



# Improvement of Dialysis Dosing Using Big Data Analytics

Syeda Leena Mumtaz<sup>1</sup>, Abdulrahim Shamayleh<sup>1,2,3</sup>, Hussam Alshraideh<sup>1,2,3</sup>, Adnane Guella<sup>4</sup>

<sup>1</sup>Biomedical Engineering Graduate Program, American University of Sharjah, Sharjah, UAE

<sup>2</sup>Engineering Systems Management, American University of Sharjah, Sharjah, UAE

<sup>3</sup>Department of Industrial Engineering, American University of Sharjah, Sharjah, UAE

<sup>4</sup>Department of Nephrology, University Hospital Sharjah, Sharjah, UAE

**Objectives:** Large amounts of healthcare data are now generated via patient health records, records of diagnosis and treatment, smart devices, and wearables. Extracting insights from such data can transform healthcare from a traditional, symptom-driven practice into precisely personalized medicine. Dialysis treatments generate a vast amount of data, with more than 100 parameters that must be regulated for ideal treatment outcomes. When complications occur, understanding electrolyte parameters and predicting their outcomes to deliver the optimal dialysis dosing for each patient is a challenge. This study focused on refining dialysis dosing by utilizing emerging data from the growing number of dialysis patients to improve patients' quality of life and well-being. **Methods:** Exploratory data analysis and data prediction approaches were performed to gather insights from patients' vital electrolytes on how to improve the patients' dialysis dosing. Four predictive models were constructed to predict electrolyte levels through various dialysis parameters. **Results:** The decision tree model showed excellent performance and more accurate results than the support vector machine, linear regression, and neural network models. **Conclusions:** The predictive models identified that pre-dialysis blood urea nitrogen, pre-weight, dry weight, anticoagulation, and sex had the most significant effects on electrolyte concentrations. Such models could fine-tune dialysis dosing levels for the growing number of dialysis patients to improve each patient's quality of life, life expectancy, and well-being, and to reduce costs, efforts, and time consumption for both patients and physicians. The study's results need to be validated on a larger scale.

**Keywords:** Data Science, Renal Dialysis, Statistical Data Analysis, Chronic Kidney Disease, Machine Learning

**Submitted:** August 26, 2022

**Revised:** February 5, 2023

**Accepted:** March 20, 2023

## Corresponding Author

Abdulrahim Shamayleh

Department of Industrial Engineering, American University of Sharjah, Sharjah, UAE. Tel: +971 6 515 4907, E-mail: ashamayleh@aus.edu (<https://orcid.org/0000-0002-0214-7052>)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

© 2023 The Korean Society of Medical Informatics

## 1. Introduction

Kidney disease is one of the most common causes of early death worldwide. The number of patients with chronic kidney disease (CKD) totaled 700 million in 2017. Around 1.2 million patients have died globally from it, a number expected to increase by 2.2–4.2 million by 2040 [1]. According to the Emirates Nephrology Society, more than 520 individuals per million in the United Arab Emirates (UAE) general population have CKD. Among the available treatment options, dialysis is the most common. The drawbacks of dialysis include its discomfort and cost. Patients must visit a healthcare facility weekly for two to four dialysis sessions, according to

their physician's recommendations. Dialysis sessions cost between USD 175 and USD 275 per session [2]. Furthermore, dialysis prescription and therapy are often complicated and rely primarily on patient data, where several attributes may affect the survival rate and treatment quality.

At the present time, machine learning (ML) and big data analytics have been integrated into the healthcare field to diagnose and treat long-term patients. Several studies have utilized big data for dialysis care. One such study, by Barbieri et al. [3], analyzed data from 766,000 patients across several years and developed a patient- and session-specific artificial neural network (ANN) model comprising 60 attributes. The model was used to predict risks from intradialytic hypotensive periods before each hemodialysis session from the heart rate and systolic blood pressure (SBP) profiles, dialysis dose (Kt/V), and post-dialysis body weight predictions.

Moreover, Dr. Brendan Bowman utilized big data from dialysis dosing records of about 3,000 patients over several months [4]. The study aimed to develop a model that helped clinicians accurately predict and control dialysis patients' red blood cell (RBC) counts, which modeled treatment outcomes to reduce treatment costs. Several dialysis dosing studies, such as the HEMO trial study [5], asserted that providing a higher dialysis dose has benefits for the RBC count stability of only some dialysis patients, not all. The HEMO trial study did not demonstrate an improved survival rate or lower morbidity when a high-flux membrane was used. This proved that larger datasets yield reliable results; however, not all problems can be addressed and deciphered by the dataset size [6]. On this note, Perl et al. [7] acknowledged a need for further research on big data analysis, which would validate the application of multidimensional measures and compute adequate dialysis dosing levels for each patient's treatment.

The number of kidney disease patients on dialysis is increasing, and their well-being relies on medical technology, which includes the necessity of proper dialysis dosing levels to ensure that patients receive the right dialysis treatment at the right time. Earlier studies offered little information regarding predictive models for treating kidney disease patients that could provide adequate dialysis dosage levels [8]. Hence, there is a need to develop a model that aids physicians and clinicians in delivering precise, optimized dialysis treatments through big data analytics.

For data modeling and analysis of dialysis patients, from a medical viewpoint, blood electrolytes are the most critical parameters and must be controlled. Understanding electrolyte parameters and predicting their outcomes to deliver the optimal dialysis dosing for each patient is a challenge [9,10].

Therefore, it is essential to understand the parameters that correlate with electrolyte levels before determining and prescribing improved dialysis dosing to patients. To understand a patient's electrolyte levels, data analytics tools may provide some insights and help interpret the significance of trends [11,12]. Furthermore, it is possible to build predictive models that determine the most significant attributes for electrolyte level prediction, which may aid in decision-making to prescribe and improve the dialysis dosing for patients [13-16]. Hence, this study aimed to analyze and predict patients' electrolyte levels to improve dialysis dosing guidance, with the ultimate goal of bringing about positive effects for patients' quality of life and well-being, hospitals, and the healthcare field.

## II. Methods

The proposed approach for analyzing and predicting patients' electrolyte levels to improve patients' dialysis dosing levels is outlined in this section. The proposed ML predictive model types for this research are also presented, with a brief description of each.

### 1. Data Collection and Pre-processing

A single-site retrospective study was conducted, and all available patients treated with kidney dialysis at University Hospital Sharjah gave consent to participate in the study. Patients were in the dialysis unit for 1–5 years. The Institutional Review Board (IRB) of University Hospital Sharjah approval was approved for this study (No. UHS-HERC-012-10062019). The established records of 45 patients undergoing several months of dialysis for kidney malfunction were obtained through the hospital's Electronic Medical Record (EMR) system. The EMR data were categorized into 117 attributes, which were converted to a format suitable for further processing by the predictive models. Data pre-processing was carried out using RapidMiner Studio 9.1 (RapidMiner, Boston, MA, USA), a data science platform that supports several artificial intelligence (AI) tools. This step involved four stages, explained below.

#### 1) Data integration and cleaning

Data integration allows predictive models to run efficiently. It was implemented herein by consolidating data from 45 dialysis patients, with each having multiple dialysis data entries across several months. Attributes lacking more than 90% of their values were dropped from the analysis; furthermore, missing numeric values were supplied by the average

attribute value, while categorical attributes were supplied by the most frequently occurring value. Further cleaning of the dataset involved manually replacing errors with a defined value as partial offset for the dataset’s “noise.”

2) Data transformation and reduction

This stage went hand-in-hand with the data-cleaning stage. Per-patient averages were calculated for attributes that were missing from some patients’ records. Several other attributes with missing values were replaced by means of proper nominal values or function expressions. If the values were missing for all months for a particular attribute in a patient’s record, then the entire row of missing values was deleted from the dataset. That is, the missing values were not input by any other means because doing so would have violated data integrity.

2. Exploratory Data Analysis and Data Algorithms

Exploratory data analysis (EDA) incorporates data knowledge and visualization tools to understand the relationships among significant patient attributes. To implement the EDA step, RapidMiner Studio 9.1 and Minitab (minitab.com) were used to create scatterplots and boxplots, which enabled

visualizing vital blood electrolyte levels per patient and analyzing their trends in the dataset. A time-lapse view of vital electrolytes was also obtained to generate further insights from the dialysis sessions on a monthly basis.

ML algorithms were utilized to predict patient electrolyte levels from the selected attributes. The chosen algorithms were decision tree (DT), neural network (NN), linear regression (LR), and support vector machine (SVM) [17,18]. Each technique was selected for its popularity in the data science field, reported accuracy in previous research, and ease of use.

3. Data Prediction, Interpretation, and Predictive Model Performance

Predictive models were then built to select several electrolytes from the big data for analysis of precise dialysis dosing. The information obtained from the EDA step was taken, following patient privacy policies of not including any personal information, and was used to build the algorithmic models. The models were trained and validated using 80% of the available data, and the remaining 20% was used for testing the models.

Each trained predictive model was then tested with the

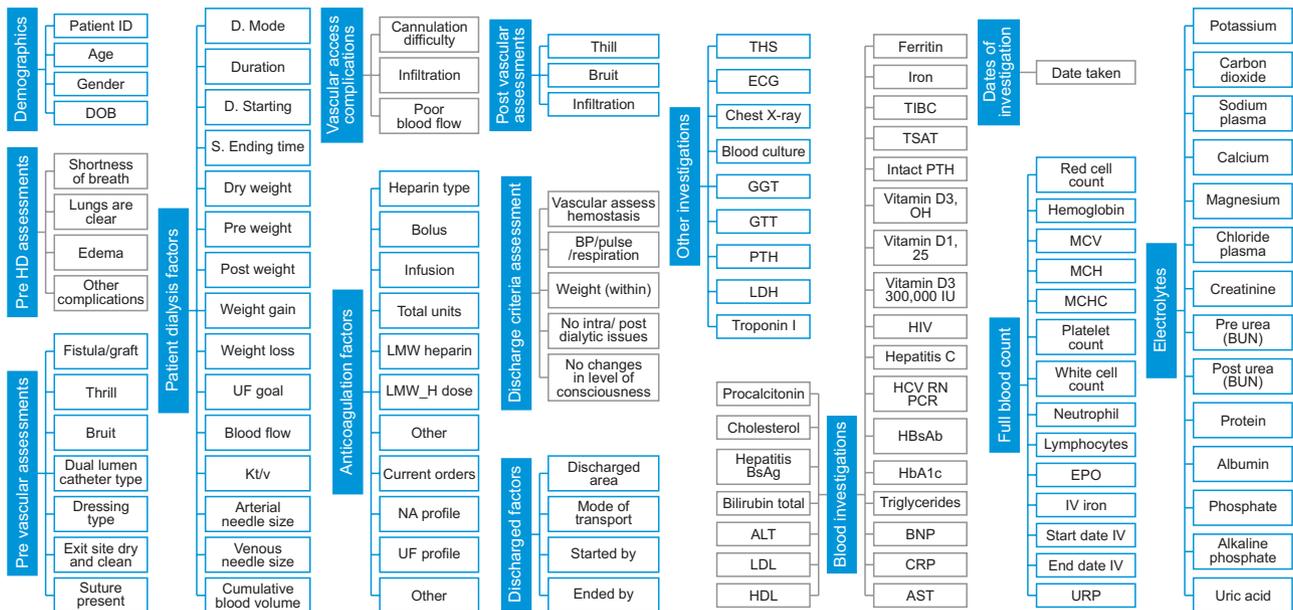


Figure 1. Definitions of categories and dialysis patients’ attributes. The figure represents the 14 categories that comprised 117 attributes, respectively. The most significant categories used for exploratory data analysis or data prediction were the demographic, electrolyte, and patient dialysis factors. DOB: date of birth, LMW: low molecular weight, TSH: thyroid-stimulating hormone, ECG: echocardiography, GGT: gamma-glutamyl transferase, GTT: glucose tolerance test, PTH: parathyroid hormone, LDH: lactate dehydrogenase, ALT: alanine transaminase, LDL: low density lipoprotein, HDL: high density lipoprotein, TIBC: total iron binding capacity, TSAT: transferrin saturation, HIV: human immunodeficiency virus, HCV: hepatitis C virus, PCR: polymerase chain reaction, BNP: B-type natriuretic peptide, CRP: C-reactive protein, AST: aspartate aminotransferase, MCV: mean corpuscular volume, MCH: mean corpuscular hemoglobin, MCHC: mean corpuscular hemoglobin concentration, EPO: erythropoietin, URR: urea reduction rate, BUN: blood urea nitrogen.

final testing dataset to predict the output attributes, which were displayed as tables, charts, and reports. The predictive models identified the most significant predictive attributes for electrolyte concentrations, which could be used to prescribe precise dialysis dosing for each dialysis patient.

Finally, the four predictive techniques were compared to evaluate the performance of each model in terms of the root

squared correlation ( $R^2$ ) and root mean square error (RMSE). The comparisons indicated which of the models would generate the best predictive results, with fewer errors and uncertainties.

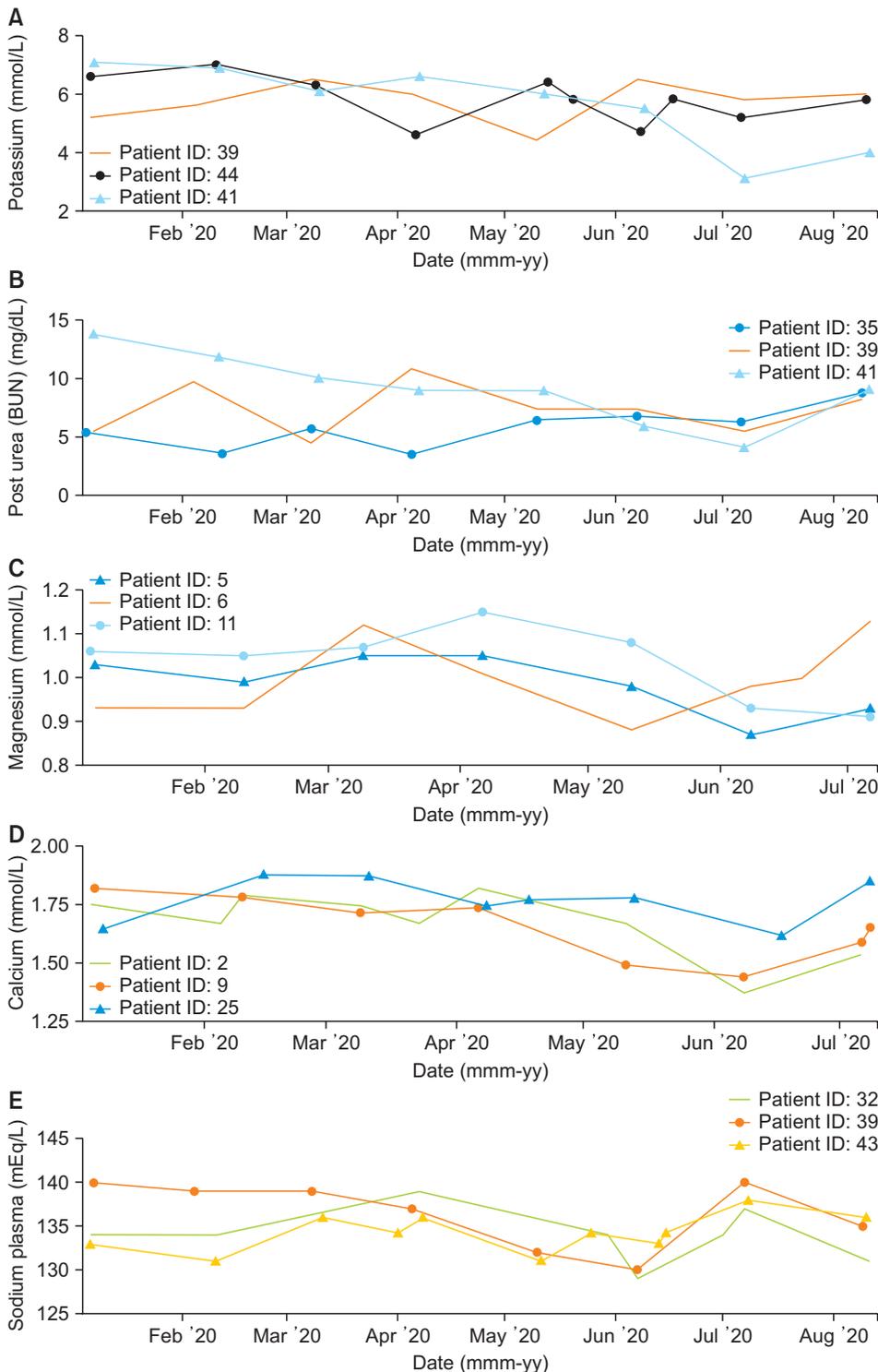


Figure 2. Electrolyte variability profiles among patients. Representation of electrolyte variability among patients over a 1-year period: (A) potassium, (B) post-urea, (C) magnesium, (D) calcium, and (E) plasma sodium.

### III. Results

With the collection of big data, the datasets were restructured. Figure 1 shows the 14 categories and their 117 attributes.

Data for potassium, calcium, plasma sodium, magnesium, and post-dialysis urea were plotted as time series to check variability among patients, as shown in Figure 2. Figure 2A displays the variability profile of potassium throughout patients' monthly checkups, highlighting the fact that an

8-month course of dialysis had some positive and negative effects on the electrolyte levels of some patients. To further study the causes of high variability in some vital electrolytes, electrolyte trends were visualized using several tools during EDA.

Boxplots of significant electrolytes were created as the first EDA approach, shown in Figure 3. Each box represents the range of the particular electrolyte for each of the 45 dialysis patients upon their monthly dialysis checkups. Within each electrolyte profile plot, the mean of every patient is shown as

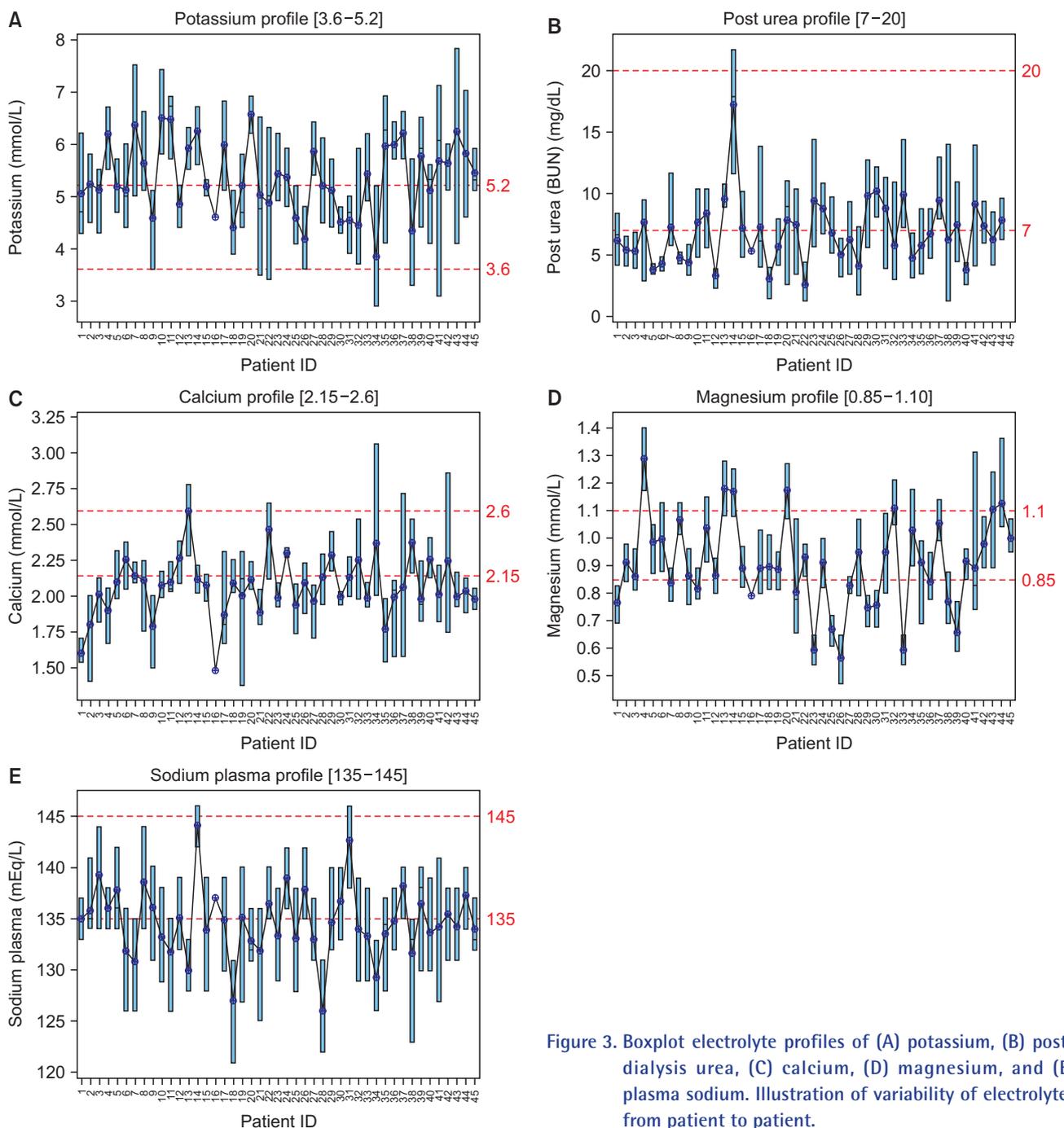


Figure 3. Boxplot electrolyte profiles of (A) potassium, (B) post-dialysis urea, (C) calcium, (D) magnesium, and (E) plasma sodium. Illustration of variability of electrolytes from patient to patient.

dark blue circles, while the electrolyte minimum and maximum ranges are plotted as dotted horizontal red lines. Black lines connect the mean from one dialysis patient to another to further emphasize variability.

The boxplots demonstrated significant variability of electrolyte levels from patient to patient, as is also seen in Figure 2. The largest degrees of variability appeared in the plasma sodium and potassium electrolyte profiles. Patients' electrolyte levels were shown to have fluctuated, which is a serious concern from a medical point of view, and it seemed possible

that significant variations could be due to the non-removal of potassium during dialysis sessions. However, the plasma sodium profile added insight, since changes in sodium levels have an inverse effect on potassium levels. Therefore, the decline in plasma sodium levels in Figure 3E corresponds to the increase in blood potassium levels in Figure 3A.

To add to the insights from important electrolytes, a time-lapse study was analyzed to view these electrolytes from the perspective of dialysis sessions over time, shown in Figure 4. The plots represented the connected mean with electrolyte

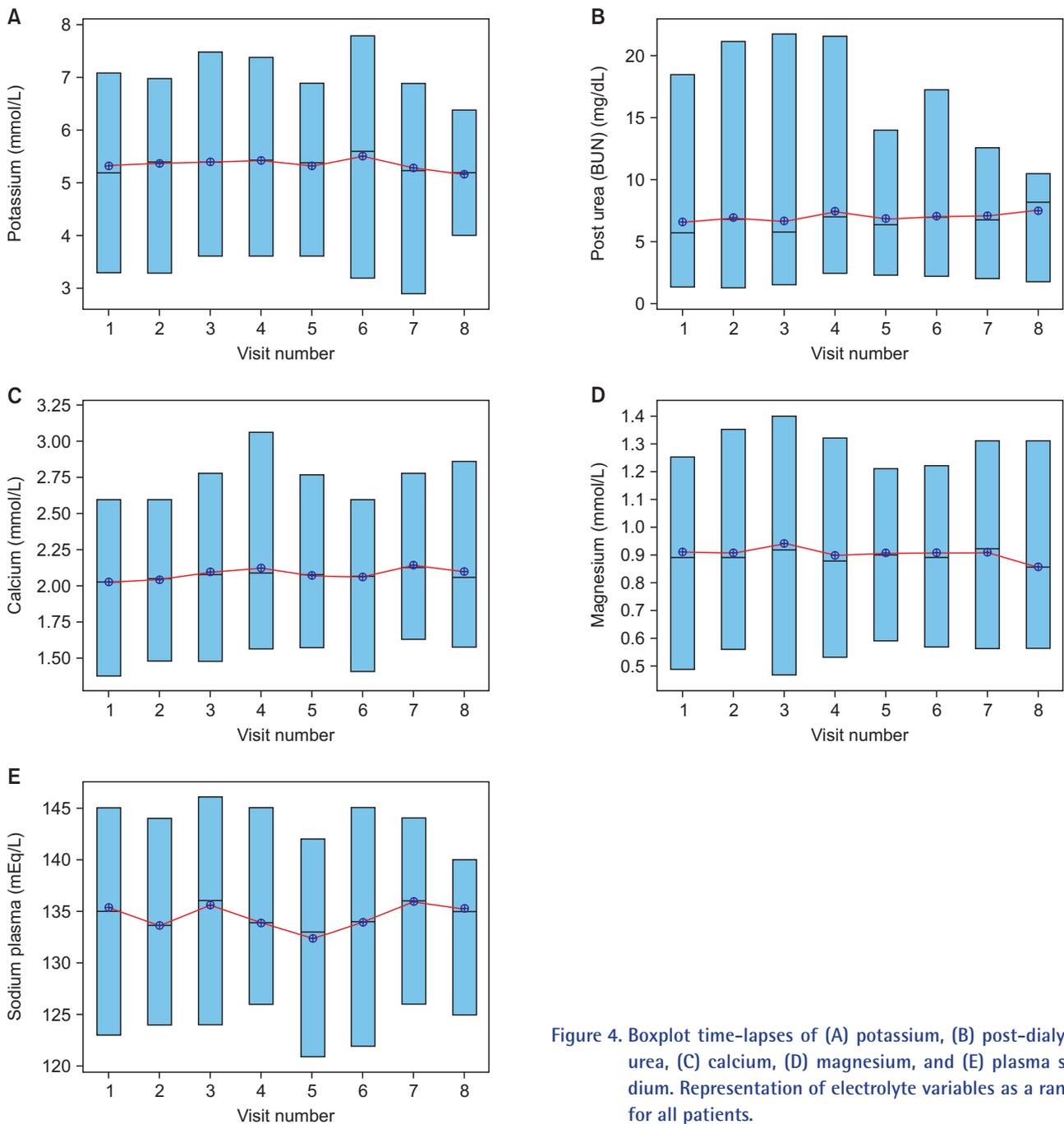


Figure 4. Boxplot time-lapses of (A) potassium, (B) post-dialysis urea, (C) calcium, (D) magnesium, and (E) plasma sodium. Representation of electrolyte variables as a range for all patients.

levels on the y-axis and visit numbers along the x-axis. Each box in the plots represents the range of a particular electrolyte for patients receiving dialysis on a monthly basis.

The electrolyte trends of dialysis patients displayed low variability with respect to the time-lapse input. Figure 3A shows high variability in potassium levels, but Figure 4A demonstrates very low variability. Because Figures 3 and 4 show consistent variability in the electrolyte distributions for different patients and visits, the time-lapse electrolyte results further support the inference that dialysis was indeed working well, with proper dialysate solutions and dialysis sessions. Hence, the analysis showed that other factors caused the high variability of levels of vital electrolytes in patients. The four predictive models mentioned earlier were used to interpret the causative factors for the data presented in Figure 4.

### 1. Data Prediction and Interpretation

Each of the four models had differently weighted predictors and predicted different electrolyte outcomes. Since the models incorporated predictions of 13 electrolytes with the output of weights, each model's most important predictor attribute was captured in Figure 5.

From Figure 5, it can be seen that various attributes affected the predictions of certain electrolytes, illustrated as multiple attributes with repeating colors and their correspond-

ing values. Level 1 (L1) defines the most important feature from the model, followed by level 2 (L2), then level 3 (L3). For further statistical details from the illustrations, Figure 6 displays the specific color code for each input attribute to the model as a weighted table for further characterization.

In other words, Figure 6 displays the sums of the weights of the input attributes at three importance levels from the four predictive models. The total sum of the attribute weights was also generated, corresponding to how much weight the attributes showed in these models for one of the top three importance levels from the prediction results. These weights were taken from the predictive model output, with reference to Figure 5.

As a result, the top five input attributes were identified as pre-dialysis blood urea nitrogen (BUN), pre-weight, dry weight, anticoagulation, and sex. Figure 6 indicates pre-dialysis BUN as the most significant attribute, with the highest analytical total weight (8.162) from all models. From a medical standpoint, each patient's BUN levels are tested monthly before dialysis is administered. The next most significant attributes are pre-weight, dry weight, anticoagulation, and sex, with total analytical weights of 4.635, 4.529, 4.216, and 4.145, respectively.

Hence, our study recommends following the DT model's decision-making support system because the results encourage physicians to ensure that patients' pre-dialysis BUN

Attribute name	LR			DT			SVM			NN		
	L1	L2	L3	L1	L2	L3	L1	L2	L3	L1	L2	L3
Sodium plasma	5	10	1	9	8	10	3	1	7	1	10	12
Potassium	12	9	7	12	7	10	1	12	3	12	13	10
Chloride plasma	5	10	12	9	10	12	1	3	12	10	5	9
Carbon dioxide	4	5	3	1	4	2	3	1	7	4	3	1
Creatinine plasma	5	12	2	12	5	9	1	3	12	12	1	2
Post urea (BUN)	9	12	3	12	5	2	3	1	12	12	1	2
Protein total	7	2	9	4	7	12	3	1	12	7	12	6
Albumin	9	3	4	5	12	6	1	3	12	4	6	13
Calcium	5	7	2	6	7	4	3	1	7	7	13	6
Phosphate	9	12	13	10	12	9	1	3	12	12	7	10
Magnesium	5	8	7	4	7	10	1	3	12	7	13	10
Alkaline phosphatase	2	5	10	10	2	4	3	1	11	2	10	13
Uric acid	1	9	2	9	12	5	1	3	12	1	12	3

- 1 Anticoagulation
- 2 UF goal
- 3 HbA1c
- 4 KT/V
- 5 Dry weight
- 6 Duration
- 7 Gender
- 8 Weight loss
- 9 Pre weight
- 10 Cumulative blood volume
- 11 Age
- 12 Pre urea (BUN)
- 13 Blood flow

Figure 5. Most significant electrolyte predictor attributes. These attributes had an impact on predicting certain electrolytes and their relative significance. LR: linear regression, DT: decision tree, SVM: support vector machine, NN: neural network.

Color code	Attribute name	Weight's sum per level			Total sum
		Level 1	Level 2	Level 3	
1	Anticoagulation	2.683	1.264	0.269	4.216
2	UF goal (L)	0.895	0.369	0.992	2.256
3	HbA1c (%)	0.775	0.945	0.533	2.252
4	KT/V	1.478	0.062	0.297	1.837
5	Dry weight (kg)	3.714	0.782	0.032	4.529
6	Duration (hr)	0.234	0.315	0.280	0.830
7	Gender	2.212	1.372	0.560	4.145
8	Weight loss (kg)	0.000	0.248	0.000	0.248
9	Pre weight (kg)	3.115	0.843	0.677	4.635
10	Cumulative blood volume	0.600	1.167	0.995	2.762
11	Age	0.000	0.000	0.055	0.055
12	Pre-urea (BUN)	4.797	2.296	1.070	8.162
13	Blood flow	0.000	0.550	0.433	0.983

Figure 6. Weighted table for input attributes according to the level of significance.

levels are within an acceptable range. To do so, physicians ought to consider each patient’s lifestyle before dialysis treatment because the five most significant attributes are related to lifestyle features. Other variables included in the models (e.g., duration and age) were not significant as predictors.

## 2. Model Performance

Tables 1 and 2 illustrate the performance of each of the four predictive models as integrated tables. Table 1 displays the performance of the predictive models from the validated and trained output, which used the training dataset, while Table 2 does the same for the tested, finalized model output and the testing dataset.

In our comparison of the models, DT achieved by far the most accurate results, with higher values of  $R^2$  and lower values of RMSE for both the training and testing models. For the training output performance in Table 1, DT achieved an average of  $0.4649 \pm 0.1639$  for  $R^2$  and an average RMSE of  $18.8783 \pm 4.74369$ . For the final tested dataset model performances in Table 2, DT had  $R^2$  and RMSE values of 0.4099 and 21.7011, respectively. The highest  $R^2$  value in the DT model was 0.752 for alkaline phosphatase, and the lowest RMSE value was 0.096 for magnesium.

## IV. Discussion

In this research, a clinical decision-making support system was developed for dialysis patients using data analytics. The results demonstrate the potential of extracting insights from big data to transform healthcare from a traditional symptom-driven practice to precisely personalized medicine.

This study utilized big data from more than 100 dialysis parameters in patients’ records, where critical electrolyte parameters were used to analyze, predict, and provide insights on how to improve a patient’s dialysis dosing. Therefore, electrolyte parameters were investigated with respect to demographic attributes and attributes typical of dialysis patients, including glycated hemoglobin and pre-dialysis BUN values. Vital electrolytes (i.e., magnesium, calcium, sodium, potassium, and post-dialysis BUN) were found to be highly variable in patients, due either to insufficient removal of electrolytes during dialysis sessions or to an inverse correlation with other electrolyte profiles. An example of inverse proportionality involved the higher levels of potassium in patients due to the lower sodium levels. However, it was also demonstrated that the electrolyte levels were consistent over time for each month when a time-lapse view of each electrolyte was analyzed. This clarified that the patients had received appropriate dialysis treatments.

After several types of predictive models (DT, NN, SVM, and LR) were built for further investigation, they predicted electrolyte levels within a range of demographic and vital dialysis parameters. Pre-dialysis BUN, pre-weight, dry weight, anticoagulation, and sex were found to be the five most significant predictive factors. Moreover, the DT model was demonstrated to be the best of the four models we evaluated because it provided a higher average  $R^2$  output and lower average RMSE values for electrolytes.

Ultimately, the predictive results achieved the goal of our study in determining important factors for the improvement of patients’ dialysis dosing levels, where the interpretations from predictors verified that the duration and frequency of

Table 1. Performance of predictive models with the training dataset

Electrolyte name	LR		DT		SVM		NN	
	RMSE	R <sup>2</sup>						
Sodium plasma	4.080 ± 0.743	0.171 ± 0.124	3.602 ± 0.642	0.367 ± 0.164	4.044 ± 0.494	0.446 ± 0.157	3.966 ± 0.821	0.234 ± 0.164
Potassium	0.701 ± 0.155	0.357 ± 0.170	0.659 ± 0.069	0.375 ± 0.116	0.632 ± 0.099	0.442 ± 0.119	0.718 ± 0.142	0.360 ± 0.145
Chloride plasma	4.249 ± 0.933	0.166 ± 0.149	3.189 ± 0.632	0.407 ± 0.240	3.808 ± 0.790	0.413 ± 0.162	3.701 ± 0.505	0.272 ± 0.164
Carbon dioxide	2.178 ± 0.309	0.248 ± 0.179	2.107 ± 0.342	0.313 ± 0.197	2.161 ± 0.207	0.417 ± 0.194	2.254 ± 0.359	0.260 ± 0.183
Creatinine plasma	172.176 ± 33.484	0.440 ± 0.248	131.578 ± 39.111	<b>0.621 ± 0.168</b>	216.273 ± 51.463	<b>0.599 ± 0.125</b>	153.821 ± 34.722	<b>0.566 ± 0.147</b>
Post-urea (BUN)	1.692 ± 0.458	<b>0.722 ± 0.103</b>	1.866 ± 0.446	<b>0.644 ± 0.177</b>	2.331 ± 0.683	<b>0.555 ± 0.135</b>	1.736 ± 0.383	<b>0.736 ± 0.112</b>
Protein total	4.559 ± 0.318	0.107 ± 0.085	4.142 ± 0.885	0.338 ± 0.197	4.453 ± 0.452	0.307 ± 0.151	4.617 ± 0.678	0.185 ± 0.157
Albumin	4.436 ± 0.925	0.148 ± 0.120	3.693 ± 0.759	0.329 ± 0.158	4.032 ± 0.791	0.335 ± 0.146	4.421 ± 0.749	0.120 ± 0.125
Calcium	0.212 ± 0.038	0.249 ± 0.110	0.173 ± 0.034	0.490 ± 0.131	0.167 ± 0.043	<b>0.525 ± 0.170</b>	0.199 ± 0.037	0.324 ± 0.106
Phosphate	0.452 ± 0.097	0.357 ± 0.139	0.495 ± 0.079	0.279 ± 0.199	0.411 ± 0.090	0.465 ± 0.132	0.418 ± 0.046	0.458 ± 0.163
Magnesium	0.158 ± 0.022	0.305 ± 0.185	0.103 ± 0.008	<b>0.668 ± 0.062</b>	0.099 ± 0.017	<b>0.699 ± 0.106</b>	0.138 ± 0.026	0.455 ± 0.197
Alkaline phosphatase	86.337 ± 21.066	0.190 ± 0.121	38.438 ± 9.977	<b>0.794 ± 0.197</b>	94.496 ± 30.171	0.304 ± 0.166	90.978 ± 23.182	0.174 ± 0.130
Uric acid	54.453 ± 8.507	0.446 ± 0.114	55.373 ± 8.684	0.419 ± 0.125	70.037 ± 5.910	2.00 ± 0.146	49.279 ± 8.477	<b>0.570 ± 0.130</b>
Average	25.668 ± 5.549	0.301 ± 0.142	18.878 ± 4.744	0.465 ± 0.164	30.996 ± 7.016	0.455 ± 0.147	24.327 ± 5.394	0.363 ± 0.148

Values are presented as mean ± standard deviation. Values more than 0.5 R<sup>2</sup> (in bold) indicate best performances of the result.

LR: linear regression, DT: decision tree, SVM: support vector machine, NN: neural network, RMSE: root mean square error.

Table 2. Performance of predictive models with the test dataset

Electrolyte name	LR		DT		SVM		NN	
	RMSE	R <sup>2</sup>						
Sodium plasma	4.210	0.123	4.048	0.247	3.818	0.467	4.203	0.121
Potassium	0.675	0.338	0.695	0.325	0.677	0.339	0.710	0.316
Chloride plasma	3.511	0.336	3.927	0.415	5.315	0.443	4.605	0.214
Carbon dioxide	2.082	0.154	1.975	0.275	2.353	0.219	2.071	0.151
Creatinine plasma	176.332	0.427	160.993	<b>0.512</b>	237.395	<b>0.772</b>	139.677	<b>0.720</b>
Post-urea (BUN)	1.951	<b>0.613</b>	1.991	<b>0.611</b>	2.530	0.343	1.927	<b>0.638</b>
Protein total	4.805	0.035	4.686	0.035	3.952	0.242	4.225	0.088
Albumin	4.075	0.234	4.333	0.204	4.144	0.382	4.508	0.164
Calcium	0.229	0.157	0.184	0.438	0.183	0.454	0.217	0.238
Phosphate	0.416	0.377	0.388	0.420	0.326	<b>0.571</b>	0.413	0.427
Magnesium	0.139	0.362	0.096	<b>0.749</b>	0.100	<b>0.727</b>	0.132	<b>0.515</b>
Alkaline phosphatase	77.871	0.128	40.216	<b>0.752</b>	80.345	0.278	81.411	0.079
Uric acid	54.093	0.442	58.582	0.346	70.881	0.323	53.316	0.463
Average	25.415	0.287	21.701	0.409	31.694	0.428	22.878	0.318

Values more than 0.5 R<sup>2</sup> (in bold) indicate best performances of the result.

LR: linear regression, DT: decision tree, SVM: support vector machine, NN: neural network, RMSE: root mean square error.

dialysis were not crucial attributes for determining patients' important electrolyte levels. Instead, the five most significant attributes were pre-dialysis BUN, pre-weight, dry weight, anticoagulation, and sex, which are related to the lifestyle features of dialysis patients.

Moreover, in clinical practice by nurses or physicians, predictive models can be carried out using multiple patients' datasets, and the three most important attributes can be noted from the models' significant levels, which will support decision-making and offer insights from the predictive models' output. Thus, from our results and an interpretation of the models' most significant attributes, physicians ought to educate patients about patterns of nutritional intake. This, likewise, provides patients with more accurate dialysis dosing using big data analytics from the growing number of dialysis patients, which will eventually help improve each patient's quality of life, life expectancy, and well-being, as well as reducing costs, efforts, and time consumption for both patients and physicians.

On that note, this study comprised 45 patients in total, with dialysis sessions recorded from January 2020 to August 2020. In the future, this study could be extended either by analyzing the records of a larger number of patients or by adding temporal data with a history of dialysis earlier than January 2020. This could further improve the predictive models and lead to more reliable results. Other dialysis machine parameters could also be added to the predictive models to

determine their correlations with patients' electrolyte levels, which may support researchers in further investigations.

The limitations of this work include the fact that it was conducted at a single site and the small number of participants. The study's results need to be validated on a larger scale. Unfortunately, to our knowledge, no previous studies have used ML models to predict electrolyte concentrations, which makes it challenging to validate our models' performance. Moreover, there is a need for further research on big data analysis, which would validate the application of multi-dimensional measures and compute adequate dialysis dosing levels for the patient's treatment. Hence, there is a need to develop a model that aids physicians and clinicians in delivering precise and optimized dialysis treatments through big data analytics.

## Conflict of Interest

No potential conflict of interest relevant to this article was reported.

## Acknowledgments

This work is part of master's thesis for Ms. Syeda Leen Mumtaz. She earned her degree from the American University of Sharjah. And the work in this paper was supported, in part, by the Open Access Program from the American University

of Sharjah; however, this paper represents the opinions of the authors and does not mean to represent the position or opinions of the American University of Sharjah.

## ORCID

Syeda Leena Mumtaz (<http://orcid.org/0000-0002-8125-1754>)  
 Abdulrahim Shamayleh (<http://orcid.org/0000-0002-0214-7052>)  
 Hussam Alshraideh (<http://orcid.org/0000-0003-1975-9861>)  
 Adnane Guella (<http://orcid.org/0000-0002-4026-4268>)

## References

1. GBD Chronic Kidney Disease Collaboration. Global, regional, and national burden of chronic kidney disease, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* 2020;395(10225):709-733. [https://doi.org/10.1016/S0140-6736\(20\)30045-3](https://doi.org/10.1016/S0140-6736(20)30045-3)
2. Chaudhary SB. Free dialysis for underprivileged kidney patients in UAE [Internet]. Dubai, UAE: Gulf News; 2019 [cited at 2023 Mar 30]. Available from: <https://gulfnews.com/uae/health/free-dialysis-for-underprivileged-kidney-patients-in-uae-1.61500736>.
3. Barbieri C, Cattinelli I, Neri L, Mari F, Ramos R, Braccaccio 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(1):28-33. <https://doi.org/10.1159/000493479>
4. Samarrai F. Dialysis drug dosing improved by using big data analytics [Internet]. Charlottesville (VA): University of Virginia; 2019 [cited at 2023 Mar 30]. Available from: <https://news.med.virginia.edu/blog/dialysis-drug-dosing-improved-by-using-big-data-analytics/>.
5. Welsh ET. Incremental dialysis may help maintain residual renal function, improve quality of life [Internet]. Thorofare (NJ): Healio News; 2020 [cited at 2023 Mar 30]. Available from: <https://www.healio.com/news/nephrology/20200225/incremental-dialysis-may-help-maintain-residual-renal-function-improve-quality-of-life>.
6. Nuemann ME. ADC keynote: big data in ESRD can be used to improve outcomes, but ties to policy may be troublesome [Internet]. Thorofare (NJ): Healio News; 2019 [cited at 2023 Mar 30]. Available from: <https://www.healio.com/news/nephrology/20190329/adc-keynote-big-data-in-esrd-can-be-used-to-improve-outcomes-but-ties-to-policy-may-be-troublesome>.
7. Perl J, Dember LM, Bargman JM, Browne T, Charytan DM, Flythe JE, et al. The use of a multidimensional measure of dialysis adequacy-moving beyond small solute kinetics. *Clin J Am Soc Nephrol* 2017;12(5):839-47. <https://doi.org/10.2215/CJN.08460816>
8. Kusiak A, Dixon B, Shah S. Predicting survival time for kidney dialysis patients: a data mining approach. *Comput Biol Med* 2005;35(4):311-27. <https://doi.org/10.1016/j.compbiomed.2004.02.004>
9. Banerjee A, Noor A, Siddiqua N, Uddin MN. Significance of attribute selection in the classification of chronic renal disease. Proceedings of 2019 2nd International Conference on Advanced Computational and Communication Paradigms (ICACCP); 2019 Feb 25-28; Gangtok, India. p. 1-6. <https://doi.org/10.1109/ICACCP.2019.8882937>
10. National Kidney Foundation. KDOQI clinical practice guideline for hemodialysis adequacy: 2015 update. *Am J Kidney Dis* 2015;66(5):884-930. <https://doi.org/10.1053/j.ajkd.2015.07.015>
11. Belle A, Thiagarajan R, Soroushmehr SM, Navidi F, Beard DA, Najarian K. Big data analytics in healthcare. *Biomed Res Int* 2015;2015:370194. <https://doi.org/10.1155/2015/370194>
12. Hueso M, de Haro L, Calabia J, Dal-Re R, Tebe C, Gibert K, et al. Leveraging data science for a personalized haemodialysis. *Kidney Dis (Basel)* 2020;6(6):385-94. <https://doi.org/10.1159/000507291>
13. Wickramasinghe MP, Perera DM, Kahandawaarachchi KA. Dietary prediction for patients with chronic kidney disease (CKD) by considering blood potassium level using machine learning algorithms. Proceedings of 2017 IEEE Life Sciences Conference (LSC); 2017 Dec 13-15; Sydney, Australia. p. 300-3. <https://doi.org/10.1109/LSC.2017.8268202>
14. Maurya A, Wable R, Shinde R, John S, Jadhav R, Dakshayani R. Chronic kidney disease prediction and recommendation of suitable diet plan by using machine learning. Proceedings of 2019 International Conference on Nascent Technologies in Engineering (ICNTE); 2019 Jan 4-5; Navi Mumbai, India. p. 1-4. <https://doi.org/10.1109/ICNTE44896.2019.8946029>
15. Amirgaliyev Y, Shamiluulu S, Serek A. Analysis of chronic kidney disease dataset by applying machine learning methods. Proceedings of 2018 IEEE 12th International Conference on Application of Information and Communication Technologies (AICT); 2018 Oct 17-

- 19; Almaty, Kazakhstan. p. 1-4. <https://doi.org/10.1109/ICAICT.2018.8747140>
16. Makino M, Yoshimoto R, Ono M, Itoko T, Katsuki T, Koseki A, et al. Artificial intelligence predicts the progression of diabetic kidney disease using big data machine learning. *Sci Rep* 2019;9(1):11862. <https://doi.org/10.1038/s41598-019-48263-5>
17. Ishan D. Demystifying support vector machine, part I [Internet]. San Francisco (CA): Medium; 2020 [cited at 2023 Mar 30]. Available from: <https://medium.com/swlh/demystifying-support-vector-machine-part-i-b5b083844c9a>.
18. Madhuresh K. Regression Analysis: linear regression, SSE, assumption of linear regression error term and best fit line [Internet]. [place unknown]: WordPress, 2015 [cited at 2023 Mar 30]. Available from: <https://madhureshkumar.wordpress.com/2015/07/21/regression-analysis-linear-regression-sse-assumption-of-linear-regression-error-term-and-best-fit-line/>.