



## OPEN Time series forecasting of weight for diuretic dose adjustment using bidirectional long short-term memory

Heejung Choi<sup>1</sup>, Yunha Kim<sup>1</sup>, Heejun Kang<sup>2</sup>, Hyeram Seo<sup>3</sup>, Minkyung Kim<sup>3</sup>, JiYe Han<sup>3</sup>, Gaeun Kee<sup>1</sup>, Seohyun Park<sup>1</sup>, Soyoung Ko<sup>1</sup>, HyoJe Jung<sup>1</sup>, Byeolhee Kim<sup>1</sup>, Jae-Hyung Roh<sup>4</sup>, Tae Joon Jun<sup>5,7</sup> & Young-Hak Kim<sup>6,7</sup>✉

Loop diuretics are prevailing drugs to manage fluid overload in heart failure. However, adjusting to loop diuretic doses is strenuous due to the lack of a diuretic guideline. Accordingly, we developed a novel clinician decision support system for adjusting loop diuretics dosage with a Long Short-Term Memory (LSTM) algorithm using time-series EMRs. Weight measurements were used as the target to estimate fluid loss during diuretic therapy. We designed the TSFD-LSTM, a bi-directional LSTM model with an attention mechanism, to forecast weight change 48 h after heart failure patients were injected with loop diuretics. The model utilized 65 variables, including disease conditions, concurrent medications, laboratory results, vital signs, and physical measurements from EMRs. The framework processed four sequences simultaneously as inputs. An ablation study on attention mechanisms and a comparison with the transformer model as a baseline were conducted. The TSFD-LSTM outperformed the other models, achieving 85% predictive accuracy with MAE and MSE values of 0.56 and 1.45, respectively. Thus, the TSFD-LSTM model can aid in personalized loop diuretic treatment and prevent adverse drug events, contributing to improved healthcare efficacy for heart failure patients.

**Keywords** Deep learning, Electronic medical records, Long Short-Term memory, Time series forecasting, Clinician decision support system

Heart failure is a global public health problem and a major cause of hospitalizations worldwide<sup>1,2</sup>. The primary symptom of heart failure is congestion with an expansion of the extracellular fluid volume referred to as volume overload<sup>3</sup>. The congestion in heart failure results in increased cardiac filling pressures<sup>4</sup>. The European Society of Cardiology (ESC) guidelines strongly recommend using loop diuretics to alleviate symptoms of fluid overload and tracking weight changes to monitor volume status, since correlations between weight loss and outcomes have been reported<sup>5,6</sup>. Intravenous loop diuretics are one of the essential agents for treating heart failure, reducing left ventricular filling, avoiding pulmonary edema, and alleviating peripheral fluid retention<sup>7,8</sup>. In general, approximately 90% of patients hospitalized associated with heart failure are administered loop diuretics<sup>9</sup>. Accordingly, it is important to deliver the optimal dose of loop diuretics to manage heart failure. However, it remains challenging to administer the optimal dose of intravenous loop diuretics because robust clinical trial evidence to guide the use of diuretics is sparse<sup>4,7</sup>. Eventually, clinicians determine the dose of loop diuretics primarily based on their own clinical experiences and the dose varies depending on the health condition of patient.

<sup>1</sup>Department of Medical Science Asan Medical Institute of Convergence Science and Technology Asan Medical Center, University of Ulsan College of Medicine, 88, Olympicro 43Gil, Songpaju, 05505 Seoul, Republic of Korea. <sup>2</sup>Division of Cardiology, Asan Medical Center, 88, Olympicro 43Gil, Songpaju, 05505 Seoul, Republic of Korea. <sup>3</sup>Department of Information Medicine, Asan Medical Center, 88, Olympicro 43Gil Songpaju, 05505 Seoul, Republic of Korea. <sup>4</sup>Department of Internal Medicine, Chungnam National University College of Medicine, Chungnam National University Sejong Hospital, 20, Bodeum 7-Ro, Sejong-Si, 30099 Sejong, Republic of Korea. <sup>5</sup>Big Data Research Center, Asan Institute for Life Sciences, Asan Medical Center, 88, Olympicro 43Gil Songpaju, 05505 Seoul, Republic of Korea. <sup>6</sup>Division of Cardiology Department of Information Medicine Asan Medical Center, University of Ulsan College of Medicine, 88, Olympicro 43Gil Songpaju, 05505 Seoul, Republic of Korea. <sup>7</sup>These authors contributed equally: Tae Joon Jun and Young-Hak Kim. ✉email: mdyhkim@amc.seoul.kr

In clinical settings, various types of hospital data—including demographics, lab orders and results, medication administration, vital signs, and physical measurements—are generated and stored in electronic medical records (EMRs). The worldwide accumulation of EMRs has led to rapid advancements in artificial intelligence (AI) in healthcare. Among AI technologies, time-series forecasting (TSF) is a crucial research tool that offers future insights and supports decision-making processes. Moreover, time-series EMRs provide more abundant information than static two-dimensional data, such as tabular formats, facilitating their secondary use. Conducting time-series analysis on EMRs is essential to minimize information loss and identify significant temporal patterns. This approach offers an opportunity to improve patient healthcare by mining significant information. In this context, several time-series studies utilizing deep learning algorithms for clinical decision support have been conducted using EMRs. These algorithms have been successfully applied to predict readmission<sup>10,11</sup>, disease detection<sup>12–17</sup>, mortality<sup>18,19</sup>, and other outcomes<sup>20,21</sup>. However, the research focused on diuretic dosage determination is deficient. Accordingly, we suggested a new decision-making support tool for objective diuretic dosing using time-series EMRs.

This study aimed to develop a novel clinician decision support system (CDSS) for the effective administration of intravenous loop diuretics to heart failure patients. We developed an attention-based deep-learning (DL) model to predict weight changes in patients receiving diuretics for heart failure using time-series EMRs. Weight changes in patients receiving diuretics for heart failure are crucial indicators, as they reflect fluid loss. Consequently, our model can guide optimal dosing of loop diuretics by providing weight changes following administration of loop diuretics. First, we designed five AI models: basic bi-directional Long Short-Term Memory (Bi-LSTM), Bi-LSTM with an attention mechanism, a transformer, an artificial neural network, and a random forest model. Then, we separated the clinical data derived from EMRs at 12-h time intervals and entered four sequences at a time into each model. The models executed time series forecasting of the weight changes of heart failure patients treated with loop diuretics using EMR data. Subsequently, we compared the model performances and selected the final model with outstanding performance. Finally, we developed a TSF model of weights for Diuretic dose using LSTM layers (TSFD-LSTM) that predicts a patient's weight after two days of loop diuretics. The TSFD-LSTM was designed based on the bidirectional LSTM with attention mechanisms that process sequential data in both forward and backward directions. The TSFD-LSTM was internally validated using EMRs from the Asan Medical Center (AMC) and achieved about 85% predictive accuracy within 1 kg. Consequently, our longitudinal data-driven AI models for EMRs can serve as a CDSS assisting clinicians in determining appropriate diuretic doses and maximizing the drug efficacy for heart failure patients.

## Methods

### Ethical approval

The Institutional Review Board of AMC approved the protocols of this study (No. 2021–0321), which were performed in agreement with the 2008 Declaration of Helsinki. Also, this study excluded the requirement for informed consent as the database used for this study consisted of anonymous, de-identified data. All experiments were performed following relevant guidelines and regulations.

### Data source

We used the EMRs database of AMC, Seoul, South Korea, between January 2000 and November 2021. The EMRs used in this study were obtained from Asan Biomedical Research Environment (ABLE) platform that Asan Medical Center has been developing a de-identification system for biomedical research<sup>22</sup>. This platform guarantees the accuracy and completeness of the data.

### Scheme of the study

This longitudinal study designed a novel LSTM framework to predict weight changes in heart failure patients treated with loop diuretics using EMRs. This study process was divided roughly into data preparation and model experiments. First, we selected acceptable patients for this study based on the inclusion criteria. Subsequently, we extracted the data in the EMRs and conducted pre-processing, such as outlier detection and data normalization. Next, data resampling into time-series sequences was carried out. In the end, we created the final dataset for time-series predictions and separated the dataset into model training (60%), test (20%), and validation (20%) datasets. Second, we designed five AI models: basic Bi-LSTM, Bi-LSTM with an attention mechanism, a transformer, an artificial neural network, and a random forest model. The training dataset served as the input for each model. The model performances of models using the validation dataset were compared using mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE), indicating the predicted errors between the actual value and the model's prediction. Then, we selected a model with the best performance as the final model.

### Data preparation

#### *Cohort selection*

We constructed inclusion criteria for patients to make the best use of the EMRs and successfully develop the LSTM model. Patients who were diagnosed with heart failure at the time of hospital admission were selected. We considered not only direct hospitalizations to general wards but also hospitalizations via the emergency department. Among them, the patients who were prescribed intravenous furosemide as the loop diuretic, and had weight measurements the following day were extracted. The entire patient flow chart is shown in Fig. 2.

#### *Feature preparation*

EMRs contain various medical variables, such as unstructured text and numerical, categorical, and temporal data, which can impact each other. However, some variables are often measured, whereas others are infrequently

measured. Accordingly, we had to select which features to use to optimize the model. Therefore, we selected 20 medicines, 18 diagnoses, and 17 lab tests clinically relevant to either heart failure or may influence loop diuretics. We used five demographic features, including sex, age, height, weight, and weight at admission, and six vital sign measurements. Also, we collected the loop diuretic dose. In particular, all the medication features were calculated as a standard conversion dose to consider the differences in the drug dose units. Then, we used the names of the drug ingredients to specify the drug categories since the names can group drugs into certain groups depending on the drug ingredients. Eventually, a total of 255 drug ingredients were categorized into 20 groups. The diagnoses were categorized based on the International Classification of Diseases, Tenth Revision (ICD10) code<sup>23</sup>. We used only three strings from the stratified characteristics of the ICD10 codes. Specifically, the ICD10 codes starting with I42, I43, and I50 were grouped as heart failure. Finally, 18 diagnoses related to heart failure or administration of loop diuretics were extracted from the 96 ICD10 codes. The complete list of the 96 diagnosis codes and 18 categories is found in Table S1. The categorical features, including diagnosis and sex, were processed to transform them into vectors that could serve as input to the model. Whether the patient had a certain disease was set as 1 or 0. For sex, male or female was encoded as 1 or 2, respectively. The rest of the features used the float value of the raw data.

#### Outlier detection

EMRs contain some entry errors since EMRs are entered manually. Therefore, we sought to detect vital sign outliers and replace them with missing values. Heart rates above 200 or below 35, respiratory rates above 50, oxygen saturation levels above 120 or below 80, body temperatures above 45 or below 35, systolic blood pressures above 150 or below 80, and diastolic blood pressures above 120 or below 50 were identified as outliers and replaced with missing values.

#### Data resampling into time series sequence

After data extraction, we reconstructed the data to make temporal sequences. First, all the data were aligned by each patient. Then, the timestamp of the first diuretic injection was used as the time index. Based on the time index, constant time intervals of 12 h were generated for each patient. For each time interval, if there were several values for the feature, the values were summed or averaged depending on the feature. Most features were averaged; however, concurrent medication features were summed. The weight was used as the most recent measurement. For example, if a patient had their albumin level examined twice during three intervals (36 h after the first diuretic injection), the average albumin level was used as the final value. Also, if a patient received 10 mg of statin twice within 12 h, the sum of the values was calculated and used as the final value.

#### Data interpolation

Our data had some missing values since the data was reconstructed every 12 h after the first diuretic injection since certain sequences did not have clinical events, such as the administration of medicines or lab tests. Eventually, missing values were created when all the sequences were organized by the patient. To successfully train the model, we had to carry out pre-processing of the missing values (Fig. 1). At first, the missing values were filled with the previous existing values, but there were still missing values. Subsequently, the missing values were filled with the next existing value.

#### Feature scaling

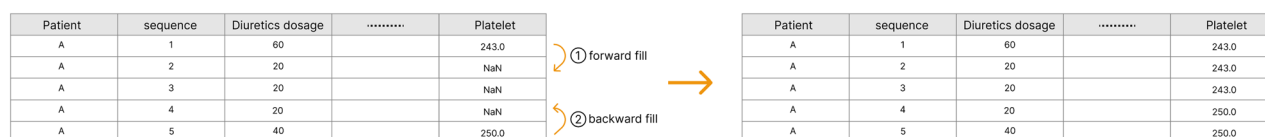
The deep learning model required feature scaling to make the varied ranges of each feature have a constant range within 0 to 1 to improve the efficacy of the model training. Specifically, we used a min–max scale method, which converts the raw values into float values from 0 to 1. The min–max scale formula used is shown in Eq. 1.

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

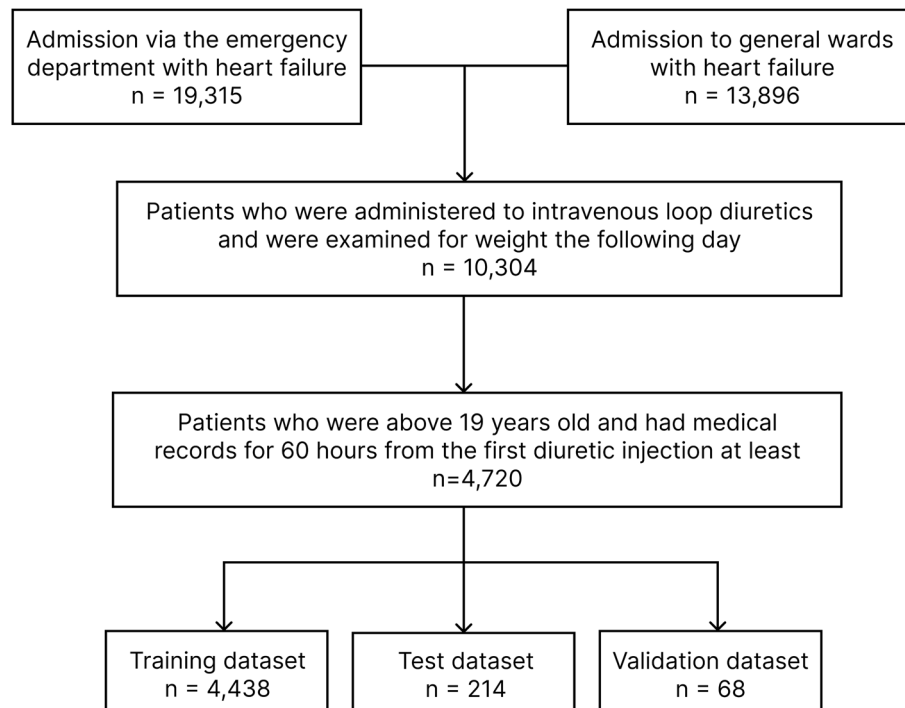
As a result, we scaled the features between 0 and 1. Finally, we created a final dataset for training and validating the model. The final dataset was made up of 182,386 rows and 66 columns for 4,720 patients. Each patient had at least five sequences since the models in this study used four sequences as inputs and predicted the following sequence.

#### Data split

The final dataset included 4,720 patients who satisfied the study criteria and was separated into three parts for model training, testing, and validation. To forestall a data leakage, we made one patient belong to one specific group among the training, testing, and validation datasets (Fig. 2). In the end, the training set accounted for 60% of the final dataset, 4,438 patients and 110,042 sequences, the testing dataset was made up of 20% of the entire



**Figure 1.** Data interpolation.



**Figure 2.** Patient flow chart and data split.

dataset, 214 patients and 37,411 sequences, and the validation dataset was made up of 20% of the final dataset, 68 patients and 34,933 sequences.

### Model development and comparison

We developed five AI models: i) the basic Bi-LSTM model, ii) the Bi-LSTM with attention model, iii) a transformer model, iv) the artificial neural network model, and v) the random forest model. The five models were compared using identical test sets and a final model was selected.

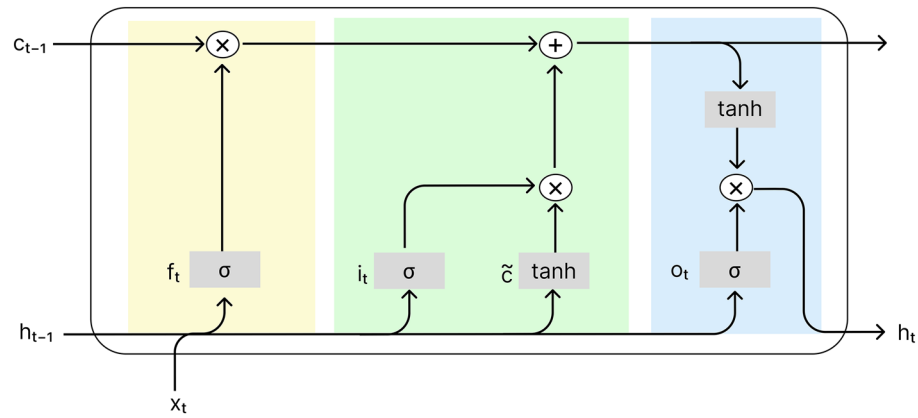
#### Sequence-to-sequence learning

Sequence-to-sequence (seq2seq) learning has been shown to effectively deal with temporal or sequential data in TSF and natural language processing (NLP) problems. The seq2seq model is comprised of an encoder and decoder module. The encoder compresses input sequences into latent context vectors and the decoder generates a target sequence using the context vectors. The inside of the model can be composed of numerous types of networks, but it primarily utilizes recurrent neural networks. Additionally, there are three formats in a seq2seq module (Figure.S2): one-to-many format, where the input is one vector and the output is multiple sequences; many-to-one format, where the input is multiple sequences and the output is one vector; and many-to-many, where both of input and output are multiple sequences. Among them, we selected a many-to-one format that produces a prediction using four sequences.

#### Long short-term memory

LSTM is one of the types of recurrent neural networks (RNN) and has been applied to several fields of modeling of sequential data such as time series and language<sup>24</sup>. The vanilla RNN can cause the network to forget significant information from earlier time steps and has exponential fading influence of older inputs, referred to as the Long-Term Dependencies or vanishing gradients problem<sup>25</sup>. An LSTM neural network alleviates the vanishing and exploding gradient problem of RNN by adding four components corresponding to cell state and input gate, output gate, and forget gate. The LSTM model takes a temporal sequence  $x_1, x_2, \dots, x_t$  as an input<sup>26</sup>. In step  $t$ , it updates the hidden state  $h_t$  by combining the current input  $x_t$  and the previous hidden state  $h_{t-1}$ . Specifically, an LSTM has a chain structure with repeat modules consisting of three gates and one cell state (Fig. 3). The LSTM unit regulates information flow by chain actions of the four layers. Three input vectors are entered into the LSTM unit. Two of them come from the previous time step  $t-1$ : cell state  $c_{t-1}$  and hidden state  $h_{t-1}$ . The remainder is vector  $x_t$  that comes from the current time step  $t$ . The four layers are entered  $h_{t-1}$  and  $c_{t-1}$  as inputs and three gates, including forget and input and output gates, use a sigmoid function. The sigmoid function makes the gates positive values since they only have values between 0 and 1.

In terms of a gating system, the forget gate initially computes  $f_t$  and determines what information is eliminated in  $c_{t-1}$  (Eq. 2). If  $f_t$  is close to zero, the gates are blocked. Whereas if  $f_t$  is close to one, the gates allow the information of  $c_{t-1}$  to pass through. In the input gate,  $i_t$  and candidate for cell state  $\tilde{c}$  are computed (Eq. 3).  $\tilde{c}$  is calculated using a tangent hyperbolic function that has values between -1 and 1 and normalizes the information



**Figure 3.** Basic LSTM unit. The yellow and green and blue boxes indicate forget, input, and output gates, respectively.  $t$  specifies a current time step. Plus sign (+) and multiplication sign ( $\times$ ) represent the addition and multiplication operation, respectively. Also,  $\tanh$  indicates a hyperbolic tangent function.  $\sigma$  denotes a sigmoid function.  $c_{t-1}$  represents cell state in previous time step and  $\tilde{c}$  indicates a candidate for cell state in the current time step.  $h_t$  denotes a hidden state in the current time step.

that will be added to  $c_t$  (Eq. 4).  $c_t$  is determined using  $f_t$ ,  $i_t$ ,  $\tilde{c}$ , and  $c_{t-1}$  (Eq. 5). Then, the output gate computes  $o_t$  and  $h_t$  is determined by  $o_t$  and  $c_t$  with a tangent hyperbolic function (Eq. 6,7). Finally,  $c_t$  and  $h_t$  are passed to the next time step  $t + 1$  as long-term memory and short-term memory, respectively. LSTM repeats these steps and implements predictions. In the end, we implemented a seq2seq model architecture by using two Bi-LSTM layers as an encoder and decoder, respectively (Fig. 4).

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (2)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (3)$$

$$\tilde{c} = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (4)$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t \quad (5)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (6)$$

$$h_t = o_t \cdot \tanh(c_t) \quad (7)$$

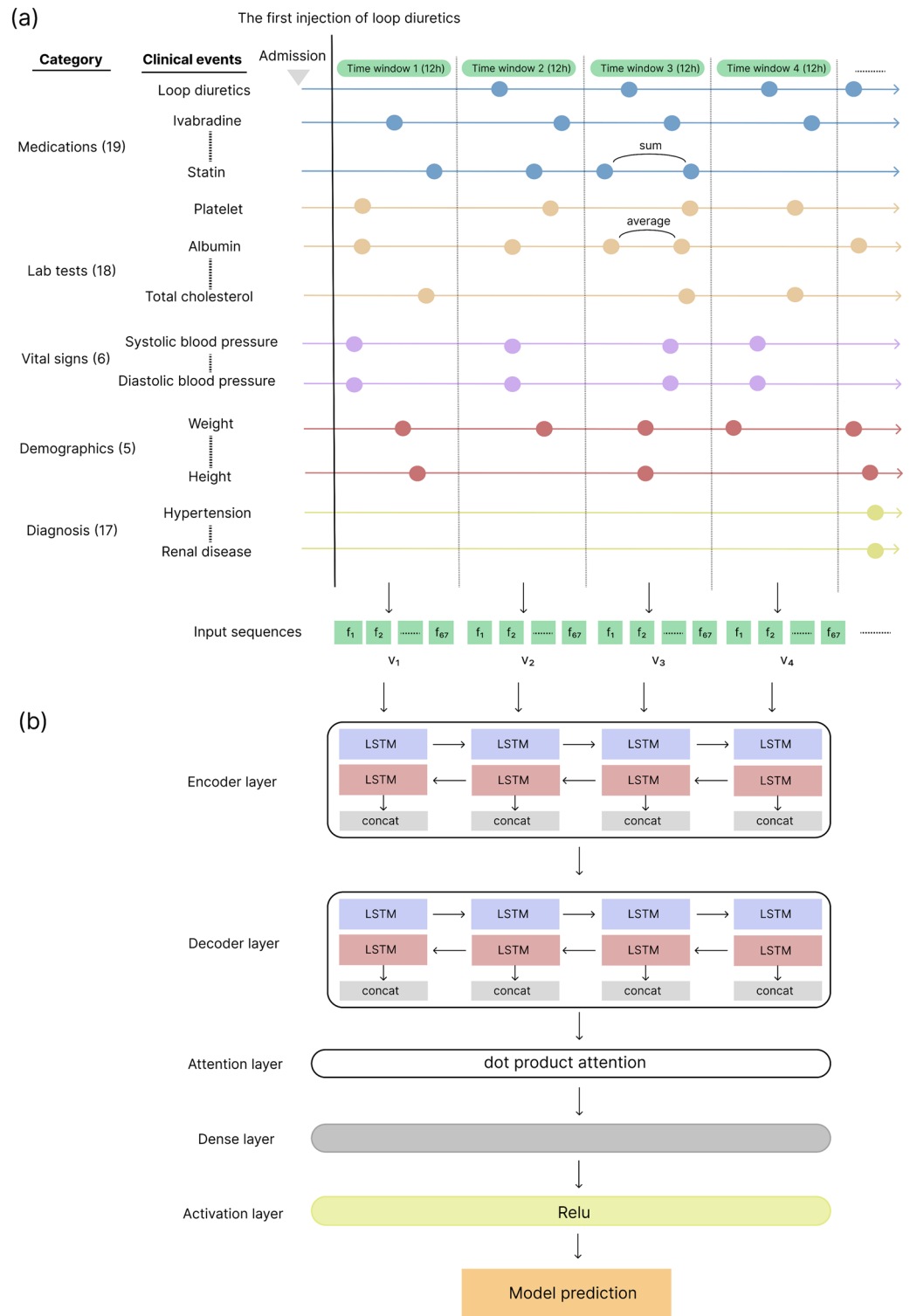
#### Attention mechanism

Numerous sequence-to-sequence tasks in the medical domain have demonstrated that the attention mechanism enhances the performance of deep learning models<sup>27–30</sup>. The idea of attention originated from the encoder and decoder structure of RNNs. The attention mechanism uses the encoders' input at every time step when the decoder predicts an output and then concentrates on notable parts of the input sequences by computing the attention similarity. In other words, the attention mechanism attends to certain parts of inputs relevant to the prediction. The similarity is calculated by the encoder's hidden state vectors at every time step and used to catch specific parts of the decoder's inputs. To obtain the attention similarity, we employed dot-product attention. The three variables, the query vector ( $Q$ ), key vector ( $K$ ) and value vector ( $V$ ), and the attention value were defined using Eq. 8. The attention value can be gained via three steps. At first, a dot-product between  $Q$ , and  $K$  is performed and used as an attention score. The attention score was scaled by dividing a square root of a dimension of  $K$  because the tremendous dimension of either  $Q$  or  $K$  makes difficult the model training. Next, the scaled attention score is normalized by using a softmax function. In the end, the attention value is computed by the dot-product of the normalized attention score and  $V$ .

$$Att(Q, K, V) = \text{softmax}\left(\frac{QK^t}{\sqrt{d}}\right)V \quad (8)$$

#### Evaluation metrics

Firstly, we measured the performance of the models using three metrics: the mean absolute error (MAE)<sup>31</sup>, the mean squared error (MSE)<sup>32</sup>, and the root mean squared error (RMSE)<sup>33</sup>. The three metrics are widely recognized as standard measures for evaluating regression models, each differing in their calculation formulas.



**Figure 4.** Scheme of the study. (a) 65 variables on the medications, lab tests, vital signs, demographics, and diagnosis were extracted and separated at 12-h intervals. If there are several values at a certain time interval, they are merged or averaged into one value. The four sequences served as input. (b) Two Bi-LSTM layers are used as encoder and decoder. The violet and red LSTM layers indicate forward and backward directions, respectively. Then, the concatenate layer combined both forward and backward outputs. Attention layer computes dot-product attention using both encoder and decoder outputs. Dense layer makes a fully connected layer to improve the model training's efficacy. Finally, activation layer with the ReLu function offers model predictions.

Specifically, the MAE calculates the average absolute differences between the predicted values and the actual target values. Unlike the MSE and RMSE, the MAE does not square the errors, thereby assigning equal weight to all errors regardless of their direction. This property makes the MAE particularly beneficial for understanding the magnitude of errors without considering either overestimations or underestimations. The MSE calculates the average of the squares of the errors, giving a higher weight to larger errors, which can be particularly useful in highlighting significant discrepancies. The RMSE is the square root of the MSE, which brings the error metric back to the original unit of measurement, making it more interpretable. We tried to interpret and demonstrate the experiment's results using the three metrics. In all these metrics, lower values indicate better model performance, as they represent smaller average errors between the predicted and actual values. Second, the predictive accuracy within 1 kg was calculated to assess how often the model's predictions fall within an acceptable range of actual values. In clinical settings, predicting patient weight within a small margin of error, such as 1 kg, can be practical for effective diuretic dose adjustments and treatment planning. Finally, we utilized the predictive accuracy within 1 kg as a performance metric.

#### *Final model selection*

We calculated the model prediction errors, and the model with the lowest error rate was selected as the final model. Finally, the LSTM with attention mechanisms model named TSFD-LSTM was adopted as the final model. The entire flow of the study including the final model architecture is shown in Fig. 4 and five models' hyperparameters are listed in S3.

## Results

### Baseline characteristics

The patients' characteristics for 4,720 heart failure patients in this study are summarized in Table 1. To maintain data integrity, we utilized the raw data for continuous variables without interpolation. Additionally, all drug-related variables were standardized to milligram units for a consistent analysis.

### Model evaluation

The performances of five models were evaluated using three metrics, namely MAE, RMSE, and MSE, and listed in Table 2. Among the five models, TSFD-LSTM was superior to other models in MAE, MSE, and accuracy. The five models' hyperparameters are listed in Table S3.

### Fluctuation in weight and diuretic dosage

We randomly selected four patients and visualized their weights and diuretic dosage fluctuations to check the correlation between weight and loop diuretics (Fig. 5). It was demonstrated that the patients had significant weights decrease during diuretic treatments and the diuretic dosages were continually adjusted. Patient A was 72-years-old female. Her height was 148.0 cm and her weight decreased from 56.9 kg to 46.4 kg during diuretic treatment. She was administered diuretics a wide of range 20 to 100. In addition to diuretics, she was administered ACE inhibitors and Beta-blockers. Patient B was 66-years-old male and had 126 sequences. His height was 163.3 cm and his weight decreased from 56.7 kg to 49.6 kg during diuretic treatment. He was prescribed a wide range of diuretic dosages from 20 to 110. He was diagnosed with atrial fibrillation and chronic ischemic heart and lung disease. Also, he had concurrent medications including ACE inhibitors and beta-blockers. Patient C 63-year-old female and had 87 sequences. Her height was 141.0 cm and her weight decreased from 65.9 kg to 51.6 kg during diuretic therapy. She had various diuretic dosage ranges of 10 to 1000. Additionally, she had various cardiovascular-related diseases, including atrial fibrillation, heart failure, hypertension, pulmonary embolism, renal disease, and valvular heart disease. She was administered direct oral anticoagulant twice. Patient D was a 79-year-old male and had a wide range of diuretic dosages from 20 to 800. His height was 157.5 cm and his weight decreased from 60.8 kg to 54.2 kg during diuretic therapy. He had atrial fibrillation and renal disease and was administered warfarin occasionally.

### Attention-based model interpretation

To enhance the interpretability of our deep learning model, we analyzed the attention weights of the attention layer. Utilizing temporal data with an attention mechanism is valuable for identifying which features and time steps are important when a deep learning (DL) model makes predictions<sup>34,35</sup>. By examining which features at specific time steps receive higher attention weights, we can gain insights into the significant factors influencing the model's predictions. Attention weights are normalized between 0 and 1, where a higher weight signifies greater importance of the corresponding feature. We visualized the attention heatmaps for two patients among those whose fluctuations in weight and diuretic dosage were previously examined (Fig. 6). For both patients, the data from the second and third time steps, particularly height, age, and comorbidities, had the most significant impact on the model's predictions. However, the contribution of these variables in TSFD-LSTM varied depending on the patient's health condition and time steps. Specifically, our model discerned the second time step as the most critical for Patient A, and the third time step as the most important for Patient B. Furthermore, for Patient A, hemoglobin (Hb) and hematocrit (HCT) measurements were significant contributors to the model's predictions. Conversely, for Patient B, the measurements of total protein and blood pressure were the most influential in the model's predictions. The heatmaps for all time steps of the two patients are shown in Supplementary Figure S4.

Variables	Cohort (N = 4,720)	Variables	Cohort (N = 4,720)
Demographics		Concurrent medications, mean (SD)	
Age (years), mean (SD)	67.9 (14.4)	ACE inhibitor	5.0 (5.4)
Female, n (%)	2,111 (44.7%)	ADP receptor inhibitor	108.7 (65.4)
Male, n (%)	2,609 (55.3%)	ARB	18.2 (30.7)
First weight (kg), mean (SD)	61.5 (13.5)	Aspirin	160.5 (201.9)
Last weight (kg), mean (SD)	59.4 (12.7)	Beta-blocker	12.3 (13.7)
Heights (cm), mean (SD)	161.5 (8.9)	Calcium channel blocker, dihydropyridine	29.2 (34.0)
Diuretic, loop, mean (SD)	76.7 (208.0)	DPP4 inhibitor	26.3 (42.9)
		Direct oral anticoagulant	35.3 (64.3)
Comorbidities, n (%)		Diuretic, thiazide	1.3 (0.8)
Angina	215 (4.6%)	Diuretic, thiazide-like	6.3 (4.9)
Arrhythmia	383 (8.1%)	Ivabradine	5.8 (4.4)
Atrial fibrillation	1,716 (36.4%)	Meglitinide	1.9 (1.4)
Cancer	761 (16.1%)	Other lipid lowering	191.1 (134.2)
Chronic ischemic heart disease	1,057 (22.4%)	PPAR agonist	7.0 (3.8)
Chronic lung disease	521 (11.0%)	SGLT2 inhibitor	13.6 (7.0)
Diabetes mellitus	1,806 (38.3%)	Statin	14.5 (9.9)
Dyslipidemia	118 (2.5%)	Sulfonylurea	5.3 (3.0)
Hypertension	2,446 (51.8%)	Warfarin	3.6 (3.0)
Intracranial bleeding	43 (0.9%)	Lab results, mean (SD)	
Liver disease	332 (7.0%)	Creatinine (mg/dL)	1.4 (1.1)
Myocardial infarction	345 (7.3%)	ALT (SGPT) (IU/L)	55.0 (195.3)
Peripheral arterial disease	96 (2.0%)	AST (SGOT) (IU/L)	65.8 (305.2)
Pulmonary embolism	545 (11.5%)	Total cholesterol (mg/dL)	123.3 (46.8)
Renal disease	1,586 (33.6%)	Total bilirubin (mg/dL)	1.9 (3.8)
Stroke / TIA	350 (7.4%)	Albumin (g/dL)	2.8 (0.6)
Valvular heart disease	1,003 (21.3%)	Total protein (g/dL)	5.9 (0.9)
Vital signs, mean (SD)		Alkaline phosphatase (IU/L)	124.7 (113.8)
Heart rate (bpm)	98.0 (25.3)	Total calcium (mg/dL)	8.5 (0.8)
Oxygen saturation (%)	97.2 (3.6)	Uric acid (mg/dL)	6.2 (3.3)
Systolic blood pressure (mmHg)	112.0 (17.5)	Hb (g/dL)	10.0 (1.9)
Diastolic blood pressure (mmHg)	67.8 (51.0)	WBC (10 <sup>3</sup> /uL)	8.7 (6.7)
Body temperature (°C)	36.7 (0.6)	Platelet (10 <sup>3</sup> /uL)	184.0 (115.2)
Tidal rate (min)	22.8 (7.4)	Hct (%)	30.6 (5.8)
		RBC (10 <sup>6</sup> /uL)	3.3 (0.7)
		Potassium (EM) (mmol/L)	4.0 (0.6)
		Sodium (EM) (mmol/L)	136.3 (5.9)

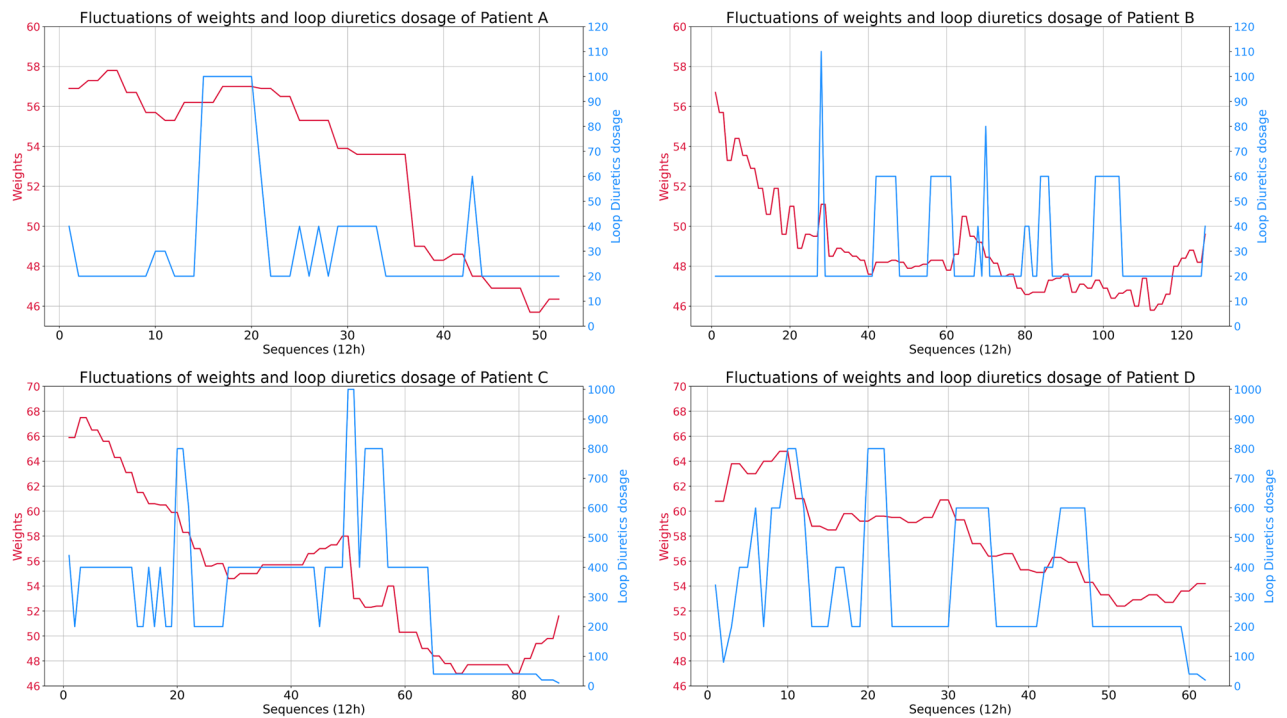
**Table 1.** Baseline characteristics.

	MAE	RMSE	MSE	Accuracy
TSFD-LSTM	0.56	1.20	1.45	85.34%
Basic BiLSTM	0.63	1.27	1.61	81.83%
Transformer	1.07	1.73	3.01	58.90%
Artificial Neural Network	1.22	1.92	3.68	52.73%
Random Forest	1.18	2.05	4.20	61.16%

**Table 2.** Summary of model performance. The predictive accuracy was calculated using a threshold within 1 kg. MAE mean absolute error, RMSE root mean squared error, MSE mean squared error.

## Discussion

Physicians face complicated decision-making processes during loop diuretic treatments because of limited guidelines on diuretic dosage. Loop diuretics reduce fluid overload, which indicates a clinical association between weight loss and diuretic response. Thus, a comprehensive understanding of individual health conditions is essential for prescribing loop diuretics. Physicians monitor the patients' weight changes after prescribing diuretics

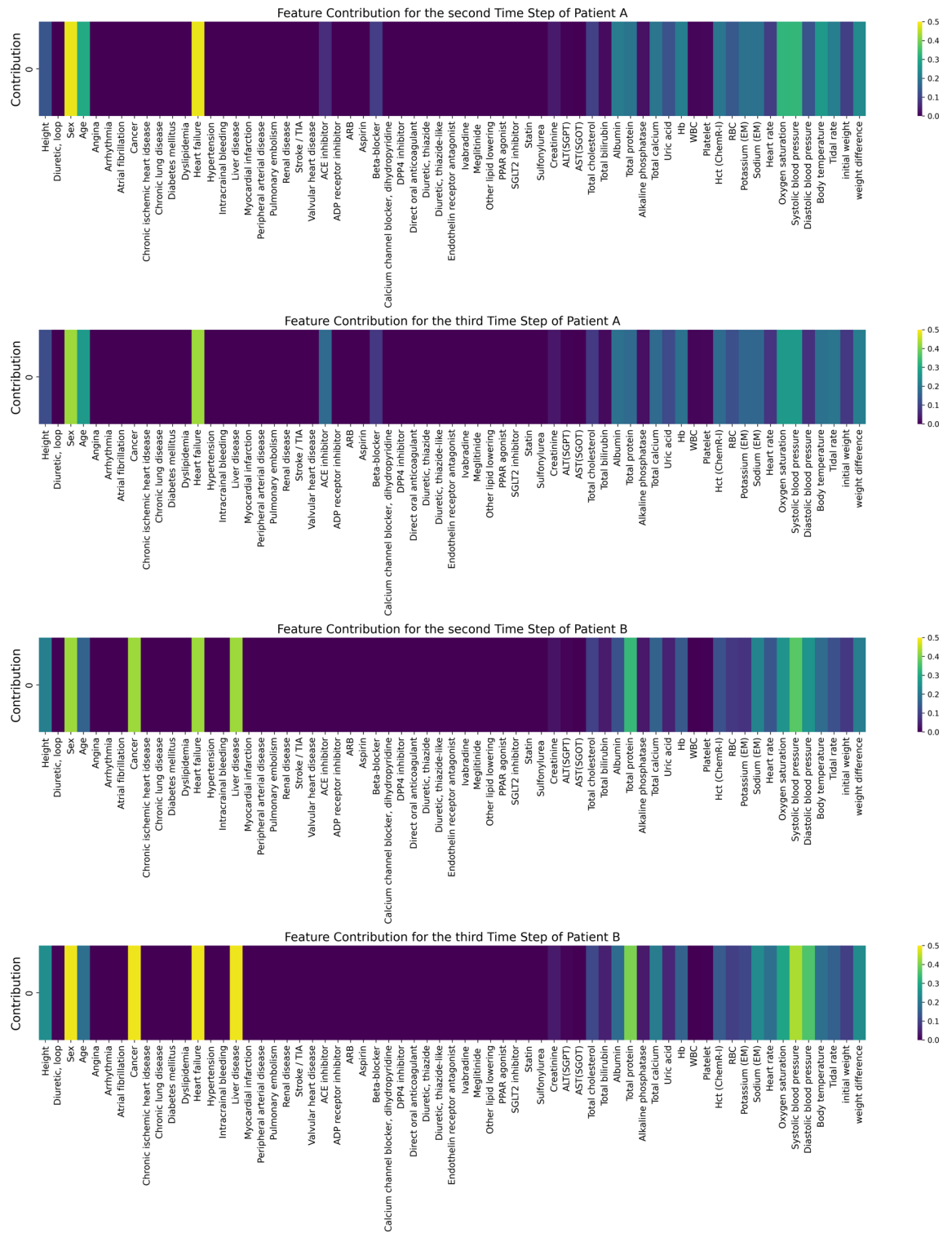


**Figure 5.** Fluctuation of weights based on the loop diuretic treatment. The x-axis indicates temporal sequences of 12 h. The left y-axis and red color represent weights. The right y-axis and blue color indicate loop diuretic dosages. The patient A and B had identical ranges of x and y axis. The patient C and D had equal ranges of x and y axis likewise.

to identify volume status. However, undiscovered diuretic interactions may exist besides weight change and unanticipated drug adverse events may be provoked. To address this, we investigated the background of loop diuretic use in a clinical setting and defined weight changes as the predictive target for developing a DL-based diuretic administration support tool. Specifically, we extracted and refined 65 clinical variables associated with cardiovascular diseases for model training. Additionally, we transformed 2-dimensional EMRs into time-series data with 12-h intervals for accurate weight change prediction. Finally, we compared five AI models: Bi-LSTM with attention, basic Bi-LSTM, a transformer, an artificial neural network, and a random forest model, selecting the best-performing model as the final TSFD-LSTM model.

This study suggests that the TSFD-LSTM model improves the management of loop diuretics treatment for heart failure patients and has the potential to be a useful CDSS. We implemented the TSFD-LSTM model, which had the seq2seq model of encoder and decoder structures for time series forecasting using Bi-LSTM algorithms. The TSFD-LSTM predicted weight changes after 48 h of heart failure patients who were prescribed loop diuretics using time-series data from their EMRs. Our results showed that continuous changes in diuretic dosage occurred based on weight fluctuations, with a gradual reduction in patients' weight (Fig. 5). Therefore, providing predictions of weight change after diuretic injection can serve as a meaningful guide for clinicians in determining diuretic dosages. Besides, we confirmed that the model accurately predicts actual weights by visualizing a prediction comparison plot (Fig.S5). This suggests that the TSFD-LSTM model can provide valuable references for diuretic dosage decisions. Accordingly, physicians may employ the TSFD-LSTM model to anticipate the potential weight change of a patient before adjusting the diuretic dosage, thereby helping to prevent excessive weight loss from incorrect dosing. Furthermore, the model can help avoid adverse drug reactions of diuretics, such as water-electrolyte imbalance, hypokalemia, and dehydration.

The TSFD-LSTM model employs a Bi-LSTM algorithm enhanced with an attention mechanism. The attention mechanism likely enhanced the TSFD-LSTM's performance since the primary difference between the TSFD-LSTM and the basic Bi-LSTM was the use of attention. The attention mechanism computes the final attention value using the encoders' inputs and reminds the decoder of specific parts it should focus on. EMRs contain vast amounts of medical data, including potentially dispensable information. The TSFD-LSTM model demonstrated that the attention function attends to important EMR information successfully. Additionally, attention heatmaps addressed the black box problem inherent in deep learning models by providing a visual interpretation of the model's decision-making process. We identified which features and time steps were most influential in the model's predictions, by analyzing attention weights. It enhanced our understanding of the model's prediction process and increased its transparency. In contrast, the transformer model showed poor performance despite using an attention mechanism. This was because transformer is specialized to capture long-range dependencies and interactions, whereas the LSTM is robust for short-term time series. In practice, the transformer model has been successfully applied to long-term time series forecasting and has shown excellent performance<sup>36–38</sup>. However, the transformer model has more parameters and is more complex than the LSTM model. Meanwhile, the gate



**Figure 6.** Individual-level attention heatmaps. The attention heatmaps for two patients (Patient A and Patient B) illustrate the importance of various features at second and third time steps in the prediction model. The attention weights, normalized between 0 and 1, indicate the relative significance of each feature, with higher weights denoting greater importance.

system of LSTM models is useful for recognizing short time-series-based patterns. Hence, the transformer model is very efficient in dealing with high-dimensional data, and the LSTM model is more conducive to limited data.

Experimental results indicate that the LSTM-based final model is more suitable for the short time series of EMRs than the transformer model.

### Limitation

This study is the first to develop a TSFD-LSTM model for use in clinical settings, but it has several limitations. Firstly, the largest limitation was that this study was a single-center study, and thereby the lack of diversity in the dataset restricted the model's generalizability and robustness. Although we attempted to acquire EMRs from other medical institutions, ethical issues related to patient privacy prevented us from obtaining additional data. This limitation resulted in a lack of external validation, which is crucial for ensuring the model's applicability across different populations and settings. Therefore, future work will focus on obtaining EMRs from multiple centers and regions to enhance the robustness and clinical utility of the TSFD-LSTM model. This will involve collaborating with other medical institutions and developing a standardized EMR framework for compatible data sharing. These steps are essential for validating the model's performance in diverse clinical environments and ensuring its generalizability. Second, our dataset had variations in sequence lengths ranging from 5 to 4,613 depending on the hospital stay period of each patient. We need to work on a hierarchical clustering of admission duration to provide more precise predictions. The model would then have to be retrained based on the disease severity of the individuals. Third, we used limited EMRs despite the effort to utilize them fully. Originally, the EMRs included more data, such as text and imaging. Initially, we tried to use chest X-ray data but realized that it was not primarily examined for diuretic therapy. Eventually, we only used structured tabular EMR data. In the future, we will use a multi-modal research approach using structured and unstructured data for the advancement of this model.

### Conclusion

We conducted a retrospective longitudinal study using EMRs from a tertiary hospital in South Korea from January 2000 to November 2021 to assist in the dosage determination of loop diuretics for heart failure patients. We extracted 65 clinical features of demographics, vital signs, lab results, diagnosis, and medication during in-hospital loop diuretic therapies. Subsequently, all the features were reconstructed as temporal sequences every 12 h after the first injection of the loop diuretic. Finally, we developed a pragmatic DL-based CDSS to predict weight changes after the prescription of loop diuretics. The model was designed to use four input sequences to create one target vector. The model predicted the weight change using the time-series data for 48 h. The bi-directional LSTM algorithm was adopted to leverage the temporal data and predict the outcome. Additionally, we implemented an encoder and decoder structure based on the seq2seq learning with an attention mechanism to elaborate the model. The MAE, MSE, and RMSE assessed the model performance. As a result, the final model, named TSFD-LSTM, surpassed the other baseline models, including the basic Bi-LSTM and transformer models. In particular, the TSFD-LSTM exceeded other baseline models, including basic BiLSTM and transformer, achieving 0.56 and 1.45 for the MAE and MSE, respectively. Additionally, it accomplished 85.35% accuracy for the weight change predictions within 1 kg. Ultimately, the TSFD-LSTM model demonstrated its potential to aid in the clinical decision-making process for adjusting loop diuretics dosage for heart failure treatment by providing physicians with weight change predictions.

### Data availability

No datasets were generated or analysed during the current study.

Received: 14 June 2024; Accepted: 26 July 2024

Published online: 31 July 2024

### References

1. Savarese, G. & Lund, L. H. Global public health burden of heart failure. *Card. Fail. Rev.* **3**(1), 7 (2017).
2. Savarese, G. *et al.* Global burden of heart failure: a comprehensive and updated review of epidemiology. *Cardiovasc. Res.* **118**(17), 3272–3287 (2022).
3. Mullens, W. *et al.* The use of diuretics in heart failure with congestion—A position statement from the Heart Failure Association of the European Society of Cardiology. *Eur. J. Heart Fail.* **21**(2), 137–215 (2019).
4. Felker, G. M. *et al.* Diuretic therapy for patients with heart failure: JACC state-of-the-art review. *J. Am. Coll. Cardiol.* **75**, 1178–1195 (2020).
5. McMurray, J. *et al.* ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure 2012: The Task Force for the Diagnosis and Treatment of Acute and Chronic Heart Failure 2012 of the European Society of Cardiology. Developed in collaboration with the Heart Failure Association (HFA) of the ESC. *Eur. Heart J.* **33**, 1787–1847 (2012).
6. McDonagh, T. A. *et al.* ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: Developed by the Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC) With the special contribution of the Heart Failure Association (HFA) of the ESC. *Eur. Heart J.* **42**, 3599–3726 (2021).
7. Felker, G. M. *et al.* Diuretic strategies in patients with acute decompensated heart failure. *N. Engl. J. Med.* **364**(9), 797–805 (2011).
8. Huxel, Chris, Avais Raja, and Michelle D. Ollivierre-Lawrence. Loop diuretics. *StatPearls [Internet]*. StatPearls Publishing, (2023).
9. Valente, M. A. E. *et al.* Diuretic response in acute heart failure: clinical characteristics and prognostic significance. *European Heart J.* **35**, 1284–1293 (2014).
10. Reddy, B. K. & Delen, D. Predicting hospital readmission for lupus patients: An RNN-LSTM-based deep-learning methodology. *Comput. Biol. Med.* **101**, 199–209 (2018).
11. Ashfaq, A. & Sant'Anna, A., Lingman, M., & Nowaczyk, S.,. Readmission prediction using deep learning on electronic health records. *J. Biomed. Inform.* **97**, 103256 (2019).
12. Lauritsen, Simon Meyer, *et al.* Early detection of sepsis utilizing deep learning on electronic health record event sequences. *Artificial Intelligence in Medicine* **104** (2020): 101820.
13. Rafiei, A., Rezaee, A., Hajati, F., Gheisari, S. & Golzan, M. SSP: Early prediction of sepsis using fully connected LSTM-CNN model. *Comput. Biol. Med.* **128**, 104110 (2021).

14. Zhang, D. *et al.* An interpretable deep-learning model for early prediction of sepsis in the emergency department. *Patterns* <https://doi.org/10.1016/j.patter.2020.100196> (2021).
15. He, Z. *et al.* Early sepsis prediction using ensemble learning with deep features and artificial features extracted from clinical electronic health records. *Crit. Care Med.* **48**(12), e1337–e1342 (2020).
16. Wu, C. *et al.* A method for the early prediction of chronic diseases based on short sequential medical data. *Artif. Intell. Med.* **127**, 102262 (2022).
17. Kim, K. *et al.* Real-time clinical decision support based on recurrent neural networks for in-hospital acute kidney injury: External validation and model interpretation. *J. Med. Inter. Res.* **23**(4), e24120 (2021).
18. Yu, K., Zhang, M., Cui, T., & Hauskrecht, M. Monitoring ICU mortality risk with a long short-term memory recurrent neural network. In PACIFIC SYMPOSIUM ON BIOCOMPUTING 2020 (pp. 103–114). (2019).
19. Thorsen-Meyer, H. C. *et al.* Dynamic and explainable machine learning prediction of mortality in patients in the intensive care unit: A retrospective study of high-frequency data in electronic patient records. *The Lancet Digital Health* **2**, e179–e191 (2020).
20. Van Steenkiste, T. *et al.* Accurate prediction of blood culture outcome in the intensive care unit using long short-term memory neural networks. *Artif. Intell. Med.* **97**, 38–43 (2019).
21. da Silva, D. *et al.* DeepSigns: A predictive model based on Deep Learning for the early detection of patient health deterioration. *Expert Syst. Appl.* **165**, 113905 (2021).
22. Shin, S. Y. *et al.* Lessons learned from development of de-identification system for biomedical research in a Korean Tertiary Hospital. *Healthcare Inform. Res.* **19**(2), 102 (2013).
23. World Health Organization. *International Statistical Classification of Diseases and related health problems: Alphabetical index* (World Health Organization, 2004).
24. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Comput.* **9**(8), 1735–1780 (1997).
25. Pascanu, Razvan, Tomas Mikolov, and Yoshua Bengio. On the difficulty of training recurrent neural networks. *International conference on machine learning*. PMLR, London.
26. Jin, Bo, *et al.* A treatment engine by predicting next-period prescriptions. *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*. (2018).
27. Ma, Fenglong, *et al.* Dipole: Diagnosis prediction in healthcare via attention-based bidirectional recurrent neural networks. *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*. (2017).
28. Zhang, Yuan. ATTAIN: Attention-based time-aware LSTM networks for disease progression modeling. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI-2019)*, pp. 4369–4375, Macao, China. (2019).
29. Song, H. *et al.* Attend and diagnose: Clinical time series analysis using attention models. *Proc. AAAI Confer. Artif. Intell.* <https://doi.org/10.1609/aaai.v32i1.11635> (2018).
30. Fridgeirsson, E. A., Sontag, D. & Rijnbeek, P. Attention-based neural networks for clinical prediction modelling on electronic health records. *BMC Med. Res. Methodol.* **23**(1), 285 (2023).
31. Qi, J., Du, J., Siniscalchi, S. M., Ma, X. & Lee, C.-H. On mean absolute error for deep neural network based vector-to-vector regression. *IEEE Signal Process. Lett.* **27**, 1485–1489 (2020).
32. Toro-Vizcarrondo, C. & Wallace, T. D. A test of the mean square error criterion for restrictions in linear regression. *J. Am. Stat. Assoc.* **63**, 558–572 (1968).
33. Chai, T. & Draxler, R. R. Root mean square error (rmse) or mean absolute error (mae)?—arguments against avoiding rmse in the literature. *Geosci. Model Devel.* **7**, 1247–1250 (2014).
34. Kaji, D. A. *et al.* An attention based deep learning model of clinical events in the intensive care unit. *PloS one* **14**(2), e0211057 (2019).
35. Gandin, I., Scagnetto, A., Romani, S. & Barbati, G. Interpretability of time-series deep learning models: A study in cardiovascular patients admitted to Intensive care unit. *J. Biomed. Inform.* **121**, 103876 (2021).
36. Zhou, T. *et al.* Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting. In *International conference on machine learning* (ed. Zhou, T.) (PMLR, 2022).
37. Wu, H., Xu, J., Wang, J. & Long, M. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Adv. Neural Inform. Process. Syst.* **34**, 22419–22430 (2021).
38. Zaheer, M. *et al.* Big bird: Transformers for longer sequences. *Adv. Neural Inform. Process. Syst.* **33**, 17283–17297 (2020).

## Acknowledgements

This research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HR20C0026). This work was supported by the Korea Medical Device Development Fund grant funded by the Korea government (the Ministry of Science and ICT, the Ministry of Trade, Industry and Energy, the Ministry of Health & Welfare, the Ministry of Food and Drug Safety) (Project Number: 1711195603, RS-2020-KD000097).

## Author contributions

H.C contributed to analysis of the study and writing of the manuscript. T.J and Y.-H.K. had contributions to the conception of this work. Y.K., H.K., H.S., M.K., J.H., G.K., S.P., S.K., H.J., B.K., J-H.R reviewed this manuscript. All authors read and approved the final version of the manuscript before submission.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-68663-6>.

**Correspondence** and requests for materials should be addressed to Y.-H.K.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024