Data collection and correlation analysis of patient-specific dialytic variables for the development of an AI assistance therapeutic tool

A. Nicosia [1], N. Cancilla [1], E. Di Liberti [1], J. D. Martín Guerrero [2], A. Ferrantelli [3], F. Iacono [3], V. M. B. Brucato [1], V. La Carrubba [1], A. Cipollina [1], I. Tinnirello [1]

[1] *Dipartimento di Ingegneria, Università degli Studi di Palermo, Viale delle Scienze Ed. 6, 90128 Palermo, Italy*

[2] *Departament d'Enginyeria Electrònica, Escola Tècnica Superior d'Enginyeria, Universitat de València (ETSE-UV), Despatx 3.2.26 Avgda, Universitat, s/n, 46100 Burjassot (València), Spain*

[3] *Dipartimento di Nefrologia, Azienda Ospedaliera di Rilievo Nazionale e di Alta Specializzazione (A.R.N.A.S.) "Civico Di Cristina e Benfratelli", 90128 Palermo, Italy*

***Abstract*—The increasing prevalence of chronic kidney disease has led to a rise in the number of patients requiring dialytic therapies. These treatments, although essential for blood purification, are often associated with complications such as hypotension, fluid overload, and patient discomfort that traditional models may not fully capture and predict. In this context, data-driven models based on the use of Artificial Intelligence can offer a solution to develop patient-specific dialysis therapies and predict potential complications. This study represents a preliminary step in developing an AI model, focusing on data collection and analysis. Clinical data were collected from 40 patients and statistical analysis, including Pearson's and Spearman's correlation coefficients, has identified significant relationships among several variables, which will be used to implement a model aimed at improving dialysis efficiency, minimizing adverse events, and increasing the overall quality of dialysis patients' care.**

***Keywords*—dialysis, statistical analysis, patient-specific study.**

I. INTRODUCTION

STUDIES aimed at improving hemodialysis and hemodiafiltration therapies for patients with critically reduced kidney function are now attracting a great deal of interest. This is particularly evident in both industry and academia, and is due to the growing number of patients affected by chronic renal insufficiency, involving reduced blood flow to the kidneys caused by renal artery disease. As evidence of this, to date more than 4 million people worldwide are undergoing hemodialysis or have already received a kidney transplant [1].

Hemodialysis is a method used to clean the blood containing waste and toxins by using a semi-permeable membrane, which separates it from the dialysate, a solution with electrolytes and glucose. The membrane allows metabolic waste, like urea and creatinine, to pass from the blood into the dialysate, while keeping proteins and blood cells in the bloodstream.

The process is carried out using cylindrical devices called hemodialyzers and filled with tiny hollow fibers.

These modules are part of complex dialysis machines, large devices that regulate and monitor the hemodialysis process through the presence of numerous sensors and pumps.

Despite the many benefits of these therapies, current research efforts on the improvement of dialytic therapies are continuously increasing due to the high number of patients undergoing treatment daily. In addition, there are many problems that can affect patients under therapy, such as volume-related adverse events, hypotension, or general discomfort as cramps and itching.

In the current scenarios, the prediction of dialysis effects on patients is based on analytic models which require assumptions and can be difficult to validate.

In this regard, *Artificial Intelligence* (AI) can serve as a valuable aid in elucidating complex problems, even when there is not a full understanding of the bio-physical mechanisms involved.

AI is a science aimed at making machines think and act like humans. To do so, this discipline makes use of algorithms that seek to create expert systems capable of making predictions or classifications based on data provided as input.

In this regard, using several AI models and clinical data from patients, many authors obtained accuracy values of over 96.00% in predicting hypotension [2] and other minor dialysis discomforts [3] at the start of therapy.

Other authors have developed AI models aimed at studying fluid overload [4] and arteriovenous (AV) fistula problems analysing audio recordings [5], patients' data [6] or images [7].

Placed on the arm and subjected to continuous stress during treatment, the AV fistula is the most common access point for therapy in hemodialysis patients. Continued stress in the region of the arm localized near the AV fistula results in the development of aneurysms, stenoses or thrombi [8] which AI models have demonstrated can predict well in advance, enabling proactive management and improved patient outcomes.

Despite the deep interest in this topic, to date the literature lacks specific studies on the comprehensive data collection and correlation analysis in hemodialysis sessions. In this regard, this study presents for the first time an extended clinical data collection on hemodialytic therapies, conducted in a dialysis outpatient clinic in Palermo.

A statistical analysis of correlation between inputs and outputs has been performed and results will be described, in order to identify the most important relationships to be consider for further investigation.

Though not yet completed here, this work represents the preliminary step for the implementation of a predictive, integrative, and dynamic *data-driven* AI model designed to guide treatment choices for each specific patient, achieve good treatment performance, and predict the occurrence of any intradialytic clinical complications.

II. METHODS

A. Data collection

The data collection was conducted at a dialysis outpatient clinic of the Azienda Ospedaliera "ARNAS Ospedali Civico Di Cristina Benfratelli" in Palermo (Italy).

Data were recorded for ~250 treatments over 60 days, with an average of 8 patients treated per day and 2-3 measurements taken for each patient. Overall, data were collected on 40 patients. Of these, 23 had extensive data available as they were regularly dialyzed at the facility, while 17 had limited data since they were generally treated at other medical centers or hospitalized for short periods due to surgery or trauma that caused significant renal failure. Data from both patient types were integrated with the aim of subsequently developing a model capable of distinguishing patients by their specific conditions.

The patients' baseline variables, the treatment's operational parameters (both time-dependent and fixed ones), as well as the patients' health metrics, were analysed and categorized according to the structure outlined in **Table I**.

TABLE I: CLASSIFICATION AND DETAILED DESCRIPTION OF THE VARIABLES INVOLVED IN THE DATA COLLECTION.

Patient's baseline variables	
Sex	*The code identifying the sex of the patient.*
Age	*The age of the patient in years.*
Dry weight	*The patient's weight without excess fluid retention post-dialysis.*
Reinfusion type	*The type of fluid reinfusion (POST/PRE) used during the dialysis session, if applicable.*
Treatment type	*The type of dialysis performed (e.g., hemodialysis, hemodiafiltration or hemofiltration).*
Access type	*The type of vascular access used for the dialysis session (e.g., fistula, catheter).*
Treatment's fixed operational parameters	
Weight loss	*The amount of weight reduced due to fluid removal during the dialysis session.*
Treatment time	*The overall duration of the dialysis session.*
Heparin volume	*The amount of heparin administered to the patient.*
Treatment's time-dependent operational parameters	
Dialysate temperature	*The temperature of the dialysate solution during the dialysis session.*
Ultrafiltration flowrate	*The flowrate at which the fluid is removed from the patient.*
Dialysate flowrate	*The flowrate of the dialysate solution entering the dialysis machine.*
Blood flowrate	*The blood flowrate entering the dialysis machine.*
Substitution flowrate	*The flowrate of the substitution fluid in treatment modalities with ultrafiltration.*
Dialysate composition and concentration	*The chemical composition and concentration of the dialysate solution (e.g., bicarbonate, Na^+, K^+).*
Patients' health metrics	
Kt/V	*A measure of dialysis adequacy, calculated with reference to sodium.*
Arterial pressure	*The blood pressure measured in the arterial line during the dialysis session.*
Venous pressure	*The blood pressure measured in the venous line during the dialysis session.*
Diastolic pressure	*The diastolic blood pressure measured during the dialysis session.*
Systolic pressure	*The systolic blood pressure measured during the dialysis session.*
Pulsations	*The patient's heart rate during the dialysis session.*
TMP	*Transmembrane pressure: the pressure gradient across the dialysis membrane.*

This framework was designed to support the model's primary objective, i.e., identifying which time-dependent operational parameters a doctor can adjust to achieve the desired patient health metrics, based on the fixed operational parameters and the patient's baseline variable values.

In particular, the values of the baseline variables were identified in order to define the range that the model should consider appropriate for each health metric, ensuring realistic and customized targets for the specific patient profile.

The secondary objective of the model will be to generate an alternative strategy when modifying only the time-dependent parameters results insufficient to reach the target outcomes. In such cases, it will reconsider operational parameters initially regarded as fixed, aiming to restore the treatment efficacy in more complex scenarios.

Based on the definitions provided in **Table I**, it is clear that weight loss, treatment time, and ultrafiltration (UF) flowrate are closely related variables, with the first two classified as fixed and the last one as time-dependent. Although this may appear contradictory, it is not because ultrafiltration flowrate can vary during the session around an average value while still ensuring that the treatment ends at the scheduled time and achieves the target weight loss.

Additionally, the operator can adjust UF flowrate during the treatment, considering that treatment time represents not a strictly constraint, but only the maximum duration allowed for that dialytic session. This means that, if conditions permit, increasing the ultrafiltration flowrate is an allowed action that can help the patient reach the target weight loss before the scheduled end time.

B. Correlation studies

A series of correlation studies were conducted to assess the relationships between treatment parameters and patient outcomes, and to select the most relevant variables to be included in the model, accordingly.

Using Pearson's and Spearman's coefficients [9], the correlation between all variables, in groups of two denoted by subscript X and Y, was calculate.

The Pearson coefficient ρ_{XY} is a correlation index that measures the strength and direction of the linear relationship between two quantitative variables.

As shown in Eq. (1), it links the covariance and standard deviation of the variables involved:

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \tag{1}$$

where σ_X and σ_Y are the standard deviation relative to *X* and *Y* variables, respectively, and σ_{XY} is their covariance.

The Spearman coefficient θ_{XY} is a non-parametric measure of the rank correlation that assesses the strength and direction of the monotonic relationship between two variables. It is calculated by applying the Pearson formula to the ranked variables *R(X)* and *R(Y)*, used to transform the raw data into an ordered scale, which helps to identify monotonic relationships between variables, without being influenced by the actual scale of values. Its formula is reported in Eq. (2):

$$\theta_{XY} = \frac{\sigma_{R(X)R(Y)}}{\sigma_{R(X)} \sigma_{R(Y)}} \tag{2}$$

where $\sigma_{R(X)}$ and $\sigma_{R(Y)}$ represent the standard deviation relative to the ranked variables *R(X)* and *R(Y)*, and $\sigma_{R(X)R(Y)}$ their covariance. Both the Pearson and Spearman coefficients take values in the range between –1 and +1, where +1 indicates a perfect positive correlation, −1 a perfect negative correlation, and 0 an absence of correlation.

A correlation threshold above 0.4 was considered relevant, the results of which will be shown in detail in the next section.

This choice was done to consider only variables with a significant relationship, an aspect that could improve the robustness and reliability of the model.

III. RESULTS

The first correlation analysis focused on determining the impact of time-dependent operational parameters on patients' health metrics for understanding their relationships and guiding real-time adjustments during treatment.

Table II presents the highest correlation values between Pearson and Spearman, selecting only those greater than 0.4.

TABLE II: CORRELATIONS BETWEEN THE TIME-DEPENDENT OPERATIONAL PARAMETERS AND THE PATIENTS' HEALTH METRICS.

	Substitution flowrate	*Blood flowrate*	*Dialysate flowrate*	*UF flowrate*	*Dialysate K^+ conc.*
Kt/V	-	0.40	0.53	-	-
Arterial pressure	-	-0.51	-	-	-
Venous pressure	-	0.67	-	0.42	-
Diastolic pressure	-	0.54	-	-	0.45
Systolic pressure	-	0.55	-	0.43	0.46
Pulsations	-	0.47	0.47	-	-
TMP	0.42	0.47	0.41	-	-

Notably, blood flowrate emerges as the most critical factor, showing important correlations with all the health metrics.

Specifically, the positive correlation with diastolic (0.54) and systolic pressure (0.55) suggests that higher blood flowrates can increase arterial pressure, which must be carefully monitored, particularly in patients prone to hypertension.

Additionally, the strong correlation between blood flowrate and venous pressure (0.67) highlights its impact on vascular dynamics, possibly indicating increased resistance or issues related to vascular access when blood flowrate is elevated.

Dialysate flowrate also plays a key role, correlating positively with Kt/V (0.53) and thus suggesting that higher dialysate flowrates enhance dialysis efficacy. However, the correlation with TMP (0.41) indicates that excessive dialysate flowrate may raise filtration pressure, potentially increasing the risk of membranes damage.

This suggests that it would necessary to balance all the values of the variables to obtain the optimal combination for each patient, a feature that the model has the aim to reach.

Similarly, UF flowrate shows a positive relationship with the venous (0.42) and the systolic pressure (0.43), reinforcing the need for precise adjustment to avoid hemodynamic instability, particularly in patients with cardiovascular vulnerability.

Another parameter of interest is the dialysate potassium concentration, which correlates with both systolic (0.46) and diastolic pressure (0.45). This suggests that the electrolyte composition of the dialysate plays a role in blood pressure regulation, making its adjustment essential for managing patients with frequent episodes of hypertension or hypotension. These initial results already underscore the complexity of dialysis management and the importance of a personalized therapy, which the proposed model could propose to ensure alignment with each patient's specific profile and clinical condition.

The second correlation study focused on assessing the impact of fixed operational parameters on time-dependent ones to understand how changes in treatment structure could influence these variables and, thus, the desired outcomes.

Table III presents the highest correlation values between Pearson and Spearman, selecting only those greater than 0.4.

TABLE III: CORRELATIONS BETWEEN THE FIXED OPERATIONAL PARAMETERS AND THE TIME-DEPENDENT ONES.

	Blood flowrate	*Dialysate flowrate*	*UF flowrate*	*Dialysate K^+ conc.*
Weight loss	-	-	(0.86)	-
Treatment time	0.50	-	0.48	0.75
Heparin volume	0.48	0.42	-	0.58

N.B. Spearman parameter between brackets

The strong positive correlation (0.86) between weight loss and UF flowrate suggests that higher flowrates are associated with greater weight loss targets, as expected. This indicates that for patients having elevated venous and systolic pressure, which are strongly influenced by UF flowrate, lowering the target weight loss could help reduce cardiovascular stress and prevent complications during therapy.

Similarly, treatment time correlates positively with blood flowrate (0.50) and dialysate potassium concentration (0.75).

Considering specifically the correlation between blood flowrate and Kt/V in **Table II**, it is clear that longer sessions are typically associated with higher dialysis efficiency.

However, it is important to note that higher blood flowrate and dialysate potassium concentration are also associated with increased systolic and diastolic pressure, an aspect that could compromise the patient's health. This highlights again the need to carefully balance all variables involved to achieve the optimal combination for each patient, which is precisely the goal the model is intended to implement.

Heparin volume shows also moderate positive correlations with blood flowrate (0.48), dialysate flowrate (0.42), and dialysate K^+ concentration (0.58). These findings suggest that increased heparin dosages can permit to use higher blood and dialysate flowrate, within safe conditions for the patient, given its capacity to prevent clot formation during the session.

Finally, the third and last analysis focused on understanding the patients' health metrics with the aim of predicting secondary effects given when adjusting a time-dependent treatment parameter that directly impacts one output, inadvertently affects others due to their intrinsic correlation.

Table IV presents the highest correlation values between Pearson and Spearman, selecting only those greater than 0.4.

TABLE IV: CORRELATIONS BETWEEN THE PATIENTS' HEALTH METRICS.

	Kt/V	*Arterial pressure*	*Venous pressure*	*Diastolic pressure*	*Systolic pressure*	*TMP*
Kt/V	*1*	-	-	-	-	*0.43*
Arterial pressure	-	*1*	*-0.48*	-	-	
Venous pressure	-	*-0.48*	*1*	*0.46*	*0.42*	*0.48*
Diastolic pressure	-	-	*0.46*	*1*	*0.88*	-
Systolic pressure	-	-	*0.42*	*0.88*	*1*	-
TMP	*0.43*	-	*0.48*	-	-	*1*

One of the strongest relationships is observed between systolic and diastolic pressure (0.88). This close correlation is expected in hemodynamic regulation and underscores the importance of monitoring both pressures simultaneously, as changes in one will almost certainly affect the other.

A moderate positive correlation exists also between venous and diastolic pressure (0.46), as well as venous and systolic pressure (0.42).

Additionally, TMP shows significant positive correlations with several variables, including Kt/V (0.43) and venous pressure (0.48). This suggests that adjustments to time-dependent treatment parameters affecting any one of these three variables may influence the others, an aspect that has to be considered when the treatment's variables are set.

Interestingly, the negative correlation between arterial and venous pressure (-0.48) highlights that operator interventions to reduce one may increase the other, and vice versa, risking damage to the AV fistula.

These results further reinforce the need to implement targeted and customized treatment strategies aimed at achieving a more comprehensive management of patient outcomes, a reduction in the risks associated and an improvement in treatment efficacy.

IV. CONCLUSIONS

The study carried out in this paper represents a first step in creating a data-driven model based on AI, aiming at improving dialysis treatments with a patient-specific approach. By analyzing data from 40 patients using correlation algorithms, significant relationships were identified between health metrics and key operational parameters.

In particular, a limited subset showed strong correlations, above the threshold of 0.4 set for relevance. While this threshold was chosen to prioritize the most significant relationships, it may have led to the exclusion of potentially important variables with weaker correlations. This suggests that lowering the correlation threshold could provide a more comprehensive understanding of the dataset for future studies.

Additionally, splitting the correlation analysis per specific subgroups based on their characteristics can certainly lead to identify subtle but clinically relevant correlations that could been hidden in this preliminary analysis.

For the sake of brevity, such developments cannot be reported in this long-abstract, but will be presented in the extended work, thus guiding toward an enhanced capability of better understanding individual patient responses to treatment.

Future studies will also focus on adding new variables and data, aimed to improve the model's predictive power.

ACKNOWLEDGEMENT

The authors acknowledge the financial support of Mozarc Medical® and Piano Nazionale di Ripresa e Resilienza (38-033-20-DOT204NJ79-3458), DM 352 M4C2 I., for funding the PhD scholarship of Ms. Alessia Nicosia.

REFERENCES

[1] A. K. Bello et al., "Epidemiology of Haemodialysis Outcomes" Nature Reviews Nephrology, vol. 18, no. 6, pp. 378–395, 2022. doi: 10.1038/s41581-022-00542-7.

[2] H. Lee et al., "Prediction of Intradialytic Hypotension Using Pre-Dialysis Features—A Deep Learning-Based Artificial Intelligence Model" Nephrology Dialysis Transplantation, vol. 38, no. 10, pp. 2310–2320, 2023. doi: 10.1093/ndt/gfad064.

[3] C. Shih, L. Youchen, C. Chen, and W. C.-C. Chu, "An Early Warning System for Hemodialysis Complications Utilizing Transfer Learning from HD IoT Dataset" in Proceedings of the 2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC), pp. 759–767, 2020. doi: 10.1109/COMPSAC48688.2020.0-168.

[4] G. F. L. Tan et al., "Automated Lung Ultrasound Image Assessment Using Artificial Intelligence to Identify Fluid Overload in Dialysis Patients" BMC Nephrology, vol. 23, no. 1, Art. no. 410, 2022. doi: 10.1186/s12882-022-03044-7.

[5] W. T. Song, C. C. Chen, Z.-W. Yu, and H.-C. Huang, "An Effective AI Model for Automatically Detecting Arteriovenous Fistula Stenosis" Scientific Reports, vol. 13, no. 1, Art. no. 17659, 2023. doi: 10.1038/s41598-023-35444-6.

[6] S. Balamuthusamy, N. Dhanabalsamy, B. Thankavel, M. S. Bala, and A. Pfaffle, "Utility of a ML Analytics on Real-Time Risk Stratification and Re-Intervention Risk Prediction on AV Access Outcomes and Cost" The Journal of Vascular Access, vol. 25, no. 4, pp. 1224–1228, 2024. doi: 10.1177/11297298231156632.

[7] H. Zhang et al., "Deep Learning to Classify Arteriovenous Access Aneurysms in Hemodialysis Patients" Clinical Kidney Journal, vol. 15, no. 4, pp. 829–830, 2021. doi: 10.1093/ckj/sfab278.

[8] G. W. Cho et al., "Quantitative Plaque Analysis with A.I.-Augmented CCTA in End-Stage Renal Disease and Complex CAD" Clinical Imaging, vol. 89, pp. 155–161, 2022. doi: 10.1016/j.clinimag.2022.06.012.

[9] M. Mukaka, "Statistics corner: A guide to appropriate use of correlation coefficient in medical research", Malawi Medical Journal, vol. 24, no. 3, pp. 69–71, 2012.