

Hybrid Regression Models for Predicting Hydration: A Case Study in Pediatric Hemodialysis

Suzana Djordjevic¹ , Mirjana Kostic² , Blerina Zana³ , Danijela Milosevic⁴ ,
Vladimir Mladenovic^{4*} 

¹Faculty of Education, University of Kragujevac, Milana Mijalkovica, 14, 35000, Jagodina, Serbia,
suzana.djordjevic@pefja.kg.ac.rs

²Department of Nephrology, University Children Hospital, Tirsova, 10, 11000, Belgrade, Serbia,
kostic1mirjana@yahoo.com

³Faculty of Economics and Agribusiness, Agriculture University of Tirana, Rruga Paisi Vodica, 1025, Tirana, Albania,
bzanaj@ubt.edu.al

⁴Faculty of Technical Sciences Cacak, University of Kragujevac, Svetog Save, 65, 32000, Cacak, Serbia,
danijela.milosevic@ftn.kg.ac.rs, vladimir.mladenovic@ftn.kg.ac.rs

Abstract: With the rapid development of artificial intelligence (AI) and machine learning (ML), the medical field is experiencing significant advances that make it easier for doctors to monitor and personalize the hemodialysis process in pediatric patients. This case study focuses on the prediction of hydration in pediatric hemodialysis using hybrid ML models with hyperparameters for prediction accuracy. Data were collected from pediatric hemodialysis at the University Children's Hospital in Tiršova, and parameters such as weight, blood pressure, lean tissue index (LTI), fat tissue index (FTI), body mass index (BMI), total body water, extracellular water (ECW), adipose tissue mass, body cell mass, and bioimpedance were adjusted for training. The model was configured for each pediatric patient and retrained after each treatment to make individualized predictions with the highest accuracy. The proposed model uses measurable parameters to estimate hydration and provide better recommendations to the physician, leading to better results than commonly used state-of-the-art competing methods. The hybrid ML models with hyperparameters represent a novel, simplified, safe, and efficient method for predicting hydration in children, making it easier for doctors to monitor the hemodialysis process for children.

Keywords: hemodialysis, support vector regression, elastic net, gradient boosting regressor, kidney diseases

1. INTRODUCTION

The kidneys are not only necessary to filter dangerous substances from the body, but also to regulate the body's acid-base balance, electrolyte balance, and blood pressure. Kidney malfunction causes mild to fatal diseases, as well as dysfunctions in other organs of the body. This is why experts around the world are devoting their time to developing techniques for the precise diagnosis and treatment of renal problems. Machine learning (ML) techniques are increasingly being used in medicine for diagnosis and as an assessment tool for physicians to make better medical decisions [1], and this also applies to various types of kidney disease.

The essence of hemodialysis therapy (HD) is to maintain a balanced fluid level in the body. It helps to prevent hypo- and hyperhydration, which can lead to long-term cardiovascular problems, reduced cardiac efficiency, and other problems. HD is a method of removing excess fluid from the body of a patient whose kidney function is impaired or absent, and is

often performed using ultrafiltration (UF). The introduction of the body composition monitor (BCM) based on bioimpedance spectroscopy (BIS) in recent years has provided an objective way to determine the fluid status of hemodialysis patients [2]. It is a promising approach to assess total body water in patients with pathologic hydration, to distinguish intracellular water (ICW) from extracellular water (ECW) and to estimate body composition in HD patients [3]. Age, body mass index (BMI), and body weight (BW) have been associated with overhydration (OH) even in healthy individuals [4]. Hemodialysis patients have difficulty maintaining sodium and water homeostasis, resulting in excess sodium and water, causing elevated blood pressure (BP) and weight gain [5]. The fat tissue index (FTI) and lean tissue index (LTI) can be affected by volume status, as polyglucose or hypertonic solutions are often used to treat fluid overload. A stable BW or BMI does not guarantee a stable body composition, as an increase in fat mass is usually associated with muscle wasting. Lower FTI at

baseline is associated with increased adiposity and/or a decrease in lean mass [6]. Therefore, hydration levels and factors related to fluid volume must be monitored during hemodialysis. Accurate estimation of hydration status in children is particularly difficult due to ongoing physiological development, individual variability, and sensitivity to fluid imbalances. Conventional methods are either invasive, unreliable, or impractical for continuous use. This case study addresses this clinical gap.

The main contributions of this paper are:

- Development of a non-invasive method to assess hydration levels in pediatric hemodialysis patients to improve quality of care and reduce discomfort with current invasive methods.
- Improve the accuracy of hydration prediction by combining the existing ML models into hybrid models that allow better tradeoffs between the linear and nonlinear parts of the medical features.

Given the complexity and variability of pediatric cases, there is a great need for personalized predictive systems tailored to individual patient profiles. This research does not aim to provide universally more advanced ML models for all disciplines. Instead, novel hybrid architectures are used and tested in a medically important and data-sensitive setting, where the accuracy of the model has a direct impact on the quality of therapy.

The paper is organized as follows: Section 2 provides an overview of studies using artificial intelligence (AI) to predict different factors during hemodialysis. Section 3 presents the data used in this study and describes the existing and proposed methodology for predicting hydration in pediatric patients. Section 4 discusses the comparison of results between the existing and the proposed hybrid model, and Section 5 presents the conclusions.

2. RELATED WORKS

A. Typically used medical metrics and methods

Conclusions regarding the medical characteristics to be included in the hydration prediction study were drawn from a comprehensive literature review. The amount of UF administered is based on pre-dialysis BP, with elevated BP assumed to be due to OH, also known as volume-dependent BP [7]. The main conclusion of Rymarz et al. [8] was that a decreasing LTI was associated with poorer survival in hemodialysis patients. According to Marcelli et al. [9], dialysis patients with an LTI and FTI within the 10th to 90th percentile (of the age- and sex-matched healthy population) had the highest survival rate. At the same time, a low FTI, a low LTI, or a combination of both were associated with higher mortality. Multiple-frequency bioimpedance analysis (MF-BIA) assesses TBW by exposing it to low and high frequency electrical currents at 50 different frequencies between 5 and 1000 kHz using bioelectrical impedance spectroscopy (BIS) [10]. BIS refines the approach by calculating the resistance at zero frequency (R_e) and the resistance at infinite frequency (R_{tot}) using the Cole model [11]. The intracellular compartment resistance (R_i) is then estimated using Kirchoff's formula for parallel circuits [12] based on R_e and R_{tot} . R_e and R_i are the largest electrical

correlates for ECW and ICW, respectively. Following this logic, the calculation techniques using body height, weight, and associated resistances as input variables, as well as constants defining body fluid resistivity, body form, and body density, are used to convert the resistance values into volumes of ECW, ICW, and TBW [13]. The OH calculation is based on a reference population of healthy age-, sex-, and weight-matched control groups and includes an assessment of ECW, ICW, and BW [14].

$$OH = 1.136 \cdot ECW - 0.430 \cdot ICW - 0.114 \cdot BW \quad (1)$$

The above-described method is a non-invasive method that allows continuous monitoring and is highly accurate in determining fluid levels in the body, but requires a high level of expertise to interpret the results. Other hydration estimation methods during HD include lung ultrasound, echocardiography, blood volume monitoring and clinical score. Lung ultrasound [15] is used to assess the presence of fluid in the lungs. An increase in the number of B-lines on ultrasound may indicate more fluid, but it does not provide quantitative data on total body fluid. Echocardiography [16] is used to assess cardiac function and fluid volume. It is useful to estimate the size of the inferior vena cava (IVC) and blood volume, but it cannot be used continuously throughout the entire hemodialysis session. Blood volume monitoring [17] measures changes in blood volume during hemodialysis. These changes may indicate fluid excess or deficiency, but they can give false positive or negative results depending on the patient's condition. The clinical score [18] is a method in which physicians perform physical examinations, such as checking for edema and assessing the patient's blood pressure, heart rate, and weight before and after dialysis. This method relies heavily on the physician's personal experience.

B. Use of machine learning in kidney-disease problems

In recent years, the expansion of AI and ML has found its application in many scientific fields, including medicine. AI techniques help medical researchers manage large amounts of patient data and allow them to analyze and interpret raw data for patient treatment. These technologies help in diagnostic processes, such as high-speed body scans, and can create 3D mapping solutions for patients [19]. Kanada E. et al. [20] have developed a revolutionary system that analyzes hemodialysis patient data, categorizes patients based on their characteristics, and identifies patients at high risk of death using ML algorithms. The sparse Laplacian regularized random vector functional link (SLapRVFL) neural network model outperforms other methods in dry weight assessment with a low prediction error [21]. Next, artificial neural networks (ANN) can be used to predict OH in pediatric patients during the hemodialysis process [22]. Study [23] proposes the random forest (RF) algorithm as the most efficient ML algorithm for data processing with high accuracy. This study also measures the performance of ML models with and without tuning hyperparameters. A significant improvement in prediction accuracy is observed, highlighting the applicability of supervised ML algorithms in bioinformatics and their compatibility with the diagnosis of fatal diseases such as chronic kidney disease.

In the RENAAL, IDNT, and ALTITUDE trials, the feed-forward neural network model predicted end-stage renal disease (ESRD) with satisfactory receiver-operator curve results [24]. The feed-forward neural network model used urine albumin to creatinine ratio, serum albumin, uric acid, and serum creatinine as significant predictors and achieved peak performance in predicting long-term ESRD. In the study [25], a robust ML model for early detection of chronic kidney disease (CKD) was developed, which achieved a high classification metric using the University of California Irvine (UCI) CKD dataset. The model's reliability and its potential for clinical application emphasize its importance in advancing the early diagnosis and management of CKD, a critical global health challenge. The SVM model was developed to accurately identify CKD in the CKD dataset, overcoming flaws and achieving a low false negative rate, demonstrating the potential for real-time diagnosis and reduced mortality rates [26]. In [27], the proposed auto-ML scheme achieved a very comprehensive model evaluation. Research [28] develops an efficient clinical diagnosis system using support vector machine (SVM) and logistic regression algorithms (LGR) for CKD. The system uses chi-square feature selection and hyperparameter tuning to increase the model's accuracy. The SVM model, which has high accuracy, helps physicians make early, accurate, and unambiguous clinical decisions. The system's usefulness is to reduce human losses by diagnosing this life-threatening renal disease.

In a previous work [29], we established the framework for combining ML models that predict hydration during hemodialysis, resulting in hybrid models. In this work, we will use this framework to increase system performance.

C. Hybrid machine learning models applied in hemodialysis setting

A hybrid ML model refers to the integration of multiple ML techniques or models that combine the strengths of different methods to improve overall prediction performance, robustness, and/or efficiency. As mentioned in [29], we have used several ML models, including elastic net (EN), support vector regression (SVR), gradient boosting regression (GBR), Gaussian Naïve Bayes (GNB), LGR, etc. We will briefly review the existing literature on hybrid models and the CKD problem.

Ren et al. [30] proposed a hybrid neural network model combining bidirectional long short-term memory (BiLSTM) and autoencoder networks to predict kidney damage in hypertensive patients. The proposed model outperforms SVM and strong neural baseline systems on a raw electronic health record (HER) data dataset. In [31], a fast hybrid model, i.e., R_FP-SVM, is developed for CKD classification problems. This diagnostic approach has two main features: fast learning with high accuracy and identification of the most important CKD risk factors. Dey et al. [32] investigate the use of ML algorithms for early detection of kidney disease. Using different algorithms, a hybrid feature selection method (Chi2-MI), focusing on correlation scores for predicting CKD was applied. The approach outperformed the other algorithms and

resulted in higher accuracy scores. Study [33] used Pearson correlation feature selection and ML classifiers such as gradient boosting (GB), GNB, decision trees, and RF to develop a stacking algorithm that predicts CKD patient status with high accuracy. The corresponding GB technique for solving regression problems is GBR, which we will use in the current study. In [34], the authors used data from the UCI Repository to store 400 instances of 26 CKD features. The results show that the neural network ensemble with Lasso model achieved the highest accuracy (99.98 %). Ratnababu and Raghava Naidu [35] confirmed that CKD can be predicted using hybrid ML classifiers, specifically k-nearest neighbor (KNN) and logistic regression. The primary objective of [35] was to evaluate different ML algorithms based on their performance accuracy. By combining these two models, the F1 score was further improved and more accurate results were obtained. In the study [36], a hybrid technique for detecting important risk factors for people with Metabolic syndrome (MetS) and CKD with moderate renal insufficiency was developed using six ML methods: RF, LGR, multivariate adaptive regression splines (MARS), extreme GB (XGBoost), GB with categorical features support (CatBoost), and a light GB machine (LightGBM). Because our databases are small and the results of our research are intended for use in hospitals on computers with limited memory capacity, the XGBoost, CatBoost, and LightGBM algorithms were not used in this study. Instead, the GBR was used as it is suitable for smaller datasets and is less memory intensive. Finally, the hybrid ANN-SVM model was shown to perform better than ANN in predicting CKD and non-CKD patients in terms of mean absolute error (MAE), root mean square error (RMSE), relative absolute error (RAE), and root relative square error (RRSE) [37].

As mentioned in Section 2. A, expertise is required to interpret BIS measurement results and to assess hydration in HD patients. As shown in [22], neural networks are good tools for hydration prediction. At the same time, [29] has shown that EN provides the most accurate estimates of fluid levels in the body compared to actual fluid levels. This raises the question of whether it is possible to further optimize EN as a linear model by creating hybrids with some nonlinear models to achieve even greater accuracy in predicting hydration. Based on the literature from Section 2. C, various combinations of existing models were tested, including combinations with models such as KNN or RF, but the results were inferior to the EN results. Since GBR and SVR as nonlinear models gave the next best results, it was our idea to create EN-GBR and EN-SVR hybrids to further improve the good EN results. These hybrid configurations will be applied and evaluated in a real clinical setting with pediatric patients undergoing hemodialysis, making this study a practical exploration of personalized hydration prediction.

The aim of this research is to predict hydration in pediatric hemodialysis patients using novel hybrid combinations. In Section 3, we explain the dataset, the basic ML models and how the data flow for the hybrid models was constructed.

3. MEASURED PARAMETERS AND BASIC METHODS

A. Experimental data: the description of measured parameters

Our measurements are based on data collected by the University Children's Hospital in Tiršova. The data were collected from pediatric patients aged 0 to 16 years in May 2022. The database consists of $n = 69$ numerical medical input features x_{ij} , which have a direct influence on the numerical medical output feature y_i , which represents hydration in liters (OH [l]). This dataset is represented as a matrix of size $m \times (n + 1)$, see Fig. 1, where m represents the total number of samples for a single patient over time.

	x_{11}	x_{12}	$\dots$	x_{1j}	$\dots$	x_{1n}	y_1
	x_{21}	x_{22}	$\dots$	x_{2j}	$\dots$	x_{2n}	y_2
	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
	x_{i1}	x_{i2}	$\dots$	x_{ij}	$\dots$	x_{in}	y_i
	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
	x_{m1}	x_{m2}	$\dots$	x_{mj}	$\dots$	x_{mn}	y_m

Fig. 1. Matrix representation of a dataset with n input parameters x_{ij} ($j = 1, \dots, n$) and output parameter y_i , measured over m time moments ($i = 1, \dots, m$).

Section 2. A. lists the necessary medical parameters (TBW, BP, ECW, ICW, R_i , R_e , BIS, etc.) used to evaluate OH. All these parameters form 69 input variables in the database. The total number of measurements m depends on the duration of hemodialysis and the number of hemodialysis sessions of each patient. The databases contain between 200 and 500 measurements per patient. The datasets are created individually for each patient, with the personalized datasets containing their specific measurements and characteristics. Hybrid ML models are then trained independently on these personalized datasets, enabling adaptation to each patient's unique characteristics and data patterns.

All measurements were taken with a BIS device, the body composition monitor (BCM). It can be used throughout the patient's life cycle, from CKD stage 1 to renal replacement therapy and transplantation. The electrodes are attached to one hand and one foot on the same side of the body while the patient is in a supine position, the patient cable is connected, the patient's height and weight are entered, and the measurement is started. The data is transferred via the PatientCard to several Fresenius Medical Care software applications for further analysis, e.g. to the fluid management tool (FMT), therapy monitor (TMon) and PatientOnline (POL). All output parameters of the BCM have been validated against the gold standard reference methods in a number of studies with more than 500 patients and healthy controls. BCM devices are typically factory calibrated and do not require routine calibration by the end user. However, periodic accuracy checks are performed by the manufacturer or authorized service personnel to ensure measurement accuracy.

B. Elastic net regression machine learning model for predicting hydration

EN regression is a mixture of the two best shrinkage regression techniques: Ridge regression (l_2 penalty) for dealing with high-multicollinearity situations and least absolute shrinkage selection operator (LASSO) regression (l_1 penalty) for feature selection of the regression coefficients [38].

The EN regression model is constructed by adding a regularization term:

$$\lambda_1 \sum (|\omega_1| + \dots + |\omega_n|) + \lambda_2 \sum (\omega_1^2 + \dots + \omega_n^2)$$

to the multiple regression model, as in:

$$y_i = \omega_0 + \omega_1 x_{i1} + \dots + \omega_n x_{in} + \varepsilon^2 + \lambda_1 \sum (|\omega_1| + \dots + |\omega_n|) + \lambda_2 \sum (\omega_1^2 + \dots + \omega_n^2) \quad (2)$$

with $\lambda_1, \lambda_2 > 0$, $\sum_{k=1}^n |\omega_k| \leq q$, $\sum_{i=1}^n \omega_i^2 \leq p$, where q and p are the shrinkage amount for the l_1 and l_2 penalties, successively. The l_1 penalty is used to construct a sparse model, and the l_2 penalty is used to stabilize the regularization of the l_1 penalty. λ_1 and λ_2 are tuning parameters that determine the regularization intensity and predictor variable selection.

The regularization term introduces a penalty into the combined regression models and reduces the sum of square errors (SSE). Consequently, the SSE is written as follows,

$$\varepsilon^2 = (y_j - \hat{y}_j) + \lambda_1 \sum (|\omega_1| + \dots + |\omega_n|) + \lambda_2 \sum (\omega_1^2 + \dots + \omega_n^2) \quad (3)$$

Here, $\hat{y}_j$ denotes the predicted value of the target variable y_j computed using the estimated coefficients ω .

EN is an effective regression model for predicting hydration because it performs feature selection by shrinking some coefficients to zero. This can be beneficial for many features as it helps to reduce noise and focus on the most important predictors. EN's regularization can help prevent overfitting, especially when many correlated features are present. It strikes a balance between Ridge and Lasso regularization and provides a good compromise between bias and variance. We found that Ridge regression alone performs less well than most other methods.

C. Gradient boosting regressor machine learning model for predicting hydration

GBR is a powerful ensemble learning method that can be particularly effective in scenarios where there are complex relationships between features and the output parameter. It works by sequentially fitting multiple decision trees, with each tree correcting the errors of the previous ones. GBR is known for its ability to capture nonlinear relationships and interactions between features, making it well-suited for datasets with complicated patterns. This ML approach makes predictions by "boosting" an ensemble of weak prediction

models, either decision trees or linear models, to create a more robust model [39]. The "boosting" technique reduces overfitting by focusing on the errors of previous models, while automatic feature selection identifies the key factors. GBR is robust to outliers and extremely flexible, allowing it to be adapted to different data types and tasks. A GBR is defined as a set of sequential approximations of y_i [40], where the initial y_i value is calculated as:

$$(y_i)_{t=0} = F_0(\mathbf{x}_i) = \underset{\rho}{\operatorname{argmin}} \sum_{i=1}^n L(y_i, \rho) \quad (4)$$

where $L(y_i, (y_i)_t) = (y - (y_i)_t)^2$ is a loss function. y_i is further improved in T successive calculations (with T trees for $t = 1, \dots, T$):

$$(y_i)_t = F_t(\mathbf{x}_i) = F_{t-1}(\mathbf{x}_i) + \rho_t h_t(\mathbf{x}_i; \mathbf{a}_t) \quad (5)$$

where $h_t(\mathbf{x}; \mathbf{a}_t)$ is the decision tree function and the GBR method calculates the gradient and other GBR coefficients as in:

$$\begin{aligned} \widetilde{\Delta y}_i &= - \left[\frac{\partial L(y_i, F(\mathbf{x}_i))}{\partial F(\mathbf{x}_i)} \right]_{F(\mathbf{x}_i)=F_{t-1}(\mathbf{x}_i)}, i = 1, \dots, n \\ \mathbf{a}_t &= \underset{\mathbf{a}, \beta}{\operatorname{argmin}} \sum_{i=1}^n [\widetilde{\Delta y}_i - \beta h_t(\mathbf{x}_i; \mathbf{a})]^2 \\ \rho_t &= \underset{\rho}{\operatorname{argmin}} \sum_{i=1}^n L(y_i, (y_i)_t). \end{aligned}$$

The GBR model used in this study is based on the GBR approach from Scikit-learn [41].

D. Support vector regression machine learning model for hydration prediction

SVR is a powerful technique for regression tasks that particularly suitable for datasets with complex relationships and nonlinear patterns. SVR finds the hyperplane that best fits the data within a certain margin of tolerance (ε) while minimizing the complexity of the model (controlled by the regularization parameter).

In SVR, the goal is to find a function $f(x)$ that has at most ε deviation from the obtained targets y_i for all training data. SVR as a kernel-based learning method uses implicit mapping φ of the input data into a high-dimensional feature and kernel function K that returns the inner product $\langle \varphi(\mathbf{x}_l), \varphi(\mathbf{x}_k) \rangle, l, k = 1, \dots, m$. The SVR function for calculating y_i can therefore be defined as [42]:

$$(y_i)_t = F_t(\mathbf{x}_i) = F_{t-1}(\mathbf{x}_i) + \rho_t h_t(\mathbf{x}_i; \mathbf{a}_t), \quad (6)$$

where we aim for solutions with small ω by minimizing the objective function:

$$\min_{\omega, b} \frac{1}{2} \|\omega\|^2$$

$$\text{subjected to } |y_i - f(\mathbf{x}_i)| \leq \varepsilon.$$

Nevertheless, it is possible that solution does not exist under these conditions or that better results can be achieved if outliers are allowed. For this reason, we introduce slack variables ξ_i^+ and ξ_i^- so that:

$$\begin{aligned} \xi_i^+ &= f(\mathbf{x}_i) - y_i > \varepsilon \\ \xi_i^- &= y_i - f(\mathbf{x}_i) > \varepsilon \end{aligned}$$

so that the objective function and the constraints for SVR are finally stated as:

$$\begin{aligned} \min_{\omega, b} \frac{1}{2} \|\omega\|^2 + C \frac{1}{2} \sum_{i=1}^n (\xi_i^+ + \xi_i^-) \\ \text{subjected to } \begin{cases} y_i - f(\mathbf{x}_i) \leq \varepsilon + \xi_i^+ \\ f(\mathbf{x}_i) - y_i \leq \varepsilon + \xi_i^- \\ \xi_i^+, \xi_i^- \geq 0, i = 1, \dots, n \end{cases} \quad (7) \end{aligned}$$

where C is a trade-off parameter between model complexity and training error.

SVR efficiently models nonlinear relationships between variables using Kernel functions, which enables better predictions in complex datasets. In addition, SVR can handle many variables by optimizing the margins to reduce error and increase accuracy. SVR is robust and can effectively manage outliers in the data and minimize their impact on the final predictions. The flexibility of SVR allows it to adapt to different types of data and tasks, making it an excellent choice for analyzing and predicting complex relationships in datasets.

4. HYBRID METHODS FOR IMPROVED PREDICTION

A. Case-specific model: elastic boosting

In this case study, we investigate a hybrid approach that combines EN and GBR to improve hydration prediction in pediatric hemodialysis patients.

EN is effective for datasets that can be approximated by linear functions. EN can reduce overfitting by shrinking irrelevant coefficients and handling correlated features. On the other hand, GBR is known for its ability to capture complex relationships and nonlinear patterns in the data. By combining these two techniques, the hybrid model can leverage EN's feature selection and regularization capabilities while benefiting from the descriptive power of GBR. The EN component can help filter out noise and focus on the most relevant features, which are then used as input to the GBR model for further refinement. This hybrid approach can be beneficial when working with datasets with many features, some of which may be indirectly or inversely correlated.

The mathematical model for the EN-GBR class can be described as a combination of EN and GBR models. First, these models are initialized with the corresponding parameters (α and λ for EN, and ν for GBR). The EN model is then fitted to the input data to obtain predictions. The EN residuals are calculated as the difference between the actual values and the predictions of the EN model. The GBR model is then fitted to the input data and these residuals. The final prediction is obtained by adding the EN prediction to the prediction of the residuals.

$$Pred_{EN-GBR} = Pred_{EN} + ResPred_{GBR} \quad (8)$$

B. Case-specific model: elastic support regressor

Similar to EN-GBR, elastic support regression (EN-SVR) uses an EN to fit the linear data part, and its residuals are predicted by SVR, as in:

$$Pred_{EN-SVR} = Pred_{EN} + ResPred_{SVR} \quad (9)$$

SVR has more nonlinear Kernel functions than GBR and is particularly useful for datasets with high dimensionality and complex patterns - it can handle a large number of features and is less sensitive to outliers. This hybrid approach can achieve high accuracy in predictive modeling tasks and is particularly useful due to its lower memory footprint and faster training, especially when dealing with complex datasets.

5. RESULTS AND DISCUSSION

In this study, 70 medical features, including OH, were recorded during pediatric hemodialysis at the University Children's Hospital in Tiršova. All these medical features are sorted into a growing database, which is updated with new measurements every 15 minutes during each hemodialysis session. This creates a comprehensive picture of a patient's health status.

A. Experimental environment

The optimization and regression were performed in Anaconda's Spyder [43] on a computer with Intel Core i5 – 10400F, 2.90 GHz, 16 GB RAM, nVidia GTX1650 4 GB DDR6, and Windows 10 operating system. The algorithm was trained by dividing the data 80-20 (80 % for training and 20 % for testing). Cross-validation is a statistical technique for evaluating the performance of a ML model in which the dataset is divided into subsets, the model is trained on some of these subsets and evaluated on the remaining subsets [44]. To overcome the overfitting limitation in standard Grid search, k -fold cross-validation is used, where the samples are randomly divided into k -folds.

Hastie et al. [48] suggested that EN can converge for $0 \leq \alpha \leq 1$ and $0 \leq \lambda \leq 1$, while GBR converges for $0 < \nu < 1$. Adjusting the parameters α and λ in the EN part of the model can help to control the complexity and select relevant variables. On the other hand, the parameter ν in the GBR part of the model can be explored to achieve optimal convergence speed and model generalization in different patients. Adjusting the ν parameter can affect how GBR uses information from previous iterations to update hydration predictions, which can be crucial for good performance.

As suggested by [49], the model parameters for SVR are as assumed to be $0 < \epsilon \leq 1$ and $1 \leq C \leq 100$. The adjustment of the ϵ and C parameters in the SVR part of the model controls the complexity and the variable selection. The exploration of the ϵ and C parameters in the SVR part of the model aims to achieve optimal generalization for each patient and performance, which affects the adjustment of the hyperplane for prediction.

When searching for the optimal parameters for hybrid models, optimizing the parameters for both components combined (EN and GBR, or EN and SVR) is a crucial but complex step. However, it leads to a model that combines the

advantages of both techniques and provides better performance than using either technique individually. To find the optimal parameters of both EN-GBR and EN-SVR models, the GridSearchCV algorithm from Scikit-learn [45] is used. The optimization process is patient-specific, reflecting the individual nature of pediatric care and allowing real-time adaptation to each patient's fluid balance trajectory.

B. Performance metrics

If $\bar{Y}$ represents the mean of m samples of real hydration values over time and Y_i and $\hat{Y}_i$ are actual and estimated hydration values, respectively, we evaluate hybrid models using the RMSE, the mean absolute percentage error (MAPE), and the coefficient of determination (R^2), [46], [47], see Table 1:

Table 1. Performance metrics.

R^2	RMSE	MAPE
$1 - \frac{\sum_{i=1}^m (\hat{Y}_i - Y_i)^2}{\sum_{i=1}^m (\bar{Y} - Y_i)^2}$	$\sqrt{\frac{1}{m} \sum_{i=1}^m (Y_i - \hat{Y}_i)^2}$	$\frac{1}{m} \sum_{i=1}^m \left \frac{Y_i - \hat{Y}_i}{Y_i} \right \cdot 100$

C. Predicted hydration using hybrid models

Fig. 2 and Fig. 3 show the performance of two hybrid models in terms of deviation from the real hydration values of a single pediatric patient.

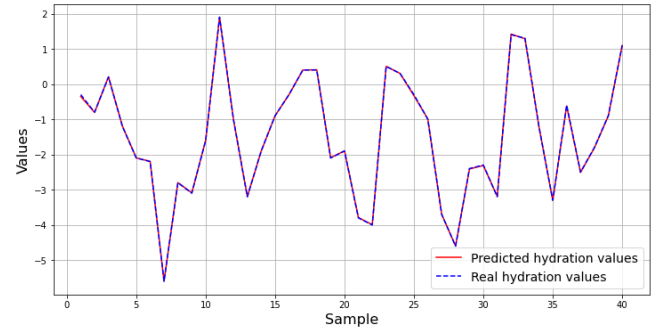


Fig. 2. EN-GBR hybrid model hydration predictions vs. real hydration values.

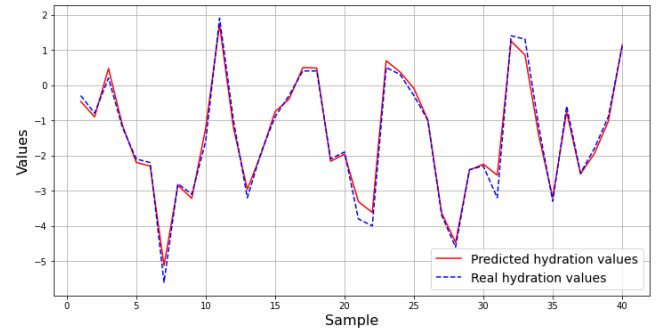


Fig. 3. EN-SVR hybrid model hydration predictions vs. real hydration values.

Comparing the two hybrid models, elastic boosting (EN-GBR) and elastic support regressor (EN-SVR), it can be seen that both models provide satisfactory prediction results. The

results of the EN-SVR are quite good (Fig. 3), with the most significant deviations occurring at the hydration peaks.

The EN-GBR model is characterized by its better performance at the peak prediction values. This advantage of the EN-GBR model can be attributed to its ability to better capture nonlinearities in the data through the combination of linear regression and a powerful nonlinear model such as the GBR.

D. Clinical relevance and case-specific interpretation

Accurate assessment of hydration in children on dialysis is a major clinical challenge due to their increased susceptibility to fluid loss or retention. Miscalculations of only a few deciliters can lead to serious complications, including hypotension, cramps, nausea, or fluid retention with edema.

The models developed in this study allow for personalized prediction of hydration status based on parameters routinely measured during each dialysis treatment. Each model is trained using the patient's personal measurements and adapted to their specific characteristics. This allows the physician to assess in real time whether the planned UF will result in excessive fluid loss or insufficient fluid removal.

For example, in patient A, an adolescent with a stable BW of 64 kg, significant oscillations in OH values were observed — from mild hyperhydration (+0.2 L) to severe dehydration (-5.6 L). Such a range may indicate a clinically unstable status in terms of fluid balance. Using the EN-GBR and EN-SVR models, predictions of OH values were on average within an error of less than 0.1 liters, even at the extreme points. This accuracy would give the clinical team additional confidence in assessing optimal UF during dialysis, reducing the likelihood of episodes of hypotension or residual hyperhydration.

Such an approach may lead to better therapeutic results, but also to a reduced need for invasive methods of hydration assessment. The introduction of such models into practice allows for continuous and automated monitoring of the patient's condition, which is particularly important in pediatrics, where tolerance limits are much narrower than in adult patients.

E. Comparison of hydration prediction of various models

In Table 2, we present the evaluation metrics of two pediatric patients with two different personalized datasets. The analysis of hybrid ML models for hydration level assessment in hemodialysis patients shows their superiority over individual models. The combined EN-GBR model achieves an exceptionally high coefficient of determination (R^2) of 0.99960 for patient A, and 0.99919 for patient B, with minimal MAPE values of 0.16218 and 0.10223 for both patients and RMSE of 0.007 and 0.011. The RMSE improves by over 60 % and the MAPE by approximately 70 %. These results indicate a greatly improved agreement between predicted and actual hydration values. The EN-SVR also performed well with a high R^2 of 0.98259 and 0.96178, respectively and very low RMSE and MAPE values. The improvement in RMSE and MAPE is 20 % and 26 %, respectively.

Table 2. Model comparisons.

	Model	R^2	RMSE	MAPE
Patient A	EN	0.97770	0.33319	0.103
	GBR	0.99969	0.50409	0.012
	SVR	0.92285	0.53131	0.095
	EN-GBR	0.99960	0.16218	0.007
	EN-SVR	0.98259	0.30489	0.092
	Ridge regression	0.91005	0.44157	0.243
	Kernel Ridge	0.90765	0.43811	0.257
	Bayesian Ridge	0.89345	0.45118	0.256
	RF	0.68556	0.46803	0.433
	LSTM	0.95325	0.38197	0.136
Patient B	EN	0.97998	0.28097	0.055
	GBR	0.99802	0.27370	0.015
	SVR	0.97227	0.54102	0.141
	EN-GBR	0.99919	0.10223	0.011
	EN-SVR	0.96178	0.37549	0.053
	Ridge regression	0.97089	0.34216	0.062
	Kernel Ridge	0.96991	0.34059	0.063
	Bayesian Ridge	0.89360	0.46269	0.055
	RF	0.93517	0.35403	0.789
	LSTM	0.84637	0.32715	0.408

Alternative models, including Ridge regression, Kernel Ridge, Bayesian Ridge, RF, and LSTM, showed significantly lower performance compared to the hybrid approaches. In particular, RF had the lowest coefficient of determination ($R^2 = 0.68556$ for Patient A) and the highest MAPE value (up to 0.789 for Patient B) for both patients, while Ridge and Kernel Ridge models achieved slightly better accuracy but remained inferior to hybrid models such as EN-GBR and EN-SVR, which consistently had the lowest error values and the highest R^2 coefficients.

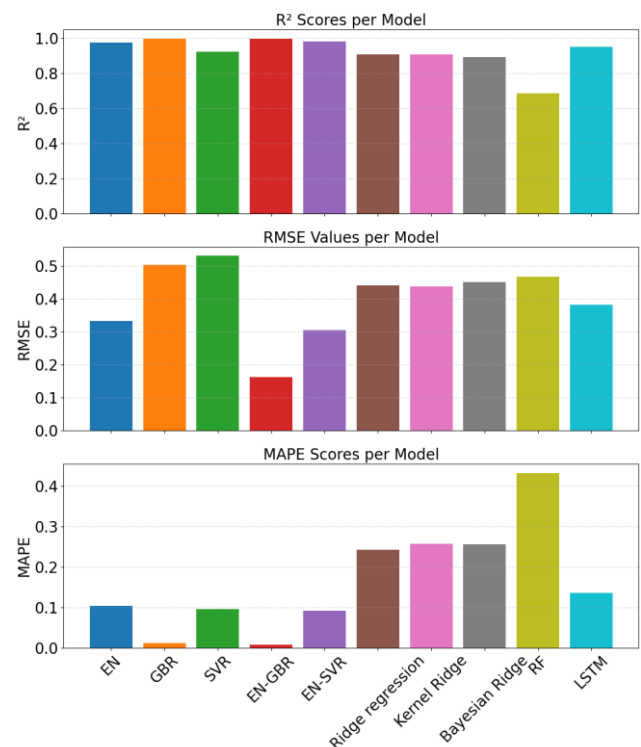


Fig. 4. Comparison between the models' metrics.

These results underline the advantage of using hybrid models, which combine the strengths of different ML algorithms, over individual models. Hybrid models exhibit higher accuracy and precision in predicting hydration levels, making them tools of choice for improving clinical practice in the monitoring and treatment of hemodialysis patients. Fig. 4 illustrates the metrics of the models.

A patient's sensitivity to the volume of fluid removed during dialysis can result in dehydration if the amount removed exceeds the optimal level, or fluid overload (edematous state) if the amount removed is insufficient. These changes are expressed quantitatively in decimal values. For example, if 2.9 liters of fluid are removed from a patient when the optimal amount was estimated at 2.1 liters, such

a discrepancy can lead to dehydration. Variations in OH are thus reflected in subtle decimal changes. The algorithm used in this context is based on a regression approach that aims to predict continuous values rather than categorical results.

To complete this analysis, we present in Table 3, Fig. 5 and Fig. 6 part of the hydration regression results of the EN, GBR, SVR, EN-GBR, and EN-SVR models on an unseen dataset.

Table 3 shows that EN and GBR achieve satisfactory results on their own, with EN-GBR hybrid as their combination showing similar or even better results than each individual model. The SVR model has the least accurate predictions, but the hybrid model obtained by combining the EN and SVR models EN-SVR improves on this, but still performs worse than EN-GBR.

Table 3. Results achieved with the EN, GBR, SVR, hybrid EN-GBR and hybrid EN-SVR models.

	Real values	EN values	GBR values	SVR values	EN-GBR values	EN-SVR values
Patient A	-0.3	-0.48129796	-0.18628015	-0.81704017	-0.3527895	-0.46503815
	-0.8	-0.91890762	-0.80598051	-1.27369152	-0.80427712	-0.90426637
	0.2	0.50873695	0.23782386	0.39220703	0.20545681	0.46932447
	-1.2	-1.11180823	-1.16581552	-1.27910948	-1.18921171	-1.13263351
	-2.1	-2.2058463	-2.09972043	-1.83435081	-2.09690962	-2.19735346
	-2.2	-2.30081989	-2.18690249	-1.91096666	-2.1958786,	-2.30016087
	-5.6	-5.05723497	-5.59310215	-3.86500347	-5.59597978	-5.12703839
	-2.8	-2.83149824	-2.79700751	-2.38790569	-2.79852713	-2.84700099
	-3.1	-3.28661257	-3.10436425	-2.90010456	-3.08728966	-3.21576016
	1.1	1.13724622	1.08674662	0.91430495	1.08914686	1.13514346
Patient B	2.3	2.05498245	2.30474957	2.07372147	2.30325372	2.32446905
	0.6	0.67973341	0.62278249	0.72046415	0.61362807	0.70735133
	-1.1	-0.98525658	-1.16755784	-1.07055426	-1.10503825	-1.14305158
	-1.3	-1.19201570	-1.40885596	-1.32889134	-1.32276658	-1.37031130
	-1.8	-1.72955661	-1.73768376	-1.84493265	-1.80881611	-1.86989233
	-1.7	-1.66446912	-1.69278792	-1.80055228	-1.73364828	-1.81502659
	0.9	0.85001765	0.93925670	0.88927059	0.91509444	0.92993130
	-3.8	-3.84511421	-3.89311761	-3.89983185	-3.79272144	-3.84172986
	-4.7	-4.97572847	-4.68874558	-4.75482693	-4.72968422	-4.79968123
	-2.2	-2.16571580	-2.17217104	-2.26907722	-2.20450492	-2.20647985

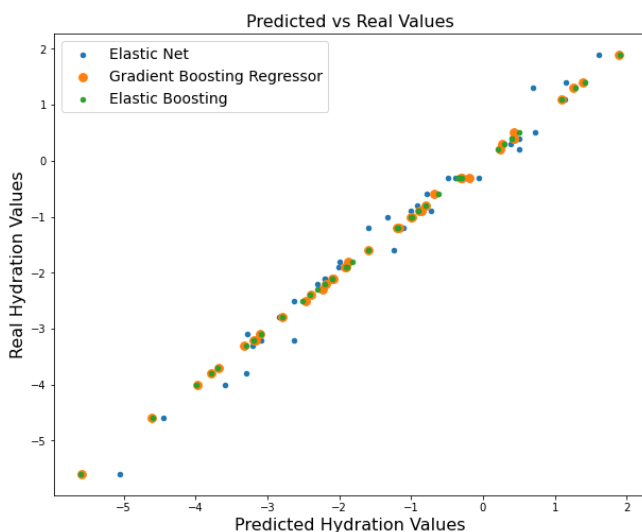


Fig. 5. Comparison between the EN, GBR, and EN-GBR hydration prediction vs. real hydration values.

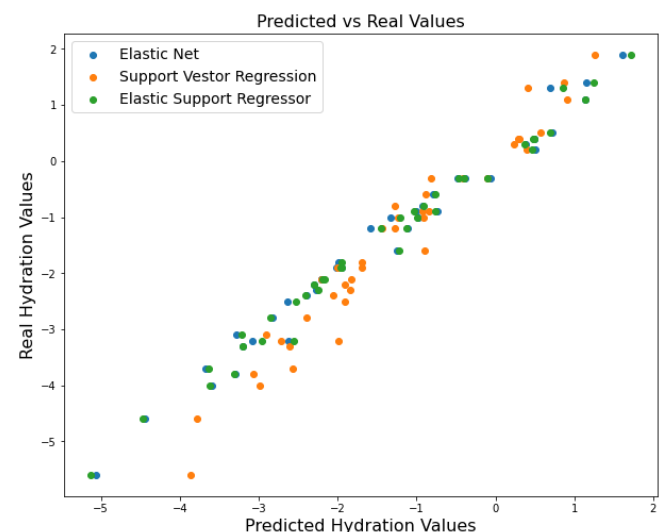


Fig. 6. Comparison between the EN, SVR, and EN-SVR hydration prediction vs. real hydration values.

Fig. 5 shows scatter plots of real vs. predicted hydration results for EN, GBR and EN-GBR hybrid. In Fig. 6 we also show a scatter plot with EN, SVR and EN-SVR. In both cases, the hybrid models improved the hydration predictions of the individual models.

In terms of computational efficiency, the EN-GBR hybrid model requires approximately 78000 ms for training, with an estimated memory usage of 51 % and CPU utilization of approximately 20 %. In comparison, the EN-SVR model completes training in approximately 17000 ms, utilizing about 50 % of memory and 12 % of CPU resources. The latest versions of the Raspberry Pi 4 have 8 GB of RAM memory and a Broadcom BCM2711 SoC with a 1.5 GHz CPU, which is quite satisfactory for hybrids. Since the interval between two measurements during the HD is 15 minutes, the hybrids have enough time for retraining.

6. CONCLUSION

Accurate prediction of hydration status is critical to improving the quality of care for hemodialysis patients. Proper hydration management can prevent complications and improve overall patient health. This paper presents a non-invasive solution for assessing hydration levels in pediatric hemodialysis patients. The development and application of hybrid models such as EN-GBR and EN-SVR show promising results in this field. These models utilize the strengths of both linear and nonlinear regression techniques and offer a more comprehensive approach to data analysis and prediction. In the case of EN-GBR, we achieve improvements of over 60 % and 70 % for RMSE and MAPE, respectively, compared to individual models. For EN-SVR, the corresponding improvements are 20 % for RMSE and 26 % for MAPE compared to individual models. The success of these hybrid models suggests that they can be valuable tools in the clinical setting, providing healthcare professionals with reliable and accurate predictions for tailoring treatments and interventions. This case study demonstrates that hybrid models, when applied to real clinical data from pediatric hemodialysis patients, can significantly improve hydration management. The individualized methodology ensures that each prediction is tailored to the patient's evolving condition, making this method highly applicable in clinical practice.

To fully exploit its potential, further improvements and adjustments are needed to explore its performance with more extensive and diverse datasets and to enable its generalizability across different patient populations and extended medical feature sets. We would also like to reduce the execution time of the EN-GBR model. One of the requirements in clinical practice is to run these algorithms on versatile mobile platforms such as the Raspberry Pi to enable portability and a personalized device per pediatric patient that allows constant monitoring of certain medical parameters, even outside of the clinical room. This option could generate notifications and alerts for both patients and hospitals to improve the quality of healthcare.

ACKNOWLEDGMENT

This study was supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia, and these results are parts of the Grant No. 451-03-66/2024-03/200132 with University of Kragujevac – Faculty of Technical Sciences Čačak.

The authors would like to thank Professor Sasa Cvetkovic for his valuable suggestions and contributions to this work.

REFERENCES

- [1] Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H., Wang, Y. (2017). Artificial intelligence in healthcare: Past, present and future. *Stroke and Vascular Neurology*, 2 (4), 230-243. <https://doi.org/10.1136/svn-2017-000101>
- [2] Chai, Y., Li, J., Sun, X., Lei, Z., Zhang, Z., Hou, G. (2020). The performance of a body composition-based equation in estimating overhydration of hemodialysis patients. *SN Comprehensive Clinical Medicine*, 2 (7), 909-913. <https://doi.org/10.1007/s42399-020-00338-5>
- [3] Pérez-Morales, R., Donate-Correa, J., Martín-Núñez, E., Pérez-Delgado, N., Ferri, C., López-Montes, A., Jiménez-Sosa, A., Navarro-González, J. F. (2021). Extracellular water/total body water ratio as predictor of mortality in hemodialysis patients. *Renal Failure*, 43 (1), 821-829. <https://doi.org/10.1080/0886022x.2021.1922442>
- [4] Nakagawa, H., Sato, Y., Toshimori, H., Fujimoto, S. (2015). Evaluation of a new bio-impedance spectroscopy device in healthy Japanese. *Bio-Medical Materials and Engineering*, 26 (3-4), 97-102. <https://doi.org/10.3233/bme-151553>
- [5] Loutradis, C., Sarafidis, P. A., Ekart, R., Papadopoulos, C., Sachpekidis, V., Alexandrou, M. E., Papadopoulou, D., Efstratiadis, G., Papagianni, A., London, G., Zoccali, C. (2019). The effect of dry-weight reduction guided by lung ultrasound on ambulatory blood pressure in hemodialysis patients: A randomized controlled trial. *Kidney International*, 95 (6), 1505-1513. <https://doi.org/10.1016/j.kint.2019.02.018>
- [6] Yajima, T., Arao, M., Yajima, K., Takahashi, H., Yasuda, K. (2019). The associations of fat tissue and muscle mass indices with all-cause mortality in patients undergoing hemodialysis. *PloS One*, 14 (2), e0211988. <https://doi.org/10.1371/journal.pone.0211988>
- [7] Zaloszc, A., Schaefer, B., Schaefer, F., Krid, S., Salomon, R., Niaudet, P., Schmitt, C. P., Fischbach, M. (2013). Hydration measurement by bioimpedance spectroscopy and blood pressure management in children on hemodialysis. *Pediatric Nephrology*, 28 (11), 2169-2177. <https://doi.org/10.1007/s00467-013-2540-6>
- [8] Rymarz, A., Gibińska, J., Zajbt, M., Piechota, W., Niemczyk, S. (2018). Low lean tissue mass can be a predictor of one-year survival in hemodialysis patients. *Renal Failure*, 40 (1), 231-237. <https://doi.org/10.1080/0886022x.2018.1456451>

- [9] Marcelli, D., Usvyat, L. A., Kotanko, P., Bayh, I., Canaud, B., Etter, M., Gatti, E., Grassmann, A., Wang, Y., Marelli, C., Scatizzi, L., Stopper, A., van der Sande, F. M., Kooman, J. (2015). Body composition and survival in dialysis patients: Results from an international cohort study. *Clinical Journal of the American Society of Nephrology*, 10 (7), 1192-1200. <https://doi.org/10.2215/cjn.08550814>
- [10] Khalil, S. F., Mohktar, M. S., Ibrahim, F. (2014). The theory and fundamentals of bioimpedance analysis in clinical status monitoring and diagnosis of diseases. *Sensors*, 14 (6), 10895-10928. <https://doi.org/10.3390/s140610895>
- [11] Jaffrin, M. Y., Morel, H. (2008). Body fluid volumes measurements by impedance: A review of bioimpedance spectroscopy (BIS) and bioimpedance analysis (BIA) methods. *Medical Engineering & Physics*, 30 (10), 1257-1269. <https://doi.org/10.1016/j.medengphy.2008.06.009>
- [12] Cole, K. S. (1940). Permeability and impermeability of cell membranes for ions. *Cold Spring Harbor Symposia on Quantitative Biology*, 8, 110-122. <https://doi.org/10.1101/SQB.1940.008.01.013>
- [13] Chamney, P. W., Wabel, P., Moissl, U. M., Müller, M. J., Bosy-Westphal, A., Korth, O., Fuller, N. J. (2007). A whole-body model to distinguish excess fluid from the hydration of major body tissues. *The American Journal of Clinical Nutrition*, 85 (1), 80-89. <https://doi.org/10.1093/ajcn/85.1.80>
- [14] Chamney, P. W., Krämer, M., Rode, C., Kleinekofort, W., Wizemann, V. (2002). A new technique for establishing dry weight in hemodialysis patients via whole body bioimpedance. *Kidney International*, 61 (6), 2250-2258. <https://doi.org/10.1046/j.1523-1755.2002.00377.x>
- [15] Zoccali, C., Tripepi, R., Torino, C., Bellantoni, M., Tripepi, G., Mallamaci, F. (2014). Lung congestion as a risk factor in end-stage renal disease. *Blood Purification*, 36 (3-4), 184-191. <https://doi.org/10.1159/000356085>
- [16] Agarwal, R., Weir, M. R. (2010). Dry-weight: A concept revisited in an effort to avoid medication-directed approaches for blood pressure control in hemodialysis patients. *Clinical Journal of the American Society of Nephrology*, 5 (7), 1255-1260. <https://doi.org/10.2215/cjn.01760210>
- [17] Agarwal, R., Kelley, K., Light, R. P. (2008). Diagnostic utility of blood volume monitoring in hemodialysis patients. *American Journal of Kidney Diseases*, 51 (2), 242-254. <https://doi.org/10.1053/j.ajkd.2007.10.036>
- [18] Wizemann, V., Schilling, M. (1995). Dilemma of assessing volume state—the use and the limitations of a clinical score. *Nephrology Dialysis Transplantation*, 10 (11), 2114-2117. <https://doi.org/10.1093/ndt/10.11.2114>
- [19] Secinaro, S., Calandra, D., Secinaro, A., Muthurangu, V., Biancone, P. (2021). The role of artificial intelligence in healthcare: A structured literature review. *BMC Medical Informatics and Decision Making*, 21 (1), 125. <https://doi.org/10.1186/s12911-021-01488-9>
- [20] Kanda, E., Epureanu, B. I., Adachi, T., Tsuruta, Y., Kikuchi, K., Kashiwara, N., Abe, M., Masakane, I., Nitta, K. (2020). Application of explainable ensemble artificial intelligence model to categorization of hemodialysis-patient and treatment using nationwide-real-world data in Japan. *PloS One*, 15 (5), e0233491. <https://doi.org/10.1371/journal.pone.0233491>
- [21] Guo, X., Zhou, W., Lu, Q., Du, A., Cai, Y., Ding, Y. (2021). Assessing dry weight of hemodialysis patients via Sparse Laplacian regularized RVFL neural network with $L_{2,1}$ -norm. *BioMed Research International*, 2021, 627650. <https://doi.org/10.1155/2021/6627650>
- [22] Djordjevic, S., Kostic, M., Milosevic, D., Cvetkovic, M., Mitrovic, K., Mladenovic, V. (2023). Prediction of overhydration in the process of pediatric hemodialysis using artificial neural network. In *2023 12th Mediterranean Conference on Embedded Computing (MECO)*. IEEE. <https://doi.org/10.1109/MECO58584.2023.10154915>
- [23] Muntasir Nishat, M., Faisal, F., Rahman Dip, R., Nasrullah, S. M., Ahsan, R., Shikder, F., Ar-Raihan Asif, M. A., Hoque, M. A. (2021). A comprehensive analysis on detecting chronic kidney disease by employing machine learning algorithms. *EAI Endorsed Transactions on Pervasive Health and Technology*, 7 (29), 170671. <https://doi.org/10.4108/eai.13-8-2021.170671>
- [24] Belur Nagaraj, S., Pena, M. J., Ju, W., Heerspink, H. L., The BEAt-DKD Consortium. (2020). Machine-learning-based early prediction of end-stage renal disease in patients with diabetic kidney disease using clinical trials data. *Diabetes, Obesity & Metabolism*, 22 (12), 2479-2486. <https://doi.org/10.1111/dom.14178>
- [25] Arif, M. S., Mukheimer, A., Asif, D. (2023). Enhancing the early detection of chronic kidney disease: A robust machine learning model. *Big Data and Cognitive Computing*, 7 (3), 144. <https://doi.org/10.3390/bdcc7030144>
- [26] Swain, D., Mehta, U., Bhatt, A., Patel, H., Patel, K., Mehta, D., Acharya, B., Gerogiannis, V. C., Kanavos, A., Manika, S. (2023). A robust chronic kidney disease classifier using machine learning. *Electronics*, 12 (1), 212. <https://doi.org/10.3390/electronics12010212>
- [27] Shih, C.-C., Chen, S.-H., Chen, G.-D., Chang, C.-C., Shih, Y.-L. (2021). Development of a longitudinal diagnosis and prognosis in patients with chronic kidney disease: Intelligent clinical decision-making scheme. *International Journal of Environmental Research and Public Health*, 18 (23), 12807. <https://doi.org/10.3390/ijerph182312807>
- [28] Swain, D., Patel, H., Patel, K., Sakariya, V., Chaudhari, N. (2022). An intelligent clinical support system for the early diagnosis of the chronic kidney disease. In *2022 IEEE 2nd International Symposium on Sustainable Energy, Signal Processing and Cyber Security (iSSSC)*. IEEE. <https://doi.org/10.1109/iSSSC56467.2022.10051517>

- [29] Mladenović, V., Kostić, M., Milošević, D., Zanjaj, E., Đorđević, S. (2024). *System for prediction and balancing excess fluid in the body during hemodialysis based on artificial intelligence*. Patent RS20240030A2.
- [30] Ren, Y., Fei, H., Liang, X., Ji, D., Cheng, M. (2019). A hybrid neural network model for predicting kidney disease in hypertension patients based on electronic health records. *BMC Medical Informatics and Decision Making*, 19 (S2), 51. <https://doi.org/10.1186/s12911-019-0765-4>
- [31] Singh, V., Jain, D. (2021). A hybrid parallel classification model for the diagnosis of chronic kidney disease. *International Journal of Interactive Multimedia and Artificial Intelligence*, 8 (2). <https://doi.org/10.9781/ijimai.2021.10.008>
- [32] Dey, S. K., Uddin, K. M. M., Babu, H. M. H., Rahman, M. M., Howlader, A., Uddin, K. M. A. (2022). Chi2-MI: A hybrid feature selection based machine learning approach in diagnosis of chronic kidney disease. *Intelligent Systems with Applications*, 16, 200144. <https://doi.org/10.1016/j.iswa.2022.200144>
- [33] Khalid, H., Khan, A., Khan, M. Z., Mehmood, G., Qureshi, M. S. (2023). Machine learning hybrid model for the prediction of chronic kidney disease. *Computational Intelligence and Neuroscience*. <https://doi.org/10.1155/2023/9266889>
- [34] Yadav, D. C., Pal, S. (2021). Performance based evaluation of Algorithmson Chronic Kidney Disease using hybrid ensemble model in machine learning. *Biomedical & Pharmacology Journal*, 14 (3), 1633-1645. <https://dx.doi.org/10.13005/bpj/2264>
- [35] Ratnababu, M., Naidu, M. R. (2019). A hybrid method for detection of kidney disease using Machine Learning. *Turkish Journal of Computer and Mathematics Education*, 11 (2), 654-669.
- [36] Jhou, M.-J., Chen, M.-S., Lee, T.-S., Yang, C.-T., Chiu, Y.-L., Lu, C.-J. (2022). A hybrid risk factor evaluation scheme for metabolic syndrome and stage 3 chronic kidney disease based on multiple machine learning techniques. *Healthcare*, 10 (12), 2496. <https://doi.org/10.3390/healthcare10122496>
- [37] Guli, U. J., Malgwi, Y. M., Adamu, I., Ezekiel, P. (2023). Hybrid diagnostic model for kidney disease prediction using data mining techniques. *African Journal of Advances in Science and Technology Research*, 11 (1), 75-89.
- [38] Zhao, Y., Ogden, R. T., Reiss, P. T. (2012). Wavelet-based LASSO in functional linear regression. *Journal of Computational and Graphical Statistics*, 21 (3), 600-617. <https://doi.org/10.1080/10618600.2012.679241>
- [39] Rao, H., Shi, X., Rodrigue, A. K., Feng, J., Xia, Y., Elhoseny, M., Yuan, X., Gu, L. (2019). Feature selection based on artificial bee colony and gradient boosting decision tree. *Applied Soft Computing*, 74, 634-642. <https://doi.org/10.1016/j.asoc.2018.10.036>
- [40] Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29 (5), 1189-1232. <https://doi.org/10.1214/aos/1013203451>
- [41] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Müller, A., Nothman, J., Louppe, G., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. In *Journal of Machine Learning Research*, 12 (85), 2825-2830. <http://jmlr.org/papers/v12/pedregosa11a.html>
- [42] Brentan, B. M., Luvizotto Jr., E., Herrera, M., Izquierdo, J., Pérez-García, R. (2017). Hybrid regression model for near real-time urban water demand forecasting. *Journal of Computational and Applied Mathematics*, 309, 532-541. <https://doi.org/10.1016/j.cam.2016.02.009>
- [43] Anaconda Inc. (2024). Unleash AI innovation and value. <https://www.anaconda.com/>
- [44] James, G., Witten, D., Hastie, T., Tibshirani, R. (2021). *An Introduction to Statistical Learning: With Applications in R. Second Edition*. Springer, ISBN 978-1071614174.
- [45] Algamal, Z. Y., Lee, M. H. (2015). Penalized logistic regression with the adaptive LASSO for gene selection in high-dimensional cancer classification. *Expert Systems with Applications*, 42 (23), 9326-9332. <https://doi.org/10.1016/j.eswa.2015.08.016>
- [46] Brentan, B. M., Luvizotto Jr., E., Herrera, M., Izquierdo, J., Pérez-García, R. (2017). Hybrid regression model for near real-time urban water demand forecasting. *Journal of Computational and Applied Mathematics*, 309, 532-541. <https://doi.org/10.1016/j.cam.2016.02.009>
- [47] Chicco, D., Warrens, M. J., Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ: Computer Science*, 7, e623. <https://doi.org/10.7717/peerj-cs.623>
- [48] Hastie, T., Tibshirani, R., Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Second Edition*. Springer. <https://doi.org/10.1007/978-0-387-84858-7>
- [49] Lin, H. T., Lin, C. J. (2003). *A study on sigmoid kernels for SVM and the training of non-PSD kernels by SMO-type methods*. Technical Report, Department of Computer Science and Information Engineering, National Taiwan Normal University, Taipei City, Taiwan.

Received July 21, 2024

Accepted July 14, 2025