Heliyon 9 (2023) e12613

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Multi-head self-attention mechanism enabled individualized hemoglobin prediction and treatment recommendation systems in anemia management for hemodialysis patients

Ju-Yeh Yang [a,b,c], Tsung-Chun Lee [d,e,f], Wo-Ting Liao [g], Chih-Chung Hsu [g,*]

[a] Institute of Health Policy and Management, College of Public Health, National Taiwan University, Taiwan
[b] Center for General Education, Lee-Ming Institute of Technology, New Taipei City, Taiwan
[c] Division of Nephrology, Department of Internal Medicine, Far Eastern Memorial Hospital, New Taipei City, Taiwan
[d] Division of Gastroenterology and Hepatology, Department of Internal Medicine, Taipei Medical University Shuang Ho Hospital, New Taipei City, Taiwan
[e] Department of Internal Medicine, School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan
[f] Good Will Clinic, Taipei, Taiwan
[g] Institute of Data Science, National Cheng Kung University, Tainan, Taiwan

ARTICLE INFO

Keywords:
Self-attention mechanism
Prediction
Recommendation
Anemia
Hemodialysis
Informer

ABSTRACT

Anemia is a critical complication in hemodialysis patients, but the response to erythropoietin-stimulating agents (ESA) treatment varies from patient to patient and is not linear across different time points. The aim of this study was to develop deep learning algorithms for individualized anemia management. We retrospectively collected 36,677 data points from 623 hemodialysis patients, including clinical data, laboratory values, hemoglobin levels, and previous ESA doses. To reduce the computational complexity associated with recurrent neural networks (RNN) in processing time-series data, we developed neural networks based on multi-head self-attention mechanisms in an efficient and effective hemoglobin prediction model. Our proposed model achieved a more accurate hemoglobin prediction than the state-of-the-art RNN model, as shown by the smaller mean absolute error (MAE) of hemoglobin (0.451 vs. 0.593 g/dL, $p = 0.014$). In ESA (including darbepoetin and epoetin) dose recommendation, the simulation results by our model revealed a higher rate of achieved hemoglobin targets (physician prescription vs. model: 86.3 % vs. 92.7 %, $p < 0.001$), a lower rate of hemoglobin levels below 10 g/dL (13.7 % vs. 7.3 %, $p < 0.001$) and smaller change in hemoglobin levels (0.6 g/dL vs. 0.4 g/dL, $p < 0.001$) in all patients. Our model holds great potential for individualized anemia management as a computerized clinical decision support system for hemodialysis patients. Further external validation with other datasets and prospective clinical utility studies are warranted.

1. Introduction

Anemia is the most critical complication of end-stage renal disease (ESRD). Even with continuous hemodialysis, ESRD patients still suffer from anemia, malaise, loss of appetite, poor quality of life, adverse cardiovascular events, and mortality. Current clinical

* Corresponding author.
E-mail address: cchsu@gs.ncku.edu.tw (C.-C. Hsu).

https://doi.org/10.1016/j.heliyon.2022.e12613
Received 21 July 2022; Received in revised form 5 December 2022; Accepted 16 December 2022
Available online 1 February 2023

guidelines for the management of anemia in ESRD patients recommend maintaining a hemoglobin (Hb) level between 10 and 11 g/dL [1, 2].

Treatment with erythropoietin-stimulating agents (ESA) to ameliorate anemia in ESRD patients has been shown to improve the quality of life [3], decrease hospitalizations [4], and reduce mortality [5]. However, optimal ESA dosing is challenging for clinicians because patient responses to ESA are "complex, nonlinear, and dynamic". At the same time, there is a lack of individualized treatment that takes into account each patient's current Hb levels and prior treatment responses to ESA. The unified standard dosage of ESA often results in either undertreatment (low Hb $<$ 10 g/dL) or overtreatment (high Hb $>$ 12 g/dL). Underdosing of ESA leads to suboptimal hemoglobin levels, resulting in symptoms of anemia and poor quality of life. Overdose of ESA is associated with hypertension, thromboembolism, retinal hyperproliferation, carcinogenesis, and higher risk of death [6, 7]. Therefore, personalized dose recommendations for ESA treatment remain a significant unmet clinical need.

Recently, a recurrent neural network (RNN)-based Hb prediction method that uses patients' historical treatment data has been proposed [8], referred to as HP-RNN. The data can be divided into three types: (1) historical data, (2) static data, and (3) future ESA and iron dosage data [8]. Here, the time-series of the historical ESA, and iron data are fed into the RNN to learn the hidden transition state for recommending the future ESA at a target time.

Yun et al. used gated recurrent unit (GRU) networks in an ESA recommendation system [9]. First, the historical data are used to indicate the Hb level at the future time point, and then the predicted Hb value is changed as the target Hb value. The rest of the historical data and the target Hb value are then treated as the input data of the GRU-based recommendation system. Finally, the ESA dose can be recommended for future timestamps. However, it is difficult for the GRU to perform parallel calculations and the inference speed may be affected. Moreover, the performance of RNN-related networks, such as GRU and LSTM, often requires heterogeneous and long-range time-series data, thus limiting their clinical applications.

In this study, we propose a novel non-recurrent approach to perform both Hb prediction and ESA dose recommendation effectively and efficiently. First, effective data processing and cleaning were carefully designed to remove outliers and select the essential features. Next, inspired by the newly developed transformer algorithm [10], we adopted the Informer algorithm with modifications to the multi-head self-attention mechanism [11] to form our Hb prediction system that better captures the intrinsic context feature of the historical data, i.e., individual patient's current Hb level, laboratory data, and previous responses to ESA.

In the following sections, we will describe the patients, overview of the proposed model, data collection and processing of missing values, the hemoglobin prediction module, the recommendation system for the ESA dosing, and the statistical analysis in the Method Section. Then we will describe Experimental results, Discussion, and the Conclusion.

2. Methods

2.1. Patients

From January 2016 to December 2020, we retrospectively collected clinical and laboratory data from the electronic health records (EHR) of 623 ESRD patients receiving maintenance hemodialysis at Far Eastern Memorial Hospital (FEMH), which is a tertiary referral medical center in the Taipei Metropolitan area. The study protocol was reviewed and approved by the FEMH Research Ethics Review Committee (FEMH IRB No.109178-E).

2.2. Overview of the proposed model

Figure 1 shows the framework of the proposed ESA dose recommendation system based on the Hb prediction module. Figure 2 shows the Hb prediction module. First, the clinical and laboratory characteristics and the target indices, such as ESA and Hb, were collected and arranged in a temporal axis to accurately predict the Hb value at a future time point. Specifically, we predicted the Hb level at timestamp t+4 (t+4:4 months later than timestamp t) using the proposed self-attention mechanism (SAM) predictor based on the historical data at timestamp t to t+3. Based on the predicted Hb y_{hb}^{t+4}, we aimed for Hb to be $\hat{y}_{hb}^{t+4}$ based on the following update Eq.

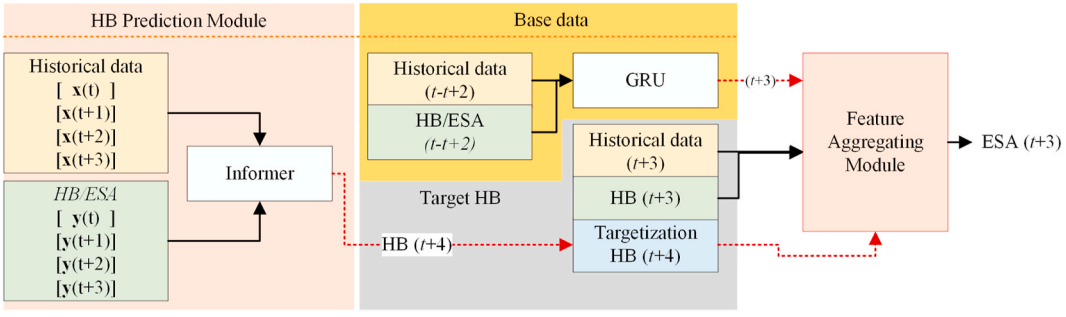

Figure 1. Framework of the proposed ESA recommendation system based on the hemoglobin prediction module. (ESA: erythropoietin stimulating agent. GRU: gated-recurrent unit networks. HB: hemoglobin. SAM: self-attention mechanism.)

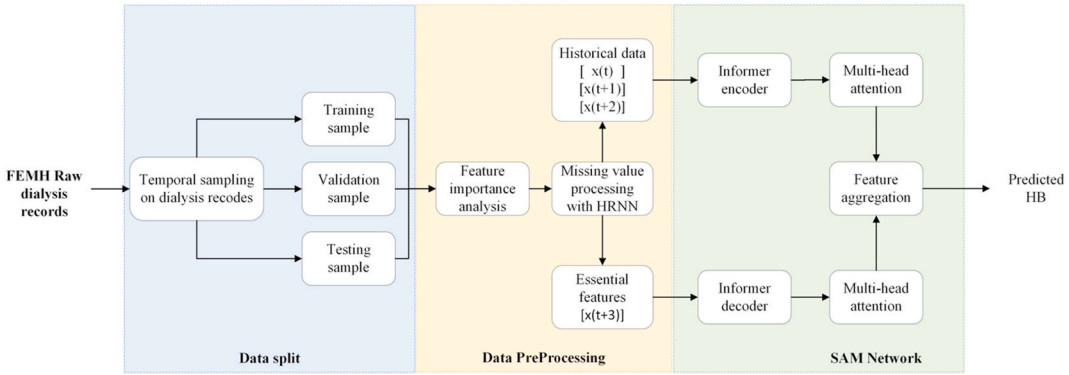

Figure 2. Framework of the proposed hemoglobin prediction module. (HB: hemoglobin. SAM: self-attention mechanism).

(1):

$$\widehat{y}_{hb}^{t+4} = \begin{cases} y_{hb}^{t+4} + 1, if\ y_{hb}^{t+4} < 10 \\ 11, otherwise \\ y_{hb}^{t+4} - 1 if\ y_{hb}^{t+4} > 12 \end{cases} \quad (1)$$

where the Hb value ranged [10, 12] is defined as expected and other Hb vales are defined as abnormal. Because our recommendation system aims to recommend the ESA dose at timestamp t+4 to satisfy the target Hb value at timestamp t+4, it is necessary to predict the future Hb value based on an accurate model rather than directly inputting the targeted Hb value. The Hb prediction model is essential because Hb value may be different every week, and the target Hb value at timestamp t+4 is different from that at timestamp t+3. Then, the historical data from timestamp t to t+2 are retrieved and fed into a two-layer GRU to obtain the predicted data at timestamp t+3. The actual historical data and the predicted Hb at timestamp t+3 were used as input data for the final ESA recommendation. A feature aggregation module (FAM) is proposed to fuse the above features. Finally, the recommended ESA dose at timestamp t+3 can be generated from the FAM that meets the target Hb value at timestamp t+4.

The number of layers of both RNN and GRU in our experiments is eight, and the number of neurons is eight. For the computational settings, we set the batch size to 32 and the starting learning rate to 0.0001 with a linear learning rate decay policy with a gamma value of 0.95. Because our treatment data are relatively complex and have few observable factors (i.e., 29 features), the inner neurons in the prob-sparse attention layer and the number of multi-heads were empirically determined to be 32 and 5, respectively. The Adam optimizer is used in the optimization of parameters, and the total epoch is 50. All the performance indices are measured in a personal computer equipped with Intel i7-9900k CPU with 64GB system memory and NVIDIA GTX-2080Ti GPU and evaluated on both GPU and CPU for comparing the effectiveness of the proposed methods.

2.3. Data collection and processing of missing values

We collected the following data from the EHR: *Hb, Albumin, ALT (SGPT), Alkaline phosphatase, BUN, Creatinine, Ca, Ferritin, IRON/TIBC, Na, Phosphorus, Age, MCV, MCH, MCHC, IRON, urea reduction rate (URR%), Delta Hb, Gender, Body weight, Dry weight, Fistula, Graft, Catheter, HBV, HCV, diabetes mellitus*, and *Intact PTH*. The historical ESA doses were calculated as *Mean drug*, indicating the average ESA dose in the last 30 days.

For missing values of data points, we investigated three processing methods: (1) Discard if the data are incomplete at any one of the P time points, termed the purification method (PUR). (2) Create missing data points by linear interpolation (LI). (3) Apply hierarchical RNN-metadata processing to the missing data points, termed the hierarchical method (HI) [12].

The hierarchical method [12] considers all historical data along the temporal axis and skips the missing values until a predefined number of historical data points is reached. A time difference feature is also proposed to calculate the length of the time interval of each sample. In this way, temporal information can be appropriately embedded in the processing to improve the performance. Compared with the conventional imputation method (LI), the hierarchical method extracts the inter-relationship of all "real" historical data points, which improves the accuracy of prediction. Specifically, each row contains historical data of {(time, features, time difference)} = {(t_k, f_k, d_k): k = 1,. . ., N}, where t, f, and d denote the timestamp, patient's features, and length of the time interval of two samples, respectively. d_k was used to obtain t_k-t_{k-1} and $d_k \leq 60$. In the historical data, the i-th patient's input $\mathbf{X}_i^{Nk} = [\boldsymbol{x}_i^N, \boldsymbol{x}_i^{N+1}, \ldots, \boldsymbol{x}_i^{N+(k-1)}]$, where $\boldsymbol{x}_i^{N+(k-1)} = (t_{k-1}, f_{k-1}, d_{k-1})$.

2.4. Hemoglobin prediction module

We propose a novel and efficient Hb prediction system based on the SAM module using the Informer algorithm [11]. A better Hb

prediction system directly leads to a better and more stable recommendation system because the predicted Hb value can be approximated to the real scenarios. The conventional Hb prediction system [8, 13] only focuses on the development of deep neural networks to improve the performance. However, it is well known that recurrent networks are difficult to parallelize on a graphics processor unit (GPU) in practice, which makes fast training and inference even more difficult. To solve this problem, we introduce the non-recurrent approach, stacked SAM, as proposed in the Informer algorithm [11], to improve performance and complexity. First, the conventional self-attention layer involves enormous computational complexity because all neurons in the attention layer are used to calculate the attention score over other neurons. To solve this problem, we refer to the efficient attention layer, prob-sparse attention, as proposed in [11], to reduce the complexity and improve the performance. The key idea in prob-sparse attention is that the attention values are usually a long-tail distribution, which means that only a few neurons are essential and needed. In this way, the computational complexity can be further reduced.

Next, we stack N_k Prob-sparse attention with a multi-head setting to form our feature encoder. In the encoder, the input of the *i*-th patient $\mathbf{X}_i^{N8} = [\boldsymbol{x}_i^N, \boldsymbol{x}_i^{N+1}, \ldots, \boldsymbol{x}_i^{N+7}] \in \mathbb{R}^{b\times t\times l}$ contains all the essential features at five timestamps, where *b, t,* and *l* denote the batch size, the number of timestamps, and the length of the feature vector, respectively. Therefore, the context feature representation $\mathbf{v}_i^N$ of $\mathbf{X}_i^N$ can be obtained as $\mathbf{v}_i^N = f_{enc}(\mathbf{X}_i^N)$, where f_{enc} represents the encoder. Meanwhile, the decoder is formed by a single prob-sparse attention module because the input of the decoder is relatively simple so that the number of stacking layers can be reduced. Since the time-series prediction in the conventional transformer is progressively predicted, we refer to the encoder-decoder architecture to solve this problem based on the additional start token [11]. Finally, a full self-attention module is used to fuse the context feature representation $\mathbf{v}_i^N$ and the start-token to learn the final Hb value in the next month (i.e., the hemoglobin value at N+30).

2.5. Recommendation system for ESA dosing

Suppose the predicted Hb level $\widehat{y}_{hb}^{t+4}$ at timestamp t+4 is based on the collected data $\mathbf{X}_i^{N4} = [\boldsymbol{x}_i^N, \boldsymbol{x}_i^{N+1}, \ldots, \boldsymbol{x}_i^{N+3}]$, the recommended ESA dose can be predicted by our model based on the aggregation of the targeted $\widehat{y}_{hb}^{t+4}$ using Eq. (1), the predicted Hb/ESA at timestamp t+3 using GRU based on the historical data, including Hb/ESA and others during the timestamp t to t+2, and the actual Hb and historical values at timestamp t+3. Specifically, denoting the historical data, Hb level, and ESA with the duration t to t+2 by $\mathbf{X}_i^{N3}, \boldsymbol{y}_{hb}^{t3} = [y_{hb}^t, y_{hb}^{t+1}, y_{hb}^{t+2}]$, and $\boldsymbol{y}_{ESA}^{t3} = [y_{ESA}^t, y_{ESA}^{t+1}, y_{ESA}^{t+2}]$, GRU is used to obtain the predicted result $\widehat{\mathbf{X}}_i^{t+3}, \widehat{y}_{hb}^{t+3}$, and $\widehat{y}_{ESA}^{t+3}$ based on the following Eqs. (2), (3), (4), and (5):

Update gate:

$$z_t = \sigma(W_z[\dot{h}_{t-1}, x_t]), \tag{2}$$

Reset gate:

$$r_t = \sigma(W_r[\dot{h}_{t-1}, x_t]), \tag{3}$$

New hidden state content:

$$\widehat{h}_t = \tanh(W[r_t \times \dot{h}_{t-1}, x_t]), \tag{4}$$

Hidden state:

$$h_t = (1 - z_t) \times h_{t-1} + z_t \times \widehat{h}_t. \tag{5}$$

where h_t in Eqs. (2), (3), (4), and (5) denotes the hidden state in which the update gate simultaneously controls what remains from the previous hidden state, and what should be updated to a new hidden state. In this way, we can extend the prediction at a given time by repeating the above GRUs.

Then, the real historical data and the Hb levels at timestamp t+3 $\mathbf{X}_i^{t+3}$ ***and*** y_{hb}^{t+3} are also used to measure the difference between the ground truth and the predicted historical and Hb levels at timestamp t+3, which can be considered as the stability of the prediction model. For example, the ideal Hb level at t+4 is between 10 and 12, but it is unreasonable to obtain an accurate ESA dose recommendation that can achieve the ideal Hb level in the future, when the current Hb level is significantly below 10 or above 12. Instead, the target Hb should be similar to the previous trends. Therefore, SAM was used to accurately predict the Hb level at t+4 so that the target Hb level would not be significantly different from the last Hb level at t+3. Finally, we have the Hb and historical data at t to t+3 and their counterparts based on the GRU predictions and the targeted Hb level at t+4 using Eq. (1). The proposed FAM is based on a fully connected layer, as follows:

$$y_{ESA}^{t+3} = \boldsymbol{\sigma}\left(\boldsymbol{GLN}\left(\boldsymbol{W}_f \boldsymbol{x}_{ensemble} + \boldsymbol{b}\right)\right) \tag{6}$$

where $\boldsymbol{x}_{ensemble} = \mathbf{concat}([\widehat{\mathbf{X}}_i^{t+3}, \widehat{y}_{hb}^{t+3}, \widehat{y}_{ESA}^{t+3}, \mathbf{X}_i^{t+3}, y_{hb}^{t+3}, \widehat{y}_{hb}^{t+4}])$ and GLN is the global normalization layer. Finally, the recommended ESA dose at timestamp t+3 y_{ESA}^{t+3} can be predicted based on Eq. (6).

2.6. Treatment for anemia and outcome parameters

Per international KDIGO guideline [1] and domestic Taiwan Society of Nephrology guideline [2], intravenous iron supplement was routinely prescribed (with the exception of iron hypersensitivity) for HD patients at FEMH HD center if all of the following three criteria were met: (1) Hb < 12 g/dL, (2) ferritin <500 ng/mL, and (3) iron saturation (IRON/TIBC) < 50%. If either one of the above three criteria was resolved by iron supplement, then the iron supplement would be discontinued for three months and the above lab data would be rechecked. Because the prescription of iron supplements was clearly defined and strictly regulated, we did not incorporate iron supplement into our treatment recommendation algorithms.

Meanwhile, erythropoietin-stimulating agents (ESA) including darbepoetin alpha (Nesp 20 μg/vial) and epoetin beta (Recormon 2,000 IU/vial) were regulated with a monthly cap per domestic National Health Insurance (NHI). (1 μg of darbepoetin alpha equals to 200 IU of epoetin alpha). Collectively, equivalent ESA doses during 30-day period was shown in IU/30 days. The ESA prescription in Taiwan was limited to 80,000 IU/30days by NHI reimbursement regulations [2].

Percentage of patients with Hb < 10 g/dL was denoted as "failure rate" among outcome parameters, while percentage of patients with Hb > 10 g/dL was denoted "non-failure rate". Hb between 10 and 12 g/dL was viewed as "on-target". The difference of Hb levels between t+3 and t+4 was denoted as "Hb difference". Smaller Hb difference suggested less Hb fluctuations.

Table 1
Clinical characteristics of the study participants.

Variables	Frequency	Data
Study participants	-	623
Follow-up duration (month), median (IQR)	-	37.3 (23.3–57.7)
Age, years	-	61.3 (12.1)
Women, n (%)	-	252 (40.4)
Weight, kg	-	64.4 (13.8)
Hemodialysis information		
Hours per session, hour	-	3.8 (0.4)
Blood flow rate, mL/min	-	249.7 (43.7)
Dry body weight, kg	Each HD	62.1 (13.5)
Pre-dialysis weight, kg	Each HD	64.4 (13.8)
Post-dialysis weight, kg	Each HD	62.1 (13.4)
Vascular access		
Arteriovenous -Fistula	Each HD	458 (73.5)
Arteriovenous -Graft	Each HD	51 (8.12)
Catheter	Each HD	215 (34.5)
Blood pressure		
Beginning-SBP, mmHg	Each HD	148.5 (24.8)
Beginning-DBP, mmHg	Each HD	78.1 (14.7)
Beginning-HR, beat/min	Each HD	80.4 (13.3)
End-SBP, mmHg	Each HD	146.1 (27.5)
End-DBP, mmHg	Each HD	77.9 (14.3)
End-HR, beat/min	Each HD	82.0 (14.1)
Laboratory results		
WBC count, 10^3/μL	Twice per month	6.7 (2.3)
Hemoglobin, g/dL	Twice per month	9.7 (2.1)
Hematocrit, %	Twice per month	29.5 (6.1)
MCV, fL	Twice per month	87.6 (7.4)
MCH, pg	Twice per month	28.9 (2.7)
MCHC, g/dL	Twice per month	33.0 (1.2)
PLT count, 10^3/μL	Twice per month	200.2 (70.7)
Calcium, mg/dL	Monthly	8.6 (0.9)
Phosphate, mg/dL	Monthly	102.5 (84.9)
Glucose, mg/dL	Monthly	148.0 (66.5)
BUN, mg/dL	Monthly	79.5 (29.2)
Creatinine, mg/dL	Monthly	10.3 (3.1)
Albumin, g/dL	Monthly	3.7 (0.5)
Sodium, mEq/L	Monthly	137.6 (4.2)
Potassium, mEq/L	Monthly	4.3 (0.7)
Serum iron, μg/dL	Quarterly	55.3 (25.2)
Ferritin, ng/mL	Quarterly	305.8 (127.9–487.8)
Parathyroid hormone, pg/mL	Quarterly	232.3 (111.6–427.3)
Urea Reduction Rate, %	Quarterly	70.7 (16.4)
ALT (SGPT)	Quarterly	15.9 (13.7)
IRON/TIBC, %	Quarterly	24.6 (11.7)
HBV, n (%)	Semi-annually	133 (21.3)
HCV, n (%)	Semi-annually	31 (5.0)
ESA dose		
Equivalent ESA dose, IU/30 days, median (IQR)	Each HD	38,000 (29,000–46,000)

Note: 1 μg of darbepoetin alpha equals to 200 IU of epoetin alpha.

2.7. Statistical analysis

Continuous data are presented as means and standard variations or medians with interquartile ranges (IQRs). Categorical data are presented as percentages. Statistical comparisons were performed using the Kruskal-Wallis test for continuous data and the chi-square test for categorical variables. Statistical calculations were performed using Python package SciPy. Kruskal-Wallis test was a nonparametric scheme for verifying whether samples are originated from the same distribution, in which the null hypothesis is that the mean ranks of the groups are the same. it is more suitable for avoiding the testing being affected by the presence of outliers or by the nonnormal distribution of data.

3. Experimental results

3.1. Baseline characteristics of study subjects

Table 1 shows the clinical characteristics of the study participants and the frequency of variables. Of the 623 patients, 252 patients (40.4 %) were women. The age was 61.3 +/- 12.1 years. The vintage of hemodialysis was 3.1 +/- 1.5 years. We then randomly divided the preprocessed data for each patient into three datasets: training (67.7 %), validation (17.4 %), and test (14.9 %) datasets. Table 2 shows the processing methods for missing values and the number of data points (an average of 55.7 datapoints per patient) of the training, validation, and test datasets. Table 3 (3A, 3 B, 3C) shows the sensitivity analysis performed to determine the settings for the proposed hemoglobin prediction model. These included (3A) previous 3 months of laboratory data, (3 B) the hierarchical method for processing of the missing values, and (3C) multiple heads of five (see Table 3).

3.2. Performance evaluation for hemoglobin prediction

The performance of the Hb prediction algorithms was evaluated using the mean square error (MSE), mean absolute error (MAE), and mean error (ME), commonly used indices in related studies [8, 9].

Table 4 shows that our proposed model (SAM) significantly outperformed the state-of-the-art model (i.e., RNN) [8]. In addition, the performance of the proposed SAM algorithms was similar across the three different processing methods, suggesting stable performance of the proposed SAM model.

3.3. Performance evaluation for ESA dose recommendation

Table 5 shows the performance evaluation for the ESA dose recommendation system with a comparison between simulation results by our proposed model and actual physician prescriptions. Overall, the proposed ESA recommendation system had a higher non-failure rate, lower failure rate, and smaller change in Hb levels.

In the subgroup analysis, in patients with low baseline Hb (Hbt+3 less than 10 g/dL), the proposed ESA recommendation system achieved a higher non-failure rate and a lower failure rate. In patients with on-target baseline Hb levels (Hbt+3:10–12 g/dL), the proposed ESA recommendation system achieved a higher non-failure rate and a lower failure rate.

4. Discussion

To the best of our knowledge, this is the first non-recurrent approach using multi-head SAM for individualized hemoglobin prediction and treatment recommendations for the management of anemia in patients with ESRD undergoing hemodialysis. Compared with state-of-the-art models, the proposed SAM has several unique advantages. First, the proposed Informer-aware Hb prediction system is state-of-the-art and can accurately estimate Hb trends given any timestamp. Second, a mixed model of GRU and Informer with multi-head SAM is proposed to effectively learn the context feature representation for better ESA dose recommendation. Third, the proposed Informer-based approach shows that the computational complexity can be significantly reduced.

Anemia management in ESRD patients undergoing hemodialysis is a major challenge, and previous studies have proposed several

Table 2
Three different methods for processing missing values. (Hierarchical method: 36,677 data points).

Group	Number of patients	Percentage among all study subjects	*Methods of processing missing values	Post-processing numbers of datapoints
Training	422	67.7%	PUR	22,536
			LI	24,959
			HI	24,372
Validation	108	17.4%	PUR	6,179
			LI	6,784
			HI	6,619
Test	93	14.9%	PUR	5,252
			LI	5,795
			HI	5,686

* PUR: purification. LI: linear interpolation. HIS [10] Hierarchical.

Table 3
Sensitivity analysis of the proposed hemoglobin prediction model. (A) different length of historical laboratory data points during the last 1, 2, and 3 months. (B) different methods for processing missing values. Table (3C) different number of multi-head processing in the algorithms. (Abbreviations: MAE: mean absolute error. RMSE: root mean square error. ME: mean error).

Length of historical laboratory datapoints	Validation set after preprocessing with HI method (n = 6,619 datapoints)			Test set after preprocessing with HI method (n = 5,686 datapoints)		
	MAE	MSE	ME	MAE	MSE	ME
Previous 1 month	0.4609	0.3874	0.0717	0.4704	0.3910	0.0237
Previous 2 months	0.4581	0.3804	0.0657	0.4563	0.3728	0.0186
Previous 3 months	0.4469	0.3639	0.0688	0.4475*	0.3607	0.0348

Model	Datapoints number			Preprocessing method for missing value			Validation Set			Test Set		
	Train	Validation	Test	Training	Validation	Test	MAE	MSE	ME	MAE	MSE	ME
Informer	22,536	6,179	5,252	PUR	PUR	PUR	0.4505	0.3675	0.0366	0.4577	0.3728	0.0027
Informer	24,959	6,784	5,252	LI	LI	PUR	0.4593	0.3801	0.1232	0.4503	0.3630	0.0994
Informer	24,372	6,619	5,252	HI	HI	PUR	0.4469	0.3639	0.0688	0.4389**	0.3473	0.0381

Multi-Head Numbers	Model	Preprocessing method for missing value	Train	Validation	Test	Validation Set			Test Set		
						MAE	MSE	ME	MAE	MSE	ME
heads = 2	Informer	HI	24,372	6,619	5,686	0.4528	0.3730	0.0908	0.4548	0.3697	0.0559
heads = 5	informer	HI	24,372	6,619	5,686	0.4469	0.3639	0.0688	0.4475***	0.3607	0.0348
heads = 8	informer	HI	24,372	6,619	5,686	0.4485	0.3676	0.0964	0.4518	0.3653	0.0641

(A): Note: * The last three months, with the lowest MAE were used for the subsequent analysis.
(B): Note: The test set was controlled using the PUR method to compare different processing methods. **The hierarchical method (HI), with the lowest MAE is used for the subsequent analysis.
(C): Note: ***The heads of five, with the lowest MAE were used for the subsequent analysis.

Table 4
Comparison of model performance in hemoglobin prediction between our proposed SAM and the published RNN [8].

No	Model	TrDP	TeDP	#Tr	#Val	#Te	ME	MAE	MSE	Comparison of MAE between models, p-value (Krusal-Wallis)	Comparison within model, p-value (ANOVA)
1	RNN	PUR	PUR	22,536	6,179	5,252	0.203	0.586	0.608	-	p value <0.005
2	RNN	LI	PUR	24,959	6,784	5,252	0.223	0.591	0.617	-	
3	RNN	HI	PUR	24,371	6,619	5,252	0.216	0.593	0.610	-	
4	SAM	PUR	PUR	22,536	6,179	5,252	0.006	**0.465**	0.617	4 vs. 1, p value <0.005	p value <0.005
5	SAM	LI	PUR	24,959	6,784	5,252	-0.003	**0.446**	0.594	5 vs. 2, p value <0.005	
6	SAM	HI	PUR	24,371	6,619	5,252	-0.002	**0.451**	0.602	6 vs. 3, p value 0.014	

Note: RNN: Hemoglobin prediction based on the recurrent neural network method [8]. SAM: self-attention mechanism, proposed method. TrDP: Training dataset processing. TeDP: test dataset processing. PUR: purification. LI: linear interpolation. HI [12]: hierarchical. ME: mean error. MAE: mean absolute error. MSE: mean standard error.

computerized clinical decision support systems (CDSS) to overcome these challenges. Fuzzy control [14, 15], support vector regression [16], Bayesian networks [17], model predictive control [18, 19], rule-based systems [20], and mathematical models [21, 22] have been utilized to address this clinically challenging issue.

Unlike other tasks of prediction in biomedical data, the prediction of hemoglobin in hemodialysis patients has unique features where the same cohort of patients contributes to dozens of data points along the chronological time axis. Such time-series data points are strikingly different from other biomedical data studies with a cross-sectional dichotomy of normal vs. abnormal data points. Recent studies have attempted to use reinforcement learning and recurrent neural networks to "learn" the nuanced pattern of time series data points. Table 6 shows a summary of studies using artificial neural networks to predict the Hb level of hemoglobin. The mean absolute error (MAE) of future (3 months) hemoglobin prediction was 0.551, 0.6178, and 0.574 g/dL, respectively. Our proposed model performed better, with an MAE of 0.451 g/dL for the future Hb after one month.

Our proposed SAM model performs better than the state-of-the-art model [9], as shown in Table 4, with an MAE of predicted Hb of 0.593 versus 0.451, respectively ($p = 0.014$). While the RNN model requires readout of previous data points by algorithms, SAM leverages the interrelation between data points and achieves effective and efficient readout of previous data points. With a more accurate Hb prediction model, patients and physicians could benefit from a more effective individualized CDSS for recommending ESA dosing, which is confirmed by the simulation results in Table 5. The proposed CDSS increased the non-failure rate (Hb > 10 g/dL) from 86.3 to 92.7 % ($p < 0.001$), reduced the failure rate (Hb < 10 g/dL) from 13.7 to 7.3 % ($p < 0.001$), and provided smaller Hb difference (from 0.6 to 0.4 g/dL). We believe our model has the potential to serve as a practical CDSS to assist physicians better manage anemia in ESRD patients.

SAM has emerged as a novel method for processing clinical data. Lee et al. used SAM to process irregular multivariate time-series data in the EHR to predict in-hospital mortality, length of stay, and phenotyping [24]. Xu Y et al. used SAM to selectively learn different positions in pathological slide images to improve the performance of colorectal cancer diagnosis [25]. Wang et al. applied SAM to the lesion segmentation network on chest CT images to diagnose COVID-19 [26].

In our patient cohort, 606 out of 623 (97.2%) patients had received intravenous iron supplement. As shown in Table 1, under the above iron supplements, our patient cohort revealed adequate iron levels: ferritin: median (IQR) 305.8 ng/mL (127.9–487.8), IRON/TIBC mean (standard deviation): 24.6 (11.7). Because the prescription of iron supplements was clearly defined and strictly regulated, and the laboratory data in Table 1 revealed that our patient cohort had adequate iron supplement responses, we did not incorporate iron supplement into our treatment recommendation algorithms.

Micro-inflammation, measured by serum C-reactive protein (CRP), was believed to be connected to ESA hypo-responsiveness [27]. However, per international KDIGO guideline [1] and domestic Taiwan Society of Nephrology guideline [2], serum CRP was not routinely checked in daily practice. At our HD center, serum CRP was checked semiannually among 505 (81%) out of 623 patients in this study, with only 2,648 (5.5%) CRP data-log among total of 48,087 data-log. After adding CRP into algorithms, Supplement Table 4S showed that algorithms with CRP exhibited higher MAE than algorithms without CRP (0.560 vs. 0.451), suggesting less accurate prediction performance. Supplement Table 5S revealed similar simulation results by algorithms with CRP than algorithms without CRP, but the simulation also suggested higher ESA doses in the former.

This study has several limitations. First, the data are from a single medical center and the performance of ESA dose recommendation in Table 5 is based on simulation, not actual prospective comparison with physicians. External validation is needed to demonstrate the generalizability of this model. Second, there are other confounding factors not captured in this model that could affect the hemoglobin levels in these patients, such as micro-inflammation status (CRP, interleukins), blood loss during hemodialysis, acute gastrointestinal hemorrhage, surgery, and nutritional status. Third, we did not incorporate intravenous iron supplement in the algorithms, and the actual prescriptions of iron and ESA may vary from hospital to hospital. Additional calibration of the different prescription patterns of iron and ESA by algorithms may be required when performing external validation. Furthermore, the clinical benefit needs to be confirmed in a prospective clinical comparison trial between physicians' prescription and model based recommendations, which trial should also consider physician adherence to the CDSS.

5. Conclusion

We propose a novel recurrent-free framework for Hb and ESA dose recommendation. First, the proposed SAM based on Informer

Table 5
Performance evaluation for ESA dose recommendation between the proposed model and actual physician prescriptions.

	Overall (validation and test sets)			Validation set			Test set		
	Physician prescriptions	Algorithm	P-value	Physician prescriptions	Algorithm	P-value	Physician prescriptions	Algorithm	P-value
ESA dose, IU/30 days. Median, (IQR)	20889 (12021–29414)	20561 (11594–29430)	0.68	20599 (11723–28896)	19436 (10740–28210)	0.46	21179 (12341–30005)	21870 (12522–30771)	0.55
Validation set and Test sets (n = 12,305 datapoints, processing of missing value: HI method)									
Non-failure (Hbt+4 > 10 g/dL) rate (%)	87.4	93.7	<0.001	88.3	94.8	<0.001	86.3	92.7	<0.001
ESA dose, IU/30 days. Median, (IQR) in patients with Hbt+4 > 10 g/dL: ESA dose, IU/30days, Median, (IQR)	19258 (8629–28424)	19249 (8825–28024)	0.87	19641 (7280–28516)	19644 (7813–28359)	0.10	18559 (9487–27804)	18557 (9529–37450)	0.23
Failure (Hbt+4 < 10 g/dL) Rate (%)	12.6	6.3	<.001	11.7	5.2	<.001	13.7	7.3	<0.001
ESA dose, IU/30 days. Median, (IQR) in patients with Hbt+4 < 10 g/dL, ESA dose, IU/30days Median, (IQR)	26387 (17224–42514)	42779 (29711–44811)	<0.001	25697 (16661–39938)	41082 (28597–44505)	<0.001	28247 (17521–43599)	43715 (32290–44983)	<0.001
Subgroup 1: Patients with low baseline Hb (Hbt+3 < 10.0 g/dL) (n = 1,507 datapoints)									
Non-failure (Hbt+4 > 10 g/dL) rate (%)	54.1	58.1	<0.001	56.2	64.1	<0.001	52.0	52.2	0.08
ESA dose, IU/30 days. Median, (IQR) in patients with Hbt+4 > 10 g/dL: ESA dose, IU/30days, Median, (IQR)	34029 (28167–38311)	33491 (27366–37886)	0.39	34028 (28320–38460)	33802 (27604–38242)	0.11	33724 (26789–38401)	33197 (27047–37484)	0.59
Failure (Hbt+4 < 10 g/dL) Rate (%)	45.9	41.9	<0.001	43.8	35.9	<0.001	48.0	47.8	0.44
ESA dose, IU/30 days. Median, (IQR) in patients with Hbt+4 < 10 g/dL, ESA dose, IU/30days Median, (IQR)	36989 (27740–42119)	38126 (29487–42731)	0.07	36573 (26763–41459)	37403 (28760–41937)	0.30	36800 (28000–41068)	38456 (30859–43307)	0.58
Subgroup 2: Patients with on-target baseline Hb (Hbt+3: 10–12 g/dL) (n = 7,959 datapoints)									
Non-failure (Hbt+4 > 10 g/dL) rate (%)	75.2	93.0	<0.001	75.8	93.2	<0.001	74.6	92.5	<0.001
ESA dose, IU/30 days. Median, (IQR) in patients with Hbt+4 > 10 g/dL: ESA dose, IU/30days, Median, (IQR)	22419 (15201–29271)	22477 (15282–29514)	0.38	22477 (15637–29088)	22510 (15557–29174)	0.64	22317 (14791–29519)	22452 (15056–29854)	0.67
Failure (Hbt+4 < 10 g/dL) Rate (%)	9.7	1.6	<0.001	9.4	1.8	<0.001	10.2	1.4	<0.001
ESA dose, IU/30 days. Median, (IQR) in patients with Hbt+4 < 10 g/dL, ESA dose, IU/30days Median, (IQR)	21016 (14431–29928)	17441 (12370–25660)	<0.001	20726 (14418–29521)	17621 (12574–26603)	<0.001	21406 (14505–30191)	15865 (12046–23825)	<0.001
Hb difference (Hbt+4 – Hbt+3), g/dL									
All patients, Median (IQR)	0.6 (0.3–1.1)	0.4 (0.2–0.7)	<0.001	0.6 (0.3–1.1)	0.4 (0.2–0.7)	<0.001	0.6 (0.3–1.1)	0.4 (0.2–0.7)	<0.001
Patients with Hbt+3 < 10 g/dL Median (IQR)	0.9 (0.4–1.5)	0.7 (0.4–1.1)	<0.001	0.9 (0.4–1.5)	0.7 (0.4–1.1)	<0.001	0.9 (0.4–1.5)	0.7 (0.4–1.1)	<0.001
Patients with Hbt+3: 10–12 g/dL Median (IQR)	0.5 (0.2–0.9)	0.3 (0.2–0.6)	<0.001	0.5 (0.2–0.9)	0.3 (0.2–0.6)	<0.001	0.5 (0.2–0.9)	0.3 (0.2–0.6)	<0.001
Patients with Hbt+3 > 12 g/dL Median (IQR)	0.7 (0.3–1.3)	0.6 (0.3–0.9)	<0.001	0.7 (0.3–1.3)	0.6 (0.3–0.9)	<0.001	0.7 (0.3–1.3)	0.6 (0.3–0.9)	<0.001

Table 6
Literature review of studies using Artificial Neural Networks (ANN) for Hb prediction and treatment recommendation systems, including the proposed model.

Author	Year	Journal	ANN method	ANN architecture	Features	Patient population	Mean Absolute Error (Hb g/dL)
Barbieri C [23]	2015	Computers in Biology and Medicine	Feed-forward neural networks	Multilayer Perceptron (MLP), 2 hidden layers, 8 neurons in each layer	Modeling taking into account the RBC lifespan and drug pharmacodynamics. Output as the single future Hb data point.	4,100 patients from Italy, Spain, Portugal	**0.574**
Lobo B [8]	2020	Artificial Intelligence in Medicine	Recurrent neural networks	RNN-Long-Short Term Memory (LSTM)	Both future ESA dosing and iron dosing for better prediction. Output as the future Hb at 1m, 2m, 3m data points	1,972 patients from 11 dialysis centers in Virgina, USA	**0.5394(1m), 0.6056(2m), 0.6178(3m)**
Pellicer-Valero OJ [13]	2020	Artificial Intelligence in Medicine	Recurrent neural networks	RNN-Gated Recurrent Units (GRUs)	Predictions at every time step since the very first day, not limited or restricted by previous 3 months Hb data. Output as the next hemoglobin delta.	110,000 patients from 12 countries	**0.551(3m)**
Yun HR [9]	2021	Computers in Biology and Medicine	Recurrent neural networks	RNN-Gated Recurrent Units (GRUs)	Prediction model for next-month Hb and recommendation model of ESA to target next-month Hb. Clinical benefits for higher Hb on-target rate, stable Hb and lower ESA dose.	446 patients from 7 tertiary hospitals in Korea	**0.59 (1m)**
Yang JY (This study)			Salf-Attention Mechanism (SAM)	SAM	Prediction model for next-month Hb and recommendation model of ESA to target next-month Hb. Clinical benefits for higher non-failure rate, lower failure rate, and stable Hb.	623 patients with 36,677 datapoints from one medical center in Taiwan.	**0.451 (1m)**

was used to accurately predict the Hb level at the future timestamp, which allows us to process Hb trends based on limited numbers of previous Hb data points. In addition, a mixed model of GRU and multi-head SAM was proposed to effectively capture the trend of historical and heterogeneous data between the correlation between a previous ESA treatment, previous Hb data points, and other laboratory data points. Extensive experiments and simulations were performed, and the proposed method achieved state-of-the-art performance in both Hb prediction and ESA dose recommendation. Our model holds great potential for individualized anemia management, such as the CDSS, in ESRD patients undergoing hemodialysis. Further external validation with other datasets and prospective clinical utility studies are warranted.

Declarations

Author contribution statement

Ju-Yeh Yang, MD, MSc: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Tsung-Chun Lee, MD, PhD; Chih-Chung Hsu, PhD: Analyzed and interpreted the data; Wrote the paper.

Wo-Ting Liao: Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

Funding statement

Dr. Ju-Yeh Yang was supported by Ministry of Science and Technology [MOST 110-2635-B-418-001], Far Eastern Memorial Hospital [FEMH-2021-C-036].

Chih-Chung Hsu was supported partly by Ministry of Science and Technology, Taiwan [111-2634-F-007-002, 110-2218-E-006-026, 110-2222-E-006-012, 111-2221-E-006-210], and partly by Higher Education Sprout Project of Ministry of Education (MOE) to the Headquarters of University Advancement at National Cheng Kung University (NCKU).

Data availability statement

Data will be made available on request.

Declaration of interests statement

The authors declare no competing interests.

Additional information

Supplementary content related to this article has been published online at https://doi.org/10.1016/j.heliyon.2022.e12613.

Acknowledgements

The authors would like to express their gratitude to the staff members of the Division of Information Technology, FEMH, and the staff members of the FEMH Hemodialysis Center.

References

[1] A.S. Kliger, et al., KDOQI US commentary on the 2012 KDIGO clinical practice guideline for anemia in CKD, Am. J. Kidney Dis. 62 (2013) 849–859.
[2] S.C. Hung, et al., Anaemia management in patients with chronic kidney disease: Taiwan practice guidelines, Nephrology 19 (2014) 735–739.
[3] J.W. Eschbach, et al., Recombinant human erythropoietin in anemic patients with end-stage renal disease. Results of a phase III multicenter clinical trial, Ann. Intern. Med. 111 (1989) 992–1000.
[4] A.J. Collins, et al., Death, hospitalization, and economic associations among incident hemodialysis patients with hematocrit values of 36 to 39, J. Am. Soc. Nephrol. 12 (2001) 2465–2473.
[5] A.J. Collins, J.Z. Ma, J. Ebben, Impact of hematocrit on morbidity and mortality, Semin. Nephrol. 20 (2000) 345–349.
[6] A.K. Singh, et al., Correction of anemia with epoetin alfa in chronic kidney disease, N. Engl. J. Med. 355 (2006) 2085–2098.
[7] Y. Sakaguchi, et al., Types of erythropoietin-stimulating agents and mortality among patients undergoing hemodialysis, J. Am. Soc. Nephrol. 30 (2019) 1037–1048.
[8] B. Lobo, et al., A recurrent neural network approach to predicting hemoglobin trajectories in patients with End-Stage Renal Disease, Artif. Intell. Med. 104 (2020), 101823.
[9] H.R. Yun, et al., Erythropoiesis stimulating agent recommendation model using recurrent neural networks for patient with kidney failure with replacement therapy, Comput. Biol. Med. 137 (2021), 104718.
[10] A. Dosovitskiy, An Image Is Worth 16X16 Words: Transformers for Image Recognition at Scale. arXiv:2010.11929, 2020, 2020.
[11] S.Z. Haoyi Zhou, et al., Informer: beyond efficient Transformer for long sequence time-series forecasting, in: The Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI), 2021, p. 2021.
[12] Y.M.a.M.B. Narayanaswamy, Hierarchical temporal-contextual recommenders, in: Workshop on Modeling and Decision-Making in the Spatiotemporal Domain, 32nd Conference on Neural Information Processing Systems (NIPS 2018), 2018. Montréal, Canada.

[13] O.J. Pellicer-Valero, et al., Enhanced prediction of hemoglobin concentration in a very large cohort of hemodialysis patients by means of deep recurrent neural networks, Artif. Intell. Med. 107 (2020), 101898.
[14] R. Bellazzi, C. Siviero, R. Bellazzi, Mathematical modeling of erythropoietin therapy in uremic anemia. Does it improve cost-effectiveness? Haematologica 79 (1994) 154–164.
[15] A.E. Gaweda, A.A. Jacobs, M.E. Brier, Application of fuzzy logic to predicting erythropoietic response in hemodialysis patients, Int. J. Artif. Organs 31 (2008) 1035–1042.
[16] J.D. Martín-Guerrero, et al., Dosage individualization of erythropoietin using a profile-dependent support vector regression, IEEE Trans. Bio Med. Eng. 50 (2003) 1136–1142.
[17] R. Bellazzi, Drug delivery optimization through Bayesian networks: an application to erythropoietin therapy in uremic anemia, Comput. Biomed. Res. 26 (1993) 274–293.
[18] A.E. Gaweda, et al., Model predictive control of erythropoietin administration in the anemia of ESRD, Am. J. Kidney Dis. 51 (2008) 71–79.
[19] A.E. Gaweda, et al., Individualized anemia management in a dialysis facility – long-term utility as a single-center quality improvement experience, Clin. Nephrol. 90 (2018) 276–285.
[20] D.C. Miskulin, et al., Computerized decision support for EPO dosing in hemodialysis patients, Am. J. Kidney Dis. 54 (2009) 1081–1088.
[21] D.H. Fuertinger, et al., Prediction of hemoglobin levels in individual hemodialysis patients by means of a mathematical model of erythropoiesis, PLoS One 13 (2018), e0195918.
[22] S. Rogg, et al., Optimal EPO dosing in hemodialysis patients using a non-linear model predictive control approach, J. Math. Biol. 79 (2019) 2281–2313.
[23] C. Barbieri, 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. 61 (2015) 56–61.
[24] Y. Lee, et al., Multi-view integrative attention-based deep representation learning for irregular clinical time-series data, IEEE J. Biomed. Health Inform. (2022).
[25] Y. Xu, et al., Dual resolution deep learning network with self-attention mechanism for classification and localisation of colorectal cancer in histopathological images, J. Clin. Pathol. (2022).
[26] X. Wang, et al., SSA-Net: spatial self-attention network for COVID-19 pneumonia infection segmentation with semi-supervised few-shot learning, Med. Image Anal. 79 (2022), 102459.
[27] H.H. Shah, et al., Inflammation and erythropoiesis-stimulating agent hyporesponsiveness: a critical connection, Kidney Med 2 (2020) 245–247.