Development of an Artificial Intelligence Model for Prediction of Dry Weight in Maintenance Hemodialysis Patients

Nataphut Boonvisuth[1], Kanitha Tiankanon[2], Raksit Raksasat[3], Sira Sriswasdi[4], Khajohn Tiranathanagul[1]

[1]*Division of Nephrology, Department of Medicine, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand*

[2]*Bhumirajanagarindra Kidney Institution Hospital, Bangkok, Thailand*

[3]*Department of Computer Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok, Thailand*

[4]*Research Affairs, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand*

Abstract

Background: The optimal dry body weight (DW) for each patient is crucial to the effectiveness of hemodialysis (HD). The traditional assessment of DW using clinical parameters has proven to be inaccurate. Although bioimpedance spectroscopy analysis using Body Composition Monitor (BCM) device demonstrated excellent accuracy but the availability is limited due to high cost. The present study introduced machine learning (ML), a branch of artificial intelligence, in the assessment of DW (ML-DW) using available clinical and laboratory parameters and compared the result with the dry weight derived from BCM (BCM-DW)

Methods: The HD treatment data between 2017 and 2022 from two dialysis centers in Bangkok, Thailand including demographic, laboratory, and intradialytic time-varying data were retrieved. The data on BCM-DW were collected on the same day as HD treatment. The data were used in the ML model development phase and performance assessment phase. There were two groups during the model development phase consisting of a training group and a validation group. The final model was externally validated on a testing group at another institution.

Results: A total of 1151 dialysis sessions accounting for 56,000 time-varying data were retrieved. The mean BCM-DW was 58.8±11.7 kgs and the mean predicted ML-DW from the model was 59.5±10. kgs. The mean difference between ML-DW and BCM-DW was -0.78 (-3.7,2.2) kilograms. The latency for running the model was less than 1 minute.

Conclusion: Despite the relatively large difference between ML-DW and BCM-DW, the present study confirmed the capability of ML in DW prediction.

Key words: dry weight; BIA; AI; body composition; machine learning; neural networks

Corresponding author: *Khajohn Tiranathanagul*
Email: *khajohn_t@hotmail.com*

Received: 12 September 2023; Revised; 3 October 2023; Accepted: 19 October 2023.

การพัฒนาโมเดลด้วยปัญญาประดิษฐ์เพื่อทำนายน้ำหนักแห้งในผู้ป่วยไตวายเรื้อรังระยะสุดท้ายที่ได้รับการฟอกเลือดด้วยเครื่องไตเทียม

ณฐพุฒิ บุญวิสุทธิ์[1], ขนิษฐา เธียรกานนท์[2], รักษิต รักษาสัตย์[3], สิระ ศรีสวัสดิ์[4], ขจร ตีรณธนากุล[1]

[1]สาขาวิชาโรคไต ภาควิชาอายุรศาสตร์ คณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

[2]โรงพยาบาลสถาบันโรคไตภูมิราชนครินทร์

[3]ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

[4]ฝ่ายวิจัย คณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

บทคัดย่อ

บทนำ: น้ำหนักแห้งที่เหมาะสมที่สุดสำหรับผู้ป่วยแต่ละคนมีความสำคัญต่อประสิทธิภาพของการฟอกเลือดด้วยเครื่องไตเทียม การประเมินน้ำหนักแห้งแบบดั้งเดิมโดยอาศัยข้อมูลทางคลินิกพบว่าขาดความแม่นยำ ในขณะที่การวัดโดยใช้หลักการประเมินจากความต้านทานของกระแสไฟฟ้า (Bioelectrical Impedance Analysis หรือ BIA) ด้วยเครื่อง Body composition monitor (BCM) ช่วยให้ได้น้ำหนักแห้ง (BCM-DW) ที่มีความแม่นยำสูง แต่อย่างไรก็ตามเครื่องมือดังกล่าวมีราคาแพงจึงเป็นข้อจำกัดของศูนย์ฟอกเลือดขนาดเล็ก จึงเป็นที่มาของความพยายามที่จะใช้ Machine learning (ML) ซึ่งเป็นโปรแกรมหนึ่งทางปัญญาประดิษฐ์ในการทำนายน้ำหนักแห้ง (ML-DW) ในผู้ป่วยที่ได้รับการฟอกเลือดด้วยเครื่องไตเทียม

ระเบียบวิธีวิจัย: การศึกษานี้อาศัยข้อมูลย้อนหลังของผู้ป่วยที่ได้รับการฟอกเลือดด้วยเครื่องไตเทียมระหว่างปี 2560 ถึง 2565 จากสองสถาบันในจังหวัดกรุงเทพฯ ได้แก่ ข้อมูลพื้นฐาน ข้อมูลทางห้องปฏิบัติการ ข้อมูลที่เกี่ยวข้องกับการฟอกเลือด และข้อมูลที่มีการเปลี่ยนแปลงในช่วงเวลาที่ได้รับการฟอกเลือด รวมไปถึงข้อมูลน้ำหนักแห้งที่วัดจาก BCM ในวันเดียวกับที่ได้รับการฟอกเลือด โดยข้อมูลทั้งหมดจะถูกนำมาใช้ใน 2 ช่วงของการพัฒนาโมเดล คือ ระยะฝึกฝนและระยะปรับค่าพารามิเตอร์ของโมเดล และโมเดลสุดท้ายจะนำไปทดสอบกับกลุ่มทดสอบที่อยู่ภายนอกสถาบัน

ผลการศึกษา: รวบรวมข้อมูลการฟอกเลือดได้ทั้งหมด 1,151 ครั้ง และมีข้อมูลส่วนที่มีการเปลี่ยนแปลงตามเวลาของการฟอกเลือด 56,000 ข้อมูล ค่าเฉลี่ยของ BCM-DW คือ 58.8±11.7 กก. และค่าเฉลี่ย ML-DW คือ 59.5±10.5 กก. โดยพบความแตกต่างเฉลี่ยของ BCM-DW และ ML-DW ที่อยู่ที่ -0.78 (-3.7,-2.2) กิโลกรัม เวลาที่ใช้ในการประมวลผลทั้งหมดน้อยกว่า 1 นาที

สรุป: ถึงแม้ว่าการศึกษานี้พบว่าค่า ML-DW ยังมีความแตกต่างจาก BCM-DW ค่อนข้างมาก แต่อย่างไรก็ตามการศึกษานี้ได้แสดงให้เห็นความเป็นไปได้ในการนำปัญญาประดิษฐ์เข้ามาใช้ทำนายน้ำหนักแห้งของผู้ป่วยฟอกเลือด

คำสำคัญ: โรคไตเรื้อรัง; ฟอกเลือด; มวลร่างกาย; น้ำหนักแห้ง; ปัญญาประดิษฐ์; โครงข่ายประสาทเทียม

ผู้ประพันธ์บรรณกิจ: *ขจร ตีรณธนากุล*
อีเมล์: *khajohn_t@hotmail.com*

รับบทความ: 12 กันยายน 2566; ปรับปรุงแก้ไข 3 ตุลาคม 2566; รับตีพิมพ์: 19 ตุลาคม 2566

Introduction

The incidence of patients receiving hemodialysis (HD) has increased steadily from 3,410 per million patients in 2007 to 7,901 per million patients in 2016[1]. Dry body weight (DW), which dictates the ultrafiltration volume for each HD session, is a crucial parameter for the efficiency of HD treatment. However, the determination of the optimum DW remains a challenge for most nephrologists. The clinical assessment has proved to be inaccurate due to the low sensitivity of parameters obtained from physical examination including blood pressure, jugular venous pulse and the degree of peripheral edema[2]. This traditional clinical-based method not only varies between different assessors but also varies within the same assessor. As a result, various tools have been introduced to assist nephrologists in obtaining more objective data in the assessment of DW. One of the most reliable tools is the body composition monitor (BCM) using the principle of bioelectrical impedance analysis, which measures the resistance caused by an electrical current passing through living tissues. BCM guided dry weight (BCM-DW) assessment is increasingly used in large HD centers. However, the availability of BCM is limited in smaller centers due to the high cost.

Machine learning (ML), a subset of artificial intelligence, has potential in assisting nephrologists in the assessment of DW. Studies in ML using input data that included time-series data during HD treatment and BCM-DW as an output for the model prediction are lacking[3-6]. The present study attempted to develop a model for DW prediction using ML.

Methods

Study design and population

The present retrospective study included the data of HD treatment from two institutions: King Chulalongkorn Memorial Hospital between January 1, 2017, to December 31, 2021; and Bhumirajanagarindra Kidney Institution Hospital between January 1, 2018 to December 31, 2022. The data were retrieved from each session of HD because each session was used separately for model learning.

The inclusion criteria were: (a) age 18-86 year; (b) receiving twice of trice weekly HD treatments; (c) dialysis vintage ≥6 months; (d) no recent history of infection, heart failure, cardiac arrhythmia, sudden cardiac arrest; (e) adequate HD according to KDIGO recommendations (weekly standard Kt/V ≥2.1 or single-pool Kt/v ≥ 1.2)[7]; (f) available of BCM-DW data on the same day as HD treatment. The exclusion criteria were: (a) active heart conditions including ischemic heart disease within or congestive heart failure within the past 6 months and moderate to severe valvular disease; (b) atrial fibrillation; (c) amputated limb; (d) cirrhosis; (e) pregnancy or breastfeeding; (f) body mass index <16 kg/m^2 or >34 kg/m^2. Dialysis sessions with the infusion of hyperosmotic agents, adjustment of dialysate electrolyte concentrations according to profiling, and incomplete data were also excluded. The study was approved by the institutional review board of King Chulalongkorn Memorial Hospital and performed according to the principles of the Declaration of Helsinki.

Data collection

The domain of parameters including the demographic data for each patient and the HD prescription, the laboratory and time-series data for each HD treatment were collected. The DW obtained by BCM on the same day as HD treatment was also collected. The data retrieving process is shown in **Figure 1**.

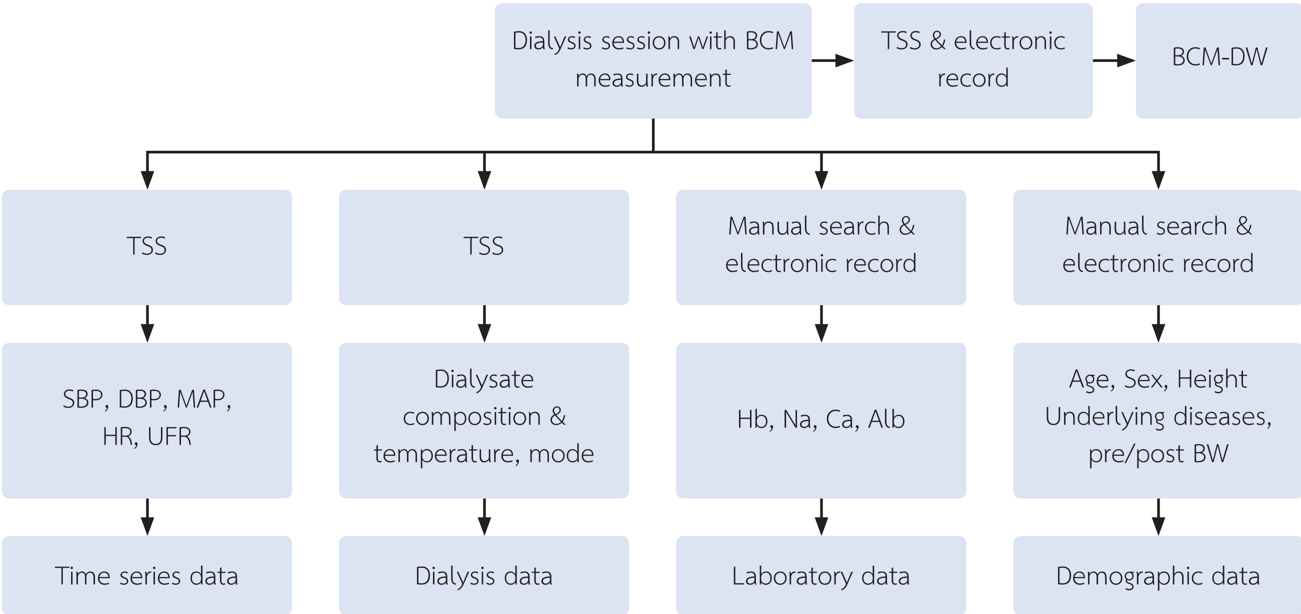

Figure 1 Data Retrieving process
BCM, Body composition monitor; TSS, Therapy support suite (Fresenius Medical Care, Bad Homburg, Germany); BCM-DW, BCM derived dry body weight; SBP, systolic blood pressure; DBP, diastolic blood pressure; MAP, mean arterial pressure; HR, heart rate; UFR, ultrafiltration rate; Hb, hemoglobin; Na, sodium; Ca, calcium; Alb, albumin; pre/post BW, pre-dialysis and post-dialysis body weight

Data preprocessing

The study first preprocessed the raw data of HD sessions by removing null values and normalizing ratio and interval data such as height, weight, and dialysate sodium using the StandardScaler command in TensorFlow. All features were then transformed to fall within the range of 0 and 1. Categorical data were marked and encoded using LabelEncoder from the 'sklearn.preprocessing' module and then encoded into binary variables using the 'get_dummies' function from the pandas module. Time-series data was scaled using the same method, and zero padding was used to create a matrix of the same size without affecting the model's performance. The output was labeled BCM-derived DW.

The data was divided into three sets for analysis: a training set (80%), a validation set (20%), and a testing set. HD sessions from the same patient were grouped together in each set to prevent the learning model from being biased by input from the same patient. The name of the institution was blinded. Two external validation configurations were used to test the robustness of the model, one using the Institution X patient database for training and validation and the Institution Y database for testing (configuration A), and the other using the Institution Y patient database for training and validation and the Institution X database for testing (configuration B).

Model development and Model evaluation

Our algorithm was built using Python v3.6.9 (https://www.python.org/) with Tensorflow backend v1.15.4 (https://www.tensorflow.org/), an open-source ML library. We deployed our application with this deep learning library based on the NVIDIA container image of TensorFlow, Release 20.11 (Nvidia Corporate, Santa Clara, CA, USA) as a virtual environment. The bidirectional Long Short-Term Memory (LSTM) model was used to capture dynamic variations in time-series data, and it was evaluated in three different approaches: LSTM regression (LSTMreg), LSTM classification (LSTMclass), and a combination of last embedding layer from time-series and non-time-series data (gtNN). For non-time-series data, multiple linear regression (LR), stacked machine learning models (STACK), and neural network models (gNN) were evaluated. The performance of each regression model was evaluated using mean squared error (MSE), while classification model was Confusion matrix, accuracy and F1 score. Bland-Altman plot was also used for determination of the performance compared with the gold standard (BCM-DW). Folded Empirical Distribution Function Curve (Folded EDFC) was plotted to compare the distribution of prediction between models. Cross-validation and hyperparameter tuning were applied to prevent overfitting. Finally, the robustness of each model was evaluated using two external validation configurations. The final model was chosen based on the smallest mean square error (MSE), lowest mean difference, and smallest limit of agreement. The final model was evaluated in terms of agreement, with upper and lower boundaries of the limit of agreement set at 1.96 standard deviation (SD). Demographic data for any samples that fell outside this range were reviewed. The importance of each feature was also assessed using SHapley Additive exPlanations (SHAP) values, which showed the contribution of each feature to the predicted outcome. Other factors that were evaluated included the computerization needs and time latency for running the model.

Statistical analysis

Data with a normal distribution are presented as mean ± SD. Data with a skewed distribution are presented as median ± interquartile range (IQR). P-value <0.05 is considered statistical significance.

Results

A total of 581 HD sessions from institution X and 570 HD sessions from institution Y were included. **Table 1** shows baseline data of HD sessions according to the institution. The two groups showed comparable ages (64.9 vs. 67.7 years). The patients from institution X were more likely to be female (62.7% vs. 52.6%) with higher proportions of cerebrovascular disease, gout, and chronic lung disease. The patients from institution Y showed higher proportions of hypertension and diabetes (93.3% vs. 91.2% and 41.4% vs. 34.1%, respectively). Hemodiafiltration was the most prevalent mode of dialysis in institution X, whereas hemodialysis was the most prevalent mode in institution Y. The average BCM-DW was slightly lower in institution X (54.9 kg vs. 58.8 kg). The average net ultrafiltration volume (2.1 L vs 2.2 L) and ultrafiltration rate (9.1 mL/kg/hr vs 9.7 L/kg/hr) were comparable between the two institutions. The average dialysate electrolyte concentrations and laboratory data were also comparable between the two institutions. The relationships between different time-series data are shown in **Figure 2**.

Table 1 Baseline parameters from each hemodialysis session

	Institution X (n = 581 sessions)	Institution Y (n = 570 sessions)
Number of patients	42	244
Age, years	64.9 ± 9.6	67.7 ± 12.3
Female, number (%)	364 (62.7)	300 (52.6)
BCM-DW, kg	54.9 ± 10.5	58.8 ± 11.7
Pre-dialysis body weight, kg	56.9 ± 10.6	61.5 ± 11.7
Post-dialysis body weight, kg	54.7 ± 10.3	59.3 ± 11.6
Height, cm	159.0 ± 8.5	160.8 ± 8.3
Body mass index, kg/m^2	21.7 ± 3.8	22.7 ± 4.0

	Institution X (n = 581 sessions)	Institution Y (n = 570 sessions)
Net ultrafiltration, L	2.1 ± 0.9	2.2 ± 0.8
Ultrafiltration rate, mL/kg/hr	9.1 ± 4.4	9.7 ± 4.0
Comorbidities, number (%)		
Hypertension	530 (91.2)	532 (93.3)
Diabetes	198 (34.1)	236 (41.4)
Cerebrovascular disease	67 (11.5)	29 (5.1)
Gout	57 (9.8)	47 (8.2)
Chronic lung disease	28 (4.8)	3 (0.5)
Mode of dialysis, number (%)		
Hemodialysis	35 (6.0)	570 (100)
Hemodiafiltration	546 (94.0)	0 (0)
Dialysate electrolyte concentrations		
Dialysate Sodium, mg/dL Median (±IQR)	137.4 ± 1.4 138 (136-138)	137.7 ± 1.1 138 (138-138)
Dialysate bicarbonate, mg/dL Median (±IQR)	32 ± 1 32 (32-32)	32.1 ± 0.8 32 (32-32)
Dialysate Potassium, mean (±SD) mg/dL Median (±IQR)	2.1 (0.3) 2 (2-2)	2.6 (0.5) 3 (2-3)
Dialysate Calcium, mg/dL Median (±IQR)	3.0 (0.3) 3.0 (2.5-3)	2.8 (0.3) 3.0 (2.5-3)
Dialysate Temperature, Celsius Median (±IQR)	36.8 (0.5) 37 (37-37)	36.5 (0.5) 36.5 (36.3-36.8)
Laboratory data		
Hemoglobin, g/dL	11 ± 1.4	10.7 ± 1.5
Calcium, mg/dL	8.7 ± 0.7	8.8 ± 0.8
Sodium, mmol/L	138.7 ± 2.7	136.5 ± 4.2
Albumin, g/dL	3.8 ± 0.3	3.9 ± 0.4

BCM-DW, Body composition monitor-Dry body weight; IQR, interquartile range

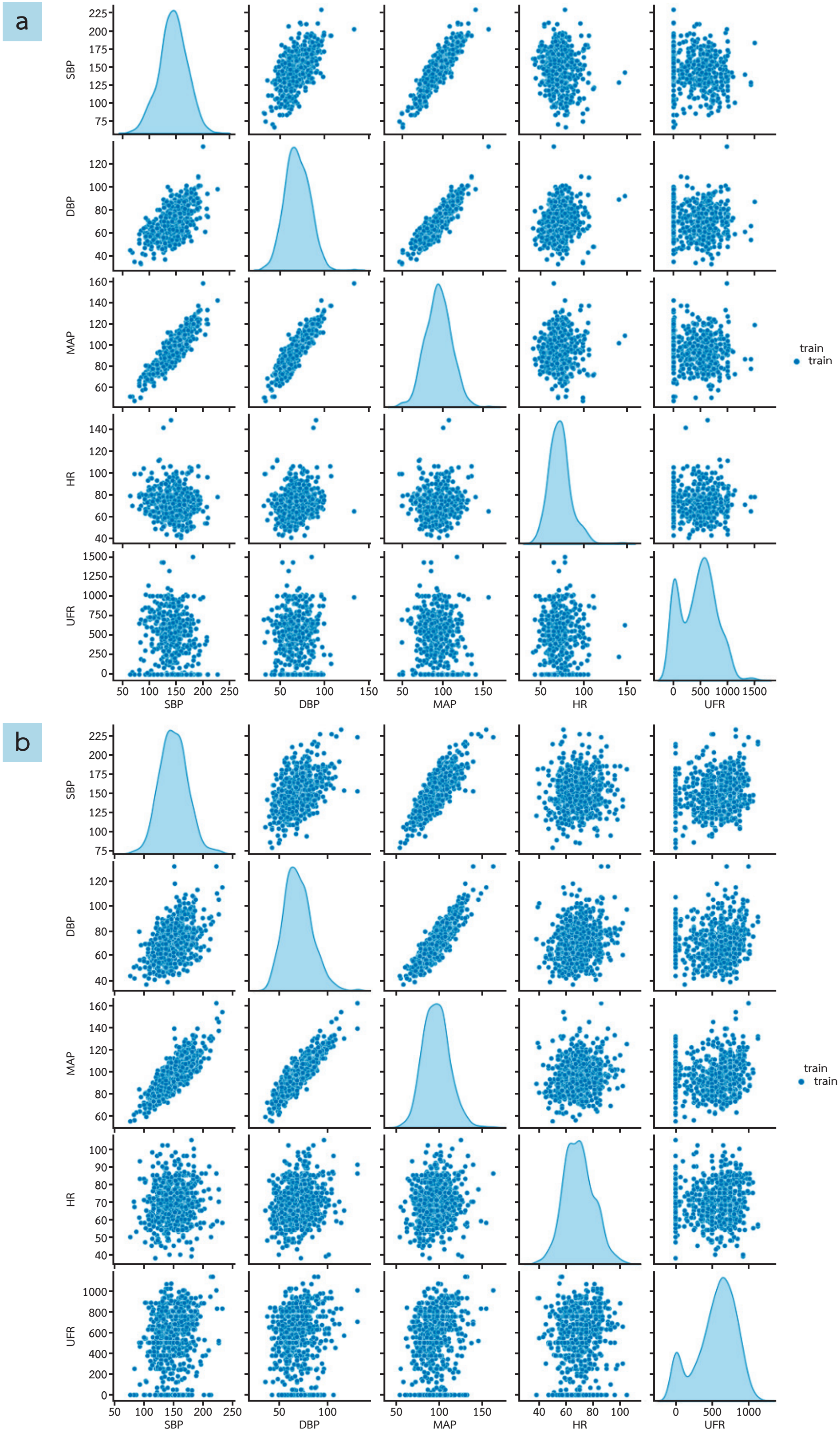

Figure 2 Scatter plots of the relationship between different time-series data
(a) Institution X; (b) Institution Y
SBP, systolic blood pressure; DBP, diastolic blood pressure; MAP, mean arterial pressure; HR, heart rate; UFR, ultrafiltration rate

The MSE varied with each model ranging from 1.68 to 151.39. The stacked machine learning model showed the lowest MSE and was selected as the final model. Folded EDFC is shown in **Figure 3**. The x-axis of the graph represented the mean difference between the ML predicted DW and the BCM-DW. The y-axis defined the probability of the samples with those values. The center of the graph featured a vertical line that represented the reference line of zero difference. The distance from this line and the peak indicated the estimated bias of each model. The shape of the graph is typically mountain-like, with the base representing the entire possible range of difference between the prediction from the new model and the BCM-DW bordered by the limits of agreement for that model. The stacked machine learning model (blue) showed a steep rise in the graph indicating a higher probability of the folded variable being small with acceptable distance from the peak to the zero line. The agreement between the new model and the BCM-DW is shown in the Bland-Altman plots in **Figure 4**. SHAP values for the parameters in the model are shown in **Figure 5**. The time latency for running the model was less than 1 minute and the size of the code was approximately 1 megabyte.

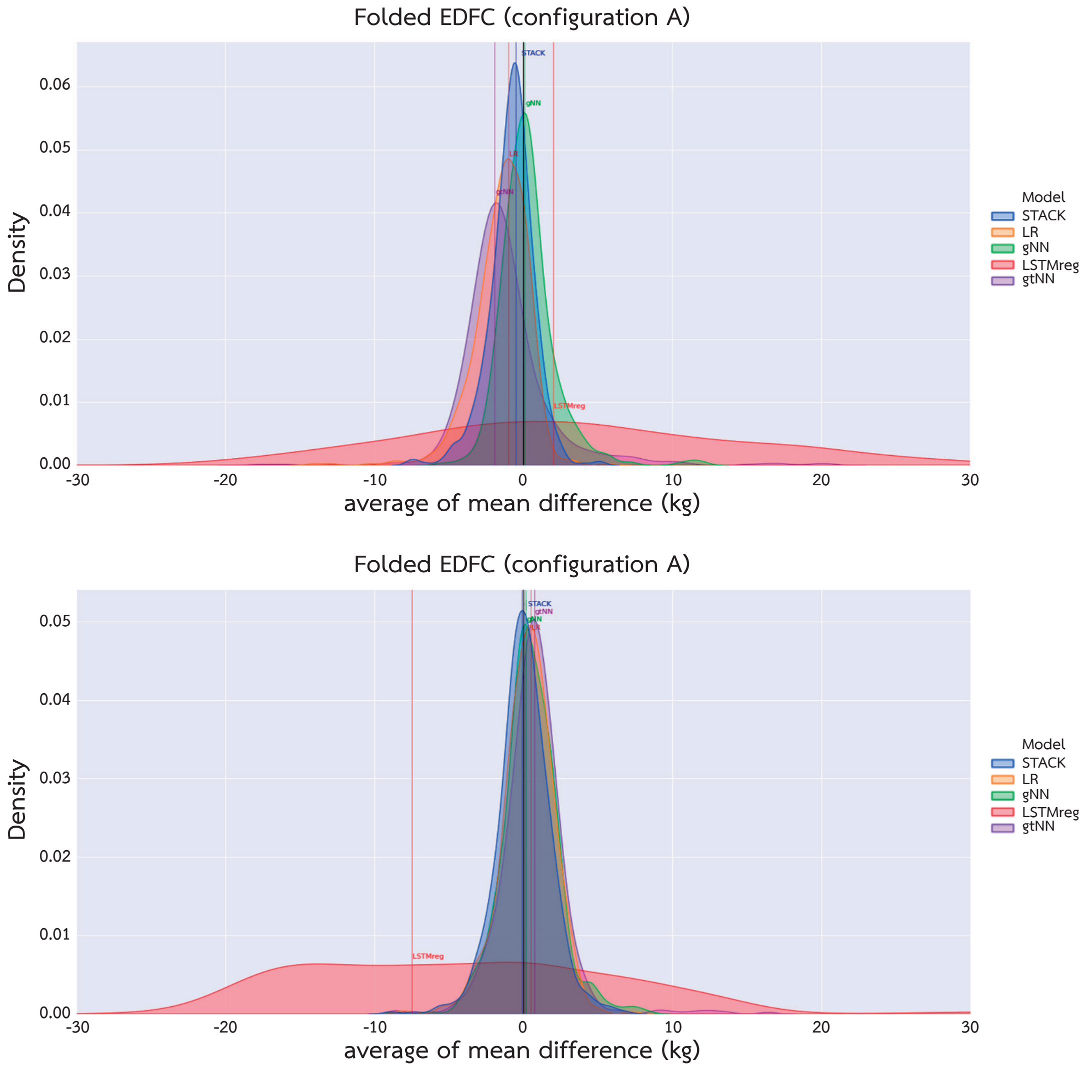

Figure 3 Folded Empirical Distribution Function Curves
Configuration A (upper); Configuration B (lower).
STACK, stacked machine learning models; LR, multiple linear regression; gNN, non-time-series data in neural network models; LSTMreg, LSTM regression; gtNN, combination of last embedding layer from time-series and non-time-series data

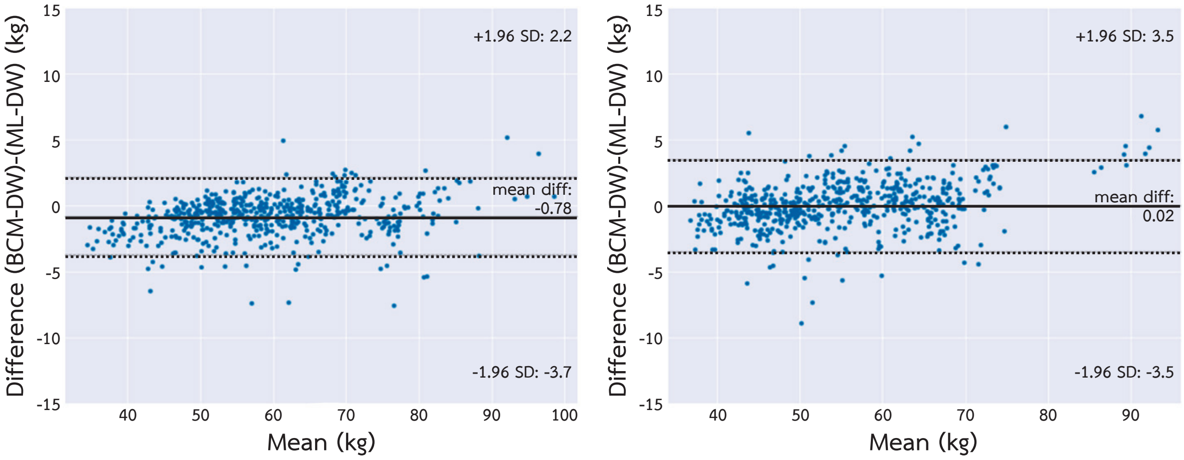

Figure 4 Bland-Altman plots showing the limits of the agreement between the dry body weight obtained from the stacked machine learning model and the body composition monitor
Configuration A (left) and configuration B (right).
BCM-DW, body composition derived dry body weight; ML-DW, machine learning derived dry body weight

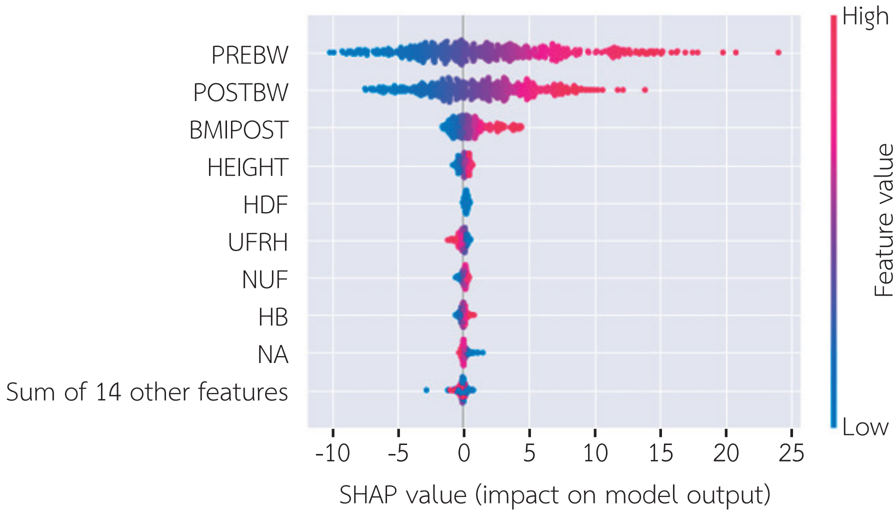

Figure 5 SHAP values of the parameters in the stacked machine learning model
PREBW, pre-dialysis body weight; POSTBW, post-dialysis body weight; BMI-POST, post-dialysis body mass index; HDF, hemodiafiltration; UFRH, ultrafiltration rate per hour; NUF, net ultrafiltration; HB, hemoglobin; NA, sodium

Discussion

The present study attempted to use various approaches including LSTMreg and LSTMclass in the prediction of DW using time-series data alone, but the results were unsatisfactory. Combining the time-series data with the non-time-series data also failed to yield positive outcomes. However, the time-series data had the potential in classifying HD sessions based on the gap between the pre-HD body weight and the post-HD body weight, but the model failed to do so possibly due to the interference by other factors such as reduced heart rate variation by the use of beta blockers or sinus node dysfunction commonly observed in older patients. The study also found that non-time-series data were able to predict BCM-DW as shown by several models such as gNN, LR, and STACK. Finally, the STACK model was selected because of its best performance.

The agreement between the STACK model (ML-DW) and the BCM-DW was acceptable with only a small bias. The mean differences of -0.78 kg in configuration A and

0.02 kg in configuration B indicated that the ML-DW was higher than that obtained by BCM in configuration A and lower in configuration B. Gradual probing of DW has been shown to have favorable impact on both blood pressure and left ventricular mass index, both of which are linked to better outcomes and reduced mortality in HD patients[8,9]. The randomized clinical trials have shown the benefit in the group of patients that was able to reach lower post-dialysis body weight up to 1.0 kgs (95% confidence interval (1.6, -0.5 kg), $p<0.001$). Thus, STACK could become a useful tool for DW assessment.

The present study showed that the limits of the agreement between the ML-DW and BCM-DW were quite wide. The maximum discrepancy was estimated at approximately 3.5 kgs, therefore, the ML-DW may not yet replace the estimated DW derived from BCM. However, when compared with the estimated DW obtained clinical assessment, the limits of the agreement were even wider ranging between 3.79 kg up to 7.21 kg[4,10]. Therefore, ML derived DW may be a better option for DW estimation compared with clinical assessment. The ML model likely offered greater objectivity, accuracy, affordability, repeatability, and generalizability compared with clinical assessment.

The present study is the first study that employed ML in the prediction of BCM-DW and used time-series data as input variables. External validation was also performed in HD patients from other centers. The present study has several limitations. The input data was collected retrospectively; therefore, it was not possible to identify all incidences and interventions that occurred during HD session that could interfere with the model prediction. The model did not incorporate cardiac parameters that could influence the blood pressure and the heart rate variability during ultrafiltration.

In conclusion, the present study served as a proof of concept study that ML could be a useful tool in DW prediction.

References

1. Annual Report Thailand Renal Replacement Therapy [Internet]. 2014. Available from: https://www.nephrothai.org/annual-report-thailand-renal-replacement-therapy-2007-2019-th/.
2. Kraemer M, Rode C, Wizemann V. Detection limit of methods to assess fluid status changes in dialysis patients. Kidney Int. 2006;69(9):1609-20.
3. Niel O, Bastard P, Boussard C, Hogan J, Kwon T, Deschenes G. Artificial intelligence outperforms experienced nephrologists to assess dry weight in pediatric patients on chronic hemodialysis. Pediatr Nephrol. 2018;33(10):1799-803.
4. Guo X, Zhou W, Yu Y, Cai Y, Zhang Y, Du A, et al. Multiple Laplacian Regularized RBF Neural Network for Assessing Dry Weight of Patients With End-Stage Renal Disease. Front Physiol. 2021;12:790086.
5. 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 L(2,1)-Norm. Biomed Res Int. 2021;2021:6627650.
6. 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(5):507-13.
7. National Kidney F. K/DOQI clinical practice guidelines for chronic kidney disease: evaluation, classification, and stratification. Am J Kidney Dis. 2002;39(2 Suppl 1):S1-266.
8. Donadio C, Halim AB, Caprio F, Grassi G, Khedr B, Mazzantini M. Single- and multi-frequency bioelectrical impedance analyses to analyse body composition in maintenance haemodialysis patients: comparison with dual-energy x-ray absorptiometry. Physiol Meas. 2008;29(6):S517-24.
9. Carter M, Morris AT, Zhu F, Zaluska W, Levin NW. Effect of body mass index (BMI) on estimation of extracellular volume (ECV) in hemodialysis (HD) patients using segmental and whole body bioimpedance analysis. Physiol Meas. 2005;26(2):S93-9.
10. Agarwal R, Bouldin JM, Light RP, Garg A. Probing dry-weight improves left ventricular mass index. Am J Nephrol. 2011;33(4):373-80.
11. Agarwal R, Alborzi P, Satyan S, Light RP. Dry-weight reduction in hypertensive hemodialysis patients (DRIP): a randomized, controlled trial. Hypertension. 2009;53(3):500-7.
12. Jian Y, Li X, Cheng X, Chen Y, Liu L, Tao Z, et al. Comparison of bioimpedance and clinical methods for dry weight prediction in maintenance hemodialysis patients. Blood Purif. 2014;37(3):214-20.