Development of Hemoglobin Prediction and Erythrocyte Stimulating Agent Recommendation Algorithm (HPERA) Using Recurrent Neural Network in End-Stage Kidney Disease Patients

Hae-Ryong Yun

Department of Medicine

The Graduate School, Yonsei University

Development of Hemoglobin Prediction and Erythrocyte Stimulating Agent Recommendation Algorithm (HPERA) Using Recurrent Neural Network in End-Stage Kidney Disease Patients

Hae-Ryong Yun

Department of Medicine

The Graduate School, Yonsei University

Development of Hemoglobin Prediction and Erythrocyte Stimulating Agent Recommendation Algorithm (HPERA) Using Recurrent Neural Network in End-Stage Kidney Disease Patients

Directed by Professor Tae–Hyun Yoo

The Doctoral Dissertation
submitted to the Department of Medicine,
the Yonsei University Graduate School of Medicine
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Hae-Ryong Yun

December 2019

This certifies that the Doctoral Dissertation
of Hae–Ryong Yun is approved.

Thesis Supervisor: Tae–Hyun Yoo

Thesis Committee Member #1: Shin–Wook Kang

Thesis Committee Member #2: Kyungsoo Park

Thesis Committee Member #3: Wooju Kim

Thesis Committee Member #4: Dong–Ryeol Ryu

The Graduate School
Yonsei University

December 2019

ACKNOWLEDGEMENTS

First of all, I would like to thank Professor Tae–Hyun Yoo for his generous guidance during my doctorate studies. In addition, I would like to express my deepest gratitude and appreciation to Professor Shin–Wook Kang, who is a mentor of life, for leading me and his endless support for both education and research. I would also like to extend my gratitude to Kyungsoo Park, Wooju Kim, and Dong–Ryeol Ryu for their consideration and sharp teaching. In addition, I would like to thank Professor Seung Hyeok Han, and Professor Jung Tak Park for their helpful advice and warm encouragements. I would like to express my deep gratitude to Gyubok Lee for his efforts in development of algorithm and to Myeong Jun Jeon. I also want to express my special gratitude to my colleagues: Dr. Hyung Woo Kim, Dr. Young Su Joo, Dr. Ki Heon Nam, and Dr. Hyoungnae Kim. Although I did not mention it here, I am deeply grateful to all those who have shown a lot of help and affection in the doctoral program. Finally, I would like to say that I love my family who have always believed in and supported me. Thank you all for your generous support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Development of Hemoglobin Prediction and Erythrocyte Stimulating Agent Recommendation Algorithm (HPERA) Using Recurrent Neural Network in End-Stage Kidney Disease Patients

Hae–Ryong Yun

Department of Medicine
Yonsei University Graduate School of Medicine

(Directed by Professor Tae–Hyun Yoo)

Rationale & Objective: The optimization of anemia management is a challenging task due to the complexities of underlying diseases and heterogeneous responses to erythropoiesis-stimulating agents (ESA) in patients with end-stage kidney disease (ESKD). Recent studies have shown that machine learning (ML) algorithms can be an effective tool to predict hemoglobin (Hb) levels and determine the ESA doses in these patients. However, most of the proposed ML approaches are not designed to handle multivariate longitudinal patient data. Thus, we developed Hb prediction and ESA doses recommendation algorithm (HPERA) using recurrent neural networks (RNN).

Method: A total of 466 participants, who underwent hemodialysis in 7 hospitals in the Republic of Korea, were included in the present study. We selected 15 variables from extreme gradient boosting (XGBoost) algorithm. The outcome of the prediction algorithm was Hb levels in next month. In the recommendation algorithm, the outcome was ESA dose for target Hb in next month. Among various types of RNN families, gated recurrent units (GRU)

were used to build both the prediction and recommendation algorithm. In addition to holding out a separate validation dataset, we used a Gaussian noise layer following each input layer to avoid overfitting. We also performed linear regression, multilayer perceptrons, and extreme gradient boosting with extensive hyperparameter search to validate our GRU-based prediction algorithm. The performances of each model were evaluated in terms of the mean absolute error (MAE).

Results: The mean age of the study population was 57.8 years, 248 (53.2%) participants are male, and the mean observation period is 30.0 months. The best result of our prediction algorithm in terms of MAE was 0.59 g/dL and was obtained by two stacked GRU layers followed by a single hidden feedforward network with 6-month follow-up patient data. The best recommendation algorithm had 43.2 μg in MAE and this was obtained by one GRU layer followed by two layers of feedforward network. The HPERA had a lower overall ESA dose (μg/months) [155 (80-240) vs. 140 (70-210), $P<0.001$], decreased Hb difference (g/dL) [0.8 (0.4-1.4) vs. 0.6 (0.3-1.0), $P<0.001$)], and had a higher success and a lower failure rates of reaching target Hb compared to those in real practice.

Conclusion: The GRU-based prediction model outperformed previous ML methodologies, though hyperparameter turning was much simpler. Using the HPERA showed the possibility of a reduced amount of ESA, decreased Hb difference, and increased the reaching rate of target Hb levels. Our study revealed a great potential direction of anemia management using ML in ESKD patients.

Key words: end-stage kidney disease, recurrent neural network, gated recurrent unit, hemoglobin prediction and erythrocyte stimulating agent recommendation algorithm

Development of Hemoglobin Prediction and Erythrocyte Stimulating Agent Recommendation Algorithm (HPERA) Using Recurrent Neural Network in End-Stage Kidney Disease Patients

Hae–Ryong Yun

Department of Medicine
Yonsei University Graduate School of Medicine

(Directed by Professor Tae–Hyun Yoo)

I. INTRODUCTION

Anemia is a common complication in patients with chronic kidney disease (CKD) and end-stage kidney disease (ESKD) that arises early in the course of the disease.[1,2] In addition, anemia is known to be a risk multiplier for CKD progression[3], cardiovascular hospitalization[4], and mortality[5,6], proper management of anemia is important for improving the prognosis in these patients. Since the first erythropoiesis-stimulating agents (ESA) was approved by the US Food and Drug Administration in 1989, the administration of ESA has been a mainstay for anemia management in CKD and ESKD patients. The correction of anemia through appropriate and timely intervention using of ESA has been improved of renal outcome[7,8], survival in patients with heart failure [9], and has reduced the unnecessary blood transfusion and transfusion-related complications. However, prospective randomized controlled studies did not show the improvement of survival in a patient with ESKD [10] and pre-ESKD.[11,12] In addition, the Trial to Reduce Cardiovascular Events with Aranesp Therapy investigators have shown that maintaining a high hemoglobin (Hb) level with the use of ESA does not reduce the risk of death, cardiovascular event but rather increase the risk of stroke and cancer-related mortality.[13] Thus, for the efficacy

and safety of the use of ESA, current guidelines recommend balancing the potential benefits of reducing blood transfusions and anemia related complication against the risk of harm in the individual level.[14,15] However, the optimization of hemoglobin control is challenging because of the complexities of an underlying disease and heterogeneous responses to ESA in dialysis patients. Moreover, there is a lack of practical guidance on how to achieve these recommendations.

To support clinical practice in anemia management in ESKD patients, a variety of machine learning (ML) approaches have been proposed to predict Hb levels in these patients and to offer a personalized ESA dose based on the predicted Hb levels. [16-18] The main objective with previous ML approaches was predicting the next Hb levels using the current and previous Hb levels, ESA dose, and other clinical variables that reflect the patient's health conditions. [19] However, these variables used in previous studies have an organic association that change over time and most of the proposed ML approaches, including multilayer perceptron (MLP) and ensemble decision trees, are not designed to handle multivariate longitudinal patient data. Thus, we developed the Hb Prediction and ESA dose Recommendation Algorithm (HPERA) using sequence-aware neural networks called recurrent neural networks (RNN). In addition, we evaluated the efficacy of HPERA and validated the effect of HPERA could reduce the deviation of Hb levels. The outlined of this study as follows. 1) data collection, 2) data preparation 3) development of the HPERA, and 5) analysis of clinical validation. (**Fig. 1**)

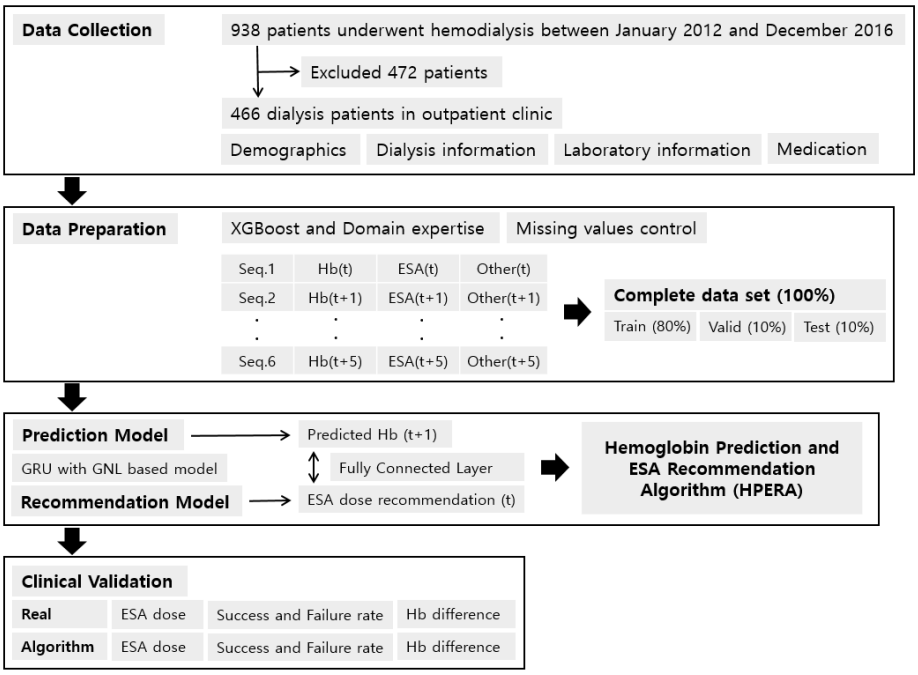

Figure 1. Overall outline of the study.

Abbreviation: XGBoost, extreme gradient boost; Seq, sequence; Hb, hemoglobin; ESA, erythrocyte-stimulating agent; t, time; GRU, gated recurrent unit; GNL, Gaussian noise layer.

II. MATERIALS AND METHODS

1. Data Collection and Preparation

The HPERA is based on the data from hemodialysis patients in seven tertiary-care general hospitals in the Republic of Korea. Patients, aged over 18 years and had at least 3 sequential clinical data, were included. The exclusion criteria were as follows: 1) Patients who underwent hemodialysis due to acute renal failure, 2) advanced heart failure (New York Heart Association class 3 or 4) or decompensated liver disease (Child-Pugh class 2 or 3), 3) past or current history of malignancy, 4) patients who received a blood transfusion during previous three months before observation period, or 5) patients with laboratory, dialysis,

and medication information for less than 3 months.

Between January 2012 and December 2016, a total of 938 patients who underwent hemodialysis in an outpatient clinic, were enrolled. Among these, 228 patients who received a blood transfusion during study period, 125 patients who underwent kidney transplantation, and 91 patients who were dead from any cause during the study period were excluded. We further excluded 28 patients with low occurrence (more than 50% of missing values). Finally, 466 participants were included in the study. **(Fig. 2)** The study was conducted in accordance with the principles of the Declaration of Helsinki, and the study protocol was approved by the Institutional Review Boards (IRB) of participating centers. Because clinical data used in the development of algorithm was collected retrospectively, informed consent was exempt.

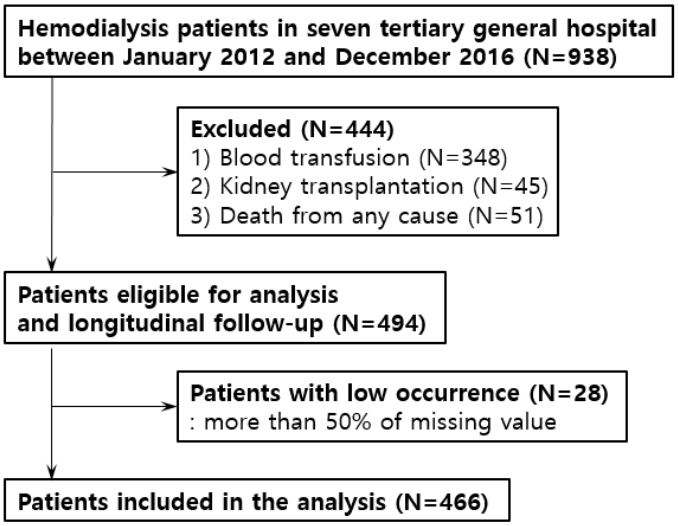

Figure 2. Flow diagram of the study subject.

We collected 116 variables of patient data, which were further divided into 4 main categories (demographics, laboratory information, dialysis information, and medications) during follow-up period. Patient's demographics were recorded at the time of the first clinic visit. Laboratory exams were performed once in a month and blood samples were drawn before the dialysis session.

Dialysis information including blood pressure, vascular access, dialysis modality, dry weight, pre-dialysis weight, etc. were recorded at the same day as well as parameters concerning the important medications, including iron supplements, type (darbepoetin-alpha, erythropoietin-alpha, or epoetin-beta), and amount of ESA. The monthly dose of erythropoietin-alpha and epoetin-beta were conversed and recalculated to darbepoetin-alpha.[20] Among these, we selected variables that are statistically associated with the Hb levels using feature importance calculated from extreme gradient boosting (XGBoost) algorithm.[21] Finally, 15 variables were selected based on the variables adopted in the previous study.[22] (**Fig. 3**) Each patient data consisted of a sequence of records, where each record stores values for 15 variables. However, since the follow-up period varies among patients, a certain period (window size) was slid over the whole longitudinal data to extract fixed time length sequence data for modeling. In this study, a window size from one to six was chosen, assuming that one window size which was a sequence, covers one month of patient health records. After generation, all data set was standardized to zero mean and unit standard deviation.

In order to obtain a robust model, the dataset was randomly divided into train, validation, and test set with 80%, 10%, and 10% of the data, respectively. All data for a single patient were assigned to exactly one of these splits. This split was used for both the prediction and recommendation models for training and performance evaluation. In addition to holding out a separate validation dataset, we used a Gaussian noise layer following each input layer to avoid overfitting and have the effect of data augmentation. This trick worked better than L1 and L2 regularization or dropout in our case. In addition, the multiple imputation by chained equation was used in order to handle missing variables.[23]

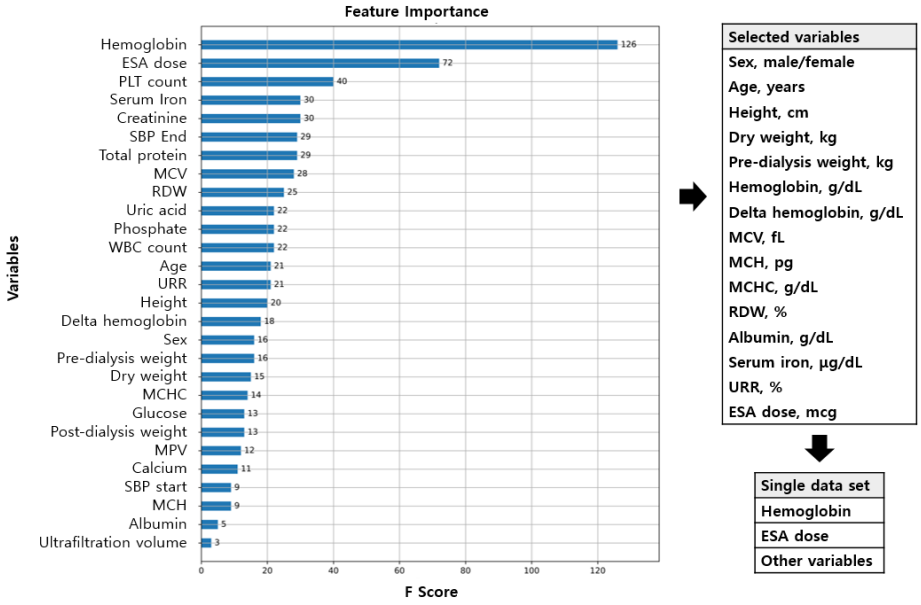

Figure 3. Variables ordered by XGBoost feature importance and domain expertise.

Note: Single data set composed of Hb, ESA dose, and other 13 variables. Each patient data consisted of a sequence of records, where each record stores values for 15 variables. **Abbreviation:** ESA, erythrocyte-stimulating agent; PLT, platelet; SBP, systolic blood pressure; MCV, mean corpuscular volume; RDW, reticulocyte distribution width; WBC, white blood cell; URR, urea reduction rate; MCHC, mean corpuscular hemoglobin concentration; MPV, mean platelet volume; MCH, mean corpuscular hemoglobin; Hb, hemoglobin.

2. Recurrent Neural Network

We used a special kind of RNN called gated recurrent unit (GRU) in order to build both the Hb prediction and ESA dose recommend models. Schematic view of the prediction and recommendation models are illustrated in **Fig. 4**. In addition, detailed architecture of the GRU and GRU with GNL were shown in **Fig. 5**. We also performed linear regression (LR), MLP, and XGBoost with extensive hyperparameter search to validate our GRU-based prediction model.

All the model was implemented in Python 3.6 with TensorFlow 2.0 and scikit-learn packages.

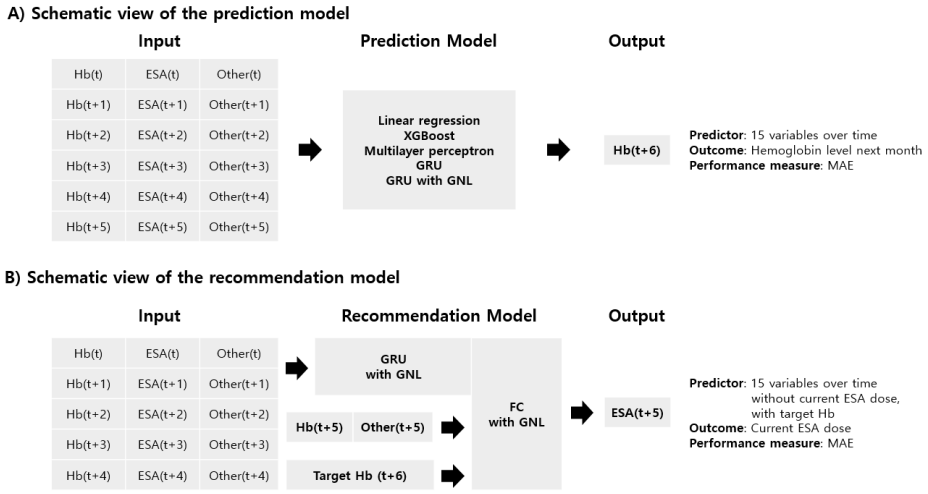

Figure 4. Schematic view of the prediction and the recommendation model.
Note: (A) Schematic view of the prediction model, **(B)** Schematic view of the recommendation model. MAE measures the difference between predicted and real values. In prediction models, MAE refers to the difference between the predicted Hb and the real estimated Hb. In recommendation models, MAE refers to the difference between the real ESA dose and the recommended ESA dose. **Abbreviation:** Hb, hemoglobin; ESA, erythrocyte-stimulating agent; t, time; XGBoost, extreme gradient boost; GRU, gated recurrent unit; GNL, Gaussian noise layer; MAE, mean absolute error; FC, fully connected network.

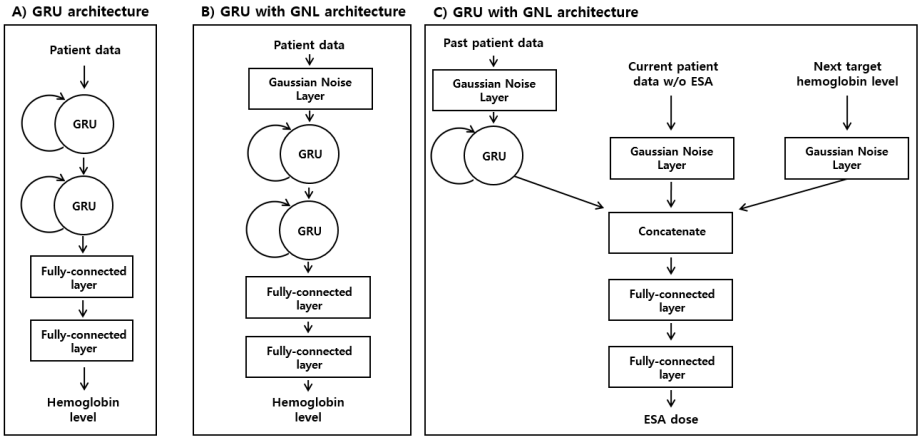

Figure 5. Detailed architecture of the GRU and GRU with GNL model in the prediction and recommendation model.

Note: (A) and **(B)** showed a GRU and GRU with GNL architecture used in the prediction model. **(C)** showed a GRU with GNL architecture used in the recommendation model. In addition, inputs and outputs were formulated differently from the hemoglobin prediction setups. **Abbreviation:** GRU, gated recurrent unit; GNL, Gaussian noise layer; ESA, erythrocyte-stimulating agent.

3. The Predictors and Primary Outcome of Interest

In the prediction model, the main exposure variables were 15 time-varying variables and the main outcome of interest was the Hb levels in next month. This prediction model aims that, if well-trained, the next Hb level as part of the recommendation model's input can be used as a placeholder for a target Hb and the output is the amount of ESA dose required to reach the input target Hb.

The ESA dose recommendation model was developed to suggest and appropriate ESA dose for individual patients to achieve a target Hb levels in next month. In the recommendation model, the main exposure variables were 15 time-varying variables except for the ESA dose in this month with a target Hb in next month. The target Hb level was defined as follows; if the current Hb below 10.0 g/dL, the target Hb level was Hb +1.0 g/dL; if Hb level within 10.0–

12.0 g/dL, the target Hb level was set to 11.0 g/dL; if the Hb level over 12.0 g/dL, target Hb levels was Hb-1.0 g/dL. The main outcome of interest in the recommendation model was the total ESA dose in this month.

The performance of the prediction and the recommendation models was evaluated in terms of model bias and accuracy. As a measure of bias, the Mean Error (ME) was used and, as measures of accuracy, the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) were used.

4. Clinical Validation

In order to evaluate the clinical implication of the prediction and recommendation model, we also compared clinical parameters such as ESA dose, Hb difference, and success/failure rate stratified by real clinical data and algorithm. The real ESA dose was the monthly dose of ESA prescribed by a doctor, and the algorithm ESA dose obtained from the recommendation model to reach the target Hb level in next month. Hb difference is defined as the absolute difference between Hb(t) and Hb(t+1). The result of real Hb difference was the difference between the real Hb(t) and the real Hb(t+1) level given the real ESA(t). In contrast, algorithm Hb difference was the difference between the real Hb(t) and the predicted Hb(t+1) calculated from the prediction model given the algorithm recommended ESA(t). We also compared success and failure rates of Hb target. In case Hb(t) < 10.0 g/dL, the success rate defined as the ratio of an increased Hb over 10.0 g/dL at the time of Hb(t+1). The failure rate was defined as the ratio of Hb(t+1) did not reach over 10.0 g/dL. In addition, in the case of Hb(t) within 10.0–12.0 g/dL, the success rate was defined as the ratio of Hb levels remaining within range to 10.0–12.0 g/dL in next month. Failure rate was defined as the Hb(t+1) level decreased below 10.0 g/dL.

5. Statistical Analysis

Continuous variables are described as means and standard deviations (SDs) or

the medians with interquartile ranges (IQRs). Categorical variables are described as numbers (%). The demographic and biochemical characteristics among hospitals were compared using one-way analysis of variance with the Scheffe's method or Kruskal-Wallis test for continuous variables, and the chi-square test for categorical variables. All statistical analyses were performed using Stata, version 15.1 (Stata Corporation, College Station, TX).

III. RESULTS

1. Baseline Characteristics of Study Participants

Detailed baseline characteristics of study participants are presented in **Table 1.** Among 466 participants, 248 (53.2%) was a male and the mean age of the study population was 57.8 ± 11.9 years. The median observation period was 30.0 (15.0–54.0) months. The mean dry weight was 60.6 ± 11.3 kg and pre-dialysis weight was 62.6 ± 11.7 kg. The mean Hb, MCV, MCH, and MCHC level was 10.8 ± 1.1 g/dL, 94.4 ± 4.9 fL, 31.3 ± 1.6 pg, and 33.2 ± 1.0 g/dL. The mean serum iron level was 71.1 ± 31.7 μg/dL. In addition, the mean urea reduction rate was 73.6 ± 6.0% at baseline. The median convert ESA dose was 160 (80–240) μg and it was significantly different across the hospitals, respectively (P<0.001).

Table 1. Baseline characteristics of study participants

Variables	Overall	A	B	C	D	E	F	G	P-value
Participants, N (%)	466	96 (20.6)	52 (11.2)	54 (11.6)	49 (10.5)	67 (14.4)	125 (26.8)	23 (4.9)	
Follow-up duration, month	30.0 (15.0–54.0)	20.5 (10.7–47.2)	41.0 (21.7–58.0)	45.0 (18.5–52.0)	28.0 (9.0–36.0)	32.0 (14.0–53.5)	39.0 (21.0–58.0)	19.0 (10.0–26.0)	<0.001
Age, years	57.8 (11.9)	59.4 (13.8)	56.2 (11.5)	58.9 (12.3)	58.6 (13.0)	58.2 (10.1)	56.8 (11.1)	56.6 (10.7)	0.08
Sex, No. (%)									0.33
Men	248 (53.2)	59 (61.4)	30 (57.7)	27 (50.0)	30 (61.2)	29 (43.3)	62 (49.6)	11 (47.8)	
Women	218 (46.8)	37 (38.6)	22 (42.3)	27 (50.0)	19 (38.8)	38 (56.7)	63 (50.4)	12 (52.2)	
Height, cm	162.0 (8.9)	161.7 (8.3)	163.5(9.1)	162.7 (8.5)	164.1 (7.3)	160.7 (9.7)	161.2 (9.1)	161.4 (8.2)	0.34
Weight, kg	61.4 (11.7)	59.3 (11.0)	62.6 (11.3)	60.8 (11.3)	65.1 (12.9)	61.8 (13.5)	60.9 (10.6)	62.2 (11.9)	0.40
BMI, kg/m^2	23.3 (3.7)	22.6 (3.2)	23.3 (3.2)	22.9 (3.4)	24.0 (4.6)	23.8 (3.9)	23.4 (3.7)	23.9 (4.3)	0.97
Dialysis information									
Hours per session, hour	3.9 (0.3)	3.8 (0.3)	3.6 (0.6)	4.0 (0.2)	3.9 (0.2)	3.9 (0.4)	4.0 (0.2)	3.9 (0.3)	<0.001
Blood flow rate, mL/min	238.5 (34.9)	220.2 (35.7)	250.0 (51.0)	254.8 (20.1)	211.8 (28.7)	249.4 (28.8)	246.6 (25.8)	234.4 (26.9)	<0.001
Dry body weight, kg	60.6 (11.3)	60.2 (10.5)	63.5 (11.7)	61.5 (11.3)	61.4 (10.9)	60.7 (12.8)	58.9 (10.4)	59.0 (13.1)	<0.001
Pre-dialysis weight, kg	62.6 (11.7)	62.5 (10.7)	65.8 (12.2)	63.1 (11.9)	63.9 (11.4)	63.1 (13.2)	60.6 (10.6)	61.3 (13.0)	<0.001
Post-dialysis weight, kg	60.6 (11.4)	60.3 (10.5)	63.8 (11.9)	61.1 (11.7)	61.7 (11.0)	60.9 (12.9)	59.0 (10.5)	59.1 (13.0)	<0.001
Dialysis mortality, N (%)									
Hemodialysis	513 (93.2)	101 (90.2)	52 (89.7)	52 (82.5)	56 (96.6)	73 (100.0)	154 (98.1)	25 (92.6)	<0.001
HDF	35 (6.4)	11 (9.8)	6 (10.3)	11 (17.5)	2 (3.4)	0 (0.0)	3 (1.9)	2 (7.4)	<0.001
Vascular access, N (%)									
AV-Fistula	339 (61.9)	37 (33.0)	42 (72.4)	44 (69.8)	41 (70.7)	49 (67.1)	113 (72.0)	13 (48.1)	<0.001
AV-Graft	64 (11.7)	12 (10.7)	5 (8.6)	7 (11.1)	4 (6.9)	5 (6.8)	26 (16.6)	5 (18.5)	0.21
Catheter	144 (26.3)	63 (56.3)	11 (19.0)	12 (19.0)	13 (22.4)	18 (24.7)	18 (11.5)	9 (33.3)	<0.001

Blood pressures									
Start-SBP, mmHg	146.4 (23.9)	146.4 (23.5)	148.7 (23.8)	148.8 (25.0)	147.7 (23.3)	144.7 (23.3)	144.2 (25.3)	143.6 (19.2)	<0.001
Start-DBP, mmHg	74.8 (13.7)	75.8 (17.1)	72.2 (11.0)	72.9 (11.7)	74.4 (14.4)	76.9 (12.0)	74.4 (13.4)	78.8 (11.4)	0.24
Start-HR, beat/min	74.5 (12.0)	74.7 (12.9)	76.3 (13.5)	72.8 (12.5)	75.6 (10.6)	72.9 (9.3)	73.7 (10.9)	79.6 (17.4)	0.13
End-SBP, mmHg	136.5 (24.0)	137.3 (22.5)	138.7 (23.7)	134.9 (24.2)	145.3 (25.3)	145.6 (19.9)	129.1 (24.1)	138.4 (18.9)	<0.001
End-DBP, mmHg	74.3 (13.8)	76.2 (15.3)	71.9 (14.1)	72.6 (14.8)	77.2 (13.8)	78.0 (10.3)	70.9 (13.3)	79.8 (11.2)	<0.001
End-HR, beat/min	76.1 (26.1)	74.7 (12.4)	75.2 (14.5)	75.7 (12.6)	77.6 (12.9)	82.1 (63.4)	74.1 (11.7)	76.3 (11.7)	0.51
Laboratory results									
WBC count, 10^3/μL	6.2 (1.9)	6.3 (1.7)	5.7 (1.6)	5.9 (2.0)	6.4 (1.7)	6.0 (1.9)	6.4 (2.1)	6.3 (2.3)	<0.001
Hemoglobin, g/dL	10.8 (1.1)	10.7 (1.1)	10.4 (1.0)	10.9 (1.2)	10.7 (1.1)	10.4 (0.9)	11.1 (1.1)	10.4 (0.9)	<0.001
Hematocrit, %	32.5 (3.5)	32.6 (3.5)	31.4 (3.1)	32.9 (3.8)	32.3 (3.4)	31.3 (2.9)	33.5 (3.4)	32.0 (3.0)	<0.001
MCV, fL	94.4 (4.9)	95.0 (5.3)	94.9 (4.7)	93.9 (4.9)	94.3 (5.3)	93.7 (4.6)	94.3 (4.7)	95.4 (5.2)	<0.001
MCH, pg	31.3 (1.6)	31.3 (1.8)	31.5 (1.7)	31.2 (1.6)	31.1 (1.8)	31.2 (1.6)	31.3 (1.5)	31.2 (1.7)	<0.001
MCHC, g/dL	33.2 (1.0)	33.0 (1.0)	33.2 (1.0)	33.2 (1.0)	33.0 (1.1)	33.3 (1.0)	33.2 (1.0)	32.7 (1.0)	<0.001
RDW, %	14.4 (1.3)	14.6 (1.4)	14.5 (1.4)	14.4 (1.3)	14.3 (1.5)	14.1 (1.3)	14.4 (1.2)	14.1 (1.3)	<0.001
PLT count, 10^3/μL	187.2 (62.8)	186.0 (63.0)	160.7 (49.2)	175.2 (58.6)	182.7 (56.6)	186.3 (49.1)	206.0 (71.0)	184.9 (56.8)	<0.001
Calcium, mg/dL	8.8 (0.8)	8.8 (0.7)	8.8 (0.6)	9.0 (0.8)	8.6 (0.6)	8.7 (0.7)	8.9 (0.8)	8.5 (0.7)	<0.001
Phosphate, mg/dL	6.0 (2.8)	10.8 (1.9)	4.6 (1.5)	5.2 (1.5)	4.9 (1.4)	5.1 (1.2)	4.5 (1.2)	4.5 (1.3)	<0.001
Glucose, mg/dL	133.8 (67.6)	157.4 (67.3)	153.2 (66.7)	119.7 (55.7)	127.5 (63.4)	137.4 (74.1)	115.6 (61.6)	171.0 (80.6)	<0.001
BUN, mg/dL	57.9 (18.8)	52.1 (18.6)	63.7 (21.2)	55.5 (19.0)	65.8 (18.8)	67.4 (18.8)	54.9 (15.4)	50.5 (13.9)	<0.001
Creatinine, mg/dL	9.4 (2.7)	9.7 (2.7)	9.8 (3.0)	9.5 (2.5)	10.3 (2.9)	10.4 (2.4)	8.5 (2.3)	7.7 (2.6)	<0.001
Total protein, g/dL	6.7 (0.5)	6.4 (0.5)	6.5 (0.4)	6.7 (0.5)	6.7 (0.5)	6.5 (0.5)	6.9 (0.5)	6.7 (0.5)	<0.001
Albumin, g/dL	3.8 (0.4)	3.4 (0.3)	4.0 (0.3)	3.9 (0.3)	3.9 (0.4)	3.7 (0.3)	4.0 (0.3)	3.9 (0.3)	<0.001
Electrolyte									
Sodium, mmol/L	137.3 (3.4)	136.9 (3.9)	139.0 (3.6)	138.1 (2.6)	137.1 (3.6)	137.9 (3.3)	136.4 (2.6)	137.7 (2.9)	<0.001
Potassium, mmol/L	4.7 (0.8)	4.5 (0.7)	4.6 (0.7)	5.0 (0.7)	4.7 (0.6)	5.0 (0.8)	4.6 (0.6)	4.6 (0.7)	<0.001
Chloride, mmol/L	100.8 (4.3)	100.3 (4.4)	102.0 (4.1)	98.5 (3.8)	100.8 (4.0)	99.4 (4.3)	102.1 (4.1)	101.4 (3.0)	<0.001

tCO2, mmol/L	22.3 (3.5)	23.8 (2.7)	20.1 (3.0)	23.3 (3.3)	22.3 (2.8)	21.4 (3.0)	21.6 (3.9)	25.0 (3.1)	<0.001
Serum iron, μg/dL	71.1 (31.7)	73.7 (32.7)	66.5 (28.6)	74.3 (32.0)	60.7 (28.4)	69.1 (30.8)	80.8 (34.0)	69.3 (22.2)	<0.001
Ferritin, ng/mL	147.0 (72.7–262.1)	162.0 (66.4–262.3)	215.1 (113.8–334.7)	120.4 (55.9–180.7)	107.0 (69.0–177.0)	164.7 (57.7–315.9)	190.9 (109.1–312.5)	122.1 (52.2–208.9)	0.03
Parathyroid hormone, pg/mL	251.0 (159.8–384.4)	265.5 (185.0–406.4)	-	249.3 (158.2–400.1)	212.8 (154.3–326.9)	367.5 (229.2–239.8)	251.2 (153.6–373.6)	239.7 (131.1–392.8)	0.02
Urea Reduction Rate, %	73.6 (6.0)	74.2 (5.8)	73.4 (5.3)	75.8 (5.4)	68.8 (6.8)	70.9 (6.2)	74.3 (5.5)	71.0 (5.7)	<0.001
Convert ESA per month, μg									
Darbepoetin-alpha, μg	160 (80–240)	200 (120–280)	140 (80–225)	140 (80–200)	150 (90–200)	120 (60–200)	150 (80–240)	160 (80–200)	<0.001

Note: Values in categorical variables indicates the number of corresponding participants (percentage) unless otherwise indicated. The monthly dose of erythropoietin-alpha and epoetin-beta were recalculated to darbepoetin-alpha. **Abbreviation:** A, National Health Insurance Service Ilsan Hospital; B, CHA Medical College Bundang Hospital; C, Yonsei University College of Medicine Severance Hospital (Shinchon); D, Yonsei University College of Medicine Severance Hospital (Gangnam); E, Ewha Womans University Hospital; F, Gachon University Gil Medical Center; G, Catholic Kwandong University; IQR, interquartile range; SD, standard deviation; BMI, body mass index; HDF, hemodiafiltration; AV, arteriovenous; SBP, systolic blood pressure; DBP, diastolic blood pressure; HR, heart rate; WBC, white blood cell; MCV, mean corpuscular volume; MCH, mean corpuscular hemoglobin; MCHC, mean corpuscular hemoglobin concentration; RDW, reticulocyte distribution width; PLT, platelet; BUN, blood urea nitrogen; ESA, erythrocyte-stimulating agent.

2. Performance of the Prediction Model

For the prediction model, we tried LR, MLPs, XGBoost, and GRU to find best performing models with varying hyperparameters and time-steps. Detailed performances of all experimented prediction models are presented in **Table 2.** The performances of overall models tended to increase as the number of time-steps increased. With 5-fold cross-validation, the best LR model had the MAE of 0.61, 0.62, and 0.62 g/dL on train, validation, and test sets, respectively. The best performing MLP model had two hidden layers of 512 and 2 nodes each and reached the MAE of 0.54, 0.61, and 0.61 g/dL, slightly better than the LR model on test data. For XGBoost models, we tried a grid search to optimize mainly four hyperparameters (the number of estimators, the minimum child weights, subsample ratio, and the maximum depth of trees). Our best XGBoost model had the MAE of 0.53, 0.60, and 0.60 g/dL, which outperformed both the LR and MLP models.

The best performing model of all was obtained by a GRU-based model consisting of four hidden layers (two stacked GRU layers followed by a single hidden layer feedforward network) and achieved 0.54, 0.59, and 0.59 g/dL on the train, validation, and test set, respectively. (**Table 3**) We also found that, when adding a Gaussian noise layer to the input layer of GRU models, the ME (or model bias) significantly decreased (all within 0.02 g/dL), while the accuracy remains almost the same.

Table 2. Performances of the LR, MLP, and XGBoost based prediction models

Models	Train set			Validation set			Test set		
	MAE	RMSE	ME	MAE	RMSE	ME	MAE	RMSE	ME
LR - Seq. 1	0.74	0.94	-	0.75	0.95	-	0.72	0.93	-
LR - Seq. 2	0.69	0.87	-	0.69	0.88	-	0.68	0.87	-
LR - Seq. 3	0.67	0.84	-	0.66	0.85	-	0.66	0.84	-
LR - Seq. 4	0.63	0.81	-	0.64	0.82	-	0.64	0.83	-
LR - Seq. 5	0.62	0.79	-	0.62	0.80	-	0.63	0.81	-
LR - Seq. 6	0.61	0.79	-	0.62	0.79	-	0.62	0.80	-
MLP (256, 8) - Seq. 1	0.64	0.84	0.011	0.73	0.93	-0.001	0.71	0.90	0.014
MLP (256, 32) - Seq. 2	0.54	0.74	0.012	0.66	0.85	-0.036	0.66	0.86	0.007
MLP (128, 8) - Seq. 3	0.57	0.76	0.009	0.65	0.84	-0.042	0.65	0.84	-0.030
MLP (1024, 8) - Seq. 4	0.44	0.62	-0.034	0.62	0.81	-0.055	0.63	0.82	0.044
MLP (128, 64) - Seq. 5	0.49	0.68	-0.014	0.61	0.78	0.001	0.63	0.82	-0.042
MLP (512, 2) - Seq. 6	0.54	0.73	0.006	0.61	0.78	0.010	0.61	0.79	-0.009
XGBoost (100, 5, 0.6, 4) - Seq. 1	0.63	0.80	0.001	0.73	0.92	0.011	0.69	0.88	-0.001
XGBoost (100, 10, 0.8, 4) - Seq. 2	0.59	0.76	-0.001	0.68	0.86	-0.022	0.66	0.84	0.001
XGBoost (100, 5, 0.6, 4) - Seq. 3	0.57	0.73	-0.002	0.65	0.84	-0.035	0.64	0.81	-0.012
XGBoost (100, 10, 0.8, 8) - Seq. 4	0.55	0.71	0.000	0.62	0.80	-0.035	0.62	0.79	-0.009
XGBoost (100, 1, 0.8, 4) - Seq. 5	0.54	0.69	-0.001	0.60	0.78	-0.017	0.61	0.79	0.002
XGBoost (100, 1, 0.8, 4) - Seq. 6	0.53	0.68	-0.001	0.60	0.77	-0.018	0.60	0.78	0.002

Note: Seq indicates how many time steps does a model concern. For example, Seq 6 refers to a model that takes as input of six

time steps, or six months follow up data. However, the method of processing such data varies depending on the choice of models. The numbers in parentheses indicate the number of hidden nodes or the cell size of neural network models. For example, MLP (256, 8) refers to a neural network that has 256 hidden nodes in the first hidden layer and 8 hidden nodes in the second hidden layer. In MLP models, we tried grid search to optimize the number nodes in two hidden layers by increasing the number of nodes with an exponent of 2 (2, 4, 8, 16, 32, 64, 128, 256, 512, and 1024). Four numbers in XGBoost models are the hyperparameter settings of the model (n_estimator, min_child_weight, subsample, and max_depth). **Abbreviation:** LR, linear regression; Seq, sequence; MLP, multilayer perceptron; XGBoost, extreme gradient boost; MAE, mean absolute error; RMSE, root mean square error; ME, mean error.

Table 3. Performances of the GRU and GRU with GNL-based prediction models

Models	Train set			Validation set			Test set		
	MAE	RMSE	ME	MAE	RMSE	ME	MAE	RMSE	ME
GRU(256, 64)+FC(4) - Seq. 1	-	-	-	-	-	-	-	-	-
GRU(256, 64)+FC(4) - Seq. 2	0.63	0.81	0.627	0.66	0.85	-0.030	0.65	0.84	-0.023
GRU(256, 64)+FC(4) - Seq. 3	0.59	0.78	0.593	0.64	0.82	-0.060	0.64	0.83	-0.066
GRU(256, 64)+FC(4) - Seq. 4	0.55	0.73	0.550	0.60	0.78	-0.011	0.62	0.81	-0.013
GRU(256, 64)+FC(4) - Seq. 5	0.56	0.74	0.564	0.59	0.76	-0.023	0.60	0.78	-0.025
GRU(256, 64)+FC(4) - Seq. 6	0.55	0.74	0.555	0.58	0.76	-0.053	0.59	0.78	-0.051
GRU (256, 64) + GNL + FC(4) - Seq. 1	-	-	-	-	-	-	-	-	-
GRU (256, 64) + GNL + FC(4) - Seq. 2	0.61	0.79	0.607	0.67	0.86	0.017	0.66	0.84	0.017
GRU (256, 64) + GNL + FC(4) - Seq. 3	0.56	0.75	0.564	0.63	0.82	0.012	0.63	0.82	0.019
GRU (256, 64) + GNL + FC(4) - Seq. 4	0.57	0.75	0.569	0.61	0.79	-0.027	0.61	0.80	0.015
GRU (256, 64) + GNL + FC(4) - Seq. 5	0.57	0.75	0.570	0.59	0.76	-0.003	0.60	0.79	-0.019
GRU (256, 64) + GNL + FC(4) - Seq. 6	0.54	0.73	0.548	0.59	0.76	0.022	0.59	0.78	0.012

Note: Seq indicates how many time steps does a model concern. GRU (256, 64) refers to a neural network that has 256 hidden nodes in the first hidden layer and 64 hidden nodes in the second hidden layer. **Abbreviation:** GRU, gated recurrent unit; FC, fully connected network; GNL, Gaussian noise layer; MAE, mean absolute error; RMSE, root mean square error; ME, mean error.

3. The Performance of the Recommendation Model

Adopting the power of GRUs, our best recommendation model was one GRU layer followed by two layers of feedforward network. The overall distribution of the real and algorithm-recommended ESA dose is illustrated in **Fig. 6**. The distribution of ESA dose between the real and the recommended dose showed similar in both continuous and binned ESA. The difference in continuous ESA was measured in MAE, and the binned categories of 0, 40, 120, 220, 310, 460, or more, were measured in accuracy. These two measures combined better describe how similarly the categorized real ESA and the algorithm recommended ESA match. As seen in the **Table 4**, the transcription accuracy was 59.3, 57.4, and 58.3% across the data set. The MAE of the recommendation model was 39.2, 43.1, and 43.2 μg on the train, validation, and test set, respectively (**Table 4**). In addition, MAE remained within 40–50 mcg in train, validation and test sets even when Hb was divided by Hb(t) levels.

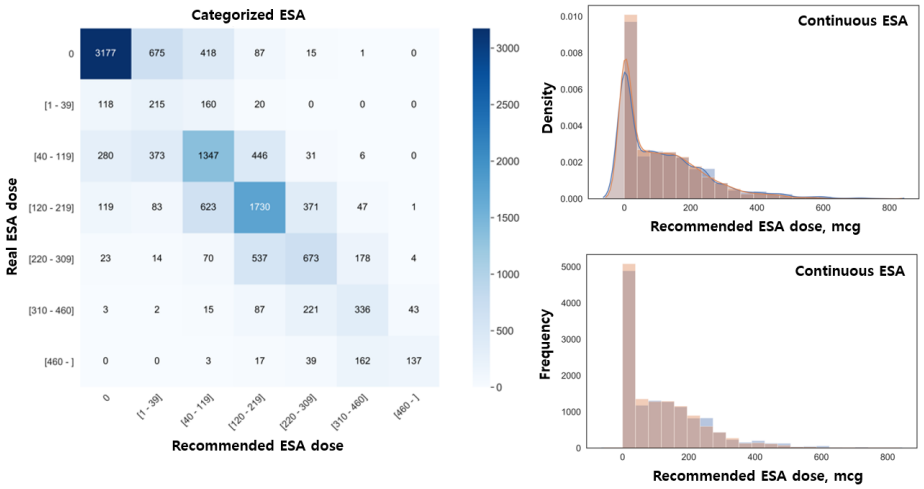

Figure 6. Comparison of distribution of ESA dose between the real and the Algorithm.

Note: Left, Confusion matrix difference between the real ESA and the recommended dose by the model in binned categories. Right, Distribution of density and frequency between the real and the recommended ESA dose in continuous values. Red bar indicated real ESA dose by doctor, Blue bar indicated ESA recommended by Algorithm. **Abbreviation:** ESA, erythrocyte-stimulating agent.

Table 4. Performance of the recommendation model

	Train set	Validation set	Test set
Transcription accuracy, %	59.3	57.4	58.3
Overall MAE, μg	39.2	43.1	43.2
Hb <10.0 g/dL	42.2	47.0	47.2
Hb 10.0–12.0 g/dL	39.2	41.8	41.2
Hb >12.0 g/dL	33.7	42.4	48.5

Note: Transcription accuracy was measured by the difference between the real ESA and the recommended dose by the model in binned categories and means that the categorized real and the AI recommended ESA dose match. MAE measured by GRU (128, 128) with GNL + FC (256, 64) - Seq. 6. MAE measures the numerical difference ESA between recommended ESA dose and real ESA dose. **Abbreviation:** MAE, mean absolute error; ESA, erythrocyte stimulating agent; GRU, gated recurrent unit; FC, fully connected network; GNL, Gaussian noise layer.

4. Result of Clinical Validation

The overall ESA dose was 155 (80−240) μg/month in real practice and 140 (70–210) μg/month in algorithm. It turned out that the amount of ESA used was significantly lower in algorithm recommendation (P<0.001) (**Table 5**). When compared the ESA dose according to the data set and Hb (t), the result was consistent. (**Fig. 7**, **Fig. 8**) In addition, algorithm shows a higher success rate in both Hb(t) <10.0 g/dL (64.0% vs. 76.9 %, P<0.001) or 10.0–12.0 g/dL (71.8 % vs. 94.7 %, P<0.001) even the amount of ESA dose used was much smaller. (**Fig. 9**) In contrast, the failure rate was significantly lower in algorithm compared to that in real practice. The algorithm recommended higher ESA dose in case of Hb(t) <10.0 g/dL. On the other hand, it was lower ESA dose in case of Hb(t) 10.0–12.0 g/dL compared to that in real practice. Finally, we compared the Hb difference between the real and algorithm. Median (IQR) Hb difference was 0.8 (0.4–1.4) g/dL in real data and 0.6 (0.3–1.0) g/dL in algorithm (P<0.001). In addition, Hb difference was significantly decreased across the data set and Hb(t) (P<0.001). (**Fig. 7**) Interestingly, Hb difference was significantly decreased compare to real Hb difference in case of Hb(t) of 10.0–12.0 g/dL. Conversely, Hb difference was similar or increased between real and algorithm when Hb(t) <10.0 or >12.0 g/dL. (**Fig. 10**)

Table 5. The clinical validation of the recommendation model

	Overall			Train set			Validation set			Test set		
	Real	Algorithm	P-value	Real	Algorithm	P-value	Real	Algorithm	P-value	Real	Algorithm	P-value
ESA dose, μg	155 (80–240)	140 (70–210)	<0.001	160 (80–240)	140 (70–210)	<0.001	150 (80–240)	140 (70–220)	<0.001	150 (80–240)	140 (70–220)	<0.001
Hb <10.0 g/dL (N = 2,723 case)												
Success Rate, %	64.0	76.9	<0.001	64.8	77.5	<0.001	58.3	67.0	<0.001	62.9	82.0	<0.001
ESA dose, μg	240 (160–320)	220 (170–280)	<0.001	240 (170–320)	220 (170–280)	<0.001	210 (150–300)	200 (155–280)	0.02	230 (180–340)	240 (180–340)	0.03
Failure Rate, %	36.0	23.1	<0.001	35.2	22.5	<0.001	41.7	33.0	<0.001	37.1	18.0	<0.001
ESA dose, μg	215 (160–300)	250 (180–340)	<0.001	210 (160–300)	250 (180–320)	0.02	240 (160–420)	365 (220–490)	0.08	210 (150–300)	210 (160–285)	0.66
Hb 10.0–12.0 g/dL (N = 8,609 case)												
Success rate, %	71.8	94.7	<0.001	71.6	94.7	<0.001	71.8	94.5	<0.001	73.0	95.7	<0.001
ESA dose, μg	130 (80–200)	120 (70–180)	<0.001	130 (80–200)	120 (70–180)	<0.001	120 (80–200)	120 (70–190)	0.01	120 (80–210)	130 (70–200)	0.04
Failure rate, %	19.8	4.6	<0.001	20.1	4.8	<0.001	19.3	4.9	<0.001	17.9	3.2	<0.001
ESA dose, μg	100 (60–160)	90 (50–150)	0.19	97 (60–160)	90 (50–155)	0.37	80 (60–160)	60 (30–150)	0.95	120 (62–160)	90 (40–130)	0.08
Hb difference, g/dL	0.8 (0.4–1.4)	0.6 (0.3–1.0)	<0.001	0.8 (0.4–1.4)	0.6 (0.3–1.0)	<0.001	0.7 (0.3–1.3)	0.6 (0.3–1.0)	<0.001	0.8 (0.4–1.3)	0.6 (0.3–1.0)	<0.001

Note: if Hb(t) <10.0 g/dL, the success rate defined as an increased Hb over 10.0 g/dL at the time of Hb(t+1). The Failure rate was defined as Hb(t+1) did not reach over 10.0 g/dL. Hb(t) 10.0–12.0 g/dL, the success rate was defined as remain within range to 10.0–12.0 g/dL in the next month. Failure rate was defined as Hb(t+1) level decreased below 10.0 g/dL. Hb difference defined as the difference between Hb(t) and Hb(t+1). **Abbreviation:** ESA, erythrocyte-stimulating agent; IQR, interquartile range; Hb, hemoglobin.

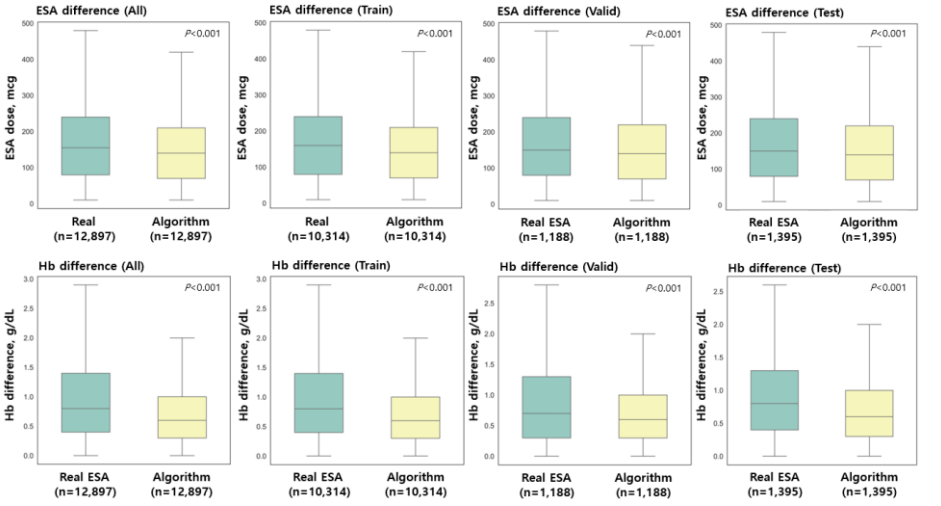

Figure 7. Comparison of ESA dose and Hb difference between the real and the Algorithm.
Abbreviation: ESA, erythrocyte-stimulating agent; Hb, hemoglobin.

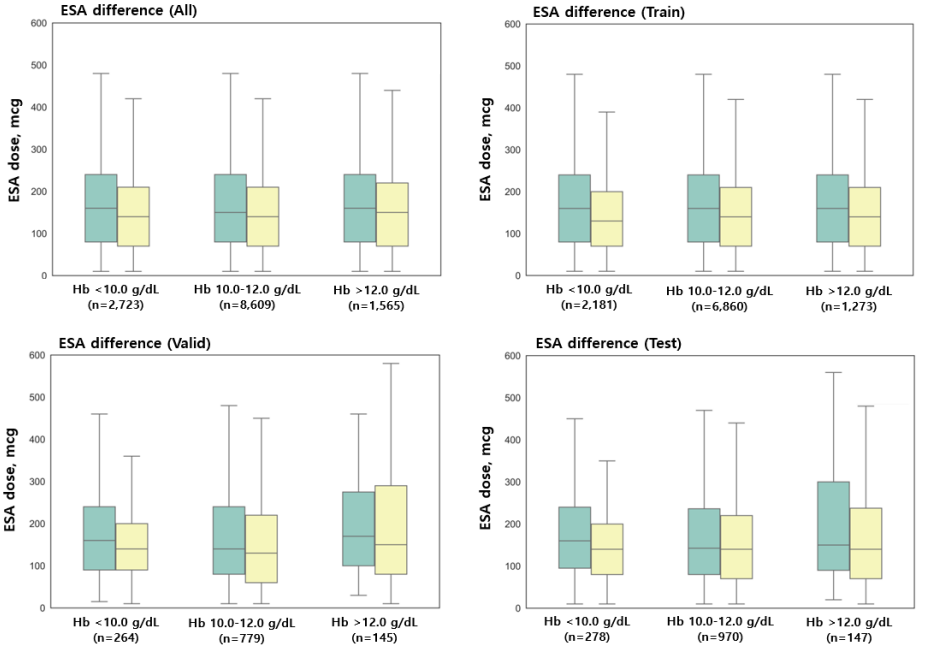

Figure 8. Comparison of ESA dose between the real and the algorithm according to Hb(t).
Abbreviation: ESA, erythrocyte-stimulating agent; Hb, hemoglobin.

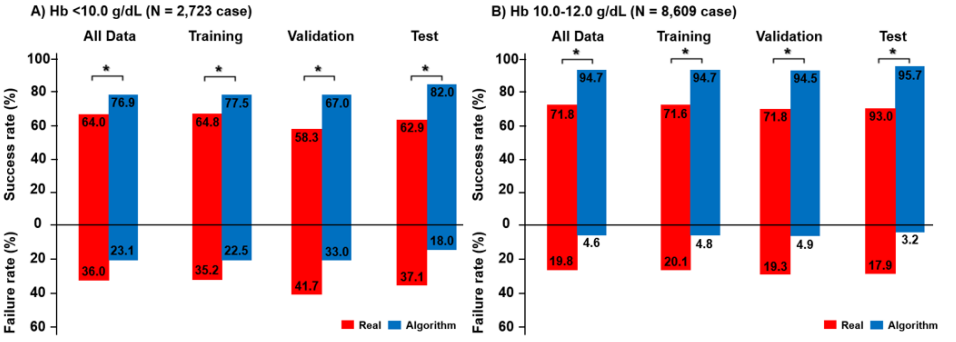

Figure 9. Comparison of success, failure rate between the real and the algorithm according to data set.
Note, **P*<0.001. **Abbreviation:** Hb, hemoglobin.

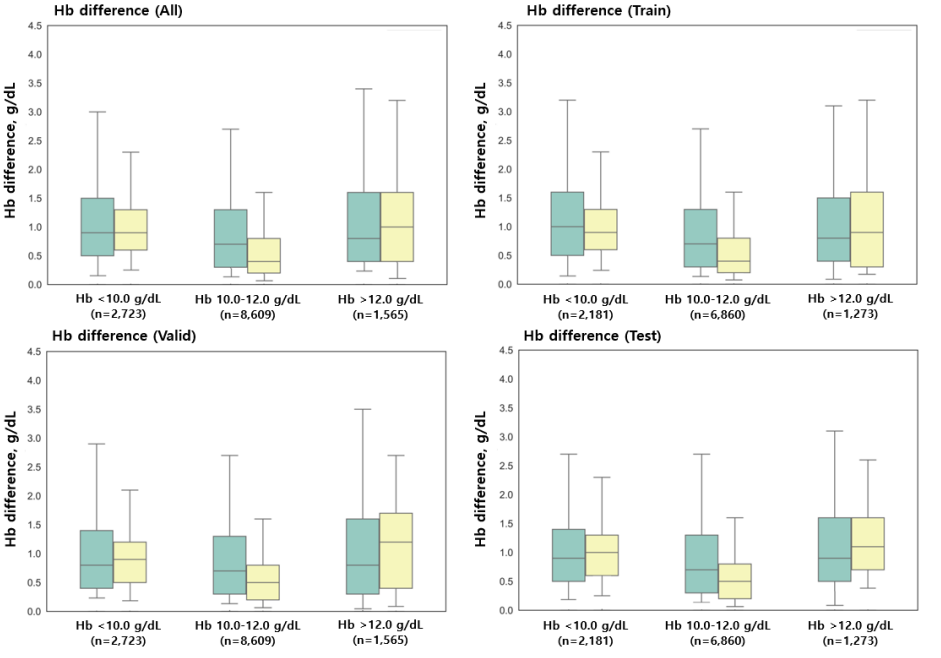

Figure 10. Comparison of Hb difference between the real and the algorithm according to Hb(t).
Abbreviation: ESA, erythrocyte-stimulating agent; Hb, hemoglobin.

IV. DISCUSSION

In this study, we developed the HPERA using advanced ML technique of GRU with GNL in ESKD patients. The overall performance of Hb prediction algorithm was 0.59 g/dL and ESA dose recommendation algorithm was 43.2 μg. In addition, we simulated different patients' response to ESA to show the validity of the HPERA.

During the last few decades, the concept of using ML method such as fuzzy logic, support vector machine, Bayesian networks, MLP, or reinforcement learning (RL) have been studied for the optimization of anemia management in ESKD patients.[16-19,24] Having the nature of longitudinally in predicting Hb in patients with ESKD, most of the previously proposed ML approaches only utilize variables at a one-time step or circumvent this issue by accumulating values of each variable over time without directly tackling limitations of models themselves.[17,18] ML models such as MLPs treat each input variable as time-independent information. For example, if there are patient records with several follow-up periods. MLPs learn their weights independent of time orders because weights are not shared over periods of time. Sometimes, however, the order of information matters if the latest period plays the most significant role in predicting the next information of interest. In contrast to MLPs, RNNs are designed to overcome such drawback. Since weights in RNNs are shared over all periods of time, the order of information is architecturally encoded in the model-recurrent inductive bias of RNNs. GRUs have a simple structure but successfully avoid the vanishing gradient problem, showing great success in handling sequential data in a wide range of tasks including natural language processing and signal processing. Therefore, the GRU used to develop the model in this study has the great advantage of being able to better reflect the relationships between the variables because it can be remembered the previous clinical variables in time sequence.

The artificial intelligence has been successfully applied at the point care to

support physicians in the decision-making process.[25] Brier *et al*. performed a single center, double-blind, randomized controlled trial using model predictive control (MPC). Although MPC did not showed the decreased in ESA dose, it showed decreased a mean absolute difference between achieved Hb compared with standard anemia management protocol.[26] In addition, Gaweda *et al.* showed the decreased in Hb variability, ESA dose using smart anemia manager algorithm based on MPC, and use of MCP might be an advantage for reducing the risk of transfusion in ESKD patients.[27,28] Most recently, Barbieri *et al*. demonstrated the promising result using anemia control model (ACM) based on MLP in a large, unselected ESKD patient.[22] During 12-month standard anemia care and followed by a 12-month ACM care, median darbepoetin level and Hb fluctuation were significantly decreased, and on-target Hb values were increased when the ACM suggestions were implemented. All these results confirmed that decision supporting system based on predictive modeling can improve of anemia management in ESKD patients. In our study, we showed the possibility in a reduced ESA dose, decreased Hb variability, and increased success rate which was reaching or maintaining target Hb compared with real clinical practice using the HPERA. Since, the analysis of clinical validation was resulted from retrospectively, however, prospective validation should be required.

This study has several strengths. In data collection, clinical variables used in the development of the HPERA was collected in seven tertiary-care general hospitals and they were significantly different across the hospitals. In addition, it indicated that HPERA is representative of the rest of the hospitals. In data preparation, the use of XGBoost in variable selection associated with Hb levels among 116 clinical parameters. In the development of algorithm, we used 15 clinical variables for 466 hemodialysis patients during a median observation period of 30.0 months and made the HPERA using a total of 271,561 variables. In general, the successful development of algorithm through ML technology depends on the quantity and quality of the training data. In terms of the number of patients used

for the development of algorithm, this is relatively small compared to ACM. However, these variables, which are used in algorithm development, have organic relationships that change over six months. In addition, GRU is suitable to handle these multivariate longitudinal patient data. In addition, the used a Gaussian noise layer to avoid overfitting to our training data and to have the effect of data augmentation. Another strength of this study is that we showed the performance of the ESA dose recommendation algorithm based on predictive models.

There are several limitations in this study, First, GRU based Hb prediction algorithm showed higher performance compared with LR, MLP, and XGBoost. However, the performance of the algorithm depends on model architecture including hyperparameters. In this study, though the best hyperparameters of MLPs and XGBoost with respect to validation data were all different across different time-steps and obtained by an extensive a grid search, the GRU model had much fewer hyperparameters to tune but outperformed other kinds models even with all the same architecture across different time-steps. Second, there was no performance comparison was made for the predictive model other ML methods such as RL. Third, we used a converted dose of darbepoetin-alpha that may cause potential bias in the algorithm development. Future studies that using RL and different ESA type should address this issue. Fourth, prospective validation should be performed in order for the validation of clinical results. Finally, our algorithm was based on a Korean ESKD population; thus, our algorithm may not be generalizable to other ethnic groups.

V. CONCLUSION

Our study provided the additional possibility to the optimization of anemia management in ESKD patients by HPERA. Considering, the complexity of anemia management and heterogeneous responses to ESA administration in these patients, this study revealed an additional direction of anemia management using the ML algorithm in ESKD patients and nephrologists.

REFERENCES

1. McFarlane SI, Chen SC, Whaley-Connell AT, Sowers JR, Vassalotti JA, Salifu MO, et al. Prevalence and associations of anemia of CKD: Kidney Early Evaluation Program (KEEP) and National Health and Nutrition Examination Survey (NHANES) 1999-2004. Am J Kidney Dis 2008;51:S46-55.
2. Stauffer ME, Fan T. Prevalence of anemia in chronic kidney disease in the United States. PLoS One 2014;9:e84943.
3. Thorp ML, Johnson ES, Yang X, Petrik AF, Platt R, Smith DH. Effect of anaemia on mortality, cardiovascular hospitalizations and end-stage renal disease among patients with chronic kidney disease. Nephrology (Carlton) 2009;14:240-6.
4. Fishbane S. Anemia and cardiovascular risk in the patient with kidney disease. Heart Fail Clin 2008;4:401-10.
5. Kovesdy CP, Trivedi BK, Kalantar-Zadeh K, Anderson JE. Association of anemia with outcomes in men with moderate and severe chronic kidney disease. Kidney Int 2006;69:560-4.
6. Locatelli F, Pisoni RL, Combe C, Bommer J, Andreucci VE, Piera L, et al. Anaemia in haemodialysis patients of five European countries: association with morbidity and mortality in the Dialysis Outcomes and Practice Patterns Study (DOPPS). Nephrol Dial Transplant 2004;19:121-32.
7. Mohanram A, Zhang Z, Shahinfar S, Keane WF, Brenner BM, Toto RD. Anemia and end-stage renal disease in patients with type 2 diabetes and nephropathy. Kidney Int 2004;66:1131-8.
8. Gouva C, Nikolopoulos P, Ioannidis JP, Siamopoulos KC. Treating anemia early in renal failure patients slows the decline of renal function: a randomized controlled trial. Kidney Int 2004;66:753-60.
9. Horwich TB, Fonarow GC, Hamilton MA, MacLellan WR, Borenstein J.

Anemia is associated with worse symptoms, greater impairment in functional capacity and a significant increase in mortality in patients with advanced heart failure. J Am Coll Cardiol 2002;39:1780-6.

10. Besarab A, Bolton WK, Browne JK, Egrie JC, Nissenson AR, Okamoto DM, et al. The effects of normal as compared with low hematocrit values in patients with cardiac disease who are receiving hemodialysis and epoetin. N Engl J Med 1998;339:584-90.
11. Singh AK, Szczech L, Tang KL, Barnhart H, Sapp S, Wolfson M, et al. Correction of anemia with epoetin alfa in chronic kidney disease. N Engl J Med 2006;355:2085-98.
12. Drueke TB, Locatelli F, Clyne N, Eckardt KU, Macdougall IC, Tsakiris D, et al. Normalization of hemoglobin level in patients with chronic kidney disease and anemia. N Engl J Med 2006;355:2071-84.
13. Pfeffer MA, Burdmann EA, Chen CY, Cooper ME, de Zeeuw D, Eckardt KU, et al. A trial of darbepoetin alfa in type 2 diabetes and chronic kidney disease. N Engl J Med 2009;361:2019-32.
14. Kliger AS, Foley RN, Goldfarb DS, Goldstein SL, Johansen K, Singh A, et al. KDOQI US commentary on the 2012 KDIGO Clinical Practice Guideline for Anemia in CKD. Am J Kidney Dis 2013;62:849-59.
15. Kdoqi. KDOQI Clinical Practice Guideline and Clinical Practice Recommendations for anemia in chronic kidney disease: 2007 update of hemoglobin target. Am J Kidney Dis 2007;50:471-530.
16. Gaweda AE, Jacobs AA, Brier ME, Zurada JM. Pharmacodynamic population analysis in chronic renal failure using artificial neural networks--a comparative study. Neural Netw 2003;16:841-5.
17. Martinez-Martinez JM, Escandell-Montero P, Barbieri C, Soria-Olivas E, Mari F, Martinez-Sober M, et al. Prediction of the hemoglobin level in hemodialysis patients using machine learning techniques. Comput Methods Programs Biomed 2014;117:208-17.

18. Barbieri C, Mari F, Stopper A, Gatti E, Escandell-Montero P, Martinez-Martinez JM, et al. A new machine learning approach for predicting the response to anemia treatment in a large cohort of End Stage Renal Disease patients undergoing dialysis. Comput Biol Med 2015;61:56-61.
19. Escandell-Montero P, Chermisi M, Martinez-Martinez JM, Gomez-Sanchis J, Barbieri C, Soria-Olivas E, et al. Optimization of anemia treatment in hemodialysis patients via reinforcement learning. Artif Intell Med 2014;62:47-60.
20. Vega A, Abad S, Verdalles U, Aragoncillo I, Velazquez K, Quiroga B, et al. Dose equivalence between continuous erythropoietin receptor activator (CERA), Darbepoetin and Epoetin in patients with advanced chronic kidney disease. Hippokratia 2014;18:315-8.
21. Ogunleye AA, Qing-Guo W. XGBoost Model for Chronic Kidney Disease Diagnosis. IEEE/ACM Trans Comput Biol Bioinform 2019.
22. Barbieri C, Molina M, Ponce P, Tothova M, Cattinelli I, Ion Titapiccolo J, et al. An international observational study suggests that artificial intelligence for clinical decision support optimizes anemia management in hemodialysis patients. Kidney Int 2016;90:422-9.
23. Zhang Z. Multiple imputation with multivariate imputation by chained equation (MICE) package. Ann Transl Med 2016;4:30.
24. Gaweda AE, Jacobs AA, Brier ME. Application of fuzzy logic to predicting erythropoietic response in hemodialysis patients. Int J Artif Organs 2008;31:1035-42.
25. Barbieri C, Cattinelli I, Neri L, Mari F, Ramos R, Brancaccio D, et al. Development of an Artificial Intelligence Model to Guide the Management of Blood Pressure, Fluid Volume, and Dialysis Dose in End-Stage Kidney Disease Patients: Proof of Concept and First Clinical Assessment. Kidney Dis (Basel) 2019;5:28-33.

26. Brier ME, Gaweda AE, Dailey A, Aronoff GR, Jacobs AA. Randomized trial of model predictive control for improved anemia management. Clin J Am Soc Nephrol 2010;5:814-20.

27. Gaweda AE, Aronoff GR, Jacobs AA, Rai SN, Brier ME. Individualized anemia management reduces hemoglobin variability in hemodialysis patients. J Am Soc Nephrol 2014;25:159-66.

28. Gaweda AE, Jacobs AA, Aronoff GR, Brier ME. Individualized anemia management in a dialysis facility - long-term utility as a single-center quality improvement experience. Clin Nephrol 2018;90:276-85.

ABSTRACT (IN KOREAN)

말기 신부전 환자에서 순환 신경망을 이용한 혈색소 수치 예측과 조혈제 용량 권고 알고리즘 개발

< 지도교수 유 태 현 >

연세대학교 대학원 의학과

윤 해 룡

배경: 혈액 투석 중인 말기 신부전 환자의 빈혈 조절은 환자의 삶의 질과 예후 향상을 위해 중요하다. 하지만, 말기 신부전 환자들은 여러 기저 질환과 조혈제에 대한 비선형적인 반응으로 인해 적정 혈중 적혈구 농도를 유지하기 어렵다.

목적: 최근 의료 분야에서 다양한 질환 군을 대상으로 인공 지능을 이용한 예측 모형의 개발이 활발하게 진행되고 있다. 따라서, 본 연구에서는 혈액 투석 중인 말기 신부전 환자에서 혈색소 수치를 예측하는 모형을 만들고 이를 기반으로 하여 적정 혈색소 수치를 유지 하기 위한 조혈제 용량 권고 알고리즘을 만들고자 하였다.

방법: 수도권의 7개 대학병원에서 외래 혈액 투석 중인 환자 466명을 대상으로 최대 5년 간의 임상적인 변수를 수집하였으며, 이 임상 자료를 기반으로 순환 신경망을 이용하여 혈색소 수치 예측과 이에 따른 조혈제 권고 알고리즘을 만들었다. 예측 모형의 경우 Linear regression, Multilayer perceptron, 그리고 Extreme gradient boosting을 이용한 예측 모형과 성능을 비교 분석 하였다. 권고 알고리즘의 경우 적정 혈색소 수치를 유지 하기 위한 조혈제 용량을

실체 투여한 조혈제 용량과 비교하여 성능을 분석하였다.

결과: 예측 모형의 경우 순환 신경망을 기반으로 한 혈색소 수치 예측 모형의 성능은 0.49 g/dL로 다른 예측 모형에 비해 뛰어남을 확인 할 수 있었다. 또한, 순환 신경망을 기반으로 한 조혈제 용량 권고 알고리즘의 경우 실제 조혈제를 투여 했을 경우에 비해 43.2 μg 정도의 편차를 보이는 것을 확인 할 수 있었다. 끝으로 혈색소 예측, 조혈제 권고 알고리즘을 이용할 경우 환자에게 투여하는 조혈제의 용량이 감소하고, 목표 혈색소 수치에 도달하거나 유지할 확률이 올라가는 것을 확인 할 수 있었으며, 혈색소의 변화량 또한 감소 할 수 있음을 확인 할 수 있었다.

결론: 순환 신경망을 이용한 예측 모형이 다른 모형에 비해 혈색소 수치 예측에 뛰어난 성능을 보인다는 것을 확인 할 수 있었다. 또한 조혈제 권고 모형을 이용하였을 때 적은 용량으로 목표 혈색소 수치 도달 및 유지를 할 수 있으며 혈색소 수치 변화량 또한 줄일 수 있음을 확인 하였다. 이상의 결과를 종합하여 볼 때, 혈액 투석 중인 말기 신부전 환자에서 적혈구 예측, 조혈제 권고 알고리즘의 사용은 적정 혈색소 수치를 유지하고, 조혈제 사용 용량을 줄여 약제 관련 부작용 및 제반 비용을 줄 일 수 있을 것으로 생각되며, 빈혈에 대한 효과적인 관리를 통해 투석 환자들의 예후를 향상 시킬 수 있을 것으로 생각된다.

핵심 되는 말: Sequential awareness neural network, 말기 신부전, 적혈구, 조혈제, 예측 모형, 권고 모형