IMPROVEMENT OF DIALYSIS DOSING

USING BIG DATA ANALYTICS

by

Syeda Leena Mumtaz

A Thesis presented to the Faculty of the
American University of Sharjah
College of Engineering
In Partial Fulfillment
of the Requirements
for the Degree of

Master of Science in
Biomedical Engineering

Sharjah, United Arab Emirates

April 2021

Declaration of Authorship

I declare that this thesis is my own work and, to the best of my knowledge and belief, it does not contain material published or written by a third party, except where permission has been obtained and/or appropriately cited through full and accurate referencing.

Signed……Syeda Leena Mumtaz……

Date……29th April, 2021.……

Approval Signatures

We, the undersigned, approve the Master's Thesis of Syeda Leena Mumtaz

Thesis Title: Improvement of Dialysis Dosing Using Big Data Analytics

Date of Defense: 22nd April, 2021.

Name, Title and Affiliation	Signature
Dr. Abdulrahim Shamayleh Assistant Professor, Department of Industrial Engineering Thesis Advisor	
Dr. Hussam Alshraideh Associate Professor, Department of Industrial Engineering Thesis Committee Member	
Dr. Mahmoud Awad Associate Professor, Department of Industrial Engineering Thesis Committee Member	
Dr. Abdulrahim Shamayleh Coordinator Biomedical Engineering Program	
Dr. Lotfi Romdhane Associate Dean for Graduate Affairs and Research College of Engineering	
Dr. Sameer Al Asheh Interim Dean College of Engineering	
Dr. Mohamed El-Tarhuni Vice Provost for Graduate Studies Office of Graduate Studies	

Acknowledgment

Firstly, all the gratitude to *Allah (S.W.T.)* for His countless blessings in advancing me to study further for biomedical engineering graduate program.

Secondly, my sincere gratefulness to my advisor and biomedical engineering graduate program coordinator *Dr. Abdulrahim Shamayleh* for his time, motivation, expedient advice, and immense support throughout the thesis.

I would also like to express my deep gratitude to my thesis committee member *Dr. Hussam Alshraideh* for his valuable time and support. My extended thanks to my thesis committee member *Dr. Mahmoud Awad* for his comments and insights.

Finally, a special and deep acknowledgement towards American University of Sharjah, *Dr. Hasan Al-Nashash*, and the entire biomedical engineering graduate program faculty, for their support in granting graduate assistantships to pursue my master's degree.

Dedication

To my parents…

Abstract

Data is transforming the healthcare sector and making it more dependent on data science. Data science is becoming a critical tool that allows looking at the data generated from various sources, such as patient health records, diagnosis, treatment, smart devices, and wearables. Extracting insights from health data has the potential to transform the healthcare from traditional symptom-driven practice into a precision personalized medicine. The dialysis treatment generates a vast amount of data that can be utilized. Data of each dialysis patient constitutes over 100 parameters that must be regulated every dialysis session. Moreover, an individual dialysis dosing may depend upon complex linkage within multiple clinical and demographical parameters, early dialysis prescriptions, medications, or other health interventions. With dialysis complications, understanding the electrolyte parameters and predicting their outcome for each patient to deliver the optimal dialysis dosing is a challenge. This research approach is intended to improve dialysis dosing from the emerging data and the rising volume of dialysis patients, with the purpose of increasing patient's quality of life and their welfare from the right dialysis treatment. Exploratory data analysis and data prediction approach were performed to provide insights on how to improve the patients' dialysis dosing. Analysis of vital electrolytes displayed high variability amongst patients, which identified the needs to improve the dialysis dosing. Four data prediction models were used to predict patient electrolytes from various parameters. The models include decision tree, neural network, support vector machine, and linear regression. The results from the prediction identified that pre urea (BUN), anticoagulation, HBA1C, gender, and cumulative blood volume, had the most significant predictor weights. The important predictors interpreted that patient's lifestyle and diet patterns are the major factors towards improper variability of the electrolytes.

Keywords: ***Big data; dialysis dosing; machine learning; data analytics; AI; data science; exploratory data analysis; data prediction.***

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AI	Artificial Intelligence
ANN	Artificial Neural Network
APD	Automated Peritoneal Dialysis
BP	Blood Pressure
CAPD	Continuous Ambulatory Peritoneal Dialysis
CKD	Chronic Kidney Disease
CRF	Chronic Renal Failures
DCT	Distal Convoluted Tubule
ECG	Electrocardiogram
EHR	Electronic Health Records
ESRD	End-Stage Renal Disease
EDA	Exploratory Data Analysis
ESA	Erythropoietin Stimulating Agents
GFR	Glomerular Filtration Rate
HD	Hemodialysis
IHHD	Intensified Home Hemodialysis
mEq	Milligram-equivalents
ML	Machine Learning
MW	Molecular Weight
mmol/L	Milli-mole per litre
nPNA	Normalized Protein Nitrogen Appearance
PCT	Proximal Convoluted Tubule
RBC	Red Blood Cell

Chapter 1. Introduction

In this chapter, a brief introduction is presented about the human kidney anatomy, physiology, related kidney disorders, and comorbidities. Then a description of big data analytics and their purpose in the medical field is presented. Also, a description of the problem investigated in this study and the research contribution is provided. Lastly, the general organization and thesis objectives are introduced.

1.1. Overview

Kidney disease is now termed as one of the most common effects of early deaths worldwide. It has been studied that the number of patients having Chronic Kidney Disease (CKD) reached up to 700 million in 2017, typically more than the number of patients with other disorders [1]. Approximately 1.2 million patients have died globally from CKD, which has been expected to increase between 2.2-4.2 million by 2040 [1]. According to the Emirates Nephrology society, more than 1040 individuals per two million in the United Arab Emirates' (UAE) general population suffer from CKD. Therefore, the only treatment options available to kidney diseases include either a kidney transplant or to retain under dialysis [2].

When undergoing dialysis, most end-stage kidney disease patients rely on three times per week dialysis sessions, which is quite costly as it ranges between 650 to 1000 AED per session in the UAE [2]. Also, dialysis prescription and therapy are often complicated, where several factors may impact the patient's rate of survival. Dialysis prescription mainly relies on patient's medical data, which is stored digitally as electronic health records (EHR). EHR consists of big sets of data with countless in-patient and out-patient information. Also, the previous decade of medicine met with an enormous volume of digital data, which are compressed in the form of EHR [3].

Recently, machine learning progresses significantly with big data analytics in the fields of pharmacokinetics, oncology, genetics, and clinical imaging. However, contemporarily there is very little information regarding predictive models of adequate dialysis dosage levels for patients having kidney disease complications [4]. Therefore, there is a need to develop a model that predicts accurate and optimized dialysis dosing using big data analytics.

1.2. Kidneys

Kidneys are organs of the urinary system, as the system is split into lower and upper portions. The upper part of the urinary tract makes kidneys and ureter, while the lower part includes urethra and urinary bladder [5]. Kidney organs relate to bean-shaped structure, weighing about 120-200 grams in humans, whereas 120-135 grams in females and 150-200 grams in males. The size of each kidney is of a closed fist, having dimensions of 10-12 cm in length and 5-7 cm in width. The thickness of each kidney varies between 3-5 cm. Kidneys are merged to ureters, which are 25-30 cm long hollow tubes passing the urine to the urinary bladder. Each kidney contains 1 million nephrons that are the basic functional and structural units of the kidneys [6]. Every minute, the kidneys receive more than one liter of blood which constitutes to cardiac output (CO) of more than 20 percent [6].

Kidneys play vital body roles that feature great significance in daily lives. The common role comprises of urination, which is the waste product excretion of toxic contents like urea, ammonia, and creatinine. Another main role of the kidney involves the regulation of electrolytes, where the body's water and salt content are balanced with the homeostatic mechanism. Also, maintaining acid-base in balance is another crucial function of the kidneys that aims to control the blood pH level of the body.

Other essential roles of the kidney involve blood pressure regulation and the conservation of certain intravascular volume by the renin-angiotensin-aldosterone system. During the process of forming urine, kidneys not only eliminate waste products but also restore the essential products into the blood by reabsorption, including calcium, amino acids, water, phosphates, glucose, and other electrolytes. Kidneys secrete vital hormones, like erythropoietin, which plays a crucial role in the red blood cell production, along with calcitriol that is released during low calcium levels in the body [6].

1.2.1. Anatomy and physiology. Kidneys are positioned overlying the posterior of the abdominal wall in the retroperitoneal location (as shown in Figure 1.1). The right kidney is posterior to the liver, whereas the left kidney is posterior to the spleen, which makes the right kidney level slightly lower than the left kidney level. Also, the left kidney is somewhat longer and less wide as compared to the right kidney [5]. Atop each kidney presents the adrenal gland, with the kidney itself being covered

by a capsule and layers of fat that protects, supports, and provides cushioning effect to the kidneys.

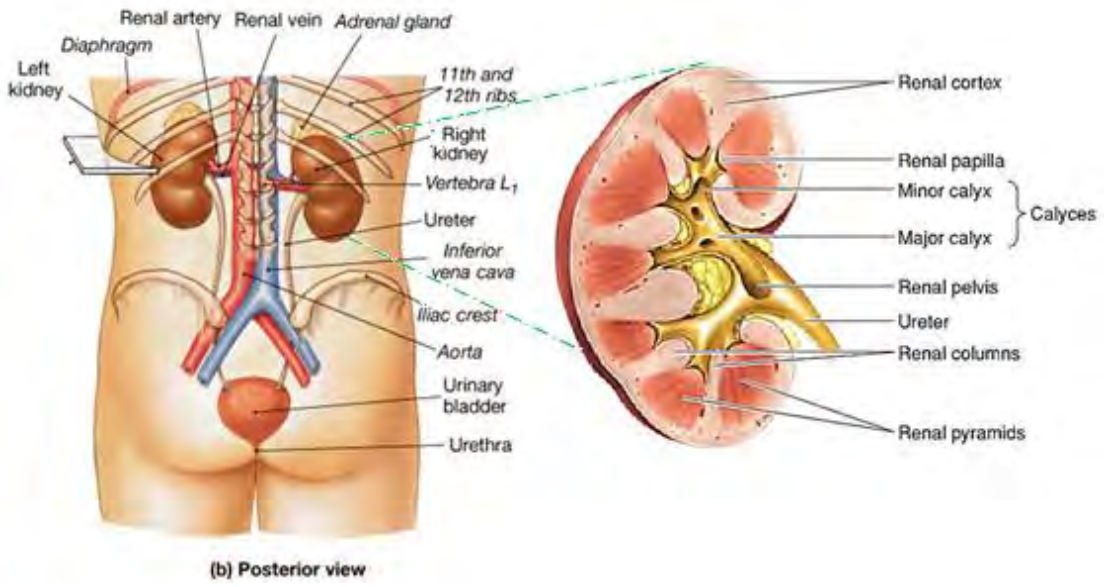

Figure 1.1: Position and parts of the kidneys [7][8].

From Figure 1.1, a kidney is divided into three main regions: the outer renal cortex, the middle renal medulla, and the inner renal pelvis. The renal cortex and medulla comprise of cone-shaped lobes of the renal pyramid. By the tip of renal pyramids, called papilla, are cup-shaped calyces (minor calyx and major calyx) that funnel the urine towards the renal pelvis.

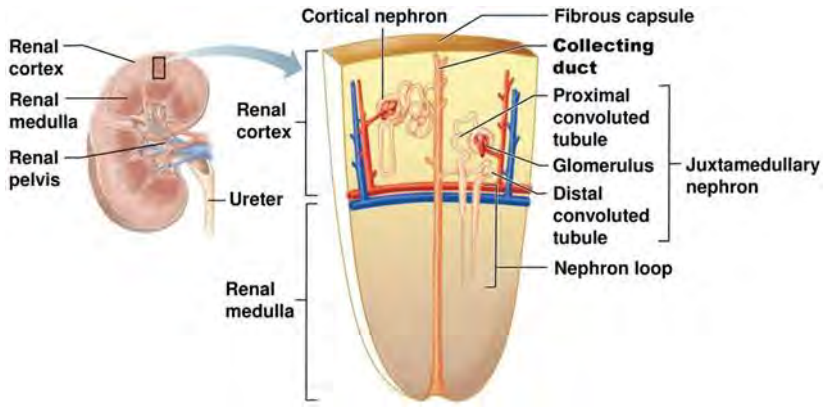

Figure 1.2: Structure of the nephron [9].

Within the renal cortex, and sometimes to the lower renal medulla, incorporate nephrons (shown in Figure 1.2). Nephrons are the main building blocks of the kidneys that form the urine. Nephrons that are located entirely in the renal cortex are termed cortical nephrons and are found to be 80 percent of one million nephrons within the

region. The rest 20 percent of the nephrons are found at the intersection of the renal cortex and medulla; they are termed as juxtamedullary nephrons.

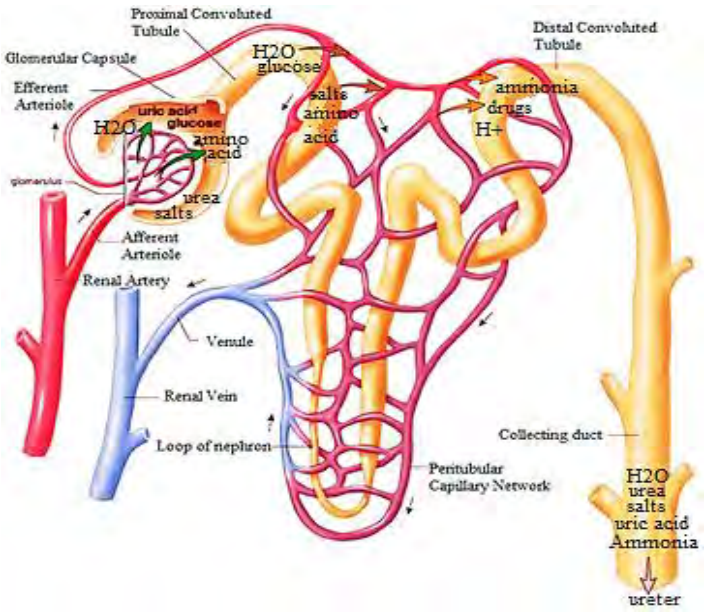

Figure 1.3: Urine formation in a nephron [10].

Urine is formed in the nephrons by the three processes: glomerulus filtration, tubular reabsorption, and tubular secretion. Glomerulus filtration is a passive, nonselective process, where the solutes (salts, nutrients, and waste material smaller than proteins) and water are pushed through the walls of the capillaries (forming filtrate), leaving blood cells and proteins as they cannot pass through the filtration membrane. This process takes place in the bowman's capsule. The filtrate is denoted by subtracting proteins and plasma proteins with blood plasma. Normally, the glomerular filtration rate (GFR) that is the amount of glomerular filtrate produced in both kidneys (in all nephrons) is 125 milliliters per minute (125ml/min) or 180 liters per day [11].

The next process is tubular reabsorption. As the name states, it reabsorbs amino acids, glucose, water, and ions from the filtrate and into the peritubular blood capillaries. Structures involved in tubular reabsorption are shown in Figure 1.3. In tubular secretion, specific molecules (excess hydrogen ions, potassium ions, and creatinine) are excreted into the DCT filtrate [12]. The remaining filtrate, known as urine, comes out into the ureter and contains toxic materials that must be eliminated from the body.

1.2.2. Diseases and treatment options. Kidney diseases are commonly caused as a result of organ strain by other chronic diseases, commonly high blood

pressure and/or diabetes. Within time, kidneys become inefficient to remove the waste products from the blood at a point where the patient needs an alternative treatment method [13]. Table 1.1 presents several types of kidney diseases along with their brief description.

Table 1.1: Types of kidney diseases [14].

Disease Type	Description
CKD	CKD results from high blood pressure or blood sugar and leads to the destruction of the glomerulus gradually due to intolerable pressure/sugar level in its vessels. This initiates kidney malfunction and progressively towards kidney failure when the body is laden with toxins.
PKD	Polycystic Kidney Disease relates to a serious genetic disorder where numerous sacs of fluid (cysts) are grown in kidneys. This hinders the functioning of the kidneys and triggers kidney failure.
Glomerulonephritis	It is an inflammatory disorder of the glomeruli (singular: glomerulus) that prevents appropriate glomerulus filtration.
UTI	Urinary Tract Infections are bacterial contaminations and usually start from the bladder or urethra that may lead to kidney failure if not treated.
Kidney Stones	A universal problem that occurs from crystallization in kidneys, forming stones of ranging sizes. Crystallization results from minerals or blood components coagulation. Small stones pass through urination, whereas massive stones cause an immense obstruction and lead to significant disorder.

Fortunately, kidney diseases can be treated, and with the technological enhances there are several treatment options where the patient himself may decide, or the doctor may recommend, the best option depending upon the kidney's situation. One of the treatment methods for kidney stones and infections, involve prescribing medication and drugs to the patient. Doctors also tend to alter the patient's lifestyle and prescribe a healthy diet routine to enhance kidney recovery and prevent them from deterioration.

However, the condition of the kidneys may worsen over a certain time, causing kidney failure. Failures occur when kidneys are hardly working, sometimes not working at all. With failures, the body is packed with waste products and water, the fluid homeostasis mechanism is impaired, and the kidney's endocrine functions are deranged, leading to various organ systems impairment. This situation may cause death if the condition of uremia (toxicity) is left untreated.

Chronic Renal (kidney) Failures (CRF) result when the kidney operates at less than 50 percent of normal kidney operating capacity, whereas End-Stage Renal Disease (ESRD) results from less than 10-15 percent of the normal functioning of the kidneys [15]. Therefore, the only options include either a complete kidney transplant or to retain under dialysis. Transplant is, however, not the best option as finding a kidney donor is not readily available, nor may it be suitable for young growing or elderly patients [16]. Another option may involve implantation of an artificial kidney, which is not certainly compatible as mimicking the organ and biocompatibility is tough and is not an appropriate treatment for elderly patients. Therefore, in such cases, the remaining option for kidney failure is undergoing a dialysis machine, which is a 'washing machine' that purifies the blood and returns it to the patient's body. The estimate of worldwide cure of ESRD patients that occurred in 2016 is shown in Figure 1.4 [17].

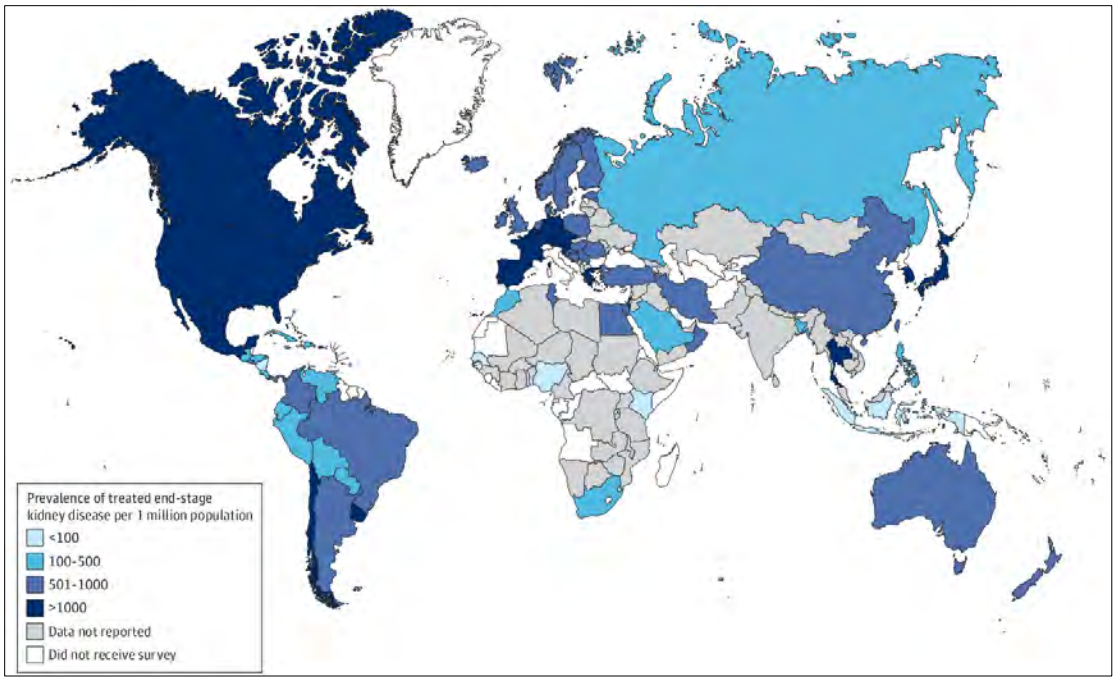

Figure 1.4: Worldwide ESRD cured patients in 2016 [17].

1.3. Dialysis

Dialysis is termed as the extra water and waste removal process from the blood. It resembles the artificial replacement to the functioning of the kidneys in cases of renal failure. Dialysis, however, cannot entirely perform the lost function of kidneys, though they still manage important tasks with ultrafiltration and diffusion. The dialysis is

certainly a treatment option for patients with CRF to excrete the toxins that are accumulated in the body [18].

According to S. Vadakedath et al. [18], dialysis may be a reason for the oxidative stress development due to disparity among toxins overproduction and the body's reduced defense mechanism. Oxidative stress disrupts abnormal cell functioning. The study in CRF [18] showed that there might be raised levels of plasma urate, further negotiating the body's defense mechanism and enhancing the oxidative stress. Overall, the indication for receiving dialysis treatment is simply the disturbed functioning of the kidneys, which includes hyperkalemia, uremic syndrome, extracellular expansion in volume, acidosis, hemorrhage diathesis, low creatinine clearance rate, or irresponsive medical treatment.

For dialysis treatment, a clinician may opt either of the two modes, hemodialysis (HD) or peritoneal dialysis (PD). Both modalities of dialysis provide patient treatment, and both have similar features that constitute to the purification of the patient's blood.

1.3.1. Peritoneal dialysis. More than 20 percent of the dialysis patient population uses this mode. It works on the exact rules of fluid ultrafiltration and solute diffusion compared to HD. However, the blood is filtered internally (within the body) rather than via an external machine, which is why the person may get their treatment while residing in their own houses [16]. With such conveniences, PD is generally advised to younger patients for its flexibility towards daily work routine.

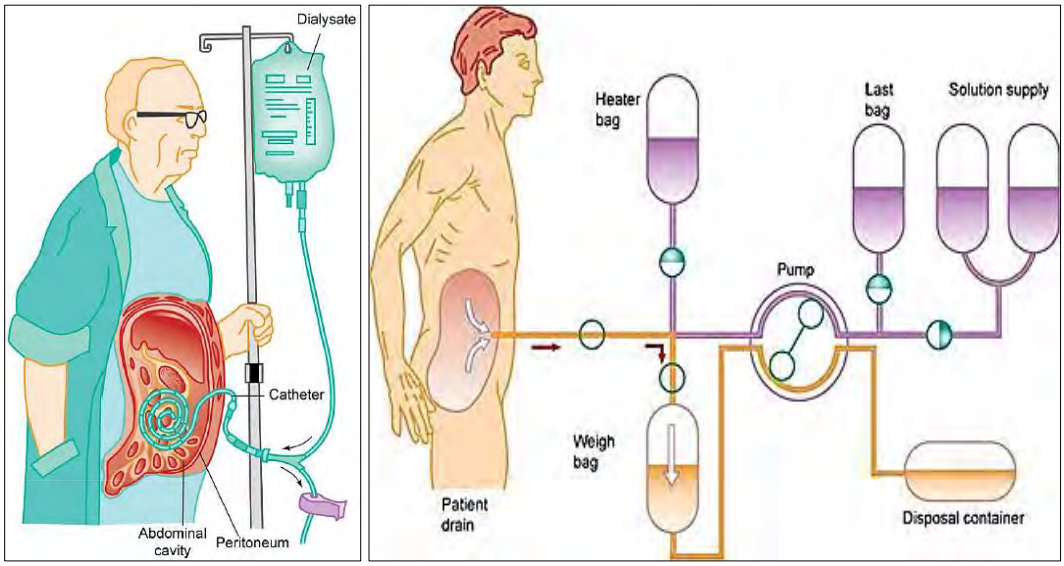

Figure 1.5: CAPD (left) and APD (right) [19].

During the PD process, as shown in Figure 1.5, a suitable dialysate is passed along the flexible tube and into the peritoneum through a few inches' incision. The contents of the dialysate enter the abdominal cavity and due to the semipermeable membrane of the organs, the blood and dialysate contents are exchanged gradually, where the blood removes waste to the dialysate. Later, the dialysate (in this case, filtrate) is flushed out of the abdominal cavity and is later disposed of.

PD, in its simplest method, is continuous ambulatory PD (CAPD), which is a manual process allowing patients who are trained to perform exchanges of the dialysis using a sterile practice. There has been significant attention given to the attachment technology to reduce the contamination risk of patients undergoing CAPD. Whereas automated PD (APD), another PD method and is also termed continuous cycling PD, brought numerous advantages permitting the overnight performance of several exchanges during sleep, likewise lessening the need for the daytime performance of filter bag exchanges. In addition to incrementing dialysis dose for certain patient groups and managing water removal difficulties, APD can be prescribed for social purposes with regards to locomotion during working hours, preventing dialysate bag exchange during daytime [20]. Some common advantages and disadvantages of PD are summarized in Table 1.2.

Table 1.2: Advantages and disadvantages of PD in general [21].

Advantages	**Disadvantages**
Patient independence and autonomy.	Occurrence of exit site infection and peritonitis.
Advantages to lifestyle, like traveling or working.	Caretaker or patient must be able to perform the technique.
Vascular access sites preservation.	Survival of the technique is a constrain.
Reduction to the blood-borne virus transmission.	Adequate dialysis becomes difficult to achieve due to the failure of residual renal function.
Improved conservation of residual renal function.	Fatigue in some patients due to exchange performances.
Less strictness towards dietary controls.	
After kidney transplantation, there is less chance of early graft dysfunction.	

1.3.2. Hemodialysis. HD is the popular dialysis mode, where 80 percent of dialysis patients use it. It is conducted within clinical settings and is delivered in a typical session of three times per week, with each session lasting three to four hours. In

this lengthy treatment, patients likely obtain hospitalization due to other infections or by the carelessness of the caregiver, which accounts for the HD patients being hospitalized in previous years [16].

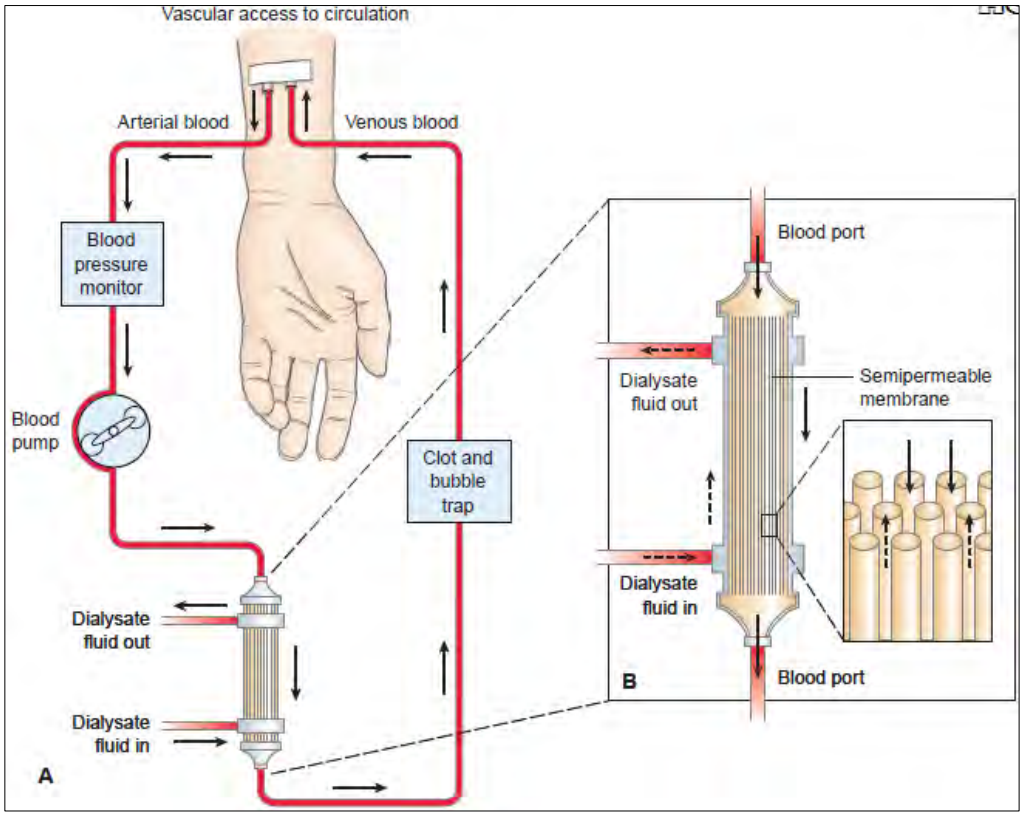

Figure 1.6: Circuit of hemodialysis [22].

In the HD process, as shown in Figure 1.6, blood flows across an extra-corporal route from the arteries, eliminating metabolites (waste material), re-establishing acid-base equilibrium, and removing excess water and salt from the incoming blood. Diffusion takes place across the semipermeable cellophane tubes (within the dialyzer unit), where the exchange of solutes and metabolites occurs, splitting the dialysate and the blood. A process using a negative pressure gradient (termed ultrafiltration) removes the water during the exchange. Transiting from the dialyzer is the filtered clean blood which is returned to the body into the veins. Regulation of the entire process takes place by the hemodialyzer [16].

In-center (within hospitals) conventional hemodialysis results in an interdialytic absence of 2 days per week, leading to increased patient morbidity and mortality. Hence, Intensive Home Hemodialysis (IHHD) gave a solution to the problem which led to better survival and improved solute clearance than in-center HD. With IHHD, treatment

is provided at least 4 times every week with a total of at least 20 hours, removing the interdialytic 2 days gap [23].

1.3.3. Parameters. The dialysis treatment is different from one patient to another because the dose depends upon several parameters. It is a challenging task as each dialysis patient differs from the other, with some patients requiring more medication, some lesser, making the clinicians' job very strenuous to fine-tune the dialysis dosing along with the monthly courses. It is necessary to continually alter the parameters to come to a cycle's sweet spot, where the patient receives adequate medication that retains the normal count of RBCs and other vital electrolytes. Disproportionate medication to the patient may lead to harmful side effects and wastage of expensive medicines. Tables 1.3 and 1.4 show some measured parameters during the dialysis sessions and their normal values for consideration.

Table 1.3: Dialysis essential clinical parameters [24][25].

Parameters	Normal Ranges
Systolic blood pressure (SBP) (pre-dialysis)	≤ 130 mmHg
Diastolic blood pressure (DBP) (pre-dialysis)	≤ 85 mmHg
Calcium	2.25-3.0 mEq/L
GFR	> 60 mL/min/1.73m^2
Hemoglobin (Hb)	Female: 10-14, male: 13-17
Kt/V urea	2.3; minimum: 2.1
spKt/V urea	1.4; minimum: 1.2
Magnesium	0.85-1.1 mmol/L
Parathyroid hormone (PTH)	1.3-6.8 pmol/L
Phosphorous	2.5-4.5 mg/dl
Potassium	3.6-5.2 mmol/L
Sodium	135–145 mEq/L (in dialysate)
Glucose	100-200 mg/dl
Bicarbonate	30-40 mEq/L
Serum ferritin	> 100 µg/L
Serum albumin	3.5-5 g/dl
Transferrin saturation (TSAT)	> 20 %
Weight (pre-dialysis)	Measured
Weight (post dialysis)	Measured
Interdialytic weight gain	Calculated
Urea reduction rate (URR)	≥ 0.65
Blood urea nitrogen (BUN)	7-20 mg/dL
Dialysate flow rate	≥ 500 mL/min
Blood flow rate	≥ 300 mL/min

Volume removal	Dependent
Serum creatinine	0.6-1.2 mg/dL
Serum uric acid	2.5-6.5 mg/dL
pH	7.35-7.45
Dialysis session time	2-4 hours
Urine volume	~ 120 ml/day
Total cholesterol	< 200 mg/dL

Table 1.4: Dialysis outcome and demographical parameters [24][39].

Demographical Parameters	**Outcome Parameters**
Age	Death date
Gender	Kidney transplant date
D.O.B. (date of birth)	Transfer to dialysis center date
Race	Transfer out of dialysis center date
Previous diseases, health conditions	Diagnosis codes (primary/secondary)

From Table 1.3, blood pressure is measured each time before and after the dialysis treatment. The amounts of sodium, glucose, calcium, potassium, phosphorous, and bicarbonates administered in the dialysate are recorded and monitored. Hb and nPNA are measured from the patient to indicate any problems with minerals, cells, bone health status, or anemia. Calcium concentration in the dialysate of 2.5 mEq/L is constituted to a level of 10 mg/dL serum calcium [26].

GFR standard precision is measured using urinary clearance or plasma of an exogenous marker of filtration. However, this procedure is not routinely performed and is complex, so GFR is estimated (termed eGFR) using cystatin C level or serum creatinine with combining demographic elements like age, gender and race using equation 1.

$$\text{eGFR} = 141 \times \min(\text{Scr}/k,1)^{\alpha} \times \max(\text{Scr}/k,1)^{-1.029} \times 0.993^{\text{Age}} \times 1.018 \text{ [if female]} \times 1.159 \text{ [if male]} \quad (1)$$

where α is -0.329 for female and -0.411 for male, k is 0.7 for female and 0.9 for male, min is minimum of Scr/k or 1 and max is the maximum of Scr/k or 1, and Scr is the serum creatinine (mg/dL) [27].

Another measure, Kt/V is the urea clearance rate or volume of urea cleared during dialysis. K is urea clearance in ml/min, t is the time in min which is multiplied

and is normalized to urea distribution volume V in ml, resulting in the unit-less parameter also called as Kt/Vurea. It is difficult to measure K and V from a person's body, and consideration of urea removal is not mainly accountable for urea generation through dialysis or during its procedure with fluid removed (altering V). Various equations have been generated for calculating Kt/V using regression analysis, where some limitations are addressed. The major limitation is the assumption of enough urea clearance along with dialysis that associates with enough entire removal of uremic toxins. This is considered a false case, as the highly grouped and protein-bound solutes are inefficiently removed, even though the removal of urea may be adequate [28].

A single pool or compartmentalized assumption for urea is termed spKt/V. It is calculated using levels of urea before and instantaneously after initiation of dialysis, as shown in equation 2.

$$\text{spKt/Vurea} = -\ln (R - 0.008 \times T) + (4 - 3.5 \times R) \times 0.55 \times \text{Weight loss/V} \quad (2)$$

where R is pre-dialysis/post-dialysis urea, T is the dialysis treatment duration in hours, Weight loss is calculated by equation 3, V is the total volume of body water (computed using the equation of Watson), and *ln* is the natural logarithm [29].

Normally the dialysate temperature is maintained at 37ºC. A lower temperature declared an association with low intradialytic hypotension risk, and a dialysate of lower than 0.5ºC lowers injurious ischemic brain. To improve hypotension, the use of the cool dialysate between 35.6-36.0 ºC should balance the patient's irritation and increased risk of dialyzer clotting incidences [28]. Moreover, a balance between levels of potassium must be maintained to prevent harm from extremely low level (hypokalemia) or high level (hyperkalemia).

URR is another calculation to define how well working dialysis treatments are in clearing blood wastes. It is the urea fraction reduction within a single dialysis therapy. It involves a blood test rather than urine collection, which is calculated using equation 3 [29].

$$\text{URR} = 1 - (\text{Post-dialysis BUN/Pre-dialysis BUN}) \quad (3)$$

Moreover, maintaining body weight is essential for the overall health of patients undergoing dialysis. A target weight, else dry weight, is measured and defines the

patient's weight after the removal of the body's excess weight by dialysis. Furthermore, interdialytic weight gain is the gained weight during dialysis treatment every day. Likewise, excessive fluid weight is gained by unfollowing salt and fluid limits within the treatments [29].

1.4. Associative Comorbidities

With improper dialysis prescriptions, kidney disease patients may further evolve to other problems such as anemia, which is the lack of red blood cells in the body. According to Barbeiri et al. [30], anemia is one of the foremost disorders encountered in ESRD (or uremic) patients. ESA and iron supplements are the preferred prescriptions to treat them. However, an adequate dose is still a concern as the dosage is primarily based on average patient behavior. Also, dosage upon the patient individuality and anemia stage variability are not deemed precisely. Hence, an AI model was developed for the precise estimation of iron and ESA therapy reaction.

Moreover, Toft et al. [31] raised concerns about the safety of cardiovascular state while handling anemia by delivering more hemoglobin (Hb) levels, particularly when ESAs are used. Such studies resulted in the practices of anemia management since 2011 to lower the targeted Hb and intensive treatment. Along with the management shift, there is a lack of perfect longitudinal real data on the anemia impact on clinical outcomes. Also, cardiovascular risks and other problematic patient comorbidities that subsequently develop to more intense anemia stay undetermined.

Barbeiri et al. [32] stated that ESRD crucial points involve BP and fluid volume management. Clinicians may control the BP by lowering the overload volume of extracellular fluid, which is done by terminating the targeted weight of post-dialysis. Therefore, the study approach exposed dialysis patients to episodes of intradialytic hypotensive and reducing BP.

Furthermore, with regards to hemodialysis patients, the dosage directed for the survival of the patient is controversial. Hence, Argyropoulos et al. [33] experimented with the hypothesis where survival analysis methods may impact inferences regarding mortality and dialysis dosage. It is crucial to connect the result of survival techniques to the dialysis dosing to understand the importance of survival methods to the effects

of ESRD. The results were found stable with the use of an analytical model for the survival of hemodialysis patients.

1.5. Big Data in Healthcare

Big data in healthcare refers to the big sets of data within EHR, having multiple information regarding products (mostly patients in healthcare systems). Also, healthcare sectors are swimming in an escalating sea of data that is either very unstructured or very voluminous to be analyzed and managed through conventional means [34]. Therefore, the task of healthcare is how to collect, find, control, and evaluate the information to create easier and healthier lives by not simply aiming to comprehend new therapies and diseases, but also to expect earlier stage outcomes and apply real-time judgments [35].

1.5.1. Data analytics. Data analytics in healthcare is described as a method of data transformation into events involving analysis and awareness from the perspective of solving problems and decision making in an organization. Several terms are interchangeably used that are often mis-conceptualized, including data analysis, data analytics, and data mining. Data analysis is a data mining superset that involves the method of cleaning, investigating, converting, and preparation of data in order to find some useful content, derive conclusions, and make current decisions. Data mining is a sequential and systematic process of discovering and identifying hidden information and patterns in a large dataset. Whereas data analytics is exploiting data, statistical analysis, machine learning, and computer models to get a clearer understanding of the problem for making proper future predictions from the data [36].

1.5.2. Big data analytics. Big data analytics corresponds to advanced logical methods that run on big sets of data. Therefore, big data analytics considers two factors: analytics and big data, along with how these factors have been collaborated to establish one of the most insightful trends [37]. The elevating big data analytics field has started to perform a crucial role in the advancement of healthcare research and practices. It has provided means to collect, manage, evaluate, and integrate huge volumes of structured, distinct, and unstructured data generated by contemporary healthcare systems [38]. Recently, big data analytics has been applied to study and predict outcomes for the kidney related diseases (mainly CKD, ESRD, or anemia) in dialysis patients. However, while making use of big data analytics, a well-defined and prescribed dialysis dosing

level is still not rendered. In this paper, we consider this as a challenge to overcome precise dialysis dosing levels with multiple parameters of kidney disease patients.

1.6. Problem Statement

Kidney disease patients that are upon dialysis are increasing worldwide and their cure mainly relies on technology, which accounts for their need of proper dialysis dosing levels to ensure patients are receiving the right dialysis treatment at the right time. Shah et al. [39] state that reducing the dialysis costs along with greater improvement towards a patient outcome is a demanding task, as dialysis care is multifaceted and patient vitality depends on multiple factors. In dialysis patients, blood electrolytes are the most important parameters considered from the medical point of view and must be controlled. Also, understanding the electrolyte parameters and predicting their outcome for each patient to deliver the optimal dialysis dosing is a challenge. Therefore, it is essential to have a comprehensive understanding of the parameters that correlate to the electrolytes before determining and prescribing improved dialysis dosing to the patients.

To understand patients' electrolytes, big data analytics tool can be used to provide some insights and interpret the significance of trends. Furthermore, data prediction models can be used to state the most significant attribute in the prediction of electrolytes, which aid as a final decision-making tool to prescribe and improve the dialysis dosing for patients. This study will help analyze and predict patients' electrolytes to provide insights to improve dialysis dosing levels for the patients, which will have a positive impact towards patients' quality of life and welfare, hospitals, and the entire healthcare.

1.7. Thesis Objectives

The objective of this research is to build a clinical decision-making support tool to analyze and predict certain dialysis electrolytes, which will determine and improve patients' dialysis dosing levels. To achieve our goal, we plan to:

1) Utilize data analytics tools for EDA, by which sifting through the dataset records will provide useful insights and certain electrolyte trends from dialysis patients.

2) Build predictive models to predict patients' electrolytes, which will identify the most significant predictor parameter from the outcomes.

3) Interpret analysis and prediction results to allow clinicians to determine factors and improve patients' dialysis dosing levels. This will provide the dialysis patients the right dialysis dosing.

4) Improve the patient's quality of life, welfare, life expectancy, and prevent further kidney disease comorbidities from the rising volume of kidney disease patients.

1.8. Research Contribution

The contribution of this research is building an AI/ML-based predictive model that will provide a clinical decision-making support system based on the electrolyte parameters of kidney disease patients under dialysis. This work is further to be implemented in healthcare facilities to generate a better understanding of patient's profile which will aid the clinicians in prescribing the right dialysis dosing at the right time. Big data on multiple parameters from multiple patients will be collected and then analyzed. The use of a data analytics tool from EDA software will facilitate the task by analyzing the electrolyte parameters, which were collected from the local source. After EDA, further predicting the electrolytes will allow clinicians to determine factors to prescribe improved dialysis dosing and prevent further disorders that evolve from kidney disease. The impact of proposed approach will majorly be reflected upon cost, time, efforts, improved quality of life, welfare, and life expectancy.

1.9. Thesis Organization

The rest of this proposal is arranged as follows: Chapter 2 reviews data analytics in healthcare and the renal system, followed by a discussion of big data in dialysis modeling as well as dialysis estimation. Chapter 3 describes the methodology of the proposed approach, along with the prediction techniques used in the process. Chapter 4 presents implemented work from the proposed model, as well as execution of the performance evaluation. Chapter 5 comprises the following results from implementation, which are analyzed and further interpreted from prediction output. Finally, Chapter 6 concludes the thesis proposal and discusses the future work to be done.

Chapter 2. Literature Review

In this chapter, we will discuss the reviews on data analytics in healthcare in general and specifically for the renal system. Furthermore, we present the recent literature of big data in dialysis modeling, along with mentioning current related studies of dialysis estimation.

2.1. Data Analytics in Healthcare

A variety of challenges faced by the healthcare sector can be addressed through data science. Different methods of big data analytics have widely been utilized for various problems and applications, such as diabetic retinopathy detection and data-exploit models for skin cancer classification. Also, machine learning has been applied in asthma, cardiovascular medicine, and other medical specialties [40]. According to M. Hueso et al. [41], the result of having several technological devices that generate the data around the patient's point of care is an advantage. This is due to the data being in a continuous electronic format that is the primary input requirement for the data analytics tool. The data analytics tool will then convert the data into ultimate knowledge which is helpful for dosing prediction and patient therapy. Therefore, AI/ML field is currently providing promising results from its tools for data analytics in healthcare [41].

Currently, the healthcare sector did not completely recognize the potential benefits that can be achieved from big data analytics. A clearer concluding knowledge from the outcomes of big data is instantly required, as there is a continuous development of the intellectual study on big data analytics, using technology. To further examine the absence of knowledge, a study by Wang et al. [42] analyzes architectural design, historical development, and component purpose of big data analytics. From the study implementation of using 26 cases of big data in healthcare, the results were able to distinguish five benefits of big data analytics, including decision support, unstructured data analytics, predictive, care patterns, and traceability capability. Therefore, the resulting benefits of big data analytics were further organized into five classifications, including: operational, strategic, organizational, IT infrastructure, and managerial benefits. However, the study results indicated that it is still an early stage for the presence of big data analytics in healthcare due to the limited benefits in the strategic, managerial, and organizational levels [42].

Along with the presence of big data in healthcare, Boukenze et al. [43] stated that various attempts were made to manage the surge of medical data, and to attain a precise knowledge from it. This encouraged researchers to utilize predictive analytics, big data analytics, deep algorithms, and machine learning (ML), for the gain of precise knowledge that will aid in the decision-making process. With promising results of big data analytics and ML algorithms, it is no longer a challenging task to predict the future, especially in healthcare medicine, as disease prediction and cure anticipation are simply possible. Furthermore, the terms 'predictive analytics' and big data analytics show impressive growth in google trends since 2011. This growth is due to the uninterrupted process of medical data analytics utilizing not only the large database management level, but surpassing this level in retrieving future knowledge that is required by many experts and researchers [43].

2.2. Data Analytics in The Renal System

Utilizing data analytics in the renal system related problems mostly includes the use of patient datasets for the diagnosis and early detection of various renal-specific conditions. These involve CKDs, renal failures, kidney stones, diabetic kidney diseases, and anemia treatment management. Diagnosis and management of renal failure or CKD is often complex due to poor health management, lack of knowledge associating with extreme risk, and improper lifestyle. With time, CKD intensifies the damage to the kidneys, which can be identified by the glomerulus filtration rate (GFR) value. Therefore, GFR can be prevented or treated if identified and diagnosed earlier, otherwise for further mishaps organ transplantation or dialysis therapy may be required.

Since CKD is complex to identify and predict, appropriate medical tests are considered necessary in which numerous attributes from the test results can disclose valuable CKD information. With greater details, the importance of these attributes should be analyzed when applying them to deliver precise cure or treatment. A study by Banarjee et al. [44] suggested classification models that utilized permutation feature, GreedyStepwise search methods, attribute inspector correlation-based feature subset, and various classifying algorithms for the data analytics of non-CKD and CKD patients. The results of the study classified and categorized the patient outcome as CKD and non-CKD with better accuracy than the conventional prediction methods [44].

Furthermore, the study of Boukenze et al. [43] also applied data analytics using a learning algorithm to the CKD medical dataset. The study was aimed for the prediction of CKD patients by using a decision tree algorithm. Big data of CKD patients, which included 400 cases with 24 attributes of two classes (non-CKD and CKD), were used from the UCI repository for machine learning. In conclusion to the study experiment, the C4.5 algorithm that was applied had performed much enhanced prediction of CKD patients, with optimized results and performance towards least execution time and greater accuracy.

In addition to the data analytics in the renal system for CKD patients, P. Sinha et al. [45] likewise implemented a predictive model using a dataset that was sourced from several clinics, hospitals, and medical labs. A testing dataset of kidney function for the kidney disease study was developed, which involved 400 cases and 25 attributes. The study employed three various ML algorithms with their comparative performance measured. From the results of testing, Bayesian algorithm showed better performance than the k-nearest neighbor and the support vector machine, and further improved the rate of the CKD patient prediction. Moreover, the ML algorithm performed efficiently for the prediction of other various diseases [45].

Another research study for the data analytics of CKD patients in renal system involved the use of a support vector machine algorithm. This study, by Amirgaliyev et al. [46], is based on data analytics of CKD datasets that had historical records, laboratory tests, and physical examinations. The results of the study experimentation then showed a success rate of over 93 percent in the prediction of kidney disease patients, with performance metrics in sensitivity, accuracy, and specificity. Precisely, the performance rate of the proper algorithm model displayed 94.602 percent sensitivity, with the value of sensitivity compared with a linear kernel of 93.1 percent. The study also demonstrated the importance on the prediction of diseases at an early stage and on promoting patient's welfare.

Moreover, Zixian et al. [47] studied the prediction of CKD patients using CKD dataset from data warehouse of UCI machine learning. Detection of CKD was done by a technique of Apriori association for 400 patients, with a validation testing of a 10-fold cross. The results of the study were compared across several ML algorithms, including OneR, J48, ZeroR, K-nearest neighbor, and naïve Bayes. Following the steps,

the initial preprocessing of the dataset was done with the completion and normalization of the missing data values. Next, specific data attributes were selected from the data analytics approach for the reduction in training time and to improve accuracy. Study results had indicated accurate detection of CKD patients of 99 percent using the Apriori ML algorithm. The study technique was later tested and performed using data samples for more patients for the prediction of other CKD patients [47].

Furthermore, a study research by Wickramasinghe et al. [48] was targeted to identify the appropriate CKD patient diet plan by applying ML algorithm on the data analytics test results. The main aim of the work was to control the disease using an appropriate diet plan, which was identified by ML algorithms. The proposed work dealt with several diet plan recommendations, also estimating the potassium zone for CKD patients according to their levels of blood potassium. Different algorithm experiments were performed, including multi-class decision forest, decision jungle, logistic regression, and neural network. The study results determined that a better result was achieved by the decision forest algorithm in comparison to the rest of the algorithms, obtaining an accuracy of 99.17 percent.

Maurya et al. [49] proposed an automated tool, where machine learning methods were used for the determination of patients with improper kidney conditions, to aid the doctors for CKD prediction and provide better treatment. The proposed method used data analytics that extracted main features for CKD data, which determined CKD patient severity using a ML algorithm. With that, the objective of the study illustrated the use of a ML algorithm for the patient's suitable diet plan using a data analytics tool. From the study, the patient's diet recommendation was prescribed by estimating the potassium zone, which was calculated from the level of blood potassium to lower the CKD progression.

Additionally, Shankar et al. [50] study proposed an approach of a learning tool and an inspired optimization data analytics model to predict CKD. The approach selected the applicable features of kidney data by Ant Lion Optimization technique that decides ideal features for the classification method. Later, data analytics using CKD dataset was done by deep neural networks ML algorithm. A comparison of the model performance in the study results indicated a better prediction, accuracy, sensitivity, and F-measures with respect to the other algorithms.

In addition to CKD prediction, as mentioned previously, data analytics have also been used for the prediction of Hb, to guide the ESA dose selection for the result of optimized anemia management. Prediction models were used and experimented on the individual patient data, which supported to optimize prediction to the numbers of Hb observations within the target range. One such model, proposed by Brier et al. [51], displayed an optimized outcome that decreased the use of ESA and the amount of transfusions to the patient. The study results also confirmed, when compared with earlier smaller studies, a greater achievement for further beneficial results.

Apart from Hb, CKD anemia is one of the major comorbidities in ESRD patients. ESA and iron supplements have become the cure choice for anemia. To identify an adequate treatment or cure is very complicated for each patient in every situation as the guidelines for dosage are in terms of average responses, so the particular response by different patients are not considered when delivering the dosage. Some drug responses may differ extremely for a different patient, or with the same patient in different anemia stages. A study by Gatti et al. [30] proposes an advancement to the previous works, confronting CKD anemia problem, by applying distinct machine learning (ML) methodologies.

The study by Gatti et al. [29] presented a consistent ML model for the prediction of Hb values of secondary stage anemia in CKD patients. This study is a result of the author's long experience of the problem and previous works that created unsatisfactory models. The newer approach to the ML model added the previous potential models (ML) to produce a precise model that exploits datasets with better information to predict RBCs lifespan before the effects of ESA and iron are displayed. The study resulted in an improvised data analytics method that surpassed any previous ML models that had encountered the same problem.

One such major challenge in providing proper patient treatment is an accurate and rapid diagnosis of the disease, which is why the healthcare uses various machine learning techniques. A study by Vinayagam et al. [52] uses 20 test data for the data analytics of having abnormal and normal kidney predictions. Magnetic resonance imaging produced patient kidney illustrations, which were then classified by the back propagation method of neural network (BPNN) technique. The BPNN output was displayed using a liquid crystal display, which was linked to the Arduino board. Fuzzy

clustering mean (FCM) algorithm was used for the accurate prediction of the stones in the kidney using the patient dataset. Comparing with various filters in the software, proper stages of kidney stones were distinguished. The study results stated accurate detection of kidney stones and proper prediction of their stages accordingly.

Using former natural language and longitudinal data, which were based upon EMR of 64,059 diabetes patients, a study of Makino et al. [4] generated an advance predictive model for diabetic kidney disease (DKD) patients. The study applied three novel methods to optimize the predictive ability for certain disease complications. First, the new predictive model was constructed for diabetic complications prior to patient symptoms or any clinical signs. Second, AI/ML algorithm was used for big data analytics without any clinical research objective. Third, time-series data was used by the artificial intelligence tool to provide predictions.

From the previous six months, AI extracted raw features as a reference period and 24 factors were selected to identify patterns of time series. This provided a linkage to the DKD aggravation of six-month by applying a convolutional auto-encoder. The applied predictive model was constructed by AI with 3,073 attributes that involved time series data formed by logistic regression analysis. The results of the model displayed accurate prediction by 71 percent. Furthermore, DKD aggravation group had shown higher hemodialysis incidence than others, for over ten years. The proposed model could also identify DKD progression and may provide more accurate and effective mediation for the reduction of hemodialysis [4].

2.3. Big Data in Dialysis Modeling

Several researched presented work in the literature on utilizing big data for dialysis care. Erickson et al. [53] stated that in 1970s or 1980s, Medicare datasets were used to detect major trends in access to the dialysis care, costs, and value of care delivered. An increase in the access to Medicare's organizational data for an investigation was made by the policymakers and research communities. Efforts were also made to address the skyrocketing expenses of injectable drugs, in 1990 and 2000, which demonstrated to an increased role of big data. While there are non-governmental and enormous governmental administrative datasets, it is crucial to understand the limitations and challenges that correlate with their use [53].

According to Dr. Bowman [13], dialysis patients endure high healthcare costs, hospitalization rates, and mortality rates than various chronic disease patients. This led to the study of improving the treatment outcome as well as cutting the treatment costs. About 650,000 ESRD disease patients undergo prolonged dialysis treatment in America. Each change of the dialysis dose requires two to three months approximately to influence the patient, although the clinicians make monthly dosage decisions. With the dosing impact, there is a tendency of change of conditions to the hemoglobin cycling, which is the alternation of RBC count values, going to either too low or too high. In too low RBC count, the patients experience symptoms of anemia, and being in a too high count leads to the drug side effects.

Adjustments and control of the problem need to be done, which is the standard engineering approach, where the historical data is used to notify predictions for the future condition to take effect. Hence Bowman's [13] study included using big data with dialysis dosing records of about 3,000 patients for over a period of care. The study aimed to develop a model that helped clinicians in the accurate prediction and control of patient's RBC counts, from huge data samples of over previous years. The big data used may be able to identify dosing levels within a large scale of conditions, that is a challenging task and would require a technique for the data analytics, which would be applicable to predict treatment outcomes for a variety of diseases [14].

Furthermore, Barbieri et al. [32] study exploited big data by developing patient and session-specific ANN model. The model was used for the prediction of heart rate and SBP (systolic BP) profiles, Kt/V, and post-dialysis body weight for each session of hemodialysis. The ANN model comprised about 60 attributes that represented characteristics of patients, physiological response from historical records, previous dialysis sessions results, data of pre-dialysis, and index session that prescribed dialysis dose. The modeling big data comprised of dataset of 760,000 patients, each demonstrating a recorded dialysis session from Spanish NephroCare center, from years 2015 until 2018. 20 percent of the records were used to test the finalized model, and the first 80 percent were used to verify the preliminary model. The study trial showed the dialysis patients with regards to intradialytic hypotensive episodes. Such modeling tool was used to anticipate the reaction of the patients through simulation, where the

appropriate strategy can be selected accordingly to qualified utility purposes or clinical judgment.

	Mean	SD	Model ME	Model MAE
Minimum SBP of the session, mm Hg	113.0	23.5	−0.16	9.3
Postdialysis heart rate, bpm	71.1	13.8	0.04	7.3
Postdialysis weight, kg	70.1	15.5	0.00	0.23
Kt/V	1.89	0.43	0.00	0.13

SD, standard deviation; ME, mean error; MAE, mean absolute error.

Figure 2.1: Observed preliminary results [32].

Preliminary results in Figure 2.1 showed that at modeling of the lowest heart rate and SBP tends to be difficult and slightly inaccurate, while post-dialysis weight prediction and Kt/V were error-free. The inaccuracy may be due to the errors of measurement by the integrated sensors of the machine, and the naturally inconsistent alterations of the hemodynamic reaction. Therefore, the study model resulted to be sufficiently precise and accurate that predicted the risk from intradialytic hypotensive before each dialysis session [32].

Estimation and proper prescription of dialysis dosing are necessary for the patients undergoing dialysis. This is because other factors are necessary to be considered when prescribing dialysis, including adequate electrolyte levels of sodium, potassium, chlorine, and calcium. Adverse effects may result if the electrolyte levels are imbalanced in the body; such cases involve high blood pressure (hypertension), weakening of the bones, dehydration, high risk of diabetes, and so on.

According to Davenport [54], increasing dialysis dosing will not be beneficial to every patient. Dialysis patients having lower than 2 mL/min residual renal urea clearance have a large gain of weight, or for the increased dialysis approaches, there should be no consideration to cardiovascular reserve. Also, treating the patients individually in prescribing dialysis is the most important consideration as one prescription point does not suit everybody, so the patients require distinct targets.

Dialysis estimation issues may also include potential biasing that are formed with the knowledge of data. Similarly, data that are derived from observational studies, as compared to the randomized studies, may also tell a different story during the control trials. Erickson declared that several dialysis dosing studies, an example of HEMO trial, have been seen which indicated that providing high dialysis dose does benefit in some

patients, but not others. The HEMO trial did not display any improved survival rate, or lower morbidity when the high-flux membrane was used. This proved so far that the large dataset results are reliable with the data; however, not all problems can be addressed by dataset size [55].

In measuring multi-dimensional parameters for dialysis adequacy, removal of urea is the crucial measure in the increase of dialysis treatment for kidney failures. Kt/V urea, i.e. the small solute removal, is widely used to quantify hemodialysis dose of three times per week. However, relying on only small solute clearance value for the measure of dialysis adequacy fails completely to achieve better clinical effects with dialysis therapy. In this case, Perl et al. [56] study aimed to present the potential wide construct for the better dialysis dosing, which considered broad goals of ESRD patient care to enhance survival of the patient, the quality of life, and to understand the strength and limitations of the small solute kinetics, which would indicate the replacement of a proper dialysis dose (as shown in Figure 2.2).

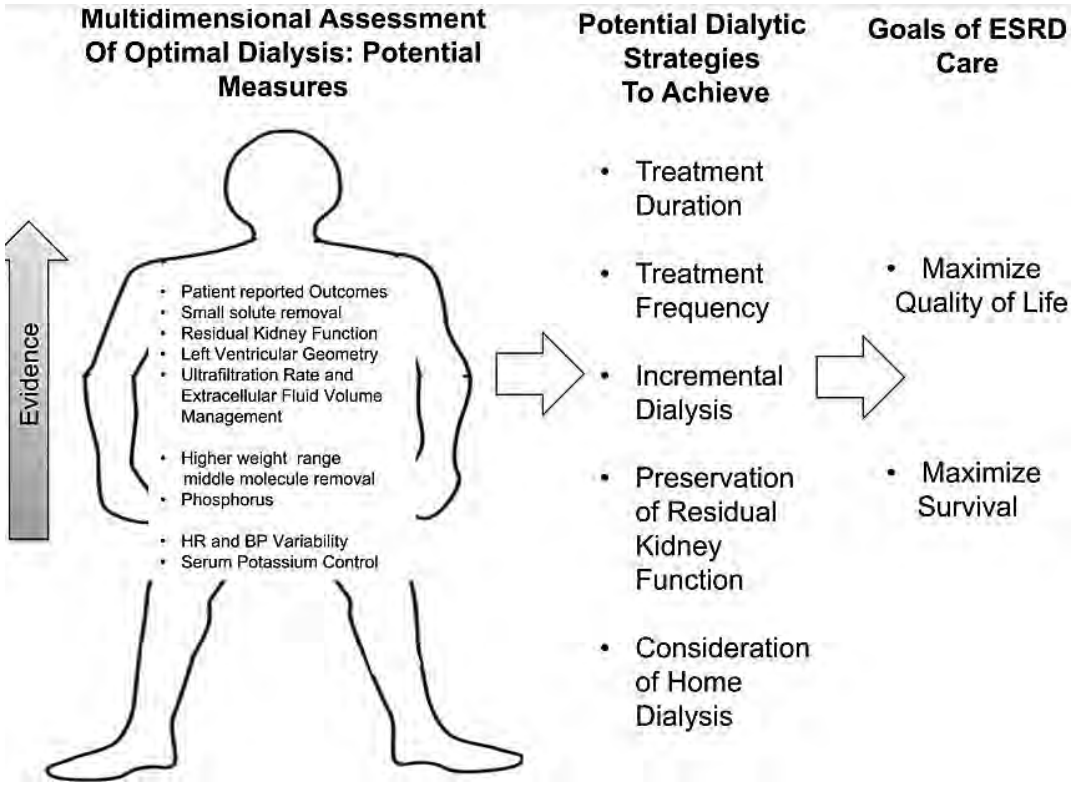

Figure 2.2: Multi-dimensional measures for optimal dialysis [56].

The study concludes that the time of dialysis (per week sessions), including both the therapy duration and the frequency-time, with the basic biochemical indices signify

a single important indicator, but not an adequate dialysis dosing. Parameters of clinical physiology, including heart rate, BP, cardiac function, and nutrition, signify the measures for the proper outcome to quantify the dialysis therapy results. Furthermore, Perl et al. [56] acknowledged the need for further research to validate the application of multi-dimensional measures, which would compute adequate dialysis dosing level for patient treatment.

Chapter 3. Methodology

This section outlines the proposed approach for analyzing and predicting patients' electrolytes, to conclude insights and provide improved patients' dialysis dosing levels. The research plan follows in collecting patient's datasets, along with utilizing data analytics tools and building AI/ML models to predict patients' electrolytes, as shown in Figure 3.1. Moreover, the proposed AI/ML prediction model types for this research are presented, with a brief description of each. The methodology steps are as follows:

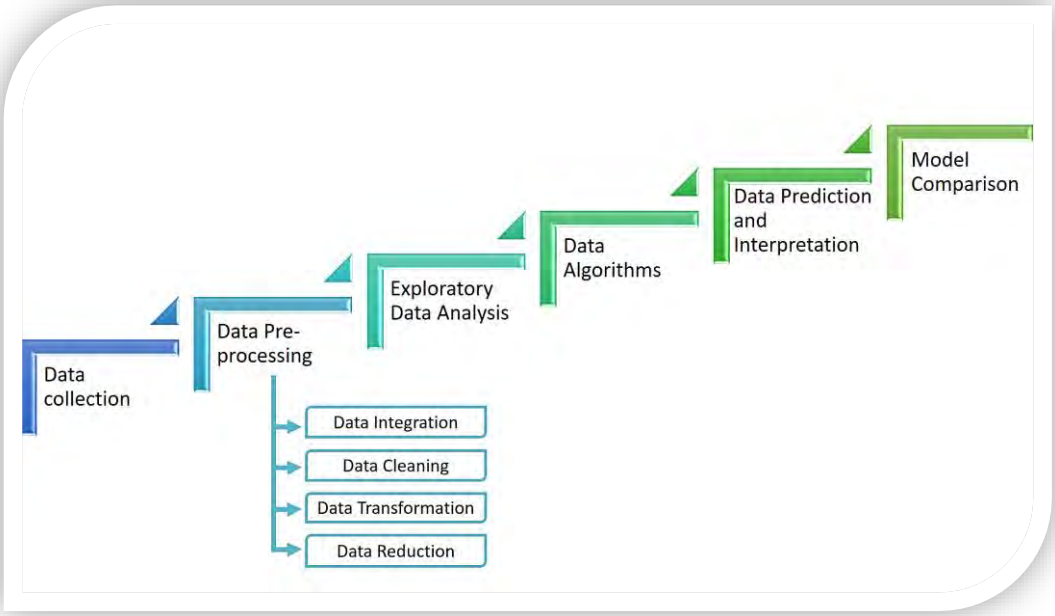

Figure 3.1: Steps of methodology.

3.1. Data Collection

The first step towards building a precise dialysis dosing approach is the data collection of pre-recorded datasets. Dataset collection is the foremost step as lack of appropriate patients in the study may make the prediction processes unreliable and inaccurate. Therefore, this study involved 45 dialysis patients into consideration, which would be suitable for data analytics and would predict precise dialysis dosing for each patient.

3.2. Data Pre-processing

The second and most important step after retrieving the raw data is to convert it into a format that can be used in further processing and within data prediction models. There are four sub-stages into this step, including data integration, data cleaning, data transformation, and data reduction of the noisy and raw dataset. These sub-stages are further explained below.

3.2.1. Data integration. Since the data is collected from several dialysis patients, data integration is the necessary stage in pre-processing. It allows prediction models to run efficiently, providing low redundancy and time consumption with consistent data, and good accuracy amongst results. Data consolidation is to be considered, where the data from several patients are merged to be used as a single big dataset in the model processes.

3.2.2. Data cleaning. The term 'cleaning' basically involves any row (or cell) amendments to be made for incomplete data or noise from the dataset. After data integration, amendments can be done by inserting values (from the average value generated per patient) to some of the incomplete data values. Further cleaning of the dataset involves replacing errors manually into a defined value for some of the noises the dataset has. This may either be done by generating attributes to define precise values (e.g., binominal class values), by generating values based upon defined attribute calculations, or by simply replacing the value into a known logical form to reduce further noise.

3.2.3. Data transformation. This stage goes hand-to-hand with the data cleaning stage. For particular attributes with patients having some missing values (not all), the data average is taken per patient. The average values are generated as new attributes per patient, which are combined with the original dataset and used in the cleaning stage to impute missing values upon the central tendency of the mean per patient. Also, normalizing data to a small range is done by dividing the dataset into training and testing parts, to be utilized in prediction models. The data is divided into two sets; 70 percent is a training dataset, whereas the remaining 30 percent is used as a testing dataset of the final prediction model.

3.2.4. Data reduction. At this stage, if the patient values are entirely missing for all the months for a particular attribute, then deletion of the entire rows of missing values is done from the dataset. This stage is executed as data prediction algorithms do not run with missing values. To a note, the missing values were not imputed by any other means as it would clutter up the dataset and will not sustain data integrity. Moreover, only certain attributes are considered during the prediction model's process, including demographics, patient dialysis factors that are measured pre-dialysis sessions (including pre-urea 'BUN' and HBA1C).

3.3. Exploratory Data Analysis

This step utilizes data acknowledge and visualization tools to understand the relationship between certain important patient attributes. Exploratory data analysis (EDA) is related to inferential type statistics, which includes visualizing and describing datasets from various perspectives and briefing them. This allows to drill down to information and gain initiatives for prediction enhancements [57].

EDA was implemented in this study using scatterplots and boxplots to visualize vital blood electrolytes per patient and decipher their trends in the dataset. Also, a time-lapse view of those electrolytes was taken for further realization upon dialysis sessions on a monthly basis. Besides, with many patients taken into consideration upon data analysis is a difficult task and requires compatible visualization tools (software), which are later mentioned and briefly described.

3.4. Data Algorithms

In this step, prediction models are then created to predict some electrolytes from the big data for the analysis of precise dialysis dosing. The information obtained from the exploratory data analysis step will be taken accordingly as per the patient policies and will be added to the building algorithm model. The trained prediction model is then added with the final testing dataset in performing predictions to generate the predicted output.

Distinct machine learning (ML) algorithms, which are the prediction techniques, are utilized to predict patient electrolytes from selected attributes. The preferred algorithms, as shown in Figure 3.2, were Decision Tree (DT), Neural Network (NN), Linear Regression (LR), and Support Vector Machine (SVM). All these techniques

were selected with regards to their popularity in AI data science field, study accuracies, and their ease of use.

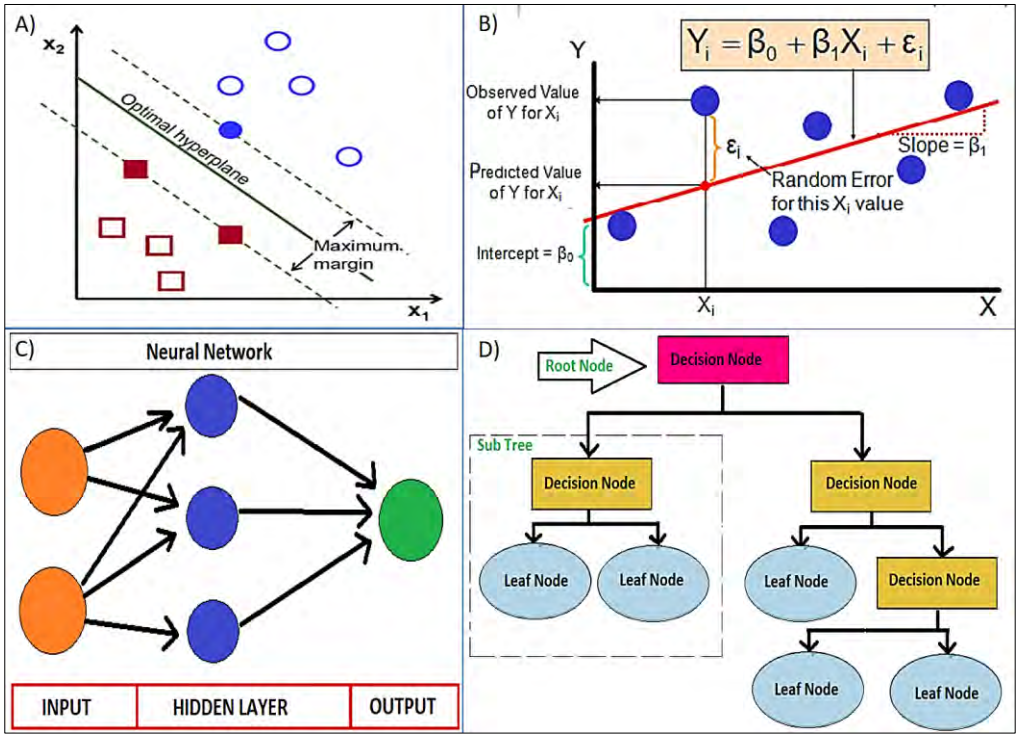

Figure 3.2: Illustration of ML algorithms; A) SVM [58], B) LR [59], C) NN, and D) DT.

A DT is a graphical illustration predictive technique that utilizes the method of branching. Nodes at the internal display the attribute tested, where the branching of the node displays the outcome, and the terminal leaf displays the decision from attribute computing. Besides, DT can adaptably be used for this study in the prediction of the patients' electrolytes.

LR is the most common model used in ML prediction technique. It includes both statistical and ML approaches to interpret the relationship between predicted output with input variables (predictors). Also, the algorithm requires numerical predictor variables that would provide the numerical output [59]. A SVM algorithm, also termed as support vector networks, tends to minimize the error by adapting a hyperplane to maximize the decision boundary while tolerating a certain error. The hyperplane is the dividing middle line to predict constant output, whereas the decision boundary are the margin lines for the demarcation of positive and negative values from the hyperplane [58].

A NN algorithm comprises numerous algorithmic layers (decision matrices) in predicting an output. It is created with hidden layers, where inputs and their respective weights are multiplied and summed within the layers, added to bias value, and inserted into an activation function to produce an outcome.

Besides, algorithms like SVM, NN, DT, and LR, can be re-processed with ensemble techniques like boosting and bagging to achieve improvised and accurate electrolyte results for different patients' predictor values.

3.5. Data Prediction and Interpretation

This step represents the output display of the predicted attributes to be later analyzed for precise dialysis dosing of each kidney disease patient. Each predicting model from the algorithms can be taken as a digital expert. Hence, a confident prediction is generated and concluded if the overall outcome is the same for all digital experts. Subsequently, the outcomes can then be displayed using tables, charts, reports, etc. The conclusions from the prediction models may then identify the most significant attribute as the predictor for electrolytes. This may further interpret our results and achieve to prescribe precise dialysis dosing for each dialysis patient.

3.6. Model Comparison

In the last step, model comparison of four different predictive techniques are evaluated and recorded according to the model performances in regards to squared correlation and root mean square error. This will represent the best model that can generate better prediction results as compared to the rest, and reduce prediction errors and uncertainties.

Squared correlation is the determination coefficient (R^2) that denotes the variance proportion between the predicted output with respect to the input predictors of the model. Root mean square error (RMSE) is also known as root mean squared deviation, which is mainly the measure of standard deviation from the output data line of regression.

Chapter 4. Implementation

This section covers the implementation of the proposed methodology. The use of the data prediction software and visualization tools are presented along with the deployment of the different prediction models.

From the first step of methodology, datasets for 45 patients were obtained, which constituted kidney malfunctioning patients undergoing several months of dialysis. The data for the patients had multiple recorded attributes, including the patient's condition before, during, and after the dialysis sessions per month. With a large set of attributes within the dataset, patient attributes were manually assorted into 14 different categories for understanding and simplification. These categories included demographics, pre-HD assessments, pre-vascular assessments, patient dialysis factors, anticoagulation factors, vascular access complications, post-vascular assessments, discharge criteria assessment, discharged factors, blood investigations, full blood count, dates of investigation, electrolytes, and other investigations.

Next, the data pre-processing step was implemented using Rapidminer software. Rapidminer is a data science platform built for the purpose of utilizing several AI tools that makes it a combined environment for the users. Since data integration was the first sub-stage in data pre-processing, it was simply implemented by the consolidation of 45 datasets of kidney dialysis patients, with each having multiple dialysis data entries of several months. Therefore, the data from several patients were physically merged altogether into a single data file to be utilized in further analysis. Further data pre-processing stages were implemented using the required operators in Rapidminer, as shown in Figure 4.1.

From Figure 4.1, the process takes in the big dataset of several dialysis patients with multiple attributes. From the dataset, attributes that had more than 80 percent of missing values were removed. Also, attributes that had just one value different from the rest of the values were also excluded. Removing such attributes was executed as they did not imply any significant meaning towards the dataset from voluminous values missed or being equivalent.

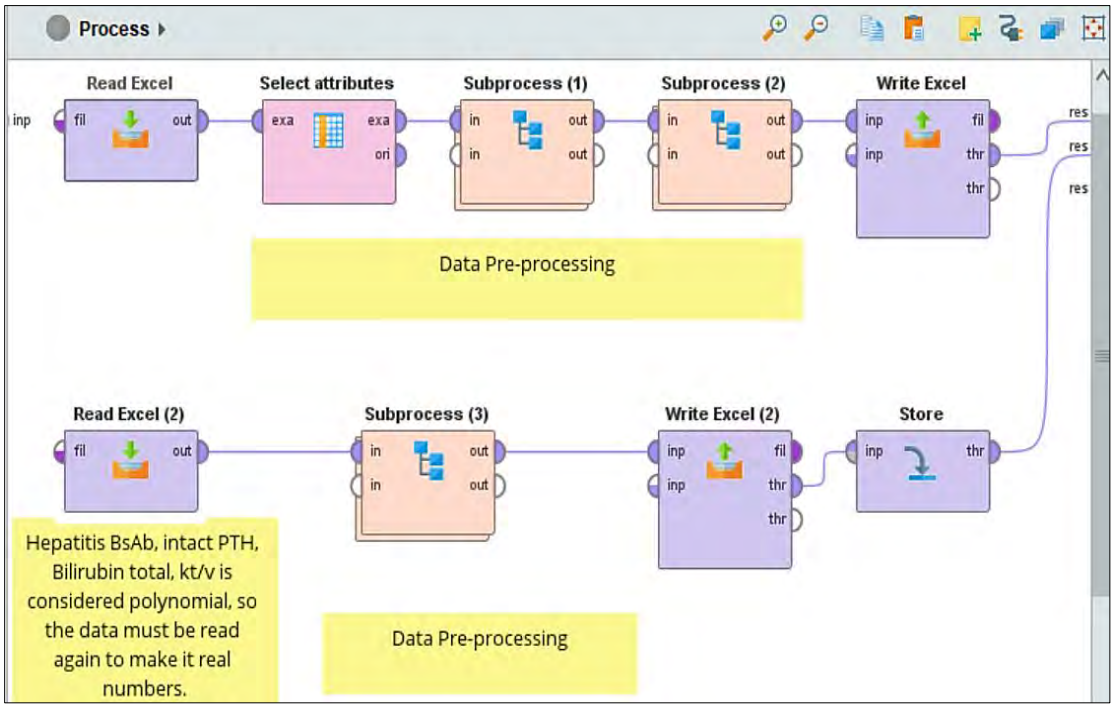

Figure 4.1: Schematic execution of data pre-processing.

Multiple attributes had noisy data, which required some values to be replaced. Therefore, Figure 4.2 shows the operators used for the raw data cleaning of some attributes, as follows:

- Replace (1): Replaces any 'NA' value into blank/missing value from two attributes, namely KT/V and Exit site clean and dry. With this, the prediction model would rely on the integrity of the valid existing data values, where noisy values are left empty to be imputed later by other means.
- Replace (2) – (9): Corrects any nominal data values into numerical attribute format, including Dry weight, Bilirubin total, Hepatitis BsAb, and Intact PTH.
- Replace (10) – (24): Corrects any nominal texts into proper and similar alphabets and words to avoid multiple indices. These attributes include Dual lumen catheter type, Fistula/graft site, and Dressing type.
- Replace Missing (1) – (3): Any missing nominal values were replaced from the attributes EPO, Suture present, and Exit site clean and dry.

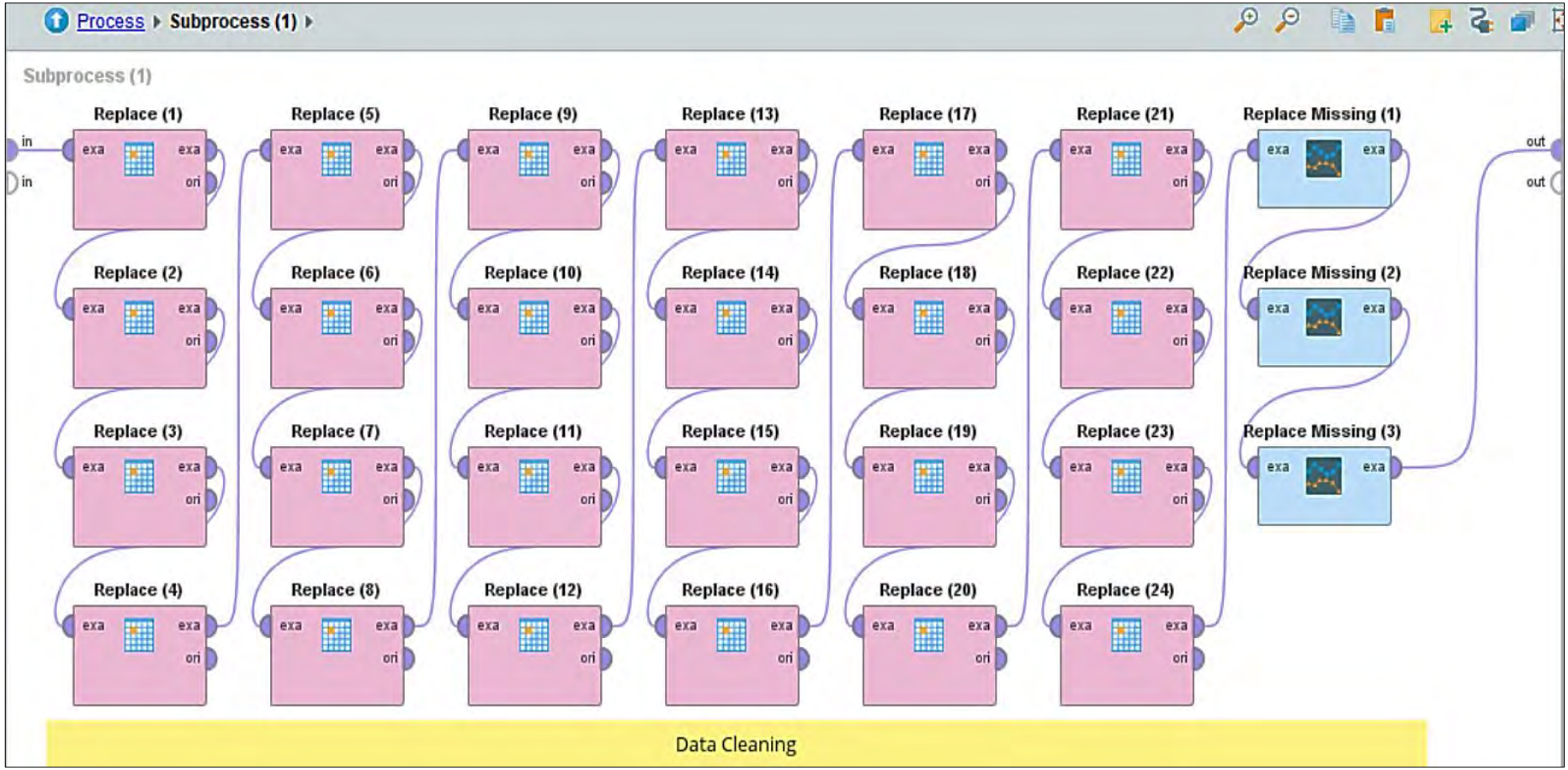

Figure 4.2: Schematic execution of data cleaning in subprocess (1).

Furthermore, Figure 4.3 shows the operators implemented for data transformation, further data cleaning, and data reduction, which are as follows:

- Aggregate (1): Used to define average values per patient from the entire dataset of certain numerical attributes with missing values.
- Join (1): The averaged attributes generated from the original dataset were merged here altogether.
- Generate Attributes (1): Any missing values from the original dataset were replaced by the average values from specific averaged attributes.
- Generate Attributes (2): Several other attributes with missing values were replaced by means of proper nominal values, or function expressions. Generated attributes included Post weight, Dry weight, HBA1C, Bruit, Thrill, Vitamin D3, Fistula_graft/Dual_lumen_catheter_type, IV iron injection, and Anticoagulation.
- Generate Attributes (3): Nominal attribute (EPO) was converted to numeral values by generating corresponding values and replacing the created numerical attribute with the original one.
- Select Attributes (3): All the average generated attributes were removed and any redundant attributes were eliminated, which had no remaining significance in the processed dataset.

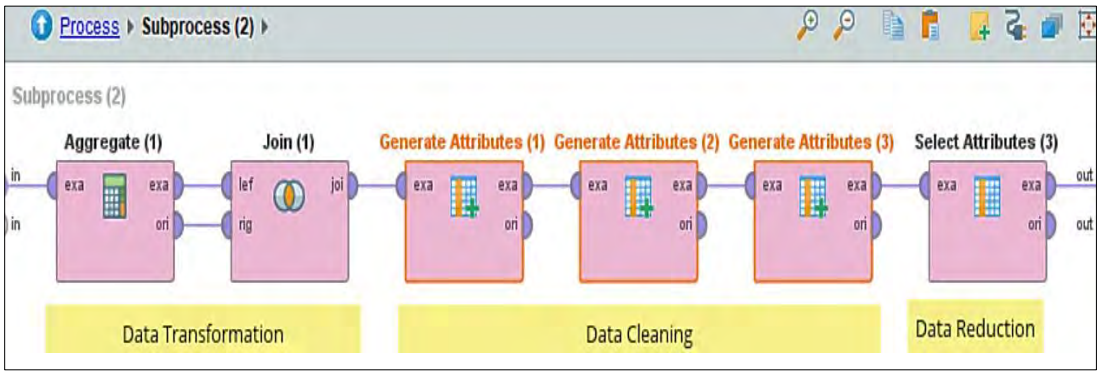

Figure 4.3: Schematic execution of data pre-processing in subprocess (2).

From earlier steps in Figure 4.3, certain attributes were unable to define the average per patient as the attribute types were nominal, including Hepatitis BsAb, Bilrubin Total, and Intact PTH. Therefore, the dataset was read again for further data

pre-processing, as shown in Figure 4.1, where the software retakes the mentioned attributes in the form of real numbers. The read dataset then follows several operators in Figure 4.4 below, which again execute the similar process from operators seen in Figure 4.3. The built process was executed and the pre-processed dataset was saved and stored for the implementation of EDA and electrolyte predictions. The saved dataset still contained some missing values, which was due to patient attributes having no data values at all to be averaged.

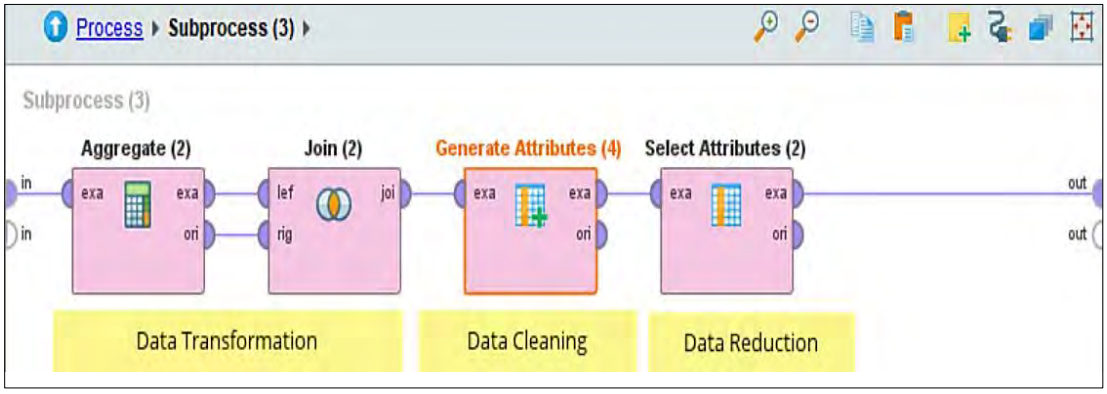

Figure 4.4: Schematic execution of data pre-processing in subprocess (3).

Next, in implementing EDA step, two software were used to best achieve our findings and understand the relationship between certain important patient attributes. One of the software was Rapidminer, where the electrolyte profiles were visually analyzed. The EDA process involved the use of several operators, as shown in Figure 4.5 below.

From Figure 4.5, the semi-cleaned dataset was inserted into the process and only certain attributes were selected from the entire dataset for the analysis purpose. These attributes included Calcium, Sodium plasma, Potassium, Post urea, Magnesium, Duration, and Date taken. The process was run, and electrolyte profiles were analyzed and reported in the following Chapter 5. Moreover, the resulting dataset from the selection of attributes in Rapidminer was saved for further EDA to be implemented in the other software.

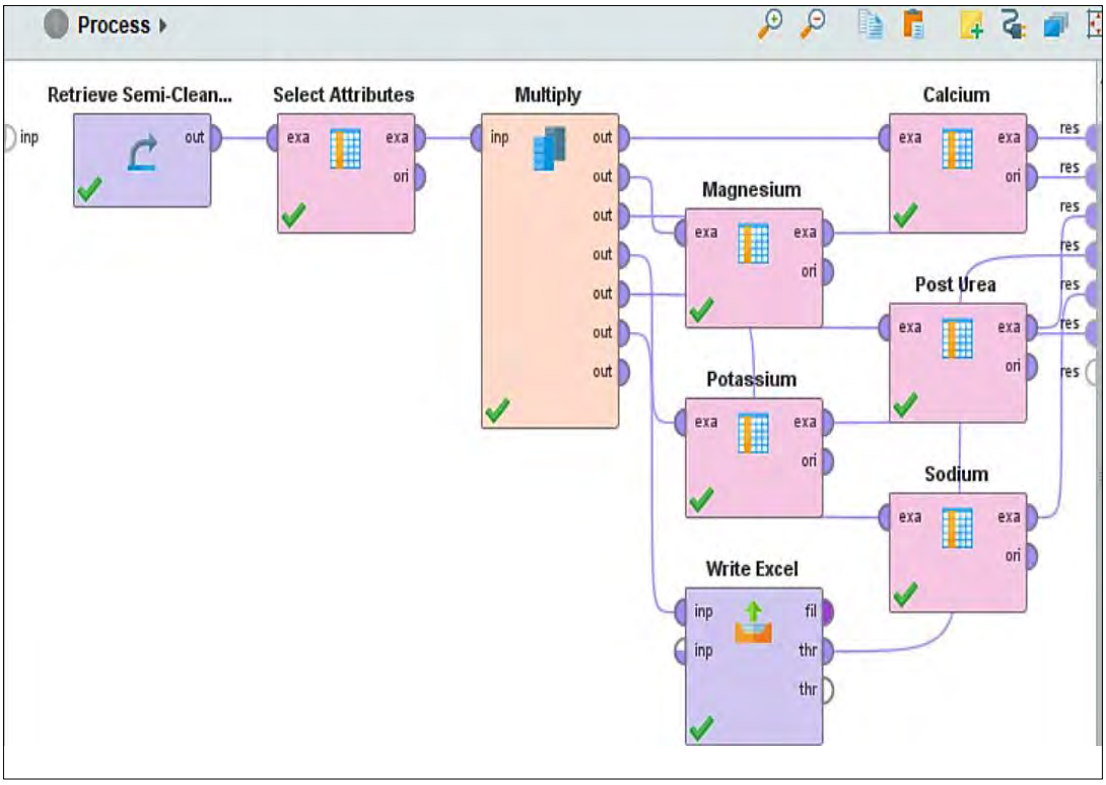

Figure 4.5: Schematic execution of EDA using Rapidminer software.

The second software used in the EDA step was Minitab. Minitab is a statistical, data analysis, and process enhancing tool, specially designed for the users and working industries to identify trends, determine important insights, and answer various problems in data. To implement the EDA, the boxplots were generated and likewise, time-lapse EDA was implemented by the same process with manually inserting the time step attribute to the dataset within the Minitab software.

After the EDA step, data prediction and interpretation step was built and executed in Rapidminer, as shown in Figure 4.6. To build the prediction models, the dataset was applied and was allowed to multiply to be used in the prediction of certain electrolytes, including post-urea (BUN), uric acid, magnesium, calcium, creatinine plasma, chloride plasma, phosphate, albumin, potassium, alkaline phosphatase, sodium plasma, carbon dioxide, and protein total. Each electrolyte prediction model had the exact same operators with the same purpose to predict the particular electrolyte from the same predictor attributes.

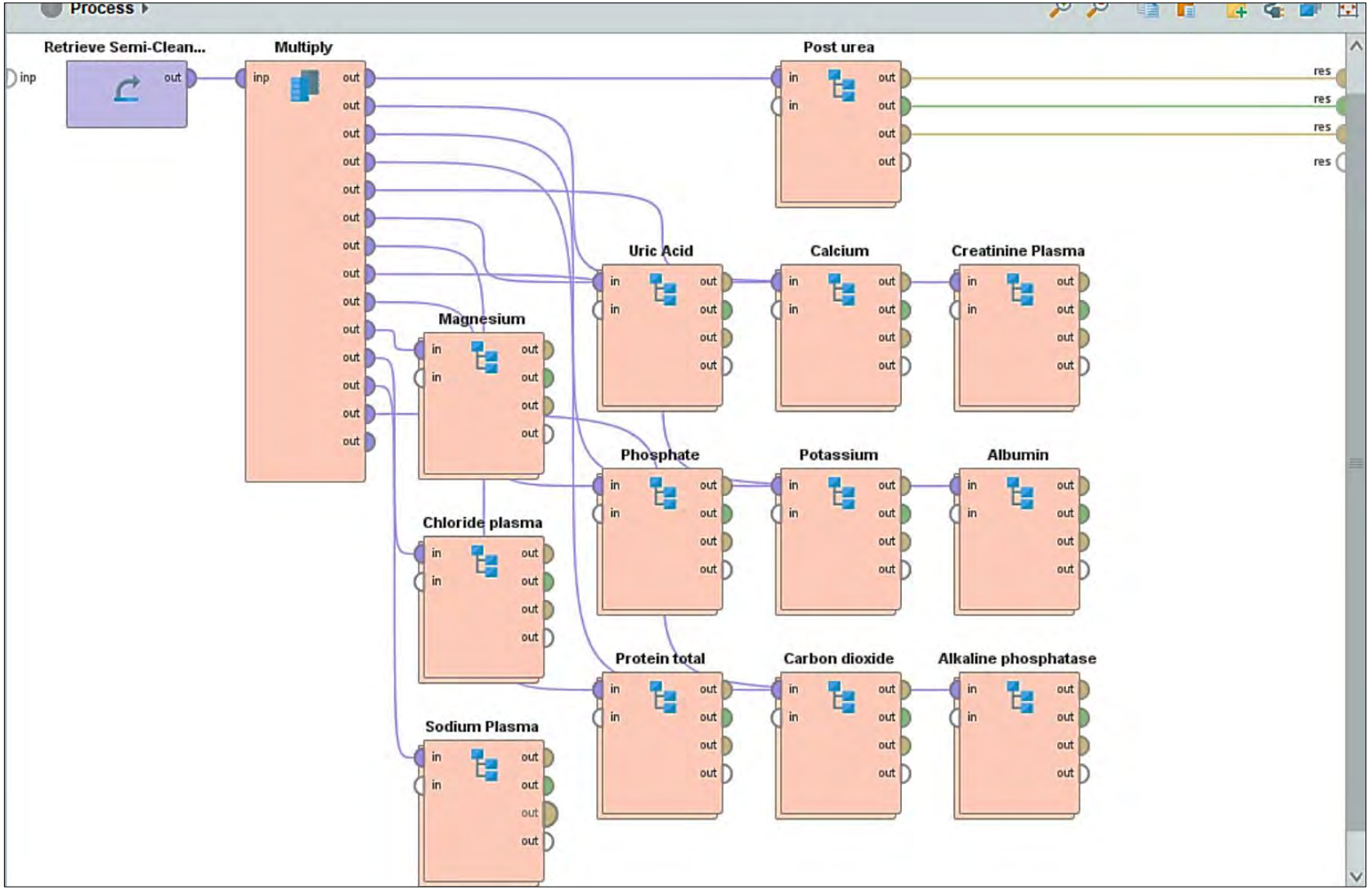

Figure 4.6: Schematic execution of prediction models for several electrolytes.

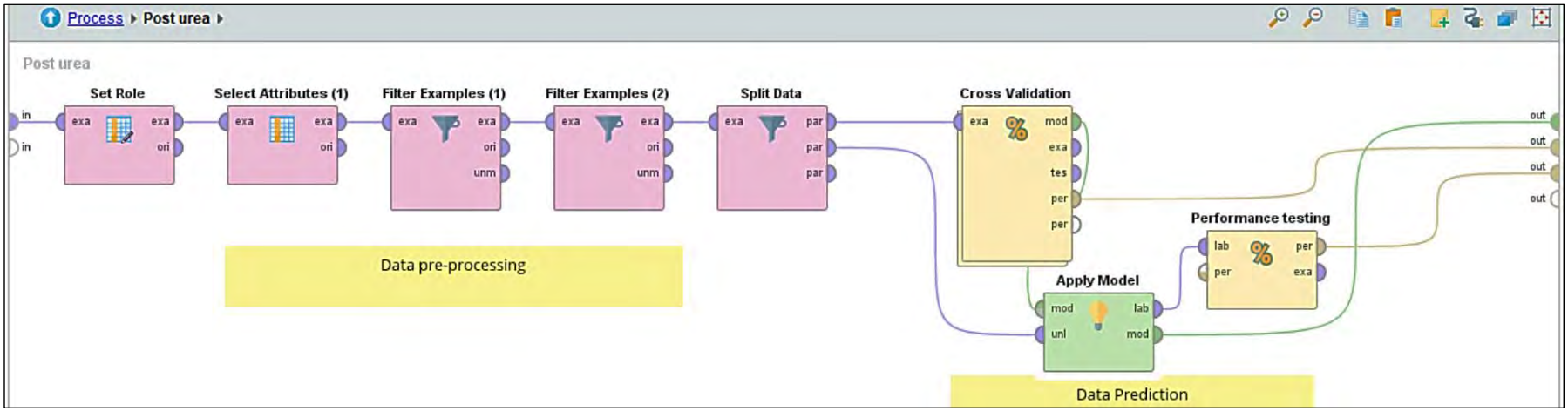

Figure 4.7: General schematic execution of data prediction.

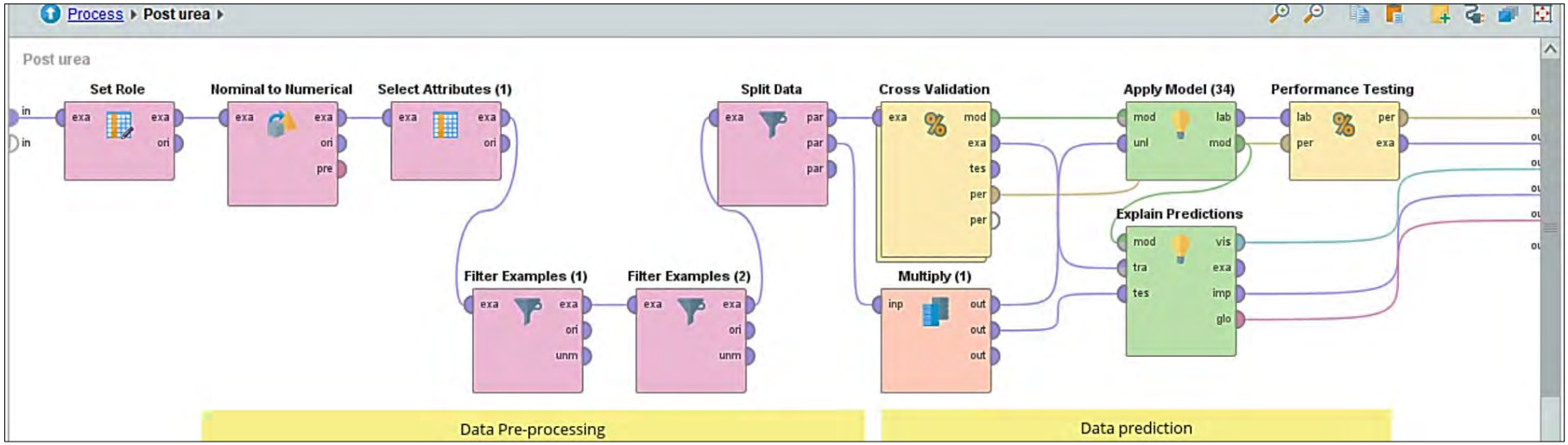

Figure 4.8: Schematic execution of data prediction with additional operators.

In each electrolyte sub-process, as shown in Figure 4.7 and 4.8, the following operators were used to build the data prediction model:

- Set Role: Electrolyte attributes were labelled here, which defined the model the desired attribute to be predicted.
- Nominal to Numerical: In some prediction models as shown in Figure 4.8, any nominal attribute is converted to numerical, including in LR, SVM, and NN prediction models, as the model algorithms do not operate on any nominal type values.
- Select Attributes: Allows the user to opt for only specific attributes in consideration for the model. In this study, demographical and other certain attributes that were inserted into the model as predictors included Age, Anticoagulation, Blood flow, Cumulative blood volume, Dry weight, Dialysis duration, Gender, HBA1C, KT/V, Pre urea, Pre weight, UF goal, and Weight loss. These attributes were selected based on relevance and correlation with respect to electrolytes from the EDA step.
- Filter Examples (1) – (2): Any missing values were removed from the dataset with selected attributes in both the predictors and the labeled attribute, as the prediction models would not execute with attributes encountering any missing values.
- Split Data: The dataset here is then split into two portions, where one part is taken as a training dataset for training the prediction model and constitutes 70 percent of the entire dataset after data pre-processing. The other part is taken as the testing dataset to test the finalized validated prediction model and constitutes to remaining 30 percent of the entire dataset. A higher percentage is used for the training dataset as the models need to be well trained and validated to achieve accuracy in their performances and to provide better prediction results.
- Cross Validation: Cross validation of 10 folds was implemented to make sure that the training dataset performs well and prevents biasing and overfitting. This operator is a sub-process operator, which is divided into two sections, training and testing, as shown in Figure 4.9.

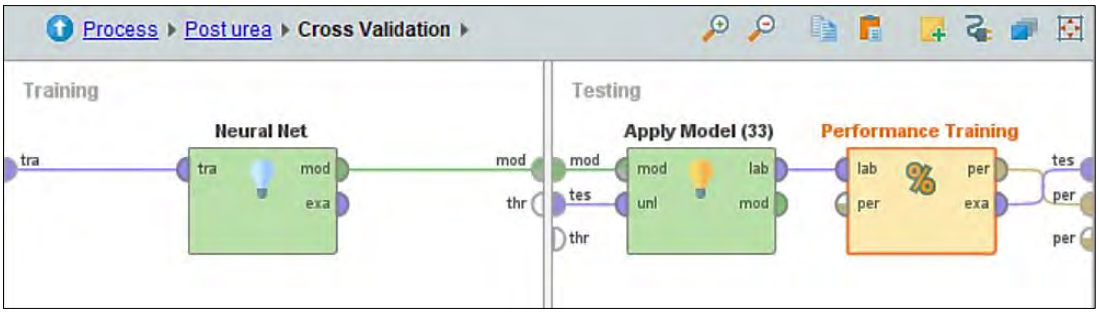

Figure 4.9: Schematic execution within cross validation operator.

In Figure 4.9, the training division included the data prediction algorithm, either DT, NN, LR, or SVM. In the SVM prediction model, the kernel type 'radial' was selected and executed. Also, in NN model, the algorithm parameters involved two hidden layers, named as 'Hidden 1' and 'Hidden 2' with hidden layer sizes of 5 and 3, respectively. Moreover, DT prediction model had the 'least_square' criterion selected, with the maximal depth of five to display the entire resulting tree from the model output.

In the testing division from Figure 4.9, the operators proceed with the selected model to give the desired electrolyte prediction outcome, where the trained model performance is evaluated in terms of selected parameters as squared correlation and root mean square error.

After the model training process, the data prediction step then takes 70 percent validated trained dataset from the predicted model output and apply it to the rest of the 30 percent testing dataset, to predict the electrolyte outcome from the testing dataset. The tested final model performance is then evaluated from each of the prediction model based on graphical or weights distribution.

Furthermore, from Figure 4.8, the algorithms were then modeled to demonstrate the most significant predictor attribute in terms of weights or figures, which would determine the most impact to predict the particular electrolyte. Therefore, an additional operator was used to state and explain the importance of the predictor attributes as the output by various means. Once the models were built for each prediction algorithm, the processes were run, where the results were produced, demonstrated, and later interpreted in Chapter 5.

Chapter 5. Results and Analysis

In this section, we present the results of the basic analysis of the information, EDA and prediction models, and we provide some discussion on the results.

5.1. Data Collection

In this study, the data is comprised of 45 distinct patients in total, for a duration of several months, i.e., from January 2020 until August 2020. From the datasets provided, the entire dataset represented 384 rows (examples) after tabulation, where each patient's dataset represented 117 distinct attributes (also known as variables, or patient parameters).

Figure 5.1 represents the 14 categories that comprised 117 attributes, respectively. The most significant categories used for EDA or data prediction were the demographics, electrolytes, and patient dialysis factors. Besides, certain studies emphasized that certain blood electrolytes are vital and the medical experts must regulate them for dialysis patients [44][48][49]. Also, patient demographical records are the necessary attributes considered for diagnostic or prediction purposes [29][32][46].

From Figure 5.1, important electrolytes in dialysis patient's body need to be regulated consistently to measure the adequacy of dialysis, as well as to maintain the function of the patient's body. Potassium, sodium plasma, chloride plasma, magnesium, and post urea, are the most considered electrolytes when it comes to kidney patients under dialysis. It is also important to know that some electrolytes associate with one another in case of an increase or decrease of a particular electrolyte. For instance, a decrease in calcium levels may cause a decrease in magnesium levels, which may further cause an increase in blood phosphorous levels. Other important patient categories involve blood investigations. Tests are usually done to make sure that the patient is not prone to other diseases or may not have encountered further disorders from kidney disease. The parameter HBA1C is another important attribute considered to measure the blood glucose (sugar) levels in the blood, which may diagnose the patient as diabetic or normal. Usually tests are done as a percentage and normally should be less than 6%, which may also state the fact that dialysis and dialysate solutions are given properly.

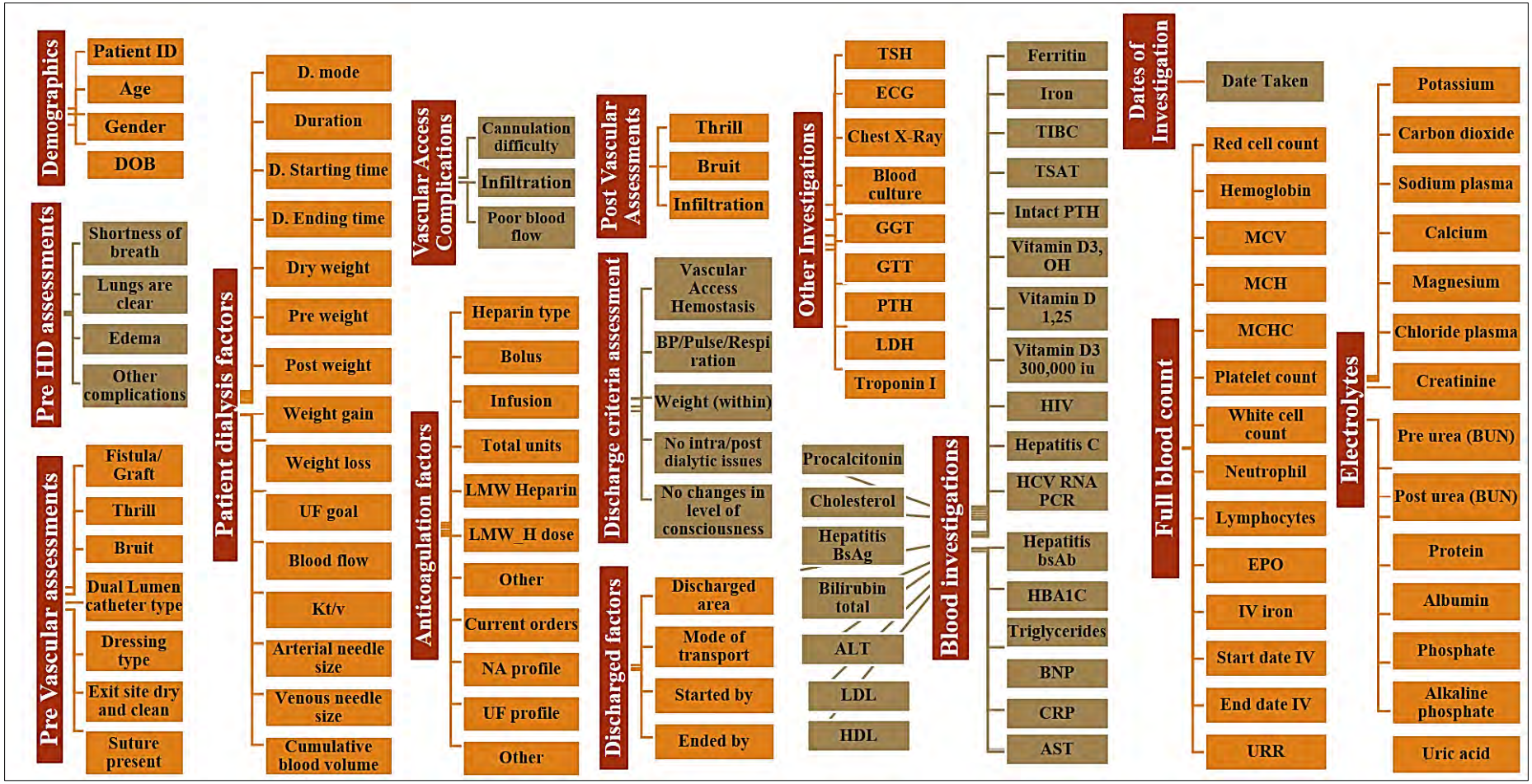

Figure 5.1: Defined categories and dialysis patient attributes.

5.2. EDA Output

In executing the EDA step using Rapidminer software, the results displayed several tabs of important electrolytes, which were mentioned during implementation. From these tabs, the output was seen in the form of connected scatterplots, with x-axis independent variable as the date taken and the y-axis dependent variable as the electrolyte attribute. Each line in the scatterplot represent data points for one particular patient. Figures 5.2-5.7 display the resulting plots of each important electrolyte profile. The plots displayed that some patients had high variability of electrolytes over a particular duration of months. Therefore, analysis results demonstrated to further check upon important electrolytes as well as the cause of such high variability from the medical point of view.

Besides, Figure 5.4 displays the highest variability of the potassium electrolyte for 45 patients throughout their monthly checkups. Hence, it elicits the fact that the course of dialysis within eight months had some positive and negative effects on some of the patients' electrolytes. To further study the cause of high variability on the vital electrolytes, electrolyte trends were visualized by other means of EDA, which were implemented using Minitab software.

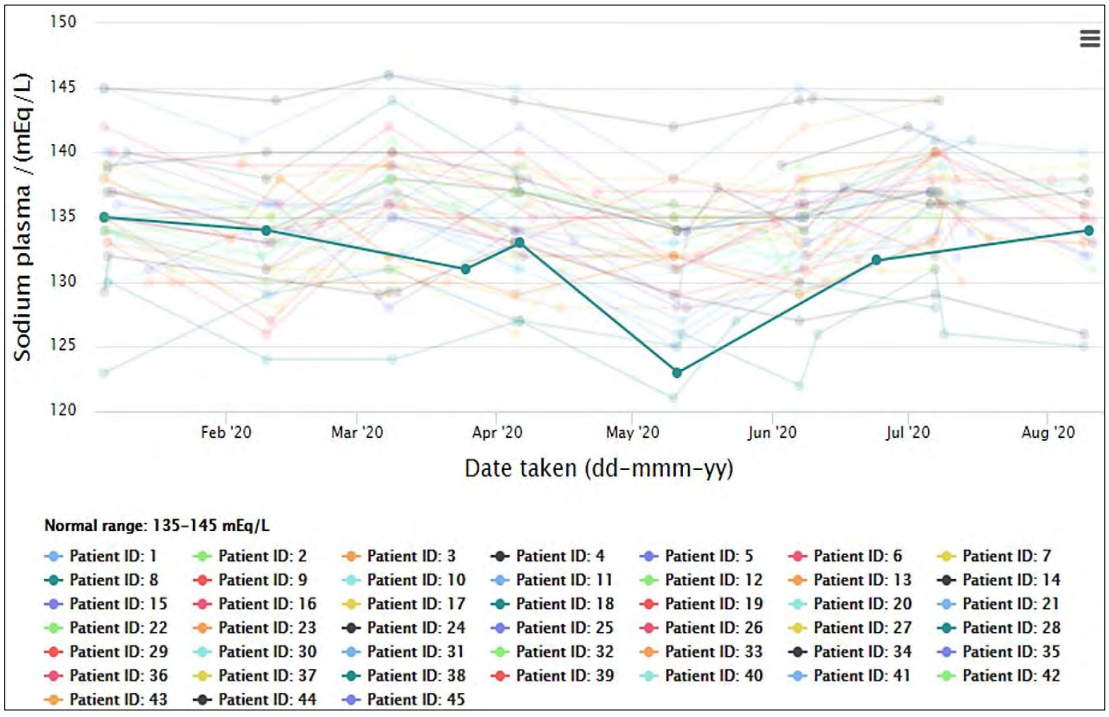

Figure 5.2: Sodium plasma profile for a single patient.

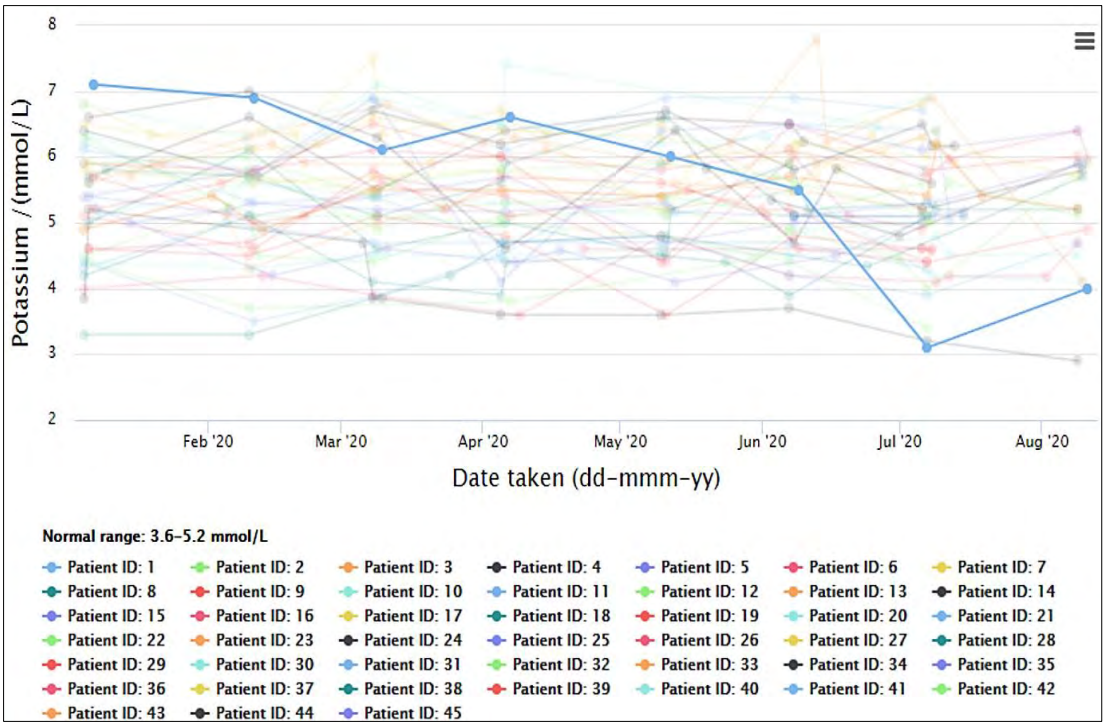

Figure 5.3: Potassium profile for a single patient.

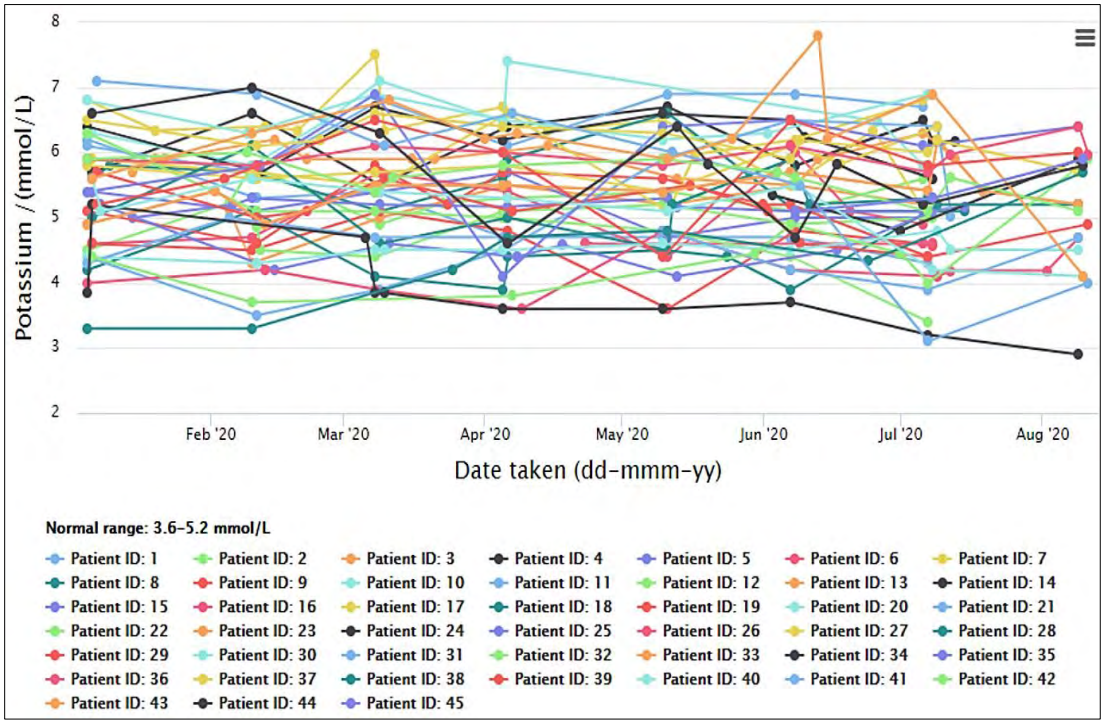

Figure 5.4: Potassium profile for all the patients.

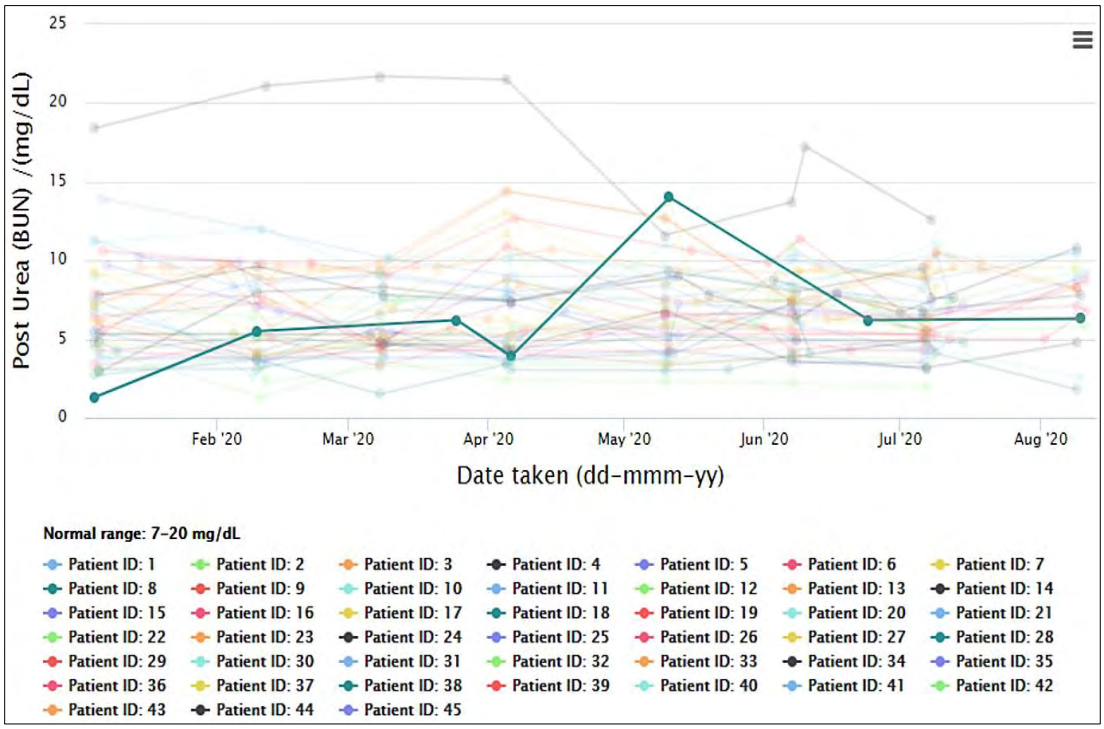

Figure 5.5: Post urea profile for a single patient.

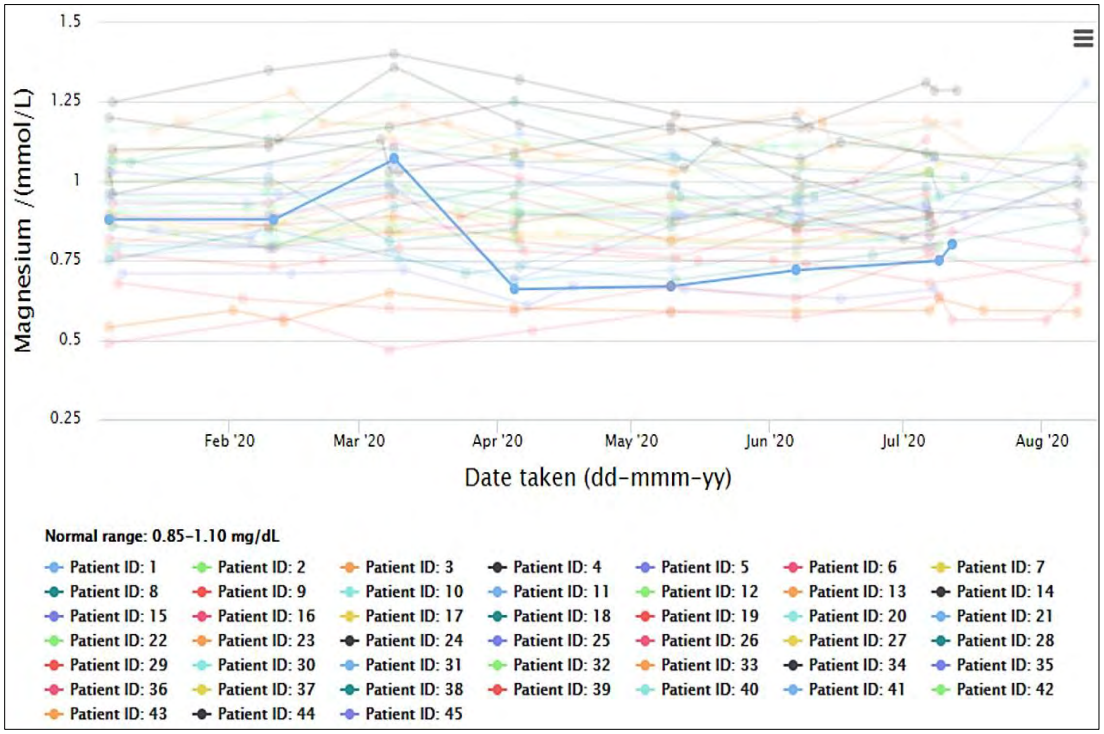

Figure 5.6: Magnesium profile for a single patient.

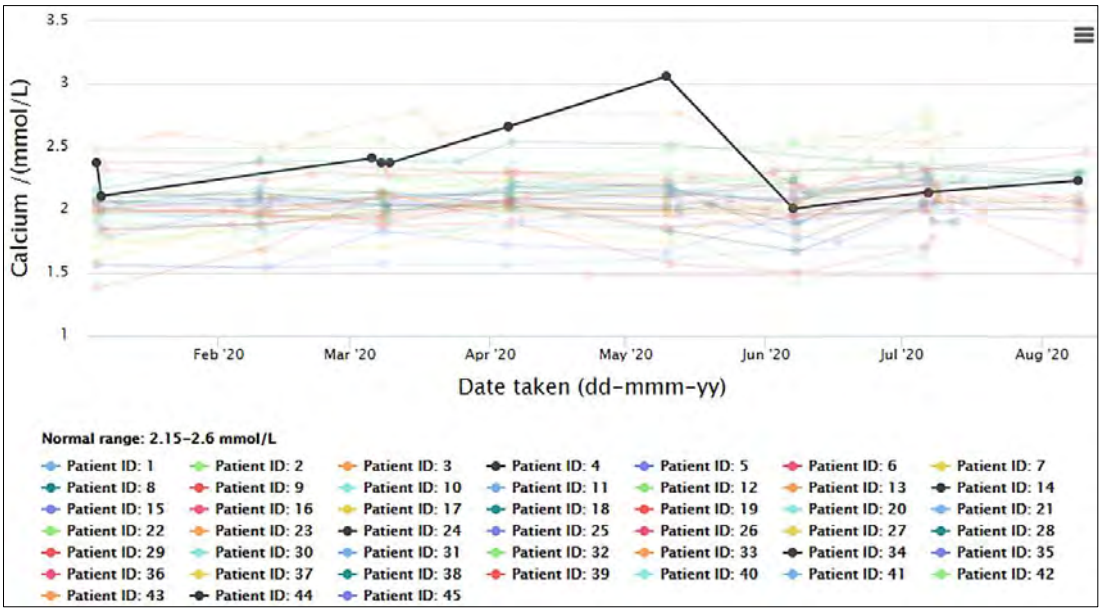

Figure 5.7: Calcium profile for a single patient.

The first alternative EDA approach on the electrolyte profiles was achieved by creating boxplot illustrations. Boxplots of significant electrolytes were created, including potassium, sodium plasma, calcium, magnesium, and post urea, which are shown in Figures 5.8-5.12. The plots displayed x-axis variable as patient ID and the y-axis displayed the electrolyte values. Each box represented the range of the particular electrolyte, for 45 dialysis patients, upon monthly dialysis checkups. The boxplots illustrated clearer insights towards specific electrolyte profiles. Within each electrolyte profile plot, mean of every patient is shown as dark blue circles, whereas the electrolyte minimum and maximum ranges are plotted as dotted horizontal red lines. Additionally, black lines displayed the connection of the mean from one dialysis patient to another, further emphasizing on the variability.

Insights from the boxplots noticeably demonstrated high variability of electrolytes from patient to patient, which can be recalled from similar insights seen earlier in Figures 5.2-5.7. Dialysis patients displayed the highest variability in sodium plasma and potassium electrolyte profiles than others, with the fact that patients' electrolytes are fluctuating and is a serious concern from the medical point of view. The insights may demonstrate that high variations could either be due to the lack of removal of potassium levels during the dialysis sessions. However, additional reasoning was associated in regards to sodium plasma electrolyte profile, as changes in blood sodium plasma levels have been acknowledged to inversely affect the blood potassium levels.

Therefore, as the blood sodium plasma levels illustrated a decrease in dialysis patients in Figure 5.12, this subsequently illustrated an increase in blood potassium levels of dialysis patients in Figure 5.8.

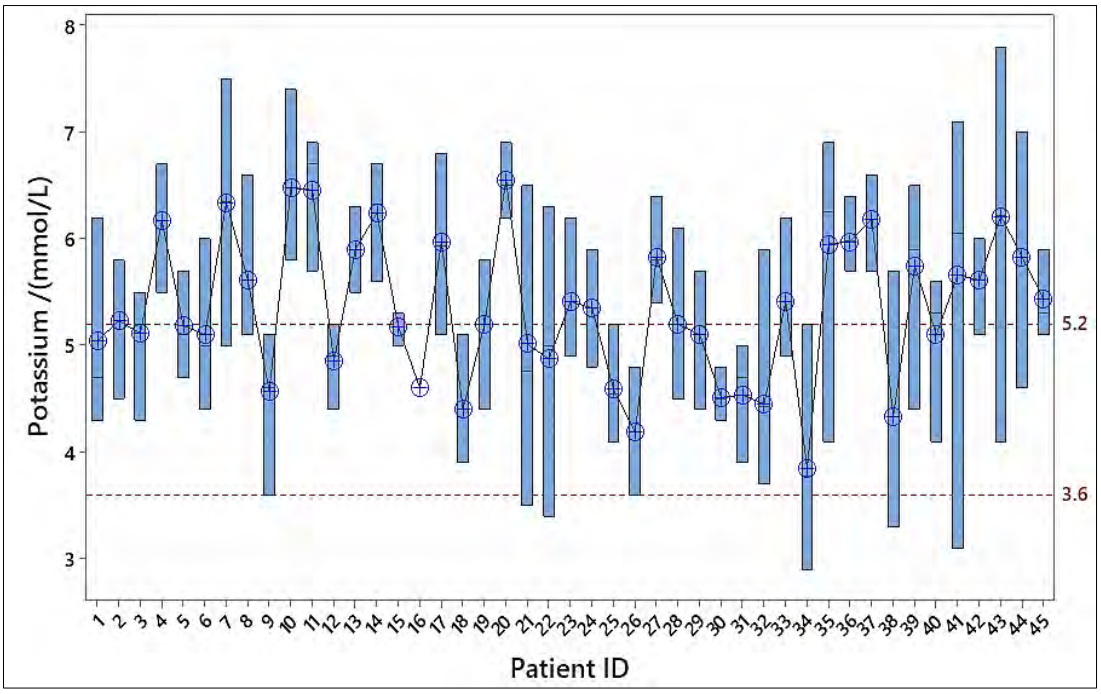

Figure 5.8: Boxplot electrolyte profile of Potassium.

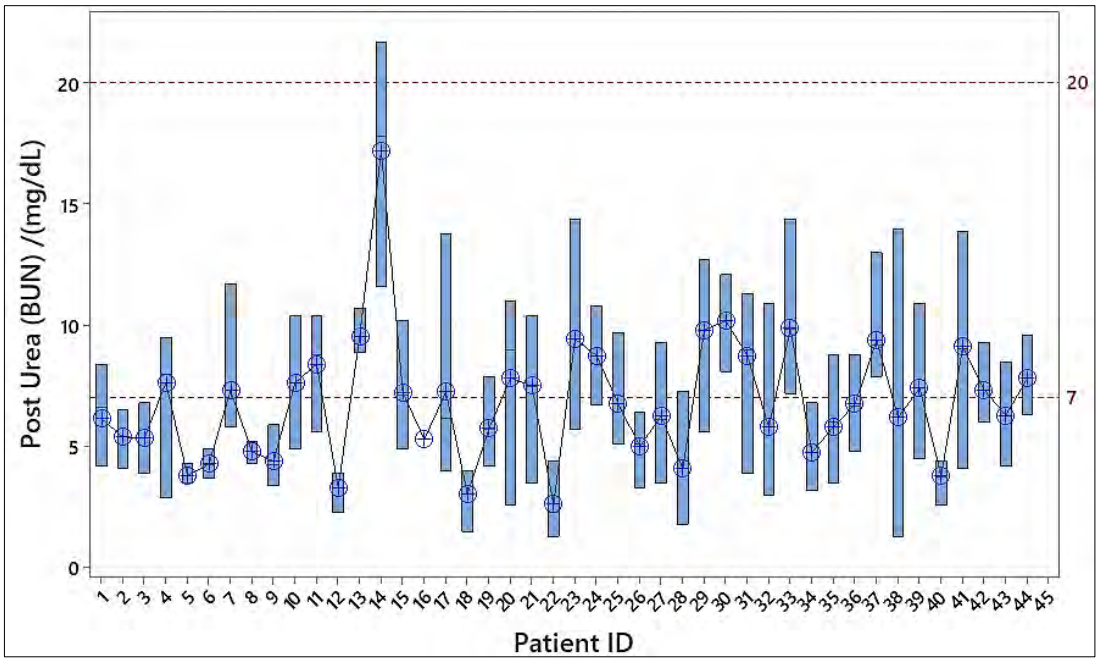

Figure 5.9: Boxplot electrolyte profile of Post urea.

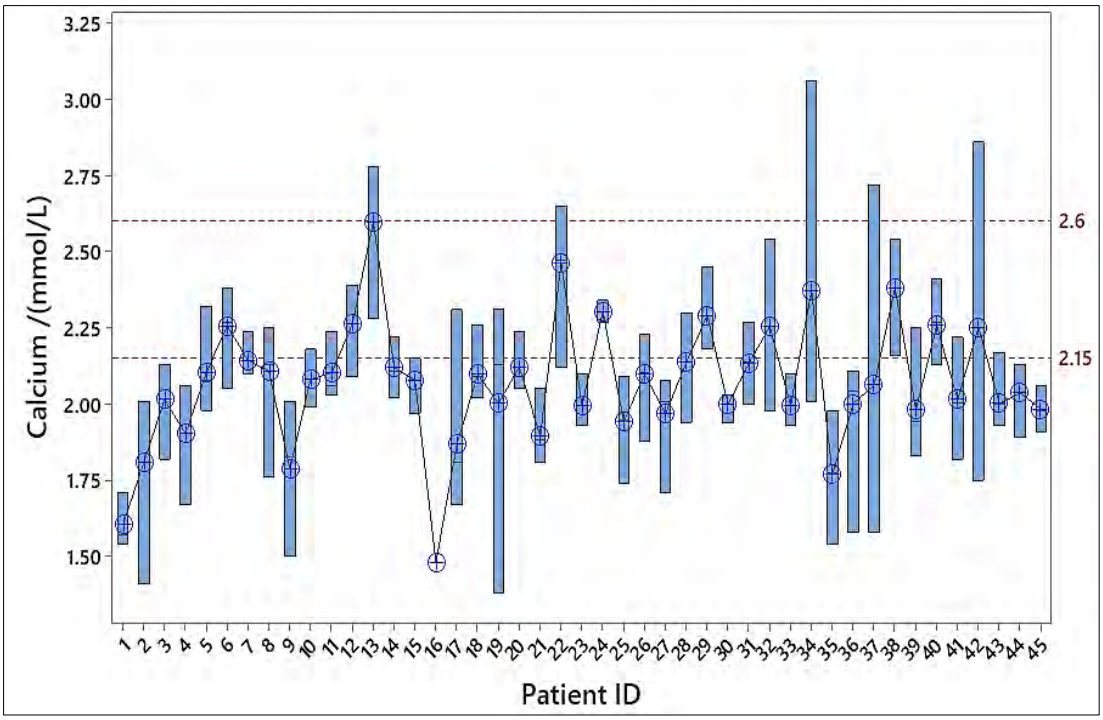

Figure 5.10: Boxplot electrolyte profile of Calcium.

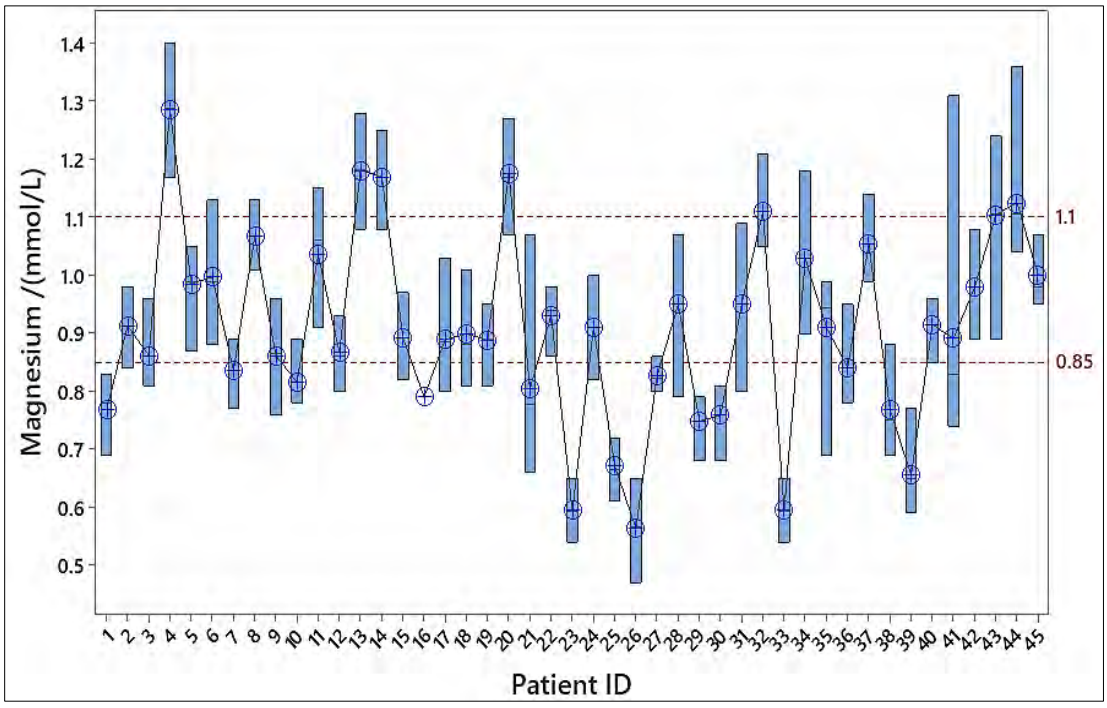

Figure 5.11: Boxplot electrolyte profile of Magnesium.

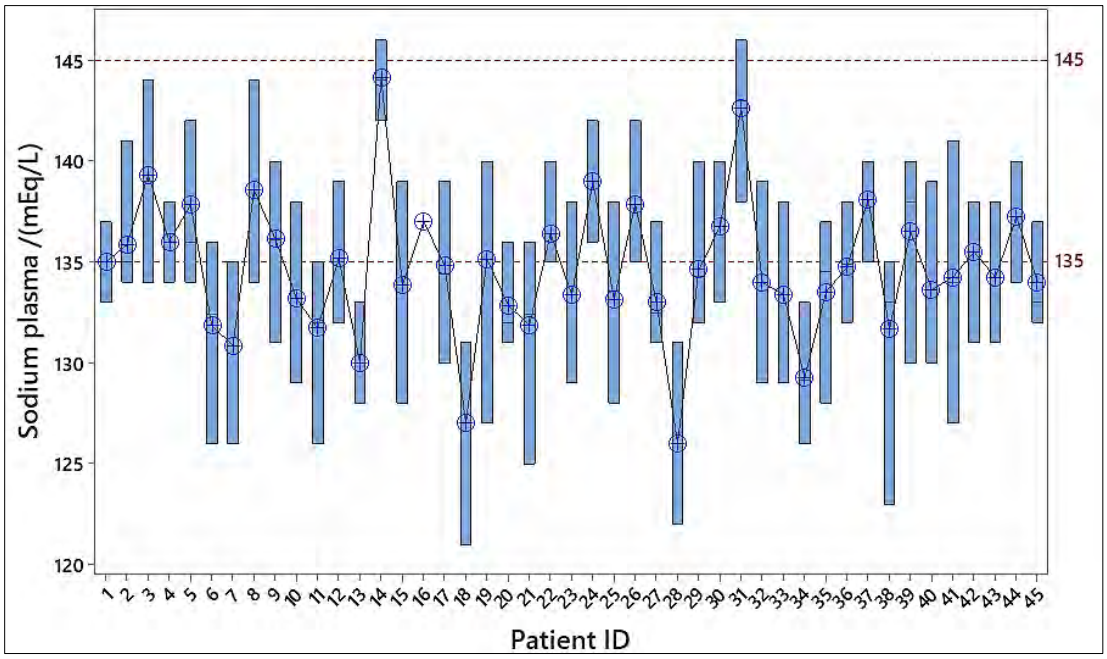

Figure 5.12: Boxplot electrolyte profile of Sodium plasma.

After achieving remarkable insights from important electrolytes, time lapse study was analyzed to further add value to the insights and view the important electrolytes from a different perspective in terms of dialysis sessions over time. Time lapse analysis is the approach to acquire and analyze electrolyte differences with changing time from dialysis patients. It is important to identify how the electrolytes change with time from a comprehensive view of the electrolyte from all the patients.

Therefore, time-lapse results were achieved and plotted as boxplots, as shown in Figures 5.13-5.17. The plots represented the electrolyte variable on the y-axis for potassium, sodium plasma, calcium, magnesium, and post-urea, with the time step variable along the x-axis. The mean was plotted as dark blue circles with red connecting lines. However, each box in the plots represented as the range of a particular electrolyte, for overall patients, upon receiving dialysis on a monthly basis.

The electrolyte trends of dialysis patients displayed low variability with respect to the time lapse input. High variability of potassium electrolyte as seen in Figure 5.8 had now demonstrated very low variability in Figure 5.13. Moreover, the time lapse electrolyte results further stated the fact that the dialysis sessions were indeed working well, with proper dialysate solutions and dialysis sessions. Hence, the analysis demonstrated that other certain factors were the cause of the high variability of vital

electrolytes in patients. Furthermore, Figures 5.8-5.12 required the need for the electrolyte prediction models to display the results and further interpret the factors that caused such high electrolytes' variability.

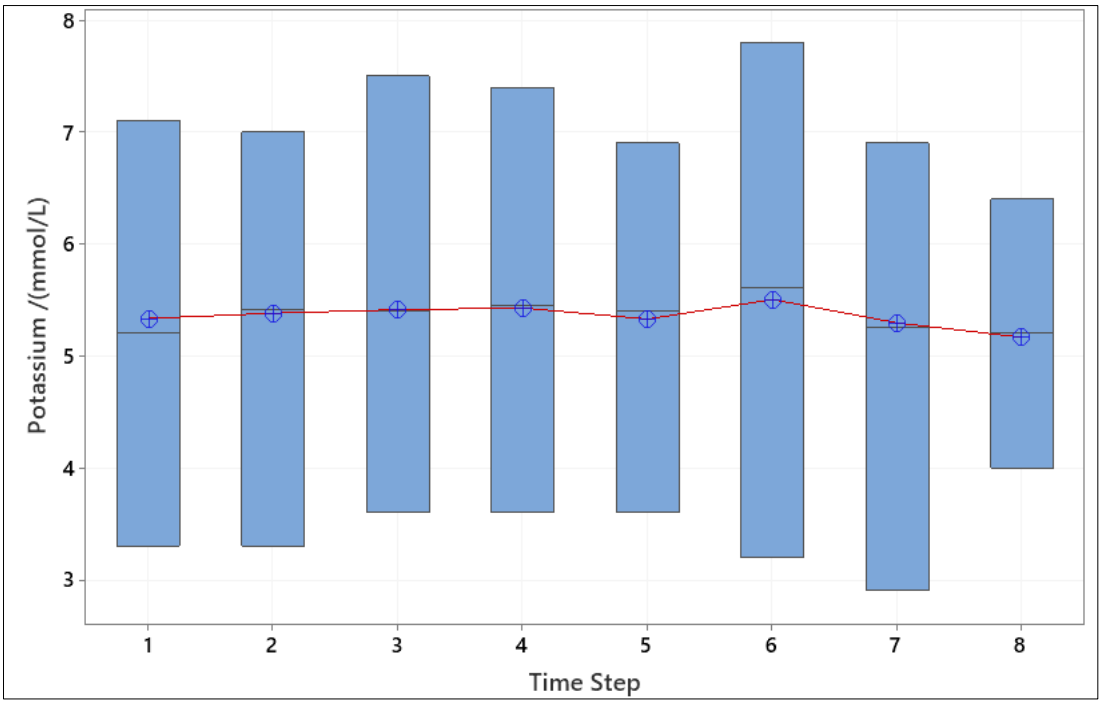

Figure 5.13: Boxplot time lapse of Potassium.

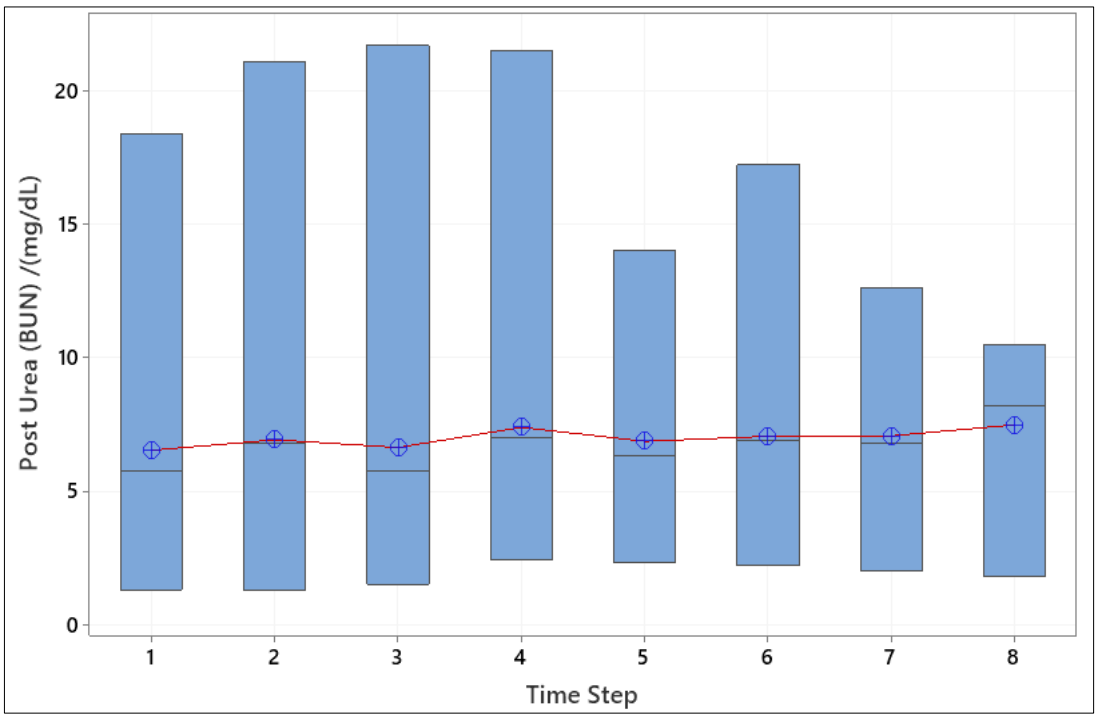

Figure 5.14: Boxplot time lapse of Post urea (BUN).

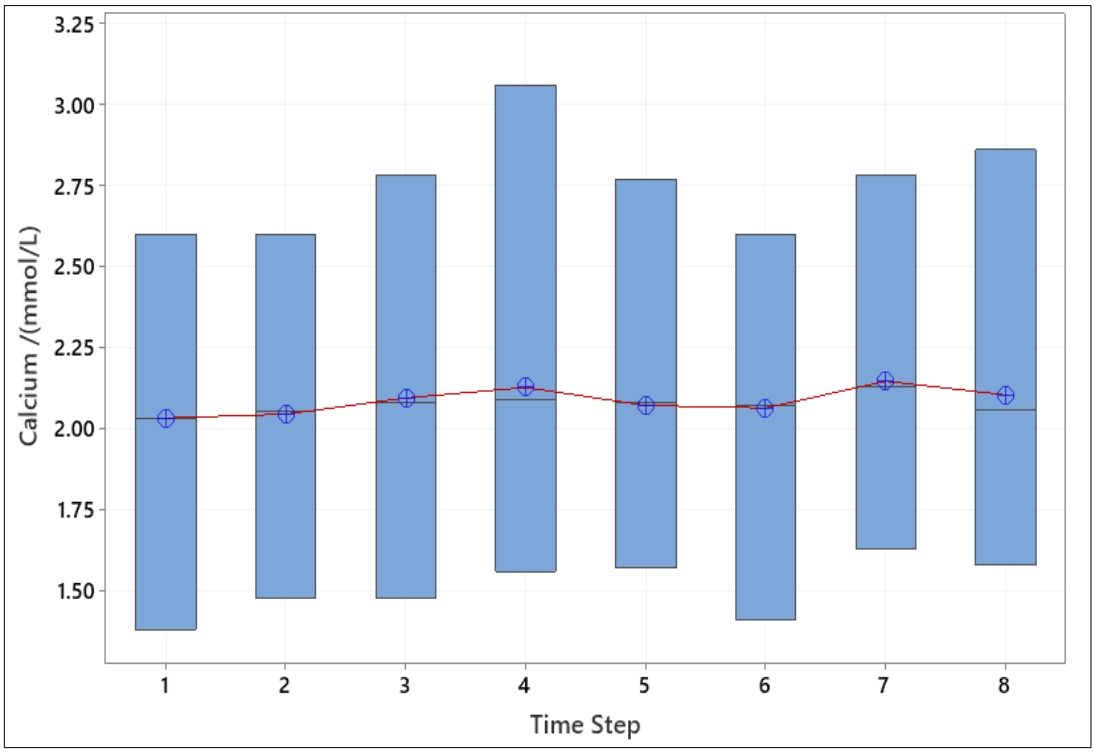

Figure 5.15: Boxplot time lapse of Calcium.

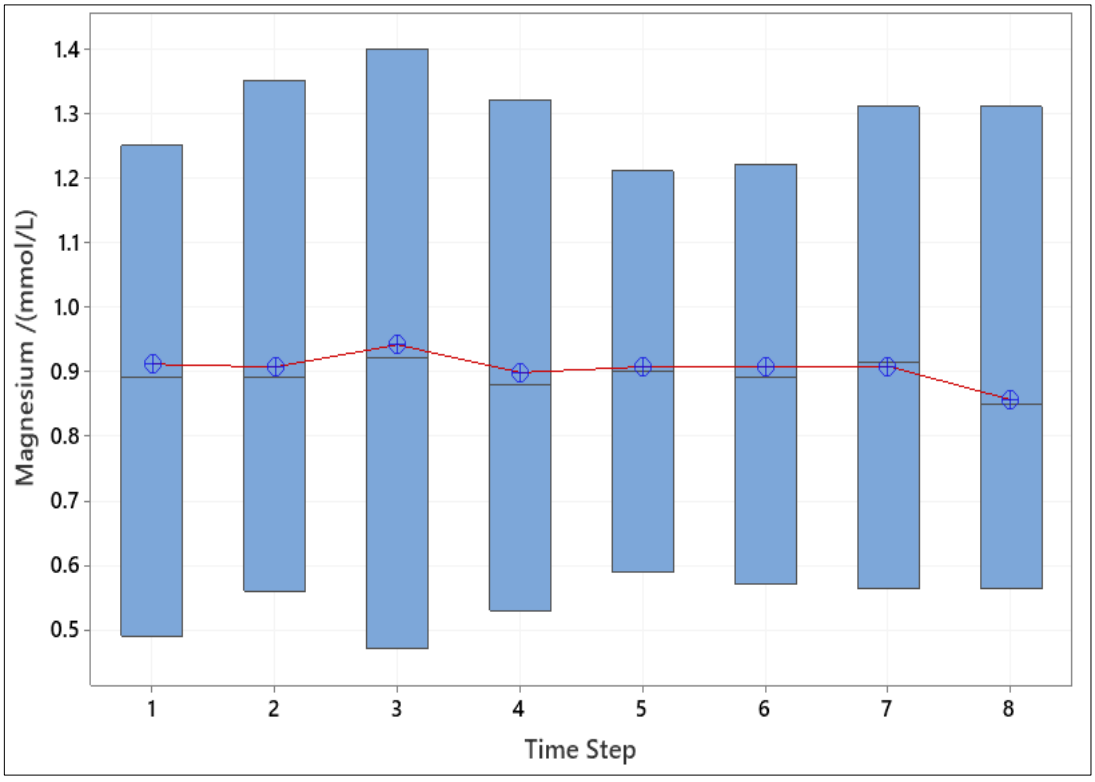

Figure 5.16: Boxplot time lapse of Magnesium.

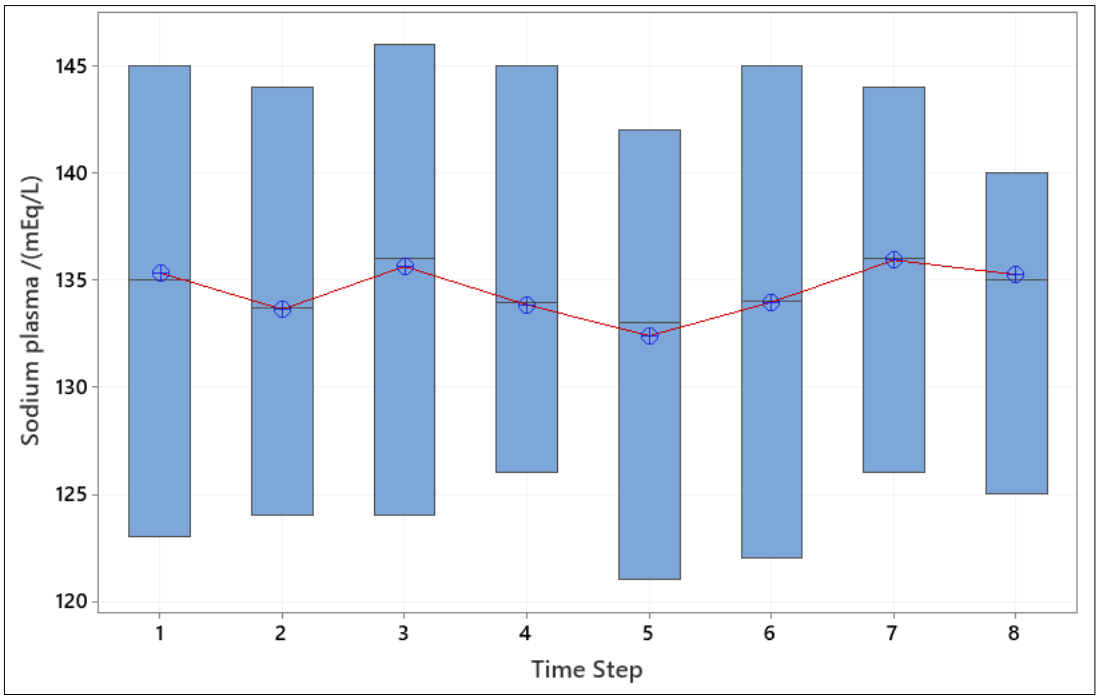

Figure 5.17: Boxplot time lapse of Sodium plasma.

Moreover, Figure 5.17 demonstrates slight variability in the sodium plasma of patients through time. The reason is that the dialysate solution tries to maintain the salt concentration of dialysis patients, where the response is only noticeable after every three months of adjusted dialysate. This allows the clinicians to regulate the dialysate contents as per the patient's body metabolism and other patient medications. Also, blood sodium plasma is an important electrolyte for muscle activity and cell's metabolism, which alters on a daily basis.

5.3. Data Prediction and Interpretation

The results from four AI/ML prediction models, Neural Network, Decision Tree, Support Vector Machine, and Linear Regression, are displayed in this sub-section. The outcome from each model is either represented as a weight or as a graphical illustration to define the most significant prediction attribute. For each of the prediction model, only one electrolyte outcome was displayed in results for illustration, analysis, and interpretation purposes. The rest are further displayed in Appendix A.

5.3.1. Decision tree outcome. In the outcome from the decision tree AI/ML model, the results demonstrated a tree-like structure output, as shown in Figure 5.18, which displays a sample output for post-urea (BUN) electrolyte prediction.

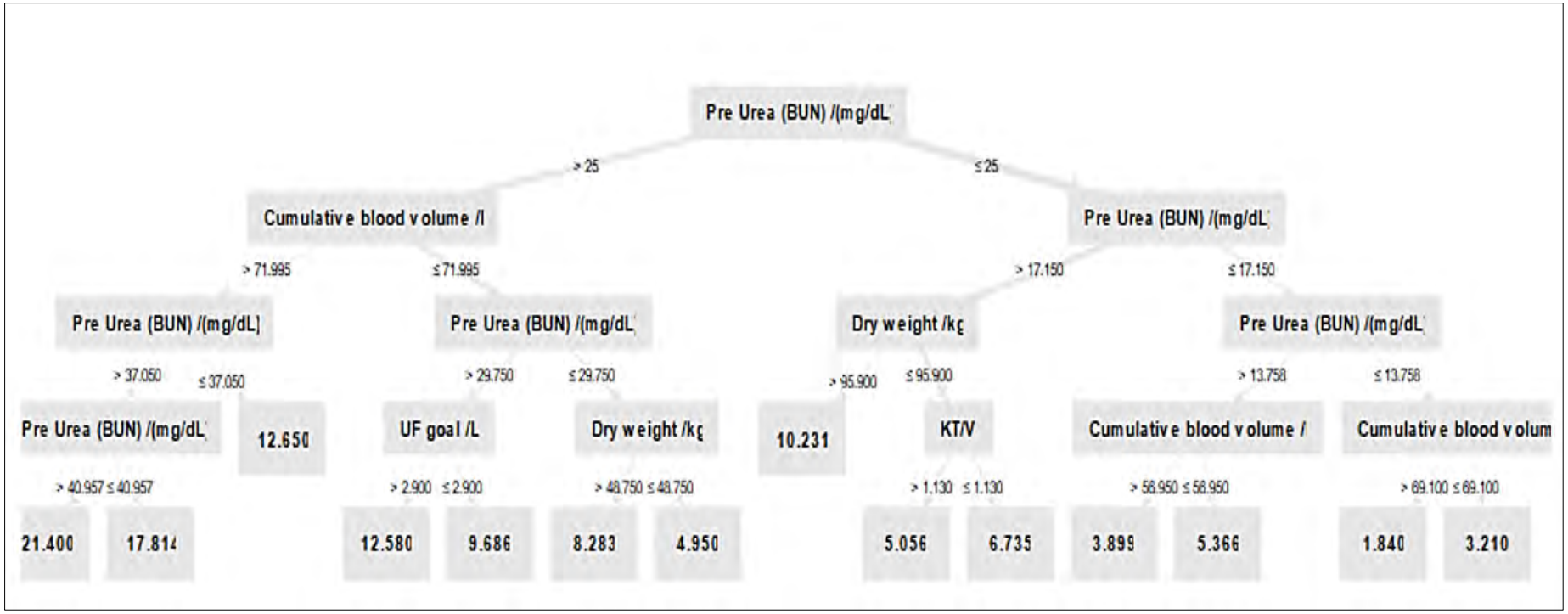

Figure 5.18: DT model output for post urea (BUN).

.

From Figure 5.18, the model displays several sub-trees with many nodes and branches, which start from the root node to the leaf node in predicting post-urea (BUN). Likewise, other electrolytes' predictions resulted similar graphical illustration with decision-making outcome from the highest weight of the predictors.

Since the model had a parameter depth of 5, the outcome model displays 5 layers of nodes represented horizontally. The root node of the decision tree usually represents the best predictor for the model. Therefore, this study visually determined the three most significant model attributes from the first three layers of the depth, which are the root node and the other two sub-tree parent nodes. In Figure 5.18, pre-urea (BUN), cumulative blood volume, and dry weight were the most significant attributes to predict post-urea (BUN). Other electrolytes represented similar results where the most significant predictor for each electrolyte was later noted and tabulated.

5.3.2. Linear regression outcome. In the outcome from the linear regression AI/ML model, the results provided a statistical measure table output, as shown in Figure 5.19, which displays a sample output for post-urea (BUN) electrolyte prediction. From Figure 5.19, the model displays several columns for the determination of the most significant attribute weight, where the rows provide the corresponding values of each predictor used in the model. To interpret the results and state the most significant predictor from the model, Coefficient and p-Value are basically considered.

Coefficient corresponds to the relationship between dependent and independent variable, whereas p-Value states whether the relationship of coefficients are statistically important. A p-Value of less than 0.05 states that coefficient is having an effect, is not equal to zero, and is a meaningful predictor to the model. On the other hand, p-Value of 0.05 or more state the predictor does not affect changes to the model and has no importance. Therefore, Figure 5.19 displays the p-Values in ascending order, where the predictor attributes were arranged according to the most significance vertically. For instance, the top three significant attributes with the lowest p-Values were pre-urea (BUN), gender, and KT/V, and demonstrated in Figure 5.19 as the most important predictor for the prediction of post-urea (BUN) electrolyte. Likewise, other electrolytes' prediction outcome from LR were noted and top significant attribute from each outcome was tabulated, which is shown later.

Attribute	Coefficient	Std. Error	Std. Coefficient	Tolerance	t-Stat	p-Value ↑	Code
Pre Urea (BUN) /(mg/...	0.314	0.019	0.685	0.813	16.819	0	****
Gender = Female	-1.157	0.290	-0.180	0.913	-3.991	0.000	****
KT/V	-3.604	0.986	-0.173	0.996	-3.656	0.000	****
Age /yrs	-0.053	0.018	-0.145	0.993	-2.931	0.004	***
Pre weight /kg	0.246	0.098	1.379	0.814	2.515	0.013	**
Dry weight /kg	-0.208	0.096	-1.164	0.837	-2.168	0.031	**
Cumulative blood vol...	0.030	0.021	0.078	0.951	1.441	0.151	
Blood flow /(ml/min)	0.025	0.018	0.069	1.000	1.365	0.174	
(Intercept)	-4.248	4.678	?	?	-0.908	0.365	
Anticoagulation = No	0.158	0.350	0.019	1.000	0.452	0.652	
Duration /hr	0.176	0.403	0.025	0.997	0.436	0.663	
Weight loss /kg	-0.221	0.722	-0.042	0.852	-0.306	0.760	
UF goal /L	-0.187	0.702	-0.037	0.869	-0.266	0.790	
HBA1C = Diabetic	0.153	0.607	0.009	1.000	0.252	0.802	

Figure 5.19: LR model output for post urea (BUN).

5.3.3. Support vector machine outcome. In the outcome from the support vector machine AI/ML model, the results provided a weighted table output, as shown in Figure 5.20, which displays a sample output for post-urea (BUN) electrolyte prediction. From Figure 5.20, the model displays a weight column that defines the corresponding set of weights for each predictor used in the model. Attributes that contribute more towards predicting the electrolyte will display the highest weight, with negative and positive signs that corresponds to their relationship in predicting the electrolyte.

The results from the prediction of post-urea (BUN) demonstrated the highest three weights for pre-urea (BUN), pre weight, and dry weight, with values of 80.338, 37.828, and 43.758, respectively. The prediction outcome verified as pre-urea (BUN) to most significant predictor towards the prediction of the electrolyte, which was noted and tabulated. Likewise, other electrolytes' prediction outcome from SVM were noted and top significant attributes from each outcome were tabulated, which is shown later.

Attribute	Weight ↓
Pre Urea (BUN) /(mg/dL)	80.338
Pre weight /kg	37.828
Dry weight /kg	34.753
Weight loss /kg	29.236
UF goal /L	25.361
HBA1C = Diabetic	9.684
Anticoagulation = No	6.001
Duration /hr	2.316
Blood flow /(ml/min)	-0.000
Age /yrs	-8.849
Cumulative blood volume /L	-12.840
Gender = Female	-13.730
KT/V	-14.875

Figure 5.20: SVM model output for post urea (BUN).

5.3.4. Neural network outcome. The outcome from the neural network AI/ML model provided a weighted table output, as shown in Figure 5.21, which was

only provided after 'explain predictions' operator was used. Figure 5.21 displays the weighted table sample output for post-urea (BUN) electrolyte prediction, with weights ranging from the positive highest (most significant contributor in prediction) to zero (no contribution in prediction). The results, from Figure 5.21 shown below, verified that pre-urea (BUN), anticoagulation, and UF goal, as the top three significant attributes to predict post-urea (BUN), with weights of 0.758, 0.327, and 0.154, respectively. Likewise, other electrolytes' prediction outcome from NN were noted and top significant attributes from each outcome were tabulated as shown later.

attribute	weight ↓
Pre Urea (BUN) /(mg/dL)	0.758
Anticoagulation = No	0.327
UF goal /L	0.154
Pre weight /kg	0.097
Dry weight /kg	0.066
Weight loss /kg	0.049
HBA1C = Diabetic	0.037
Blood flow /(ml/min)	0.019
Cumulative blood volu...	0.018
Gender = Female	0.000
KT/V	0
Age /yrs	0
Duration /hr	0

Figure 5.21: NN output from explain prediction operator for post urea (BUN).

The weights of predictors demonstrated different results for the four AI/ML prediction model outcomes. Moreover, each electrolyte prediction outcome varies from another. Since the model incorporated predictions of 13 electrolytes with the output of weights, the most important predictor attribute from each model was tabulated in Table 5.1.

Table 5.1: List of top five important predictor attributes for patient electrolytes.

1	2	3	4	5	6	7	8	9	10	11	12	13
Anti-coagulation	UF goal /L	HBA1C %	KT/V	Dry weight /kg	Duration /hr	Gender	Weight loss /kg	Pre weight /kg	Cumulative blood volume /L	Age /yrs	Pre Urea (BUN) /(mg/dL)	Blood flow /(ml/min)

Attribute Name	**LR**			**DT**			**SVM**			**NN**		
	level 1	level 2	level 3	level 1	level 2	level 3	level 1	level 2	level 3	level 1	level 2	level 3
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

From Table 5.1, it can be seen that various attributes had an impact on predicting certain electrolytes. Comprehensively, Table 5.1 displays the specific color code for each input attribute to the model, where the values were also added and were characterized. The levels show the most important feature from the prediction models, where level 1 is the most important feature outcome from the model, following level 2, then level 3.

Based on the prediction models, only the corresponding attributes at the top three levels from each electrolyte model were noted, which was mentioned earlier. Therefore, the resulting table illustrated multiple attributes with many repeating colors and values seen in some of the prediction outcomes. In Table 5.1, the most repeating color and its corresponding value can be seen as light green with a value of 12, which

represented the attribute pre urea (BUN). For further statistical details from the illustrations, a frequency table was arranged as shown in Table 5.2.

Table 5.2: Frequency table for the predictor attributes in each importance level.

Color code	Attribute Name	Frequency			Total count
		Level 1	Level 2	Level 3	
1	Anticoagulation	11	8	2	21
2	UF goal /L	2	2	7	11
3	HBA1C %	6	8	4	18
4	KT/V	5	1	3	9
5	Dry weight /kg	6	5	1	12
6	Duration /hr	1	1	3	5
7	Gender	4	6	5	15
8	Weight loss /kg	0	2	0	2
9	Pre weight /kg	6	2	4	12
10	Cumulative blood volume /L	3	5	7	15
11	Age /yrs	0	0	1	1
12	Pre Urea (BUN) /(mg/dL)	8	9	12	29
13	Blood flow /(ml/min)	0	3	3	6

In simpler means, Table 5.2 displays the frequency of the input attribute at three importance levels from the four prediction models. Also, the total count of the attribute was generated that corresponds to how many times the attributes showed on these models in one of the top three importance levels from the prediction results. For instance, the first 'anticoagulation' attribute showed that it was used as the level 1 most important predictor 11 times. Besides, it had been the second most important attribute (in level 2) only eight times and the third most important attribute (in level 3) only two times in all the prediction models. Next, the second attribute as 'UF goal' showed a total count of 11 times to predict the electrolytes, regardless of the prediction model. Likewise, the frequency of other important input attributes in the prediction models was taken from Table 5.1 and were summed up for their total count in Table 5.2.

From Table 5.2, the top five input attributes are identified as pre urea (BUN), anticoagulation, HBA1C, gender, and cumulative blood volume. It is obvious from this table that pre urea (BUN) input is the first and foremost important attribute, because it is showing a total count of 29 times from all the models. The following important attributes are anticoagulation, HBA1C, gender, and cumulative blood volume, with the

total counts of 21, 18, 15 and 15, respectively. Upon the understanding from the medical point of view, pre urea (BUN) electrolyte is tested on a monthly basis from each dialysis patient before the dialysis is being administered.

Hence, this study expresses in following the decision-making support system, as the results interpret and state the physicians to get the dialysis patient's post urea levels within the acceptable range. To achieve adequate blood post urea levels from dialysis patients' blood tests, the physicians need to pay attention to the patient's lifestyle in advance to the dialysis. Such interpretation is declared because the pre urea (BUN), anticoagulation, HBA1C, gender, and cumulative blood volume, attributes were seen as the most important predictor parameters, which are basically the lifestyle features of dialysis patients. The other variables (like duration, age, etc.) included in the model were not much of significance as predictors.

5.4. Models' Performances

Upon achieving the electrolyte prediction results from each built prediction model, the models were then evaluated based on their squared error (R^2) and root mean square error (RMSE). The output values of R^2 and RMSE were tabulated for each predicted electrolyte. All four AI/ML prediction models' performances were noted and created as integrated tables for a better illustration, as shown in Tables 5.3 and 5.4 below. Table 5.3 displays the performances of the prediction models from the validated and trained output, using the training dataset. On the other hand, Table 5.4 displays the performances of the prediction models from the tested finalized model output, using the testing dataset.

In comparison to the prediction models, Support vector machine results demonstrated the least accuracy in achieving much lower values of R^2 and higher values of RMSE. On the other hand, DT achieved the best and far most accurate results by getting higher values of R^2 and lower values of RMSE for both the training and testing model performances. For the training output performances, DT achieved R^2 an average of 0.4649 ± 0.1639 and RMSE of an average of 18.8783 ± 4.74369, respectively. Besides, the testing finalized model performances for R^2 and RMSE on average were 0.4099 and 21.7011, respectively. NN and LR model outcome performances were acceptably higher than SVM, but were not better than the DT model's performance.

Table 5.3: Performances of prediction models from the training dataset.

Electrolyte Name	LR		DT		SVM		NN	
	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2
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	0.621 ± 0.168	216.273 ± 51.463	0.599 ± 0.125	153.821 ± 34.722	0.566 ± 0.147
Post Urea (BUN)	1.692 ± 0.458	0.722 ± 0.103	1.866 ± 0.446	0.644 ± 0.177	2.331 ± 0.683	0.555 ± 0.135	1.736 ± 0.383	0.736 ± 0.112
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	0.525 ± 0.170	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	0.668 ± 0.062	0.099 ± 0.017	0.699 ± 0.106	0.138 ± 0.026	0.455 ± 0.197
Alkaline Phosphatase	86.337 ± 21.066	0.190 ± 0.121	38.438 ± 9.977	0.794 ± 0.197	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	0.402 ± 0.146	49.279 ± 8.477	0.570 ± 0.130
Average	25.6679 ± 5.5497	0.3005 ± 0.1421	18.8783 ± 4.74369	0.4649 ± 0.1639	30.9956 ± 7.0162	0.4545 ± 0.1468	24.3266 ± 5.3944	0.3626 ± 0.1479

Table 5.4: Performances of prediction models from the testing dataset.

Electrolyte Name	LR		DT		SVM		NN	
	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2
Sodium plasma	4.21	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	0.71	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	0.512	237.395	0.772	139.677	0.72
Post Urea (BUN)	1.951	0.613	1.991	0.611	2.53	0.343	1.927	0.638
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	0.42	0.326	0.571	0.413	0.427
Magnesium	0.139	0.362	0.096	0.749	0.1	0.727	0.132	0.515
Alkaline Phosphatase	77.871	0.128	40.216	0.752	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	25.4145	0.2866	21.7011	0.4099	31.6937	0.4276	22.8780	0.318

From the performance evaluation of the four prediction models, DT model is the most recommended outperforming model to predict patient electrolytes. DT model precisely predicted the electrolytes based on the same input parameters used for other models and displayed the five most important predictor attributes, including pre urea (BUN), anticoagulation, HBA1C, gender, and cumulative blood volume, in its top three importance levels in the model outcome. The model also verified pre urea as the top most important attribute, along with anticoagulation the second most important attribute; following the next three important attributes as HBA1C, gender, and cumulative blood volume. From Table 5.4, the highest R^2 value of the DT model is shown for alkaline phosphatase electrolyte of 0.752, along with the lowest RMSE value for magnesium electrolyte of 0.096.

Chapter 6. Conclusion and Future Work

In this thesis, a clinical decision-making support system was developed for dialysis patients using data analytics. The study results realized the potential of extracting insights from the big dataset in the transformation of the healthcare from traditional symptom-driven practice into a precision personalized medicine.

This study utilized big data of more than 100 of several patients' pre-recorded dialysis parameters, where some critical electrolyte parameters were taken into consideration to analyze, predict, and provide insights on how to improve patient's dialysis dosing. Therefore, the electrolyte parameters were investigated with respect to demographical and dialysis patient factor attributes, including HBA1C and pre-urea (BUN) parameters. Based on the understanding from the results of analysis, the findings concluded that vital electrolytes (i.e., magnesium, calcium, sodium, potassium, and post urea) displayed high variability in patients, which can either be due to insufficient electrolytes removed during dialysis sessions or with respect to the inverse correlation with other electrolytes' profile. One such inverse proportionality involved the increasing levels of potassium in patients due to the lower sodium levels, as illustrated in results. On the other hand, it was also demonstrated that the electrolyte levels were consistent with time for each month when time-lapse view of each electrolyte was analyzed. This clarified that the patients were receiving proper dialysis sessions and the dialysis was undeniably working well.

Furthermore, AI/ML prediction models of the decision tree, neural network, support vector machine, and linear regression were built for further investigations and predicted the electrolytes within a range of demographical and vital dialysis parameters. Upon the interpretation of predictions, pre urea (BUN), anticoagulation, HBA1C, gender, and cumulative blood volume, were stated as the five most important predictors in predicting the electrolytes. Moreover, the best AI/ML model from the evaluation of four prediction models demonstrated to be the decision tree, which provided higher squared correlation average output and achieved a lower root mean square error average values for certain electrolytes.

Ultimately, the prediction 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 dialysis were not that of significant attributes for determining patient's important electrolyte levels. The five most important attributes were, therefore, pre urea (BUN), anticoagulation, HBA1C, gender, and cumulative blood volume, which are the lifestyle features of the dialysis patients.

Moreover, in practical use to the nurses or physicians, the prediction models can be executed with multiple patients' datasets and the three most important attributes noted from the models' significant levels, which will support the decision making and provide implications based upon the results of the prediction models. Thus, from the results and the interpretation of the most important attributes from the prediction models, physicians need to educate and prescribe the patients based on their adequate diet patterns, what patients need to eat and drink, and the amount of nutrition in each diet the patient needs to intake. This, likewise, provides patients improved dialysis dosing using big data analytics from the rising volume of dialysis patients, to increase the patient's quality of life, life expectancy, welfare, and conversely reduce costs, efforts, and time consumption, for patients and the physicians from the right dialysis treatment.

On the note, this study comprised of 45 patients in total, with dialysis sessions recorded from January 2020 to August 2020. In future, this study can further be enhanced by either utilizing a much higher number of patients, or by having temporal data with a history of dialysis earlier than January 2020. This could further improve the prediction models and declare better prediction results. Also, other dialysis machine parameters can be added to the built prediction models, to achieve additional interpretation in determining their correlation with the patient electrolytes, which may opt the researchers for an advanced series of investigations.

References

[1] B. Bikbov, C. A. Purcell, A. S. Levey, M. Smith, A. Abdoli, M. Abebe *et al.*, "Global, regional, and national burden of chronic kidney disease, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017," *The Lancet*, vol. 395, no. 10225, pp. 709–733, Feb. 2020, doi: 10.1016/S0140-6736(20)30045-3.

[2] S. B. Chaudhary, "Free dialysis for underprivileged kidney patients in UAE | Health – Gulf News," Jan. 18, 2019. https://gulfnews.com/uae/health/free-dialysis-for-underprivileged-kidney-patients-in-uae-1.61500736 (accessed Feb. 10, 2020).

[3] B. Shickel, P. J. Tighe, A. Bihorac, and P. Rashidi, "Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 5, pp. 1589–1604, Sep. 2018, doi: 10.1109/JBHI.2017.2767063.

[4] M. Makino, R. Yoshimoto, M. Ono, T. Itoko, T. Katsuki, A. Koseki, M. Kudo, K. Haida, J. Kuroda, R. Yanagiya, E. Saitoh, K. Hoshinaga, Y. Yuzawa, and A. Suzuki, "Artificial intelligence predicts the progression of diabetic kidney disease using big data machine learning," *Scientific Reports*, vol. 9, no. 1, pp. 11862, Dec. 2019, doi: 10.1038/s41598-019-48263-5.

[5] V. Mahadevan, "Anatomy of the kidney and ureter," *Surgery (United Kingdom)*, vol. 37, no. 7, pp. 359–364, Jul. 01, 2019, doi: 10.1016/j.mpsur.2019.04.005.

[6] S. S. Bhimji and S. W. Leslie, "Anatomy, Abdomen and Pelvis, Kidneys," StatPearls [Internet], Ncbi.nlm.nih.gov, 2018. https://www.ncbi.nlm.nih.gov/books/NBK482385/ (accessed Feb. 11, 2020).

[7] R. Kumar, "Urinary System," Mar. 13, 2010. https://www.slideshare.net/rajud521/urinary-system-3419307 (accessed Feb. 11, 2020).

[8] S. M. Sulivan, "Calyx Anatomy - Anatomy Drawing Diagram," Feb. 11, 2019. https://sen842cova.blogspot.com/2019/09/calyx-anatomy.html (accessed Feb. 15, 2020).

[9] J. Tyler, "Urinary System © 2018 Pearson Education, Inc. - ppt download," 2020. https://slideplayer.com/slide/14263970/ (accessed Feb. 20, 2020).

[10] R. Gilbert, "KIDNEY FUNCTIONS URINE FORMATION - ppt video online download," 2016. https://slideplayer.com/slide/5701963/ (accessed Feb. 20, 2020).

[11] J. G. Edwards, "The formation of urine," *Archives of Internal Medicine*, vol. 65, no. 4, pp. 800–824, Apr. 1940, doi: 10.1001/archinte.1940.00190100141006.

[12] E. A. Lawrence, D. Doherty, and R. Dhanda, "Function of the nephron and the formation of urine," *Anaesthesia and Intensive Care Medicine*, vol. 19, no. 5, pp. 249–253, May 2018, doi: 10.1016/j.mpaic.2018.03.001.

[13] F. Samarrai, "Dialysis Drug Dosing Improved by Using Big Data Analytics | News | News," Jun. 12, 2019. https://news.med.virginia.edu/blog/dialysis-drug-dosing-improved-by-using-big-data-analytics/ (accessed Feb. 25, 2020).

[14] S. Johnson, “Kidney Health and Kidney Disease Basics: Causes and Questions,” May 05, 2017. https://www.healthline.com/health/kidney-disease (accessed Feb. 20, 2020).

[15] J. Y. Yeh, T. H. Wu, and C. W. Tsao, “Using data mining techniques to predict hospitalization of hemodialysis patients,” *Decision Support Systems*, vol. 50, no. 2, pp. 439–448, Jan. 2011, doi: 10.1016/j.dss.2010.11.001.

[16] A. R. Sehgal, “What is the best treatment for end-stage renal disease?,” *American Journal of Medicine*, vol. 112, no. 9, pp. 735–736, Jun. 15, 2002, doi: 10.1016/S0002-9343(02)01110-5.

[17] A. K. Bello, A. Levin, M. Tonelli, I. G. Okpechi, J. Feehally, D. Harris *et al.*, “Assessment of global kidney health care status,” *JAMA - Journal of the American Medical Association*, vol. 317, no. 18, pp. 1864–1881, May 2017, doi: 10.1001/jama.2017.4046.

[18] S. Vadakedath and V. Kandi, “Dialysis: A Review of the Mechanisms Underlying Complications in the Management of Chronic Renal Failure,” *Cureus*, vol. 9, no. 8, pp. 1-8, Aug. 2017, doi: 10.7759/cureus.1603.

[19] V. Harrington, “Care of Patients with Acute Kidney Injury and Chronic Kidney Disease - ppt download,” 2020. https://slideplayer.com/slide/10174454/ (accessed Feb. 25, 2020).

[20] L. Gillis and M. Wilkie, “Peritoneal dialysis,” *Medicine (United Kingdom)*, vol. 47, no. 9, pp. 603–608, Sep. 01, 2019, doi: 10.1016/j.mpmed.2019.06.003.

[21] P. K.-T. Li, C. C. Szeto, B. Piraino, J. D. Arteaga, S. Fan, A. E. Figueiredo, D. N. Fish, E. Goffin, Y.-L. Kim, W. Salzer, D. G. Struijk, I. Teitelbaum, and D. W. Johnson, “ISPD peritonitis recommendations: 2016 update on prevention and treatment,” *Peritoneal Dialysis International*, vol. 36, no. 5, pp. 481–508, Sep. 01, 2016, doi: 10.3747/pdi.2016.00078.

[22] R. Mkahal, “extracopreal ciculation,” Sep. 23, 2015. https://www.slideshare.net/rayyanmkahal/extracopreal-ciculation (accessed Feb. 26, 2020).

[23] A. G. Nishio-Lucar, S. Bose, G. Lyons, K. T. Awuah, J. Z. Ma, and R. S. Lockridge, “Intensive Home Hemodialysis Survival Comparable to Deceased Donor Kidney Transplantation,” *Kidney International Reports*, vol. 5, no. 3, pp. 296–306, Mar. 2020, doi: 10.1016/j.ekir.2019.12.019.

[24] A. H. al Elq, “Symptomatic hypocalcemia associated with zoledronic acid treatment for Osteoporosis: A case report,” *Oman Medical Journal*, vol. 28, no. 2, pp. 1-3, 2013, doi: 10.5001/omj.2013.42.

[25] M. J. Lee, F. M. Doh, C. H. Kim, H. M. Koo, H. J. Oh, J. T. Park, S. H. Han, T.-H. Yoo, Y.-L. Kim, Y. S. Kim, C. W. Yang, N.-H. Kim, and S.-W. Kang, “Interdialytic weight gain and cardiovascular outcome in incident hemodialysis patients,” *American Journal of Nephrology*, vol. 39, no. 5, pp. 427–435, 2014, doi: 10.1159/000362743.

[26] B. N. Becker and G. Schulman, “Technical Aspects of Hemodialysis,” 2008. https://www.sciencedirect.com/topics/medicine-and-dentistry/dialysis-fluid (accessed Feb. 28, 2020).

[27] A. S. Levey and L. A. Stevens, “Estimating GFR Using the CKD Epidemiology Collaboration (CKD-EPI) Creatinine Equation: More Accurate GFR Estimates,

Lower CKD Prevalence Estimates, and Better Risk Predictions," *American Journal of Kidney Diseases*, vol. 55, no. 4, pp. 622–627, Apr. 2010, doi: 10.1053/j.ajkd.2010.02.337.

[28] T. Shafi, "Hemodialysis: Prescription and Assessment of Adequacy - Renal and Urology News," 2017. https://www.renalandurologynews.com/home/decision-support-in-medicine/nephrology-hypertension/hemodialysis-prescription-and-assessment-of-adequacy/ (accessed Feb. 28, 2020).

[29] J. T. Daugirdas, T. A. Depner, J. Inrig, R. Mehrotra, M. V. Rocco, R. S. Suri, D. E. Weiner *et al.*, "KDOQI Clinical Practice Guideline for Hemodialysis Adequacy: 2015 Update," *American Journal of Kidney Diseases*, vol. 66, no. 5, pp. 884–930, Nov. 2015, doi: 10.1053/j.ajkd.2015.07.015.

[30] C. Barbieri, F. Mari, A. Stopper, E. Gatti, P. Escandell-Montero, J. M. Martínez-Martínez, and J. D. Martín-Guerrero, "A new machine learning approach for predicting the response to anemia treatment in a large cohort of End Stage Renal Disease patients undergoing dialysis," *Computers in Biology and Medicine*, vol. 61, pp. 56–61, Jun. 2015, doi: 10.1016/j.compbiomed.2015.03.019.

[31] G. Toft, U. Heide-Jørgensen, H. V. Haalen, G. James, K. Hedman, H. Birn, C. F. Christiansen, and R. W. Thomsen, "Anemia and clinical outcomes in patients with non-dialysis dependent or dialysis dependent severe chronic kidney disease: a Danish population-based study," *Journal of Nephrology*, vol. 33, no. 1, pp. 147–156, Feb. 2020, doi: 10.1007/s40620-019-00652-9.

[32] C. Barbieri, I. Cattinelli, L. Neri, F. Mari, R. Ramos, D. Brancaccio, B. Canaud, and S. Stuard, "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 Diseases*, vol. 5, no. 1, pp. 28–33, 2019, doi: 10.1159/000493479.

[33] C. Argyropoulos, C. C. H. Chang, L. Plantinga, N. Fink, N. Powe, and M. Unruh, "Considerations in the statistical analysis of hemodialysis patient survival," *Journal of the American Society of Nephrology*, vol. 20, no. 9, pp. 2034–2043, Sep. 2009, doi: 10.1681/ASN.2008050551.

[34] T. H. Davenport, P. Barth, and R. Bean, "How 'Big Data' Is Different," 2012. https://sloanreview.mit.edu/article/how-big-data-is-different/ (accessed Feb. 20, 2020).

[35] H. Asri, H. Mousannif, H. A. Moatassime, and T. Noel, "Big data in healthcare: Challenges and opportunities," *2015 International Conference on Cloud Technologies and Applications (CloudTech)*, Marrakech, Morocco, 2015, pp. 1-7, doi: 10.1109/CloudTech.2015.7337020.

[36] P. Bisht, "Clarifying Differences between Data Analysis, Data Mining, Data Science, Machine Learning, and Big Data," May 14, 2019. https://www.houseofbots.com/news-detail/11973-1-clarifying-differences-between-data-analysis-data-mining-data-science-machine-learning,-and-big-data (accessed Feb. 20, 2020).

[37] P. Russom, "Introduction to Big Data Analytics | Big Data Analytics," 2011. https://vivomente.com/wp-content/uploads/2016/04/big-data-analytics-white-paper.pdf (accessed Feb. 25, 2020).

[38] A. Belle, R. Thiagarajan, S. M. R. Soroushmehr, F. Navidi, D. A. Beard, and K. Najarian, “Big data analytics in healthcare,” *BioMed Research International*, vol. 2015, no. 370194, pp. 16, 2015, doi: 10.1155/2015/370194.

[39] A. Kusiak, B. Dixon, and S. Shah, “Predicting survival time for kidney dialysis patients: A data mining approach,” *Computers in Biology and Medicine*, vol. 35, no. 4, pp. 311–327, May 2005, doi: 10.1016/j.compbiomed.2004.02.004.

[40] A. Burlacu, A. Iftene, E. Busoiu, D. Cogean, and A. Covic, “Challenging the supremacy of evidence-based medicine through artificial intelligence: The time has come for a change of paradigms,” *Nephrology Dialysis Transplantation*, vol. 35, no. 2, pp. 191–194, Feb. 01, 2020, doi: 10.1093/ndt/gfz203.

[41] M. Hueso and A. Vellido, “Artificial Intelligence and Dialysis,” *Kidney Diseases*, vol. 5, no. 1, pp. 1–2, 2019, doi: 10.1159/000493933.

[42] Y. Wang, L. A. Kung, and T. A. Byrd, “Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations,” *Technological Forecasting and Social Change*, vol. 126, pp. 3–13, Jan. 2018, doi: 10.1016/j.techfore.2015.12.019.

[43] B. Boukenze, H. Mousannif, and A. Haqiq, “Predictive Analytics in Healthcare System Using Data Mining Techniques,” *The Fourth International Conference on Database and Data Mining*, Settat, Morocco, 2016, pp. 1–9, doi: 10.5121/csit.2016.60501.

[44] A. Banerjee, A. Noor, N. Siddiqua and M. N. Uddin, "Significance of Attribute Selection In The Classification of Chronic Renal Disease," *2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP)*, Gangtok, India, 2019, pp. 1-6, doi: 10.1109/ICACCP.2019.8882937.

[45] Parul Sinha and Poonam Sinha, “Comparative Study of Chronic Kidney Disease Prediction using KNN and SVM,” *International Journal of Engineering Research and*, vol. 4, no. 12, pp. 608–612, Dec. 2015, doi: 10.17577/ijertv4is120622.

[46] Y. Amirgaliyev, S. Shamiluulu and A. Serek, "Analysis of Chronic Kidney Disease Dataset by Applying Machine Learning Methods," *2018 IEEE 12th International Conference on Application of Information and Communication Technologies (AICT)*, Almaty, Kazakhstan, 2018, pp. 1-4, doi: 10.1109/ICAICT.2018.8747140.

[47] Z. Wang, J. Won Chung, X. Jiang, Y. Cui, M. Wang, and A. Zheng, “Machine Learning-Based Prediction System For Chronic Kidney Disease Using Associative Classification Technique,” *International Journal of Engineering & Technology*, vol. 7, no. 4.36, pp. 1161, Dec. 2018, doi: 10.14419/ijet.v7i4.36.25377.

[48] M. P. N. M. Wickramasinghe, D. M. Perera, and K. A. D. C. P. Kahandawaarachchi, “Dietary prediction for patients with Chronic Kidney Disease (CKD) by considering blood potassium level using machine learning algorithms,” in *2017 IEEE Life Sciences Conference,* Sydney, NSW, Australia, 2017, pp. 300-303, doi: 10.1109/LSC.2017.8268202.

[49] A. Maurya, R. Wable, R. Shinde, S. John, R. Jadhav and R. Dakshayani, "Chronic Kidney Disease Prediction and Recommendation of Suitable Diet Plan by using Machine Learning," *2019 International Conference on Nascent*

Technologies in Engineering (ICNTE), Navi Mumbai, India, 2019, pp. 1-4, doi: 10.1109/ICNTE44896.2019.8946029.

[50] K. Shankar, P. Manickam, G. Devika and M. Ilayaraja, "Optimal Feature Selection for Chronic Kidney Disease Classification using Deep Learning Classifier," *2018 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, Madurai, India, 2018, pp. 1-5, doi: 10.1109/ICCIC.2018.8782340.

[51] M. E. Brier and A. E. Gaweda, "Artificial intelligence for optimal anemia management in end-stage renal disease," *Kidney International*, vol. 90, no. 2, pp. 259–261, Aug. 01, 2016, doi: 10.1016/j.kint.2016.05.018.

[52] P. Vinayagam, M. Sreemathi, K. Jeevitha, and S. Sandhya, "Kidney Stone Detection Using Neural Network," *International Journal of Applied Engineering Research*, vol. 14, no. 0973–4562, pp. 67–70, Nov. 2019, doi: 10.37622/000000.

[53] K. F. Erickson, S. Qureshi, and W. C. Winkelmayer, "The Role of Big Data in the Development and Evaluation of US Dialysis Care," *American Journal of Kidney Diseases*, vol. 72, no. 4, pp. 560–568, Oct. 2018, doi: 10.1053/j.ajkd.2018.04.007.

[54] E. T. Welsh, "Incremental dialysis may help maintain residual renal function, improve quality of life," Feb. 25, 2020. https://www.healio.com/news/nephrology/20200225/incremental-dialysis-may-help-maintain-residual-renal-function-improve-quality-of-life (accessed Mar. 8, 2021).

[55] N. E. Neumann, "ADC keynote: Big data in ESRD can be used to improve outcomes, but ties to policy may be troublesome," Mar. 29, 2019. 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 (accessed Mar. 8, 2021).

[56] J. Perl, L. M. Dember, J. M. Bargman, T. Browne, D. M. Charytan, J. E. Flythe, L. T. J. Hickson, A. M. Hung, M. Jadoul, T. C. Lee, K. B. Meyer, H. Moradi, T. Shafi, I. Teitelbaum, L. P. Wong, and C. T. Chan, "The use of a multidimensional measure of dialysis adequacy—moving beyond small solute kinetics," *Clinical Journal of the American Society of Nephrology*, vol. 12, no. 5, pp. 839–847, May 08, 2017, doi: 10.2215/CJN.08460816.

[57] R. Singh, "Exploratory Data Analysis(EDA) from Scratch | With Pythin Implementation," Aug. 27, 2020. https://www.analyticsvidhya.com/blog/2020/08/exploratory-data-analysiseda-from-scratch-in-python/ (accessed Mar. 12, 2021).

[58] I. Dixit, "Demystifying Support Vector Machine Part I | by Ishan Dixit | The Startup | Medium," Dec. 07, 2020. https://medium.com/swlh/demystifying-support-vector-machine-part-i-b5b083844c9a (accessed Mar. 12, 2021).

[59] M. Kumar, "Regression Analysis – linear regression, SSE, Assumption of linear regression Error term and best Fit line « Madhuresh Kumar," Jul. 21, 2015. https://madhureshkumar.wordpress.com/2015/07/21/regression-analysis-linear-regression-sse-assumption-of-linear-regression-error-term-and-best-fit-line/ (accessed Mar. 12, 2021).

Appendix A

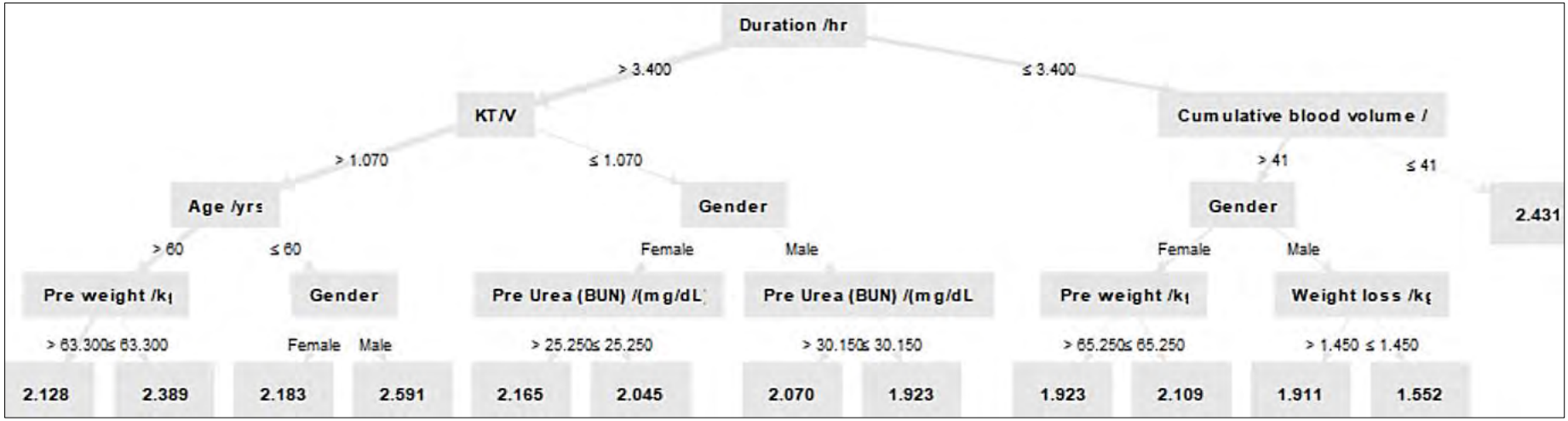

Figure A.1: DT model output for calcium.

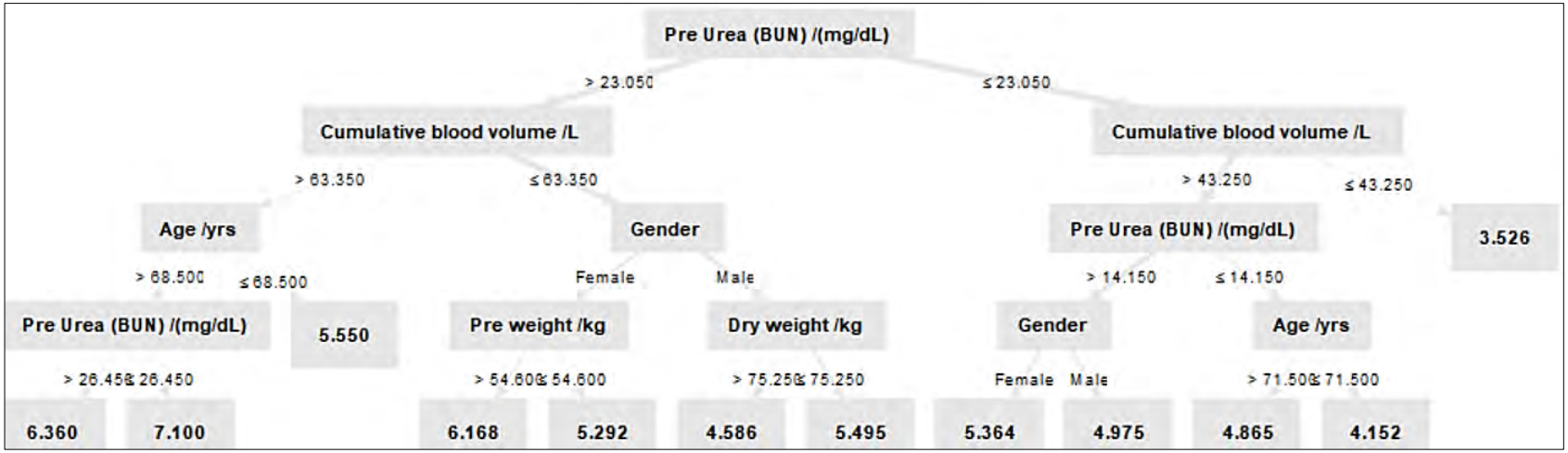

Figure A.2: DT model output for potassium.

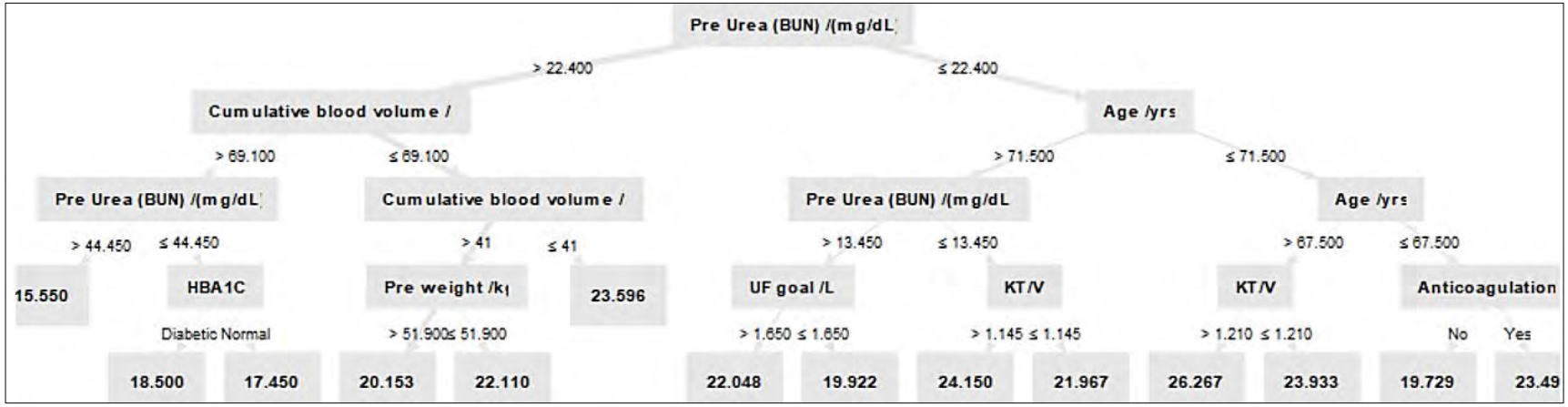

Figure A.3: DT model output for carbon dioxide.

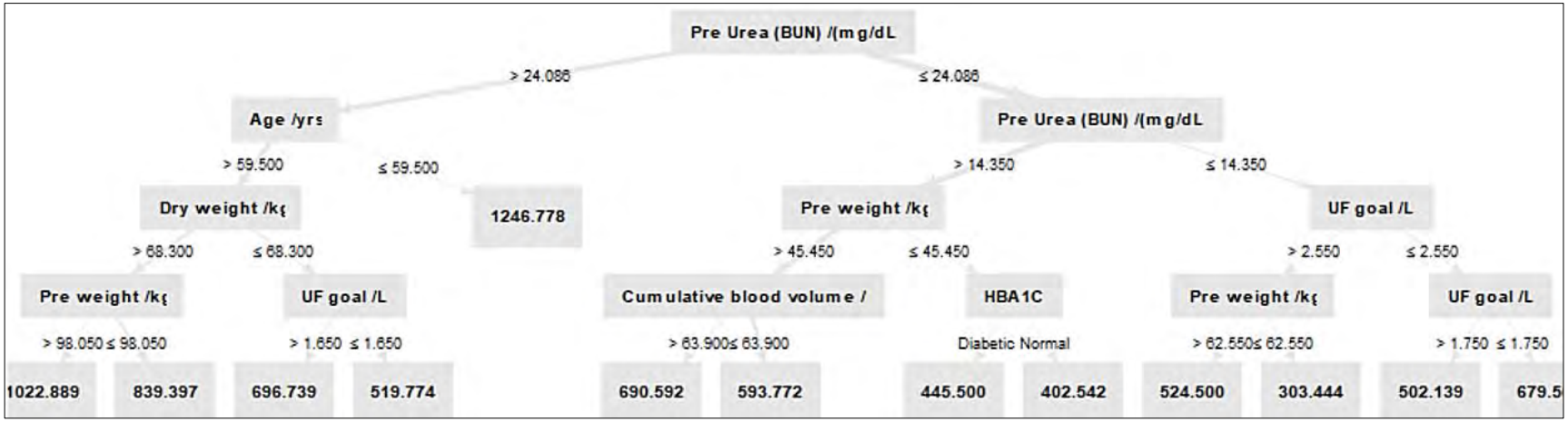

Figure A.4: Decision Tree model output for creatinine plasma.

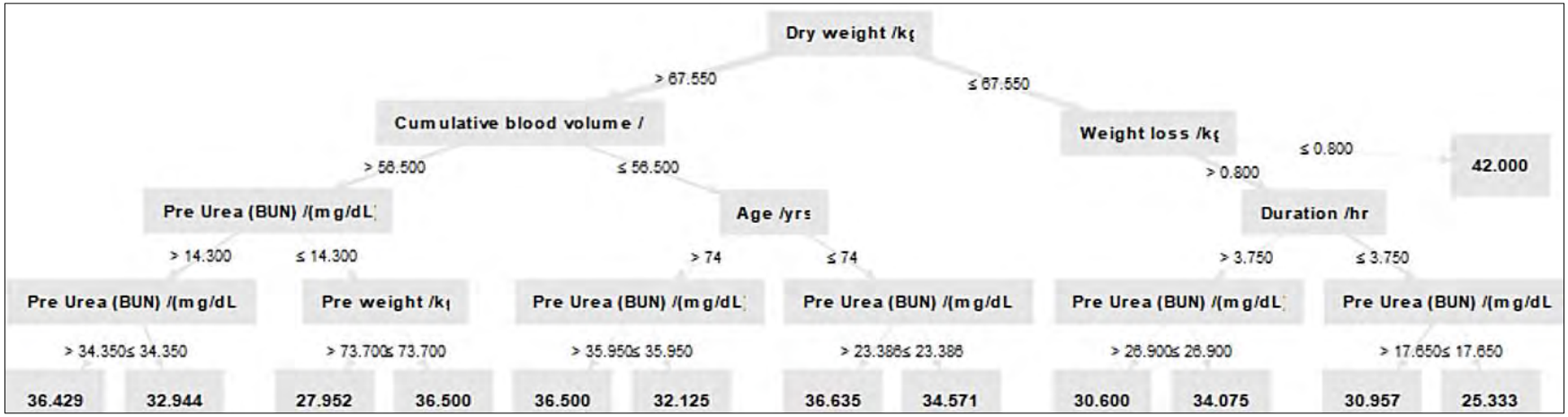

Figure A.5: DT model output for albumin.

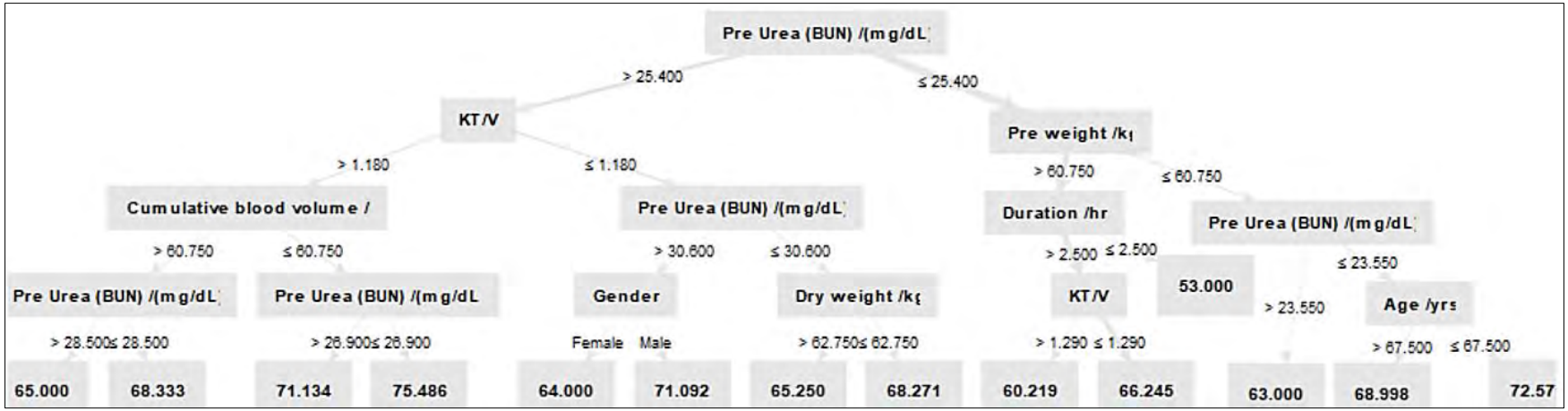

Figure A.6: Decision Tree model output for protein total.

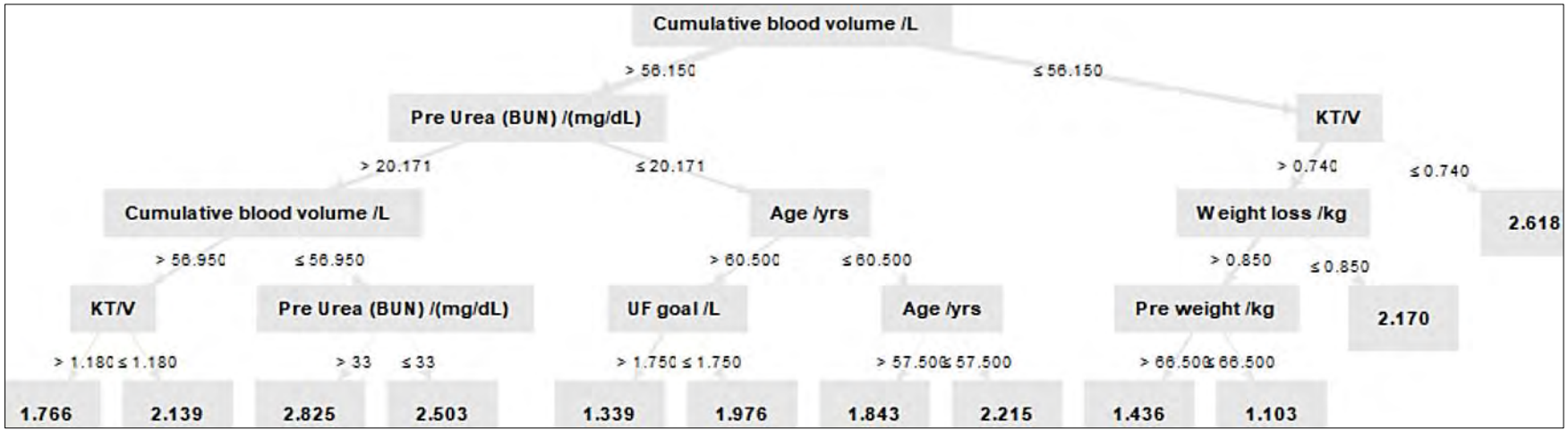

Figure A.7: DT model output for phosphate.

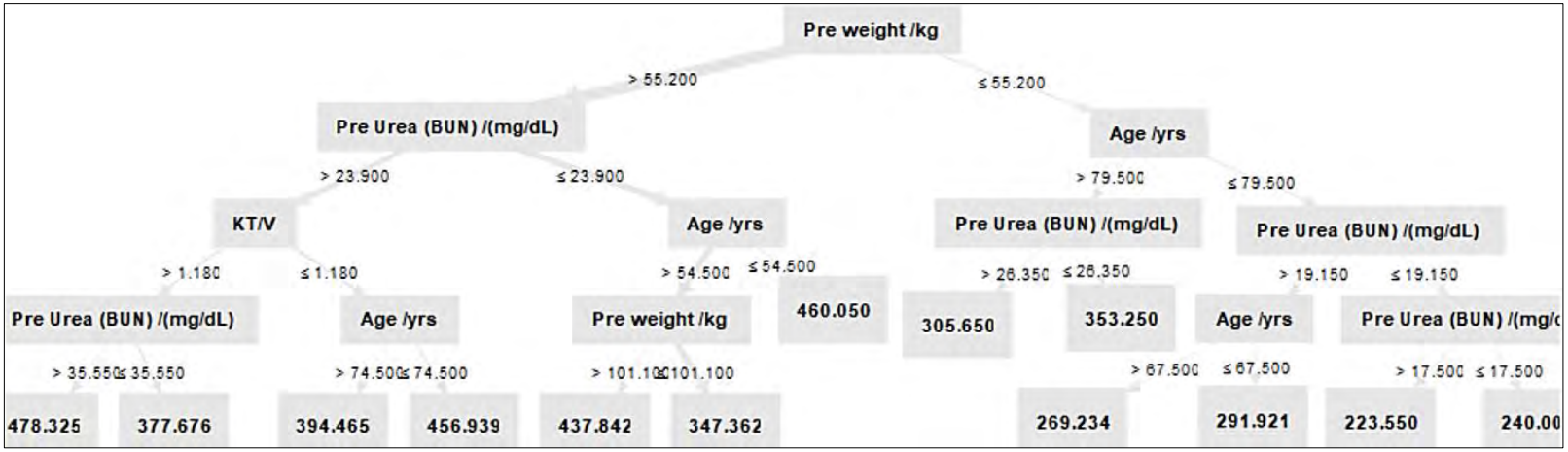

Figure A.8: Decision Tree model output for uric acid.

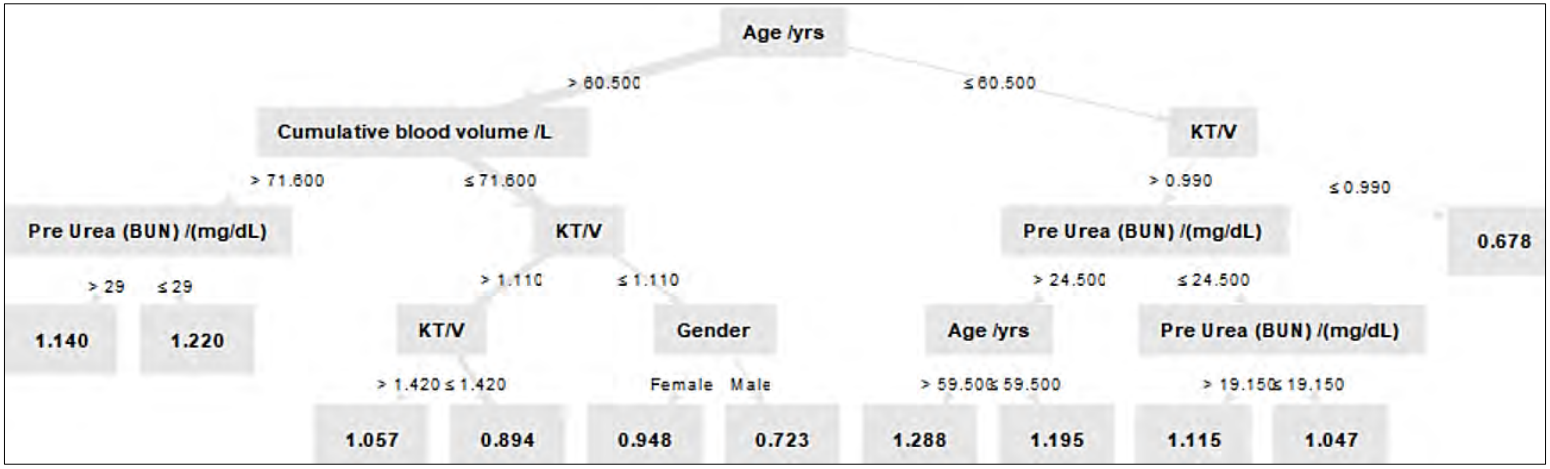

Figure A.9: DT model output for magnesium.

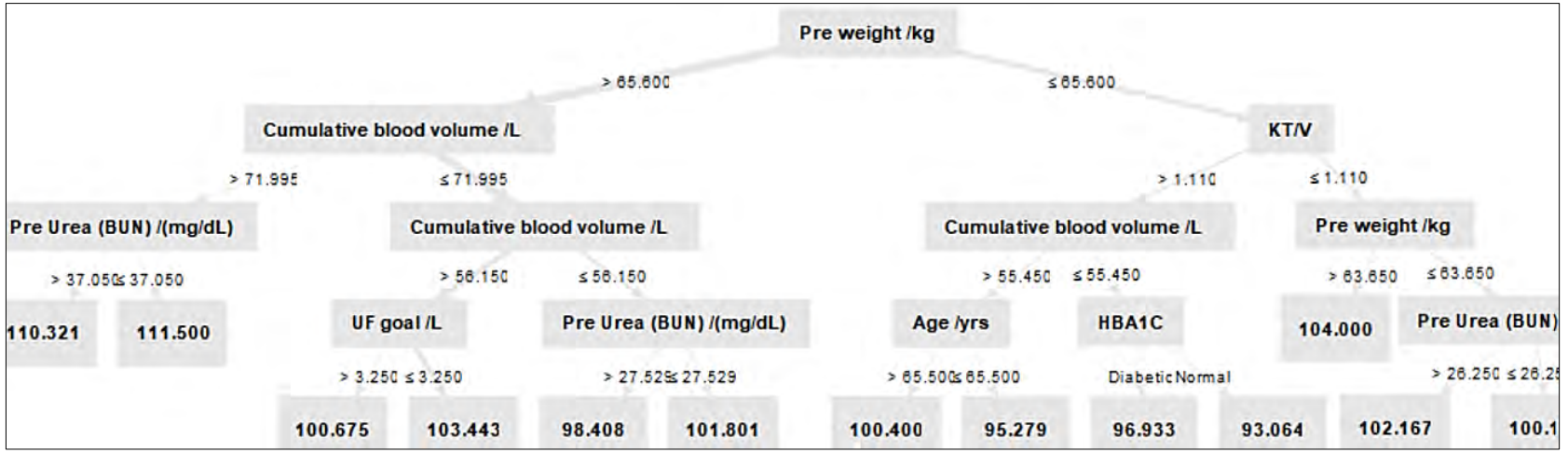

Figure A.10: Decision Tree model output for chloride plasma.

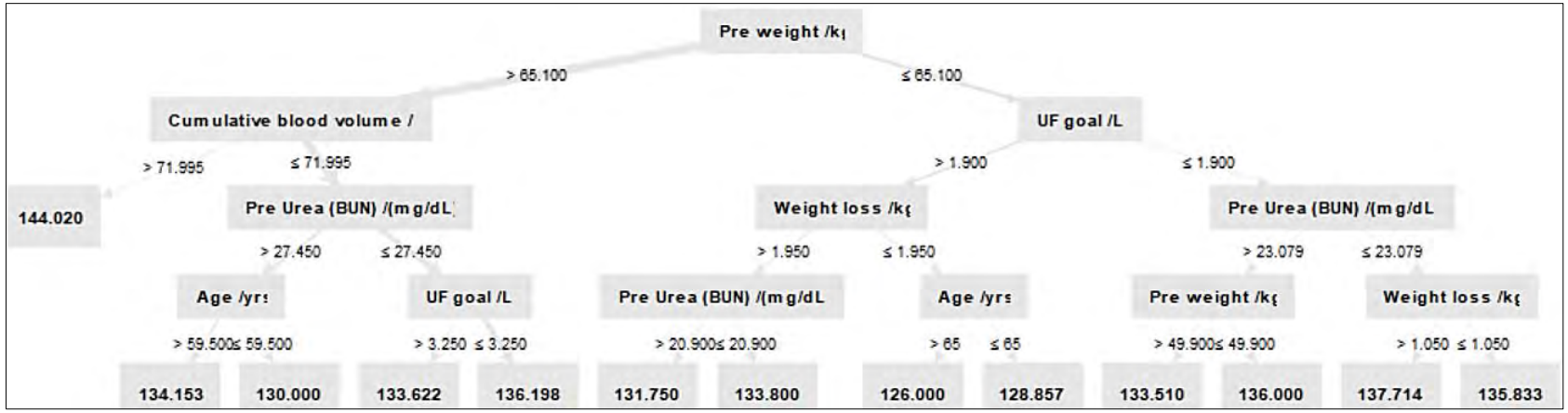

Figure A.11: DT model output for sodium plasma.

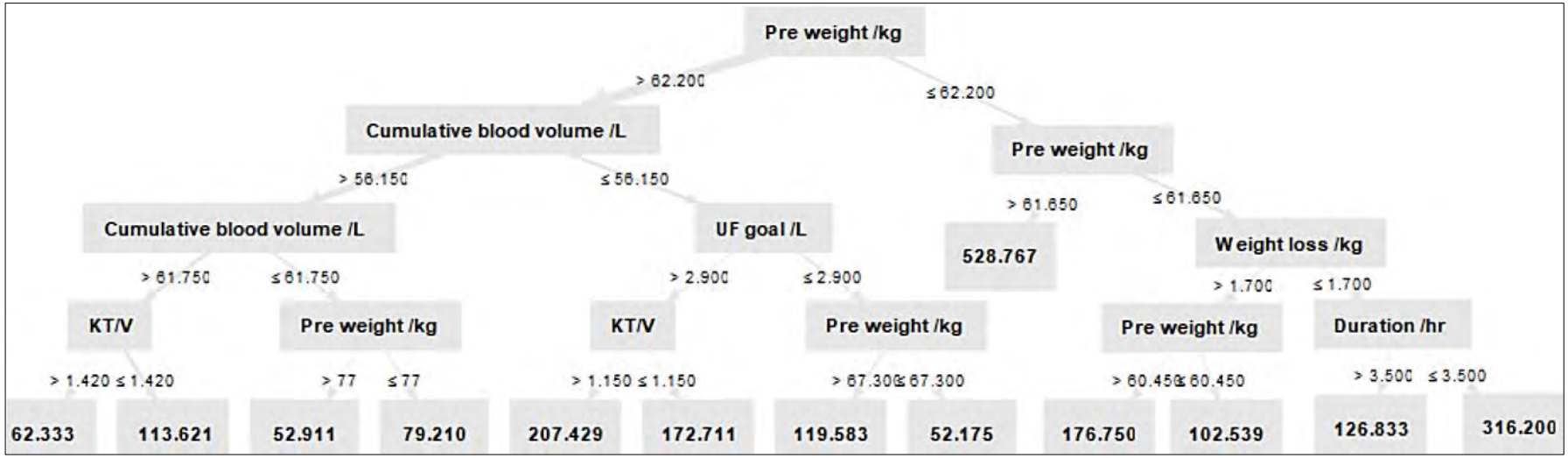

Figure A.12: DT model output for alkaline phosphatase.

Attribute	Coefficient	Std. Error	Std. Coefficient	Tolerance	t-Stat	p-Value ↑	Code
Pre Urea (BUN) /(mg/...	0.010	0.002	0.361	0.912	5.818	0.000	****
Pre weight /kg	-0.045	0.009	-5.121	0.998	-5.257	0.000	****
Dry weight /kg	0.043	0.008	4.918	0.995	5.101	0.000	****
Cumulative blood vol...	0.008	0.002	0.351	0.963	4.176	0.000	****
Age /yrs	-0.006	0.002	-0.282	0.999	-3.983	0.000	****
Gender = Female	0.088	0.026	0.236	0.920	3.423	0.001	****
Blood flow /(ml/min)	0.007	0.002	0.214	1.000	3.319	0.001	***
Weight loss /kg	0.190	0.062	0.659	0.990	3.086	0.002	***
Duration /hr	-0.097	0.035	-0.244	0.905	-2.742	0.007	***
KT/V	0.197	0.083	0.167	0.983	2.383	0.018	**
(Intercept)	-1.178	0.571	?	?	-2.063	0.040	**
UF goal /L	-0.109	0.059	-0.385	0.996	-1.845	0.066	*
Anticoagulation = No	-0.017	0.030	-0.036	0.922	-0.561	0.575	
HBA1C = Diabetic	0.012	0.047	0.015	0.997	0.253	0.801	

Figure A.13: LR model output for magnesium.

Attribute	Coefficient	Std. Error	Std. Coefficient	Tolerance	t-Stat	p-Value ↑	Code
(Intercept)	78.042	12.590	?	?	6.199	0.000	****
Cumulative blood vol...	0.215	0.052	0.377	0.936	4.113	0.000	****
KT/V	-5.700	2.208	-0.198	0.987	-2.581	0.011	**
Pre Urea (BUN) /(mg/...	0.105	0.045	0.160	1.000	2.312	0.022	**
Gender = Female	-1.409	0.699	-0.154	0.982	-2.018	0.045	**
HBA1C = Diabetic	-1.749	1.531	-0.072	0.998	-1.143	0.255	
Duration /hr	-0.873	0.980	-0.087	0.839	-0.891	0.374	
Dry weight /kg	0.203	0.237	0.802	0.564	0.857	0.392	
Blood flow /(ml/min)	0.040	0.047	0.066	0.992	0.853	0.395	
Anticoagulation = No	0.716	0.850	0.060	0.991	0.843	0.400	
UF goal /L	1.212	1.639	0.171	0.962	0.739	0.460	
Weight loss /kg	-1.029	1.674	-0.142	0.902	-0.615	0.539	
Pre weight /kg	-0.140	0.242	-0.553	0.564	-0.579	0.563	
Age /yrs	0.019	0.041	0.037	0.963	0.459	0.647	

Figure A.14: LR model output for chloride plasma.

Attribute	Coefficient	Std. Error	Std. Coefficient	Tolerance	t-Stat	p-Value ↑	Code
(Intercept)	125.125	12.407	?	?	10.085	0	****
Cumulative blood vol...	0.129	0.052	0.239	0.841	2.478	0.014	**
Pre Urea (BUN) /(mg/...	0.082	0.044	0.131	0.997	1.858	0.065	*
Age /yrs	0.062	0.039	0.126	0.995	1.578	0.116	
Anticoagulation = No	0.991	0.829	0.084	0.999	1.195	0.233	
Gender = Female	-0.799	0.681	-0.090	0.962	-1.173	0.242	
KT/V	-1.871	2.396	-0.068	0.991	-0.781	0.436	
Dry weight /kg	0.149	0.225	0.662	0.381	0.662	0.509	
HBA1C = Diabetic	-0.890	1.368	-0.042	0.987	-0.651	0.516	
Blood flow /(ml/min)	-0.023	0.046	-0.039	0.986	-0.505	0.614	
Weight loss /kg	-0.838	1.822	-0.118	0.932	-0.460	0.646	
UF goal /L	0.602	1.695	0.089	0.983	0.355	0.723	
Pre weight /kg	-0.080	0.230	-0.354	0.397	-0.347	0.729	
Duration /hr	0.138	1.017	0.015	0.742	0.135	0.893	

Figure A.15: LR model output for sodium plasma.

Attribute	Coefficient	Std. Error	Std. Coefficient	Tolerance	t-Stat	p-Value ↑	Code
(Intercept)	66.874	13.694	?	?	4.884	0.000	****
Pre Urea (BUN) /(mg/...	0.159	0.047	0.245	0.986	3.382	0.001	****
Duration /hr	2.840	1.034	0.280	0.999	2.746	0.007	***
Age /yrs	-0.097	0.046	-0.173	0.948	-2.111	0.036	**
Gender = Female	1.487	0.747	0.156	1.000	1.992	0.048	**
KT/V	-4.843	2.475	-0.155	0.967	-1.957	0.052	*
Weight loss /kg	-3.053	1.765	-0.416	0.997	-1.730	0.085	*
UF goal /L	2.406	1.700	0.330	0.984	1.415	0.158	
Anticoagulation = No	-1.193	0.909	-0.095	1.000	-1.312	0.191	
HBA1C = Diabetic	1.039	1.387	0.050	0.999	0.749	0.455	
Cumulative blood vol...	-0.038	0.054	-0.068	0.995	-0.706	0.481	
Dry weight /kg	-0.104	0.246	-0.399	0.925	-0.423	0.673	
Pre weight /kg	0.064	0.252	0.245	0.926	0.254	0.799	
Blood flow /(ml/min)	0.012	0.052	0.019	0.986	0.232	0.817	

Figure A.16: LR model output for protein.

Attribute	Coefficient	Std. Error	Std. Coefficient	Tolerance	t-Stat	p-Value ↑	Code
Pre Urea (BUN) /(mg/...	0.038	0.005	0.469	0.997	7.736	0.000	****
Cumulative blood vol...	0.035	0.005	0.534	0.996	6.636	0.000	****
KT/V	-0.907	0.236	-0.281	0.999	-3.844	0.000	****
Blood flow /(ml/min)	0.019	0.005	0.258	0.998	3.815	0.000	****
Age /yrs	-0.016	0.004	-0.252	0.998	-3.501	0.001	****
(Intercept)	-4.496	1.346	?	?	-3.340	0.001	****
Pre weight /kg	0.070	0.025	2.273	0.950	2.767	0.006	***
Dry weight /kg	-0.068	0.025	-2.196	0.945	-2.712	0.007	***
Gender = Female	0.191	0.076	0.172	0.952	2.521	0.012	**
UF goal /L	-0.252	0.176	-0.290	0.985	-1.436	0.152	
Duration /hr	-0.052	0.103	-0.045	0.995	-0.509	0.611	
Weight loss /kg	-0.047	0.182	-0.052	0.987	-0.257	0.797	
Anticoagulation = No	0.016	0.092	0.010	0.937	0.170	0.866	
HBA1C = Diabetic	0.011	0.130	0.005	1.000	0.083	0.934	

Figure A.17: LR model output for phosphate.

Attribute	Coefficient	Std. Error	Std. Coefficient	Tolerance	t-Stat	p-Value ↑	Code
(Intercept)	931.496	153.742	?	?	6.059	0.000	****
Pre Urea (BUN) /(mg/...	2.863	0.571	0.275	0.923	5.017	0.000	****
KT/V	-150.277	30.355	-0.326	0.996	-4.951	0.000	****
Anticoagulation = No	49.015	10.839	0.270	0.997	4.522	0.000	****
Blood flow /(ml/min)	-2.017	0.580	-0.228	0.979	-3.475	0.001	****
Gender = Female	-29.602	8.723	-0.205	0.968	-3.394	0.001	****
UF goal /L	52.022	20.735	0.474	0.966	2.509	0.013	**
Weight loss /kg	-51.316	21.786	-0.454	0.897	-2.355	0.019	**
Duration /hr	11.736	11.361	0.074	0.991	1.033	0.303	
HBA1C = Diabetic	15.400	15.380	0.052	0.995	1.001	0.318	
Pre weight /kg	1.625	3.052	0.416	0.656	0.533	0.595	
Age /yrs	-0.167	0.515	-0.021	0.928	-0.324	0.746	
Dry weight /kg	-0.243	2.976	-0.062	0.666	-0.082	0.935	

Figure A.18: LR model output for uric acid.

Attribute	Coefficient	Std. Error	Std. Coefficient	Tolerance	t-Stat	p-Value ↑	Code
Pre Urea (BUN) /(mg/...	-0.168	0.024	-0.474	0.927	-7.125	0.000	****
(Intercept)	40.926	5.811	?	?	7.043	0.000	****
Duration /hr	-1.231	0.521	-0.228	0.964	-2.363	0.019	**
KT/V	2.539	1.168	0.167	0.990	2.174	0.031	**
Blood flow /(ml/min)	-0.043	0.023	-0.151	0.972	-1.887	0.061	*
Gender = Female	-0.406	0.370	-0.082	0.971	-1.098	0.274	
Cumulative blood vol...	-0.018	0.026	-0.059	0.989	-0.670	0.504	
Age /yrs	-0.009	0.022	-0.032	0.995	-0.405	0.686	
Weight loss /kg	0.242	0.891	0.062	0.914	0.271	0.786	
HBA1C = Diabetic	0.159	0.725	0.013	0.999	0.220	0.826	
Anticoagulation = No	0.057	0.446	0.009	0.998	0.127	0.899	
Pre weight /kg	-0.010	0.122	-0.072	0.908	-0.080	0.936	
UF goal /L	-0.054	0.875	-0.014	0.901	-0.062	0.951	
Dry weight /kg	0.006	0.120	0.043	0.918	0.049	0.961	

Figure A.19: LR model output for carbon dioxide.

Attribute	Coefficient	Std. Error	Std. Coefficient	Tolerance	t-Stat	p-Value ↑	Code
Pre Urea (BUN) /(mg/...	0.059	0.007	0.512	0.979	8.059	0.000	****
Duration /hr	-0.631	0.156	-0.365	1.000	-4.035	0.000	****
Cumulative blood vol...	0.034	0.008	0.337	0.952	3.992	0.000	****
Gender = Female	0.248	0.111	0.149	0.990	2.243	0.026	**
Age /yrs	0.015	0.007	0.156	0.995	2.161	0.032	**
UF goal /L	0.356	0.282	0.264	0.842	1.262	0.208	
(Intercept)	2.196	2.091	?	?	1.050	0.295	
Pre weight /kg	0.021	0.038	0.494	0.985	0.564	0.573	
Dry weight /kg	-0.017	0.037	-0.403	0.994	-0.464	0.643	
HBA1C = Diabetic	-0.090	0.206	-0.025	0.996	-0.435	0.664	
KT/V	-0.131	0.378	-0.025	0.968	-0.346	0.730	
Weight loss /kg	-0.098	0.291	-0.069	0.797	-0.336	0.737	
Anticoagulation = No	-0.047	0.152	-0.019	1.000	-0.305	0.760	
Blood flow /(ml/min)	-0.000	0.008	-0.001	0.937	-0.015	0.988	

Figure A.20: LR model output for potassium.

Attribute	Coefficient	Std. Error	Std. Coefficient	Tolerance	t-Stat	p-Value ↑	Code
Anticoagulation = No	-0.192	0.042	-0.317	1.000	-4.541	0.000	****
Gender = Female	0.137	0.036	0.283	0.999	3.760	0.000	****
Duration /hr	0.143	0.049	0.289	0.823	2.896	0.004	***
UF goal /L	0.174	0.086	0.468	0.883	2.025	0.044	**
Pre weight /kg	-0.021	0.012	-1.717	0.995	-1.758	0.080	*
Dry weight /kg	0.020	0.012	1.605	0.992	1.669	0.097	*
Blood flow /(ml/min)	0.003	0.002	0.122	0.984	1.494	0.137	
Age /yrs	-0.003	0.002	-0.100	0.992	-1.226	0.221	
Weight loss /kg	-0.102	0.087	-0.276	0.956	-1.172	0.243	
(Intercept)	0.671	0.576	?	?	1.165	0.245	
Pre Urea (BUN) /(mg/...	0.002	0.002	0.069	0.993	1.055	0.293	
Cumulative blood vol...	-0.000	0.003	-0.017	0.862	-0.185	0.853	
HBA1C = Diabetic	0.007	0.066	0.006	0.999	0.105	0.916	
KT/V	0.010	0.112	0.007	0.862	0.089	0.930	

Figure A.21: LR model output for calcium.

Attribute	Coefficient	Std. Error	Std. Coefficient	Tolerance	t-Stat	p-Value ↑	Code
Pre Urea (BUN) /(mg/...	15.943	1.852	0.522	0.849	8.607	0.000	****
Gender = Female	-109.311	28.876	-0.245	0.948	-3.786	0.000	****
Blood flow /(ml/min)	5.589	1.844	0.205	0.999	3.030	0.003	***
(Intercept)	-1198.122	485.577	?	?	-2.467	0.014	**
Age /yrs	-3.364	1.758	-0.131	1.000	-1.913	0.057	*
Cumulative blood vol...	3.378	2.203	0.123	0.991	1.533	0.127	
UF goal /L	98.807	65.247	0.281	0.863	1.514	0.131	
KT/V	-124.585	90.411	-0.095	0.995	-1.378	0.170	
Pre weight /kg	-13.462	10.247	-1.155	0.888	-1.314	0.190	
Dry weight /kg	13.051	10.001	1.128	0.907	1.305	0.193	
Anticoagulation = No	43.640	35.140	0.074	0.982	1.242	0.216	
HBA1C = Diabetic	-63.902	55.352	-0.063	0.993	-1.154	0.250	
Weight loss /kg	-47.215	68.470	-0.133	0.883	-0.690	0.491	
Duration /hr	2.519	40.178	0.005	0.955	0.063	0.950	

Figure A.22: LR model output for creatinine plasma.

Attribute	Coefficient	Std. Error	Std. Coefficient	Tolerance	t-Stat	p-Value ↑	Code
Pre Urea (BUN) /(mg/...	0.109	0.048	0.160	0.912	2.278	0.024	**
Cumulative blood vol...	-0.121	0.054	-0.209	0.993	-2.237	0.026	**
Age /yrs	-0.092	0.046	-0.176	0.957	-2.001	0.047	**
Duration /hr	2.348	1.184	0.224	0.963	1.984	0.049	**
Blood flow /(ml/min)	0.087	0.052	0.136	0.999	1.671	0.096	*
KT/V	3.865	2.553	0.130	0.924	1.514	0.132	
Pre weight /kg	0.345	0.252	1.441	0.830	1.370	0.172	
Dry weight /kg	-0.319	0.248	-1.336	0.822	-1.285	0.200	
UF goal /L	-2.162	1.834	-0.300	0.993	-1.179	0.240	
Anticoagulation = No	-1.118	0.960	-0.084	0.989	-1.164	0.246	
Gender = Female	-0.641	0.748	-0.067	0.938	-0.857	0.392	
HBA1C = Diabetic	1.165	1.385	0.055	0.986	0.841	0.401	
(Intercept)	6.384	13.711	?	?	0.466	0.642	
Weight loss /kg	0.602	1.846	0.082	0.999	0.326	0.745	

Figure A.23: LR model output for albumin.

Attribute	Coefficient	Std. Error	Std. Coefficient	Tolerance	t-Stat	p-Value ↑	Code
Weight loss /kg	-154.809	36.091	-1.008	0.871	-4.289	0.000	****
UF goal /L	145.494	34.881	0.988	0.845	4.171	0.000	****
Duration /hr	-60.916	20.345	-0.302	0.972	-2.994	0.003	***
Cumulative blood vol...	2.890	1.075	0.260	0.921	2.688	0.008	***
(Intercept)	410.245	224.805	?	?	1.825	0.069	*
KT/V	-83.979	48.740	-0.140	0.998	-1.723	0.086	*
Pre Urea (BUN) /(mg/...	-1.127	0.960	-0.083	0.974	-1.174	0.242	
Anticoagulation = No	-19.380	18.590	-0.076	0.943	-1.043	0.298	
Gender = Female	-14.596	14.616	-0.076	0.961	-0.999	0.319	
Pre weight /kg	-2.669	4.992	-0.500	0.634	-0.535	0.593	
HBA1C = Diabetic	-13.774	28.328	-0.031	0.961	-0.486	0.627	
Age /yrs	-0.393	0.887	-0.037	0.996	-0.443	0.658	
Dry weight /kg	2.020	4.886	0.376	0.645	0.413	0.680	
Blood flow /(ml/min)	-0.181	0.880	-0.018	0.995	-0.206	0.837	

Figure A.24: LR model output for alkaline phosphatase.

attribute	weight ↓
Pre Urea (BUN) /(mg/dL)	0.547
Anticoagulation = No	0.261
UF goal /L	0.142
Blood flow /(ml/min)	0.138
KT/V	0.123
Gender = Female	0.108
HBA1C = Diabetic	0.081
Cumulative blood volume /L	0.056
Duration /hr	0.025
Dry weight /kg	0.021
Pre weight /kg	0.009
Age /yrs	0.008
Weight loss /kg	0

attribute	weight ↓
KT/V	0.448
Duration /hr	0.315
Blood flow /(ml/min)	0.236
Pre weight /kg	0.223
Dry weight /kg	0.115
Pre Urea (BUN) /(mg/dL)	0.086
Gender = Female	0
HBA1C = Diabetic	0
Anticoagulation = No	0
Age /yrs	0
Weight loss /kg	0
UF goal /L	0
Cumulative blood volume /L	0

attribute	weight ↓
UF goal /L	0.350
Cumulative blood volume /L	0.234
Blood flow /(ml/min)	0.081
Gender = Female	0
HBA1C = Diabetic	0
Anticoagulation = No	0
KT/V	0
Age /yrs	0
Duration /hr	0
Pre weight /kg	0
Weight loss /kg	0
Pre Urea (BUN) /(mg/dL)	0
Dry weight /kg	0

Figure A.25: NN model weight output for creatinine plasma (left), albumin (middle), and alkaline phosphatase (right).

attribute	weight ↓
Anticoagulation = No	0.575
Cumulative blood volume /L	0.181
Pre Urea (BUN) /(mg/dL)	0.168
Age /yrs	0.160
Pre weight /kg	0.109
Weight loss /kg	0.101
Dry weight /kg	0.094
KT/V	0.069
Duration /hr	0.055
Gender = Female	0
HBA1C = Diabetic	0
UF goal /L	0
Blood flow /(ml/min)	0

attribute	weight ↓
Gender = Female	0.597
Pre Urea (BUN) /(mg/dL)	0.269
Duration /hr	0.082
Blood flow /(ml/min)	0.069
UF goal /L	0.050
Dry weight /kg	0.046
Pre weight /kg	0.001
HBA1C = Diabetic	0
Anticoagulation = No	0
KT/V	0
Age /yrs	0
Weight loss /kg	0
Cumulative blood volume /L	0

attribute	weight ↓
KT/V	0.342
HBA1C = Diabetic	0.147
Anticoagulation = No	0.061
Weight loss /kg	0.019
Pre weight /kg	0.018
UF goal /L	0.009
Gender = Female	0
Age /yrs	0
Duration /hr	0
Blood flow /(ml/min)	0
Cumulative blood volume /L	0
Pre Urea (BUN) /(mg/dL)	0
Dry weight /kg	0

Figure A.26: NN model weight output for sodium plasma (left), protein total (middle), and carbon dioxide (right).

attribute	weight ↓
Cumulative blood volume /L	0.302
Dry weight /kg	0.193
Pre weight /kg	0.156
Pre Urea (BUN) /(mg/dL)	0.125
Duration /hr	0.018
Blood flow /(ml/min)	0.012
Gender = Female	0.004
Age /yrs	0.003
Anticoagulation = No	0.001
HBA1C = Diabetic	0
KT/V	0
Weight loss /kg	0
UF goal /L	0

attribute	weight ↓
Pre Urea (BUN) /(mg/dL)	0.598
Gender = Female	0.357
Cumulative blood volume /L	0.278
HBA1C = Diabetic	0.203
Weight loss /kg	0.186
Blood flow /(ml/min)	0.184
Anticoagulation = No	0.171
Pre weight /kg	0.045
Duration /hr	0.008
UF goal /L	0.003
Dry weight /kg	0.002
Age /yrs	0.000
KT/V	0

attribute	weight ↓
Pre Urea (BUN) /(mg/dL)	0.763
Gender = Female	0.268
Cumulative blood volume /L	0.241
Weight loss /kg	0.170
Blood flow /(ml/min)	0.138
Age /yrs	0.133
UF goal /L	0.120
Dry weight /kg	0.040
Pre weight /kg	0.022
HBA1C = Diabetic	0.000
Anticoagulation = No	0
KT/V	0
Duration /hr	0

Figure A.27: NN model weight output for chloride plasma (left), phosphate (middle), and potassium (right).

attribute	weight ↓
Gender = Female	0.664
Blood flow /(ml/min)	0.141
Cumulative blood volume /L	0.093
KT/V	0.082
Pre Urea (BUN) /(mg/dL)	0.052
Dry weight /kg	0.019
UF goal /L	0.002
HBA1C = Diabetic	0
Anticoagulation = No	0
Age /yrs	0
Duration /hr	0
Pre weight /kg	0
Weight loss /kg	0

attribute	weight ↓
Anticoagulation = No	0.460
Pre Urea (BUN) /(mg/dL)	0.431
HBA1C = Diabetic	0.315
Pre weight /kg	0.103
Dry weight /kg	0.084
Cumulative blood volume /L	0.065
Duration /hr	0.049
UF goal /L	0.039
Gender = Female	0.037
Weight loss /kg	0.024
Age /yrs	0.005
Blood flow /(ml/min)	0.005
KT/V	0

attribute	weight ↓
Gender = Female	0.637
Blood flow /(ml/min)	0.142
Duration /hr	0.111
UF goal /L	0.107
Dry weight /kg	0.042
Pre Urea (BUN) /(mg/dL)	0.026
HBA1C = Diabetic	0.015
Anticoagulation = No	0
KT/V	0
Age /yrs	0
Pre weight /kg	0
Weight loss /kg	0
Cumulative blood volume /L	0

Figure A.28: NN model weight output for magnesium (left), uric acid (middle), and calcium (right).

Attribute	Weight ↓
Pre Urea (BUN) /(mg/dL)	2.271
KT/V	2.024
Cumulative blood volume /L	1.473
Blood flow /(ml/min)	1.461
Gender = Female	1.279
Weight loss /kg	0.818
Duration /hr	0.801
UF goal /L	0.664
Dry weight /kg	0.252
Pre weight /kg	0.190
Anticoagulation = No	-0.004
Age /yrs	-0.613
HBA1C = Diabetic	-1.313

Attribute	Weight ↓
Pre weight /kg	52.628
Dry weight /kg	51.490
Pre Urea (BUN) /(mg/dL)	51.072
Weight loss /kg	20.452
UF goal /L	15.592
Duration /hr	12.602
Anticoagulation = No	12.242
HBA1C = Diabetic	11.563
Age /yrs	-9.998
Cumulative blood volume /L	-12.021
Gender = Female	-19.186
KT/V	-26.533
Blood flow /(ml/min)	-30.833

Attribute	Weight ↓
Duration /hr	5.754
UF goal /L	3.474
KT/V	3.027
Cumulative blood volume /L	1.925
Blood flow /(ml/min)	1.816
Weight loss /kg	1.557
Gender = Female	1.354
HBA1C = Diabetic	1.057
Pre Urea (BUN) /(mg/dL)	0.622
Dry weight /kg	-0.072
Pre weight /kg	-0.078
Age /yrs	-2.532
Anticoagulation = No	-3.299

Figure A.29: SVM model output for magnesium (left), uric acid (middle), and calcium (right).

Attribute	Weight ↓
Dry weight /kg	37.101
Pre weight /kg	36.959
Cumulative blood volume /L	35.425
Duration /hr	25.539
Pre Urea (BUN) /(mg/dL)	18.354
Weight loss /kg	14.746
UF goal /L	11.002
Age /yrs	8.219
Blood flow /(ml/min)	-0
Anticoagulation = No	-5.238
KT/V	-8.092
HBA1C = Diabetic	-8.498
Gender = Female	-11.762

Attribute	Weight ↓
Pre Urea (BUN) /(mg/dL)	16.179
Cumulative blood volume /L	8.970
Pre weight /kg	8.204
Blood flow /(ml/min)	7.888
Dry weight /kg	7.799
Anticoagulation = No	5.125
Duration /hr	3.989
Weight loss /kg	3.086
UF goal /L	2.231
Age /yrs	1.381
Gender = Female	1.159
HBA1C = Diabetic	-1.971
KT/V	-2.218

Attribute	Weight ↓
Pre Urea (BUN) /(mg/dL)	24.194
Weight loss /kg	10.891
UF goal /L	10.246
Pre weight /kg	5.711
Blood flow /(ml/min)	5.223
Age /yrs	5.130
Dry weight /kg	4.760
Anticoagulation = No	1.998
Gender = Female	1.775
Cumulative blood volume /L	1.703
Duration /hr	-1.441
KT/V	-3.292
HBA1C = Diabetic	-8.270

Figure A.30: SVM model weight output for chloride plasma (left), phosphate (middle), and potassium (right).

Attribute	Weight ↓
Dry weight /kg	35.279
Pre weight /kg	35.236
Cumulative blood volume /L	32.143
Duration /hr	27.784
Weight loss /kg	18.840
Pre Urea (BUN) /(mg/dL)	15.535
UF goal /L	14.444
Age /yrs	6.447
HBA1C = Diabetic	4.604
Blood flow /(ml/min)	0.000
Anticoagulation = No	0
KT/V	-0.604
Gender = Female	-6.640

Attribute	Weight ↓
Pre Urea (BUN) /(mg/dL)	19.658
Gender = Female	7.243
UF goal /L	3.063
KT/V	1.916
Duration /hr	1.498
Weight loss /kg	-4.058
Cumulative blood volume /L	-5.564
Anticoagulation = No	-8.161
HBA1C = Diabetic	-8.367
Blood flow /(ml/min)	-8.661
Age /yrs	-13.065
Pre weight /kg	-16.126
Dry weight /kg	-16.495

Attribute	Weight ↑
Pre Urea (BUN) /(mg/dL)	-51.573
Duration /hr	-22.649
Pre weight /kg	-21.823
Dry weight /kg	-21.131
UF goal /L	-19.119
Blood flow /(ml/min)	-17.167
Weight loss /kg	-16.704
Cumulative blood volume /L	-14.716
Age /yrs	-6.056
HBA1C = Diabetic	-4.604
Anticoagulation = No	-0.000
KT/V	0.192
Gender = Female	0.922

Figure A.31: SVM model weight output for sodium plasma (left), protein total (middle), and carbon dioxide (right).

Attribute	Weight ↓
Pre Urea (BUN) /(mg/dL)	67.906
UF goal /L	33.386
Pre weight /kg	32.504
Dry weight /kg	31.131
Weight loss /kg	26.719
Duration /hr	26.671
Blood flow /(ml/min)	18.500
Cumulative blood volume /L	11.221
KT/V	0.774
Anticoagulation = No	0
Age /yrs	-2.680
HBA1C = Diabetic	-13.139
Gender = Female	-13.838

Attribute	Weight ↓
Pre Urea (BUN) /(mg/dL)	18.639
Dry weight /kg	11.637
Pre weight /kg	11.421
KT/V	10.012
Blood flow /(ml/min)	6.740
Duration /hr	4.947
HBA1C = Diabetic	0
Anticoagulation = No	-1.226
UF goal /L	-3.621
Weight loss /kg	-4.208
Gender = Female	-5.016
Age /yrs	-8.505
Cumulative blood volume /L	-12.865

Attribute	Weight ↑
Dry weight /kg	-40.288
Pre weight /kg	-38.028
Duration /hr	-31.606
Anticoagulation = No	-26.406
KT/V	-19.991
HBA1C = Diabetic	-17.519
Cumulative blood volume /L	-11.471
Pre Urea (BUN) /(mg/dL)	-8.908
Weight loss /kg	-5.459
Gender = Female	-3.789
Blood flow /(ml/min)	-0.000
UF goal /L	0.462
Age /yrs	13.482

Figure A.32: SVM model weight output for creatinine plasma (left), albumin (middle), and alkaline phosphatase (right).

Vita

Syeda Leena Mumtaz was born in the year 1996 in Karachi, Pakistan. She moved to United Arab Emirates in year 1998, where she received her primary and secondary education in Sharjah, UAE. She earned her Bachelor of Science degree in Biomedical Engineering from Ajman University in the year 2018. She received scholarships throughout her four-year academic performance and received awards and certificate of Honors.

In September 2019, Ms. Leena enrolled for Biomedical Engineering Master's program at American University of Sharjah and was granted to work as a graduate teaching assistant along with graduate studies. Her research interests include healthcare operations management, human anatomy and physiology, and biomedical devices.