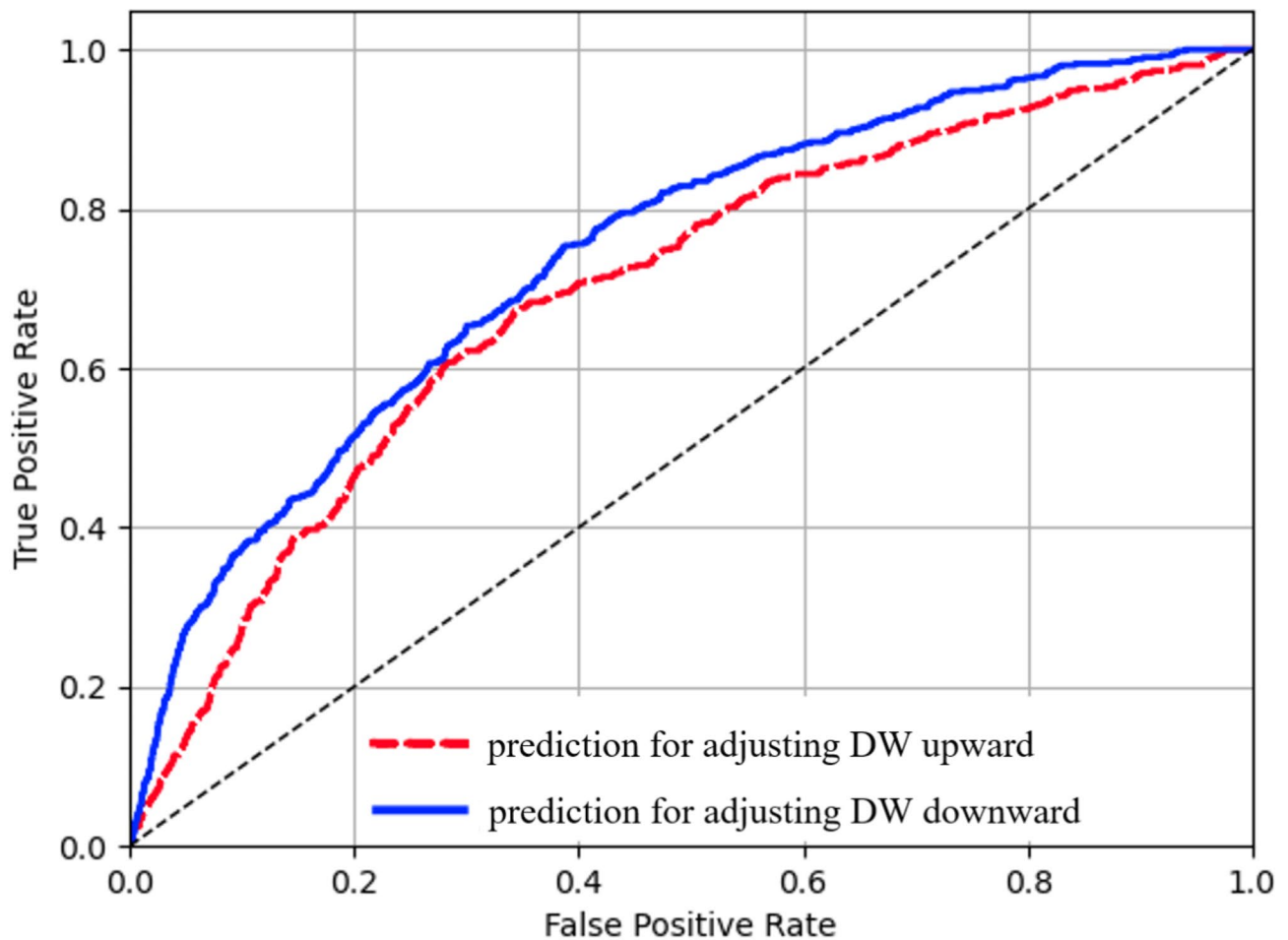
AUC = 0.70, accuracy = 0.656, precision = 0.045, recall = 0.676, F1 score = 0.084
for DW up model
AUC = 0.74, accuracy = 0.618, precision = 0.055, recall = 0.751, F1 score = 0.102
for DW down model

**Fig. 2** ROC curves of the ML models for the test data set
The dashed line and solid line indicate the ROC curves of the models for the prediction of an upward and downward change in DW, respectively
ROC, receiver operating characteristic; ACU, area under the curve; DW, dry weight

slightly from 148 mm Hg to 150 mm Hg on average over 30 dialysis sessions.

Discussion

In this study, we developed novel methods to predict whether DW should be adjusted at each dialysis session using an RF classifier. The AUCs of the models were 0.70 or more, indicating high reliability. By analyzing approximately 150 variables, our approach revealed the most important input factors for each decision-making. These models may enable medical staff to determine the correct timing for adjusting DW more efficiently.

Several studies have used machine learning to predict DW in patients undergoing dialysis. Guo et al. predicted DW using a neural network model and compared it to the DW predicted by a body composition monitor; their model's root mean square error was 1.316 [26]. Kim et al. used XGBoost machine learning to predict DW and compared it with DW based on bioimpedance spectroscopy [27]. If the difference between the two groups was between 1 and 2 kg, the average accuracies were 72–83%. Both results suggest the usefulness of machine learning. However, challenges remain, with significant errors in the prediction of the DW itself. Therefore, we focused on
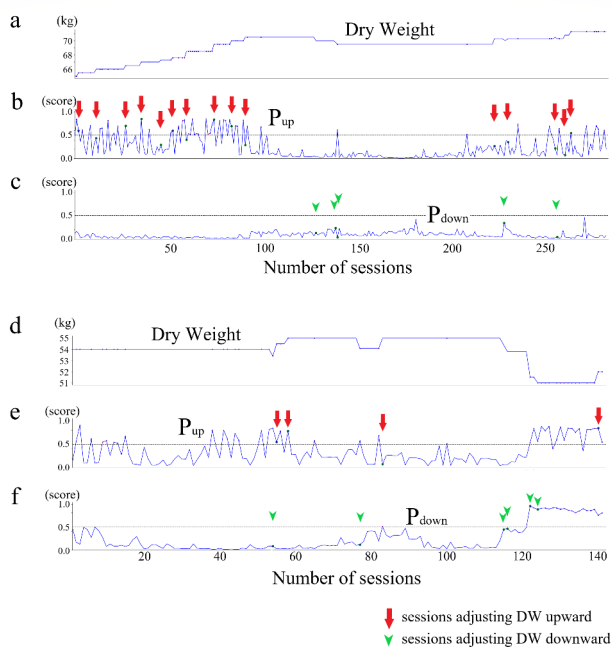
Fig. 3 Representative examples of prediction

(a–c) and **(d–f)** shows the representative examples of two patients from the test data set

(a) and **(d)** show the actual DW trend; **(b)** and **(e)** show the P_{up} scores; **(c)** and **(f)** show the P_{down} scores

Arrows and arrowheads indicate sessions where the DW was adjusted upward and downward

Horizontal axis: the number of dialysis sessions along the time

Vertical axis: weight in kilograms in **(a)** and **(d)**, the probability value from 0 to 1 in **(b)**, **(c)**, **(e)**, and **(f)**

adjusting the DW upward or downward, rather than estimating the value of DW.

DW could be acceptable in a particular range for stable maintenance hemodialysis patients with few cardiac problems. Usually, DW is not immediately determined as a constant value but is probed and adjusted gradually based on various information gathered [20]. Thus, we believe our model helpful in informing us when the current dry weight may be out of the appropriate range and our approach has the advantage of immediate clinical application.

We have also shown how the predicted scores varied over time, another point that has yet to be addressed. Our models can capture even small changes in the input variables that reflect the patient's condition.

The average P_{up} score showed a quick rise, precisely detecting the timing at which the DW should be adjusted upward. However, the average P_{down} score gradually changed. A possible reason is that the model could not catch up with real-time changes in medical conditions because the frequency of laboratory tests was only twice a month. Furthermore, a gradual reduction of DW is recommended in clinical settings [33], corresponding to the gentle, slow slope of the average P_{down} score.
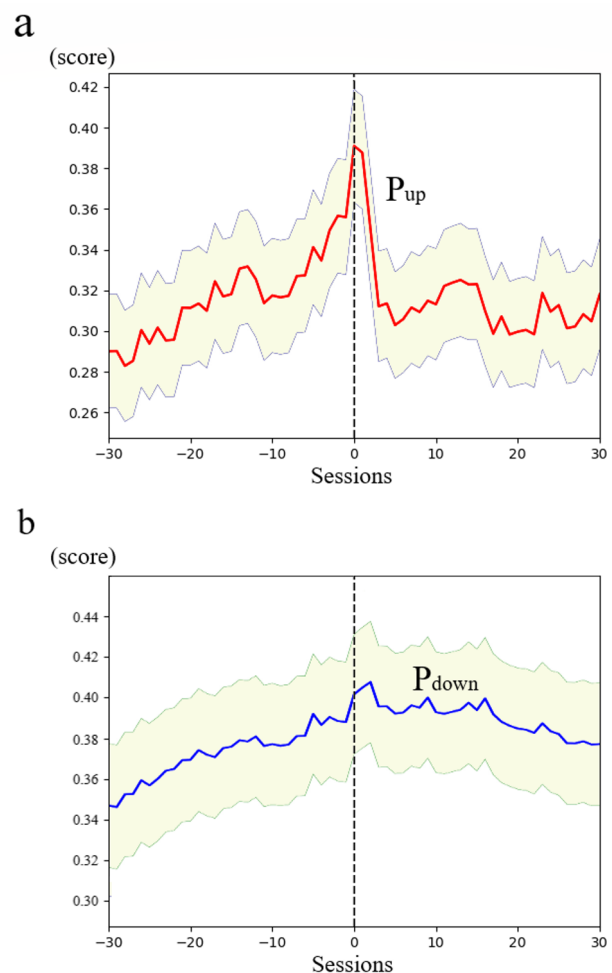
Fig. 4 The behavior of P_{up} and P_{down} scores around the actual DW changes in the test dataset

(a) Average P_{up} score and **(b)** average P_{down} score 30 sessions before and after the actual DW changes, represented as session 0. Even if a patient's DW was changed in successive sessions, each change was treated separately. Shaded areas indicate 95% confidence intervals

Horizontal axis: the number of dialysis sessions along the time

Vertical axis: the probability value from 0 to 1

Importance analysis revealed that serum levels of CRP and albumin, rather than blood pressure, showed the highest importance in predicting DW reduction. In addition to the elevation of CRP, hypoalbuminemia also indicates an inflammatory status [34, 35], which is related to a decrease in muscle mass [36] and an increase in extracellular volume [37, 38] in patients. Many studies have reported a strong association between inflammation and overhydration [39–41], which can lead to heart failure. Additional studies have shown that inflammation is associated with mortality [42–44]. Although these are notable findings, using ML clearly indicates that an inflammatory status is the most important in adjusting DW downward.
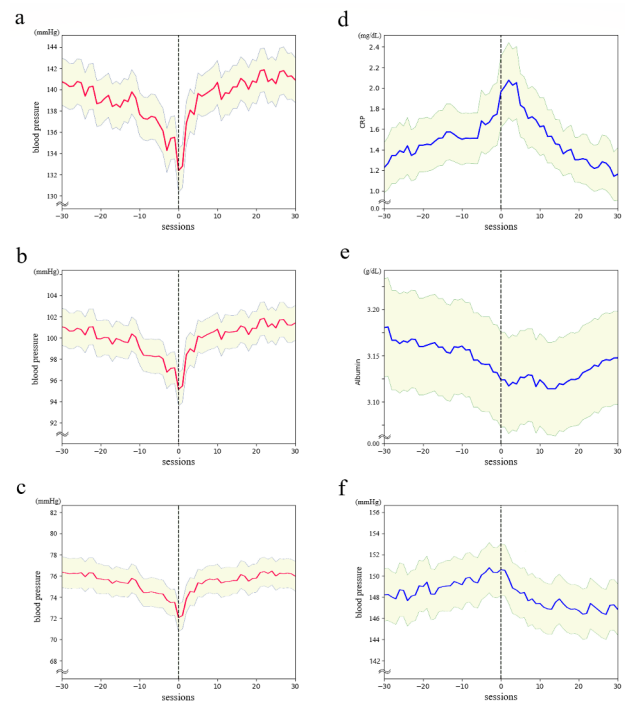
Importance analysis also showed that declining blood pressure was related to predicting upward changes in
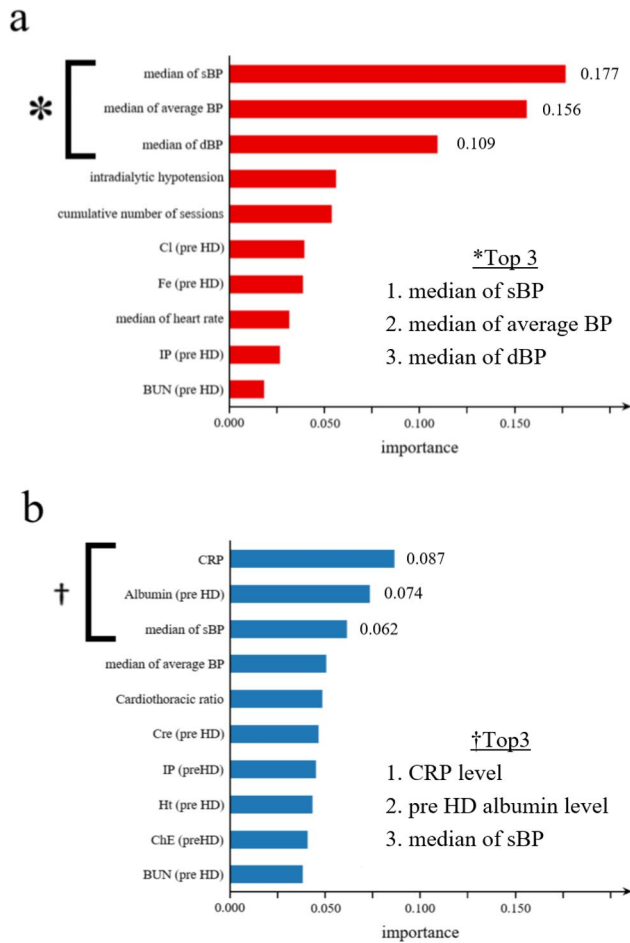
Fig. 5 Variable importance of the two ML models
 The relative importance of input variables **(a)** in the model to predict an upward DW change and **(b)** in the model to predict a downward DW change
 Horizontal axis: relative importance
 Vertical axis: input variables listed in order of importance values
 sBP, systolic blood pressure; dBP, diastolic blood pressure; CRP, C-reactive protein; HD, hemodialysis

Fig. 6 The trend in the most important input variables around the actual DW changes
 The trends of the top three most important input variables in the model to predict an upward DW change (a–c) and predict a downward DW change (d–f)
(a) median systolic blood pressure, **(b)** median average blood pressure, and **(c)** median diastolic blood pressure declined at the time of the up-ward DW change
(d) CRP level was elevated, **(e)** pre-dialysis albumin level declined, **(f)** me-dian systolic blood pressure was slightly elevated at the time of the down-ward DW change. Shaded areas indicate 95% confidence intervals
 Horizontal axis: the number of dialysis sessions along the time
 sBP, systolic blood pressure; dBP, diastolic blood pressure; CRP, C-reactive protein; HD, hemodialysis

DW. Several studies have shown that hypotension during dialysis is a significant poor prognostic factor [45, 46]. Decreased blood pressure can lead to a higher risk of falls, stroke, arteriovenous shunt occlusion, and so on [47–49]. Therefore, to prevent hypotension during dialysis, the ability to notify medical staff of the appropriate time that the DW should be adjusted upward is useful.

In our data, there were only small changes in the average levels of CRP, albumin, and blood pressure in the sessions where DW was actually adjusted, compared to 30 previous sessions. Even if these changes are gradual and difficult for medical staff to recognize, our ML models can detect slight changes and inform them by raising their scores. Using our models, DW can be adjusted to prevent adverse events and improve a patient's quality of life and prognosis.

BNP has been used to indicate fluid volume status, cardiac function, and cardiovascular disease. Many reports have shown an association between the biomarker and heart failure and prognosis in hemodialysis patients, but the threshold for this association is not consistent [50]. In the current analysis, the importance of BNP was not ranked high. This was most likely because BNP is measured too infrequently to be used to predict DW adjustment at every dialysis session. Furthermore, cardiac function and cardiovascular disease can significantly change the BNP value even in the same fluid volume status, so BNP may not be effectively used in our model.

For each patient, the predicted scores were elevated around the time of the actual DW changes in most cases; however, in some instances, both the P_{up} and P_{down} scores were elevated simultaneously. Most of them were as follows: the elevation of a P_{down} score caused the reduction of DW and was immediately followed by a drop in blood pressure and hence an increasing P_{up} score, as well as

Inoue *et al. BMC Nephrology*   (2023) 24:196

Page 9 of 10

elevated inflammatory states and low blood pressure in critical conditions. Our models can detect these unstable situations.

In a real clinical setting, based on inputs such as vital signs, clinical findings and blood test results of each patient, our model could calculate the probabilities that DW should be raised and lowered respectively. If the probability exceeds thresholds, alerts will be issued, prompting the doctor to consider changing the DW. The alerts are also helpful for non-specialist doctors, other medical staff, and trainees. Furthermore, our model can be presented to patients as a rationale for changing their DW.

Our study has several limitations. Due to the retrospective nature of this study, the description of some findings, such as edema and pleural effusion, seemed inadequate in some cases. In addition, because of the low frequency of laboratory tests, the models do not reflect real-time changes. Therefore, we plan to start collecting data prospectively, using templates to enter findings and devices to monitor blood concentration during dialysis to improve the accuracy of our models. Furthermore, the findings of this work are not immediately generalizable because it was a single-center study. In the future, we plan to collect data from other dialysis facilities and apply transfer learning to them.

Conclusions

In summary, we developed novel models to predict whether the DW of hemodialysis patients should be adjusted using an RF classifier. Our analysis showed the importance of declining blood pressure during dialysis for predicting an upward change in the DW and the impact of elevated CRP levels and hypoalbuminemia on predicting a downward change in the DW. Further studies are required to evaluate the clinical effectiveness of our models.

Abbreviations

ML	Machine learning
DW	Dry weight
BNP	Brain natriuretic peptide
RF	Random forest
SMOTE	Synthetic Minority Over-sampling Technique
ROC	Receiver operating characteristic
AUCs	Area under the receiver operating curves
CRP	C-reactive protein

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12882-023-03248-5.

Supplementary Material 1

Declarations

Ethics approval and consent to participate

The study was performed in accordance with the Declaration of Helsinki and Ethical Guidelines for Medical and Health Research Involving Human Subjects. The study was reviewed and approved by the ethical review committee of Seirei Sakura Citizen Hospital (#2019021) and Chiba University (#3390). The research plan was posted on the website and at the dialysis center, and informed consent was obtained from all the participants through the opt-out approach.

Consent for publication

Not Applicable.

Competing interests

K.A. is the secretary of the Japanese Society of Nephrology, and N.H. is the Japan Society for Blood Purification in Critical Care councilor. All other authors declare no conflicts of interest for this study.

References

1. Deo RC. Machine learning in Medicine. Circulation. 2015;132:1920–30.
2. Dong J, Feng T, Thapa-Chhetry B, Cho BG, Shum T, Inwald DP, Newth CJL, Vaidya VU. Machine learning model for early prediction of acute kidney injury (AKI) in pediatric critical care. Crit Care. 2021;25:288.
3. Xue B, Li D, Lu C, King CR, Wildes T, Avidan MS, Kannampallil T, Abraham J. Use of Machine Learning to develop and evaluate models using preoperative and intraoperative data to identify risks of postoperative complications. JAMA Netw Open. 2021;4:e212240.
4. Churpek MM, Carey KA, Edelson DP, Singh T, Astor BC, Gilbert ER, Winslow C, Shah N, Afshar M, Koyner JL. Internal and External Validation of a machine learning risk score for acute kidney Injury. JAMA Netw Open. 2020;3:e2012892.
5. Kang MW, Kim J, Kim DK, Oh KH, Joo KW, Kim YS, Han SS. Machine learning algorithm to predict mortality in patients undergoing continuous renal replacement therapy. Crit Care. 2020;24:42.
6. Belur Nagaraj S, Pena MJ, Ju W, Heerspink HL, BEAt-, DKD Consortium. Machine-learning-based early prediction of end-stage renal disease in patients with diabetic kidney disease using clinical trials data. Diabetes Obes Metab. 2020;22:2479–86.
7. Schena FP, Anelli VW, Trotta J, Di Noia T, Manno C, Tripepi G, D'Arrigo G, Chesnaye NC, Russo ML, Stangou M, Papagianni A, Zoccali C, Tesar V, Coppo R. Members of the VALIGA study. Development and testing of an artificial intelligence tool for predicting end-stage kidney disease in patients with immunoglobulin A nephropathy. Kidney Int. 2021;99:1179–88.

8.   Pellicer-Valero OJ, Cattinelli I, Neri L, Mari F, Martín-Guerrero JD, Barbieri C. Enhanced prediction of hemoglobin concentration in a very large cohort of hemodialysis patients by means of deep recurrent neural networks. Artif Intell in Med. 2020;107:101898.

9.   Lobo B, Abdel-Rahman E, Brown D, Dunn L, Bowman B. A recurrent neural network approach to predicting hemoglobin trajectories in patients with end-stage renal disease. Artif Intell Med. 2020;104:101823.

10. Barbieri C, Molina M, Ponce P, Tothova M, Cattinelli I, Ion Titapiccolo J, Mari F, Amato C, Leipold F, Wehmeyer W, Stuard S, Stopper A, Canaud B. An international observational study suggests that artificial intelligence for clinical decision support optimizes anemia management in hemodialysis patients. Kidney Int. 2016;90:422–9.

11. Ohara T, Ikeda H, Sugitani Y, Suito H, Huynh VQH, Kinomura M, Haraguchi S, Sakurama K. Artificial intelligence supported anemia control system (AISACS) to prevent anemia in maintenance hemodialysis patients. Int J Med Sci. 2021;18:1831–9.

12. Gómez-Pulido JA, Gómez-Pulido JM, Rodríguez-Puyol D, Polo-Luque ML, Vargas-Lombardo M. Predicting the Appearance of Hypotension during Hemodialysis Sessions using machine learning classifiers. Int J Environ Res Public Health. 2021;18:2364.

13. Huang JC, Tsai YC, Wu PY, Lien YH, Chien CY, Kuo CF, Hung JF, Chen SC, Kuo CH. Predictive modeling of blood pressure during hemodialysis: a comparison of linear model, random forest, support vector regression, XGBoost, LASSO regression and ensemble method. Comput Methods Programs Biomed. 2020;195:105536.

14. Kang MW, Kim S, Kim YC, Kim DK, Oh KH, Joo KW, Kim YS, Han SS. Machine learning model to predict hypotension after starting continuous renal replacement therapy. Sci Rep. 2021;11:17169.

15. Chiu JS, Chong CF, Lin YF, Wu CC, Wang YF, Li YC. Applying an artificial neural network to predict total body water in hemodialysis patients. Am J Nephrol. 2005;25:507–13.

16. Nitta K. 2019 Annual Dialysis Data Report, JSDT Renal Data Registry. Article in Japanese. J Jpn Soc Dial Ther. 2020;53:579–632.

17. Pinter J, Chazot C, Stuard S, Moissl U, Canaud B. Sodium, volume and pressure control in hemodialysis patients for improved cardiovascular outcomes. Nephrol Dial Transplant. 2020;35(Suppl 2):ii23–ii30.

18. Hirakata H, Nitta K, Inaba M, Shoji T, Fujii H, Kobayashi S, Tabei K, Joki N, Hase H, Nishimura M, Ozaki S, Ikari Y, Kumada Y, Tsuruya K, Fujimoto S, Inoue T, Yokoi H, Hirata S, Shimamoto K, Kugiyama K, Akiba T, Iseki K, Tsubakihara Y, Tomo T, Akizawa T. Japanese society for Dialysis Therapy. Japanese society for Dialysis Therapy Guidelines for Management of Cardiovascular Diseases in patients on chronic hemodialysis. Ther Apher Dial. 2012;16:387–435.

19. Agarwal R, Weir MR. Dry-weight: a concept revisited in an effort to avoid medication-directed approaches for blood pressure control in hemodialysis patients. Clin J Am Soc Nephrol. 2010;5:1255–60.

20. Joseph MS, Palady M, Bhave NM. Management of heart failure in patients with end-stage kidney disease on maintenance dialysis: a practical guide. Rev Cardiovasc Med. 2020;21:31–9.

21. Yang L, Wu H, Jin X, Zheng P, Hu S, Xu X, Yu W, Yan J. Study of cardiovascular disease prediction model based on random forest in eastern China. Sci Rep. 2020;10:5245.

22. Uchida K, Kouno J, Yoshimura S, Kinjo N, Sakakibara F, Araki H, Morimoto T. Development of Machine Learning Models to predict probabilities and types of stroke at Prehospital Stage: the Japan Urgent Stroke triage score using machine learning (JUST-ML). Transl Stroke Res. 2022;13(3):370–81.

23. Lin D, Chen J, Lin Z, Li X, Zhang K, Wu X, Liu Z, Huang J, Li J, Zhu Y, Chen C, Zhao L, Xiang Y, Guo C, Wang L, Liu Y, Chen W, Lin H. A practical model for the identification of congenital cataracts using machine learning. EbioMedicine. 2020;51:102621.

24. Kawakami E, Tabata J, Yanaihara N, Ishikawa T, Koseki K, Iida Y, Saito M, Komazaki H, Shapiro JS, Goto C, Akiyama Y, Saito R, Saito M, Takano H, Yamada K, Okamoto A. Application of Artificial Intelligence for Preoperative Diagnostic and Prognostic Prediction in Epithelial Ovarian Cancer based on blood biomarkers. Clin Cancer Res. 2019;25:3006–15.

25. Breiman L. Random forests. Mach Learn. 2001;45:5–32.

26. Guo X, Zhou W, Lu Q, Du A, Cai Y, Ding Y. Assessing Dry Weight of Hemodialysis Patients via Sparse Laplacian Regularized RVFL Neural Network with L2,1-Norm. Biomed Res Int 2021. 2021;6627650.

27. Kim HR, Bae HJ, Jeon JW, Ham YR, Na KR, Lee KW, Hyon YK, Choi DE. A novel approach to dry weight adjustments for dialysis patients using machine learning. PLoS ONE. 2021;16:e0250467.

28. K/DOQI Workgroup. K/DOQI clinical practice guidelines for cardiovascular disease in dialysis patients. Am J Kidney Dis. 2005;45(4 Suppl 3):1–S153.

29. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. J Artif Intell Res. 2002;16:321–57.

30. Zadrozny B, Elkan C. Obtaining calibrated probability estimates from decision trees and naive bayesian classifiers. Proc Int Conf Mach Learn. 2001;609–16.

31. Zadrozny B, Elkan C. Transforming classifier scores into accurate multiclass probability estimates. KDD. 2002;694–9.

32. Niculescu-Mizil A, Caruana R. Predicting good probabilities with supervised learning. Proc Int Conf Mach Learn. 2005;625–32.

33. Ohashi Y, Sakai K, Hase H, Joki N. Dry weight targeting: the art and science of conventional hemodialysis. Semin Dial. 2018;31:551–6.

34. Evans DC, Corkins MR, Malone A, Miller S, Mogensen KM, Guenter P, Jensen GL, ASPEN Malnutrition Committee. The Use of visceral proteins as Nutrition markers: an ASPEN position paper. Nutr Clin Pract. 2021;36:909.

35. Friedman AN, Fadem SZ. Reassessment of albumin as a nutritional marker in kidney disease. J Am Soc Nephrol. 2010;21:223–30.

36. Costamagna D, Costelli P, Sampaolesi M, Penna F. Role of inflammation in muscle homeostasis and myogenesis. Mediators Inflamm. 2015;2015:805172.

37. Wiig H, Swartz MA. Interstitial fluid and lymph formation and transport: physiological regulation and roles in inflammation and cancer. Physiol Rev. 2012;92:1005–60.

38. Reed RK, Rubin K. Transcapillary exchange: role and importance of the interstitial fluid pressure and the extracellular matrix. Cardiovasc Res. 2010;87:211–7.

39. Zsom L, Faludi M, Fülöp T, Dossabhoy NR, Rosivall L, Tapolyai MB. The association of overhydration with chronic inflammation in chronic maintenance hemodiafiltration patients. Hemodial Int. 2019;23:384–91.

40. Hassan MO, Duarte R, Dix-Peek T, Vachiat A, Naidoo S, Dickens C, Grinter S, Manga P, Naicker S. Correlation between volume overload, chronic inflammation, and left ventricular dysfunction in chronic kidney disease patients. Clin Nephrol. 2016;86:131–5.

41. Jacobs LH, van de Kerkhof JJ, Mingels AM, Passos VL, Kleijnen VW, Mazairac AH, van der Sande FM, Wodzig WK, Konings CJ, Leunissen KM, van Dieijen-Visser MP, Kooman JP. Inflammation, overhydration and cardiac biomarkers in haemodialysis patients: a longitudinal study. Nephrol Dial Transplant. 2010;25:243–8.

42. Dekker MJ, Marcelli D, Canaud BJ, Carioni P, Wang Y, Grassmann A, Konings CJ, Kotanko P, Leunissen KM, Levin NW, van der Sande FM, Ye X, Maheshwari V, Usvyat LA, Kooman JP, MONDO Initiative. Impact of fluid status and inflammation and their interaction on survival: a study in an international hemodialysis patient cohort. Kidney Int. 2017;91:1214–23.

43. Qureshi AR, Alvestrand A, Divino-Filho JC, Gutierrez A, Heimbürger O, Lindholm B, Bergström J. Inflammation, malnutrition, and cardiac disease as predictors of mortality in hemodialysis patients. J Am Soc Nephrol. 2002;13(Suppl 1):28–S36.

44. Cobo G, Lindholm B, Stenvinkel P. Chronic inflammation in end-stage renal disease and dialysis. Nephrol Dial Transplant. 2018;33(Suppl 3):iii35–iii40.

45. Hörl MP, Hörl WH. Hemodialysis-associated hypertension: pathophysiology and therapy. Am J Kidney Dis. 2002;39:227–44.

46. Chou JA, Kalantar-Zadeh K, Mathew AT. A brief review of intradialytic hypotension with a focus on survival. Semin Dial. 2017;30:473–80.

47. Correa S, Pena-Esparragoza JK, Scovner KM, Mc Causland FR. Predictors of intradialytic symptoms: an analysis of Data from the Hemodialysis Study. Am J Kidney Dis. 2020;76(3):331–9.

48. Khan A, Khan AH, Adnan AS, Syed Sulaiman SA, Gan SH, Khan I. Management of Patient Care in Hemodialysis while focusing on Cardiovascular Disease events and the atypical role of Hyper- and/or hypotension: a systematic review. Biomed Res Int. 2016;2016:9710965.

49. Chang TI, Paik J, Greene T, Desai M, Bech F, Cheung AK, Chertow GM. Intradialytic hypotension and vascular access thrombosis. J Am Soc Nephrol. 2011;22:1526–33.

50. Yang WL, Fahim M, Johnson DW. Pathophysiology and significance of natriuretic peptides in patients with end-stage kidney disease. Clin Biochem. 2020;83:1–11.

Publisher's Note