Optimizing Hemodialysis Machine Performance: A Comparative Study of DL vs ML for Patient Weight Prediction

Intissar Zaway[1], Lassaad Zaway[2], and Kaouthar Mansour[3]

[1] CEM-Laboratory, National School of Engineering of Sfax, University of Sfax, Tunisia,

[2] CES-Laboratory, National School of Engineering of Sfax, University of Sfax, Tunisia,

[3] IResCoMath-Laboratory, National School of Engineering of Gabes, University of Gabes , Tunisia,

Abstract

Accurate evaluation of post-hemodialysis weight is critical for improving fluid management dear patient with renal disease, since both fluid overload and excessive ultrafiltration are linked with increased morbidity and death. Despite its clinical importance, existing procedures to determine weight after hemodialysis session frequently rely on empirical clinical experience and can be subjective and inaccurate. This study compares the efficacy of different machine learning and deep learning models for predicting post-hemodialysis weight in patients with renal illness. Five alternative architectures Multilayer Perceptron (MLP), Support Vector Regression (SVR), Gradient Boosting (GB), Long Short-Term Memory (LSTM), and Bidirectional LSTM (Bi-LSTM) were implemented and analyzed using numerous performance criteria.

While all models had excellent correlation coefficients (> 0.99), the MLP model displayed higher prediction accuracy, with the lowest Mean Absolute Error (MAE: 0.274 kg) and Root Mean Square Error (RMSE: 0.383 kg). Despite their theoretical advantage in handling temporal data, sequential models (LSTM and Bi-LSTM) demonstrated greater error rates, with the Bi-LSTM (MAE: 0.239 kg) surpassing the regular LSTM (MAE: 0.856 kg).

Visualization of prediction patterns across patient weight ranges (40-110 kg) revealed the MLP's constant performance to effectively track both slow weight transitions and rapid weight changes. These findings indicate that simpler structures may be more successful for clinical weight prediction in dialysis settings, thereby enhancing fluid management decisions..

Keywords : Hemodialysis, Nephrologists, Dry weight, Clinical decision support, Machine learning and Deep learning.

1 Introduction

The management of dry weight (DW) in patients with end-stage kidney disease (ESKD) remains a critical and persistent challenge in hemodialysis care. Precise estimation of DW is essential to ensure proper fluid balance, prevent cardiovascular complications, and enhance overall treatment outcomes. Inaccurate DW assessment can result in severe consequences such as hypotension, fluid overload, or shock, thereby compromising patient safety and the effectiveness of dialysis sessions. Over the past two decades, DW estimation methods have significantly progressed from conventional bio impedance techniques to advanced, data-driven approaches powered by artificial intelligence (AI) and machine learning (ML) that leverage multimodal clinical data [1, 3, 13].

Bioimpedance analysis (BIA) has long served as a foundational tool for evaluating body composition and hydration status. Chamney et al. [3] introduced whole body BIA for estimating extracellular fluid volume, establishing a baseline methodology for DW estimation. This was later validated by Basile et al. [2], who confirmed the clinical utility of BIA based prediction equations (mean absolute error [MAE] 0.8 kg). Comparative studies, such as that by Jian et al. [**?**], further demonstrated BIA's superior accuracy (sensitivity: 92%, specificity: 88%) over traditional clinical assessment. More recently, Mussnig et al. [8]

underscored the necessity of sex-specific BIA parameters to address inter individual variability in body composition, linking optimal survival rates to lean tissue index (LTI 12.0 kg/m^2) and fat tissue index (FTI 8.5 kg/m^2).

AI and Machine Learning Approaches have revolutionized DW prediction through improved accuracy and adaptability. Barbieri et al. [1] introduced one of the first AI models combining blood pressure, fluid volume, and dialysis dose, achieving a mean absolute percentage error (MAPE) of 3.1% in clinical validation. Building on this, Guo et al. [5, 10] employed advanced models including multiple kernel support vector regression (RMSE: 1.38 kg) and sparse Laplacian-regularized RVFL neural networks (accuracy: 94.7%), demonstrating high predictive performance. Boonvisuth [4] confirmed that ML models could rival BIA accuracy (MAE: 0.72 kg) while reducing dependence on manual measurement. Recent advancements by Inoue et al. [9] incorporated machine learning to dynamically adjust DW, achieving area under the curve (AUC) values of 0.70 and 0.74 for upward and downward adjustments, respectively. Yang et al. [13] further optimized DW assessment using reinforcement learning, reducing prediction errors by 21% compared to static models.

Clinical Integration via EHR Systems has enabled scalable implementation. Bi et al. [6] designed a time-series regression model using EHR data from 1,852 patients, achieving 95.44% accuracy in predicting post-dialysis weight within a 0.5 kg margin. Kim et al. [12] developed an ML-based decision-support tool that minimized inter-clinician variability (Cohens improvement: from 0.32 to 0.67), while Kim et al. [15] demonstrated the applicability of ML in hemodialysis athletes, achieving 89% precision in DW prediction despite unique body composition challenges.

Limitations and Future Directions: Heterogeneity among patient populations, variability in clinical practices, and the absence of standardized predictive frameworks hinder broader implementation. Germain et al. [14] highlighted discrepancies between estimated DW and actual post-transplant weight (mean deviation: 2.1 $\pm$ 1.3 kg), advocating for dynamic recalibration. In response, de Miranda Guimarães et al. [11] introduced adiposity-focused equations ($R^2 = 0.82$) to refine DW estimates by incorporating fat mass as a critical factor.

Our research is driven by the clinical importance of accurately predicting post-dialysis weight, a key factor for tailoring treatment and ensuring patient safety. We present a comprehensive study employing both classical ML models (MLP, SVR, Gradient Boosting) and deep learning architectures (LSTM, Bi-LSTM) to identify the most accurate and computationally efficient approach. Our goal is to develop a robust, real-time solution that supports clinical decision-making and contributes to improved patient outcomes in dialysis care.

2 Data preprocessing

The following section provides a summary of the characteristics and analytical algorithms employed in this research. Multivariate time-series data collected during hemodialysis sessions are included in the collection. Below is a collection of the raw and derived features:

2.1 Dataset Description

2.2 Basic Features

The dataset includes multivariate time-series data collected during hemodialysis sessions, capturing physiological and operational parameters critical for monitoring patient stability (Table 1).

Table 1: Measured clinical parameters

Feature	Description	Unit Format
Patient_ID	Unique patient identifier	Number
Hour	Time elapsed in dialysis session	Hours (1-5)
SYS_BP	Systolic blood pressure	mmHg
DIA_BP	Diastolic blood pressure	mmHg
UF_Rate	Prescribed ultrafiltration rate	mL/kg/hr
UF_Real	Actual ultrafiltration achieved	mL
Pump_Speed	Blood pump speed	mL/min
Weight_Before	Patient weight pre-dialysis	kg
Weight_After	Patient weight post-dialysis	kg

The database consists of nine direct observations from dialysis equipment and patient monitoring devices. It includes vital indicators (systolic and diastolic blood pressure), machine characteristics (ultrafiltration rate, pump speed), and weight measurements taken before and after treatment. These raw measurements serve as critical inputs for all future extracted features.

2.3 Extracted Features

To enhance the predictive capacity of machine learning models significant characteristics are derived from raw hemodialysis data. The following features are calculated:

2.3.1 Hemodynamic Features

Five Cardiovascular markers are developed from blood pressure measurements (Table 2). While pulse pressure (PP) indicates arterial stiffness, the mean arterial pressure (MAP) reflects cardiac cycle dynamics. Moreover, blood pressure stability during therapy is measured by the three temporal features: trend, decline, and variability.

Table 2: Blood Pressure Derived Feature

Feature	Formula	Clinical Relevance
MAP	$DIA\_BP + \frac{SYS\_BP - DIA\_BP}{3}$	Perfusion pressure
PP	$SYS\_BP - DIA\_BP$	Vascular stiffness indicator
BP_Variability	$\sigma(SYS\_BP)$	Blood pressure stability
BP_Drop	$SYS\_BP_{initial} - SYS\_BP_{current}$	Hypotension detection
BP_Trend	$\frac{\Delta SYS\_BP}{\Delta time}$	Progressive BP changes

2.3.2 Ultrafiltration Features

The second features type characterizes fluid removal efficiency (Table 3). The system normalizes UF rates by patient weight, calculates cumulative fluid extraction, and compares actual versus theoretical removal rates. Advanced metrics like UF variability and acceleration detect abnormal patterns and the second derivative of fluid removal indicates the accelerating fluid loss. These features enable real-time machine calibration and treatment adjustment.

Table 3: Fluid removal metrics

Feature	Formula	Purpose
UF_Rate_per_Weight	$\frac{UF\_Rate}{Weight\_Before}$	Normalized UF rate
UF_Cumulative	$\sum UF\_Real$	Total fluid removed
UF_Cumulative_Rate	$\frac{UF\_Cumulative}{Hour \times Weight\_Before}$	Time-adjusted efficiency
UF_Efficiency	$\frac{UF\_Real}{UF\_Rate \times Hour \times 250}$	Actual vs. expected UF
UF_Variability	$\sigma(UF\_Real)_{per\ session}$	Treatment consistency
UF_Acceleration	$\frac{\Delta^2 UF\_Real}{\Delta time^2}$	UF rate changes

2.3.3 Weight and Temporal Features

Three time-dependent weight metrics complete the feature set as it is depicted in table 4. The weight loss rate (kg/hr) and its acceleration (kg/hr^2) provide safety monitoring, with thresholds triggering alerts when exceeding or showing positive acceleration. Session duration plays a quality control measure, identifying prematurely terminated treatments. Together, these features create a comprehensive safety that addresses the most critical risks in dialysis management.

Table 4: Weight dynamics and time-dependent metrics

Feature	Formula	Purpose
Session_Duration	$\max(Hour)_{per\ session}$	Total treatment duration
Weight_Loss_Rate	$\frac{Weight\_Before - Weight\_After}{Hour}$	Hourly weight loss velocity
Weight_Loss_Acceleration	$\frac{\Delta^2(Weight\_Loss\_Rate)}{\Delta time^2}$	Non-linear weight change detection

These engineered features capture hemodynamic stability, fluid kinetics, and patient-specific treatment responses, enabling robust modeling of post-dialysis outcomes.

3 Proposed Algorithms

The methodlogy process is described in Figure 1. It shows the extensive machine learning process for dialysis weight prediction.
Initially, the framework evaluates essential physiological variables. The feature importance analysis reveals that baseline weight, normalized ultrafiltration rates, and cumulative fluid removal metrics are the most significant predictors of patient outcomes.

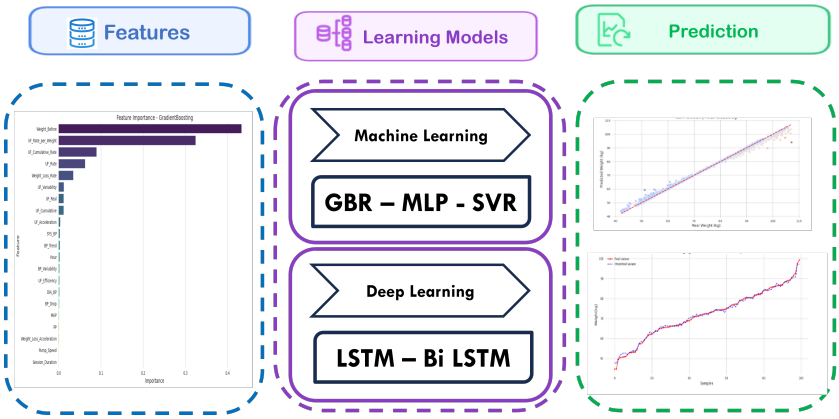

Figure 1: Methodology Schema

The process then employs a multi-model comparative approach, assessing both classic machine learning algorithms (Gradient Boosting, MLP, and SVR) and specific deep learning architectures (LSTM and Bi-LSTM) built to capture temporal relationships during dialysis sessions. The final prediction stage validates the model's therapeutic value through rigorous validation, revealing a good connection between anticipated and real post-dialysis weights across patient samples, allowing for accurate weight forecasting throughout therapy.
This integrated analytical approach gives medical practitioners a dependable tool for forecasting patient responses to dialysis procedures, which could improve clinical decision-making and therapy personalization.
The models are categorized into two groups: traditional machine learning-based and deep learning-based approaches.

3.1 Machine Learning Models

The Gradient Boosting Regressor (GBR): The GBR model combines 300 decision trees of maximum depth 5, using mean squared error (MSE) as the splitting criteria as it is showed in table 5. GBR uses a learning rate of 0.05 for each split. Regularization requires a minimum of two leaf samples and five split samples.
Training includes 10 boosting rounds with little progress.

Table 5: Gradient Boosting Regression (GBR) Specifications

Category	Description
Architecture	• 300 decision trees • Maximum depth: 5 levels • Splitting criterion: MSE
Learning	• Learning rate: 0.05 • Feature subsampling ($\sqrt{p}$) at each split • Minimum 5 samples per split
Regularization	• min_samples_leaf=2 • Stopping after 10 iterations without improvement

Support Vector Regression (SVR): The radial basis function (RBF) kernel is implemented with penalty parameter C=10 and epsilon=0.1. This configuration provides robust performance against the moderate noise levels expected in dialysis monitoring data. The RBF kernel's non-linear transformation enables effective modeling of complex physiological relationships while maintaining computational efficiency through the kernel trick. The carefully tuned epsilon-insensitive band creates an appropriate tolerance for typical measurement variations in blood pressure and ultrafiltration rates. Table 6 summarizes all the parameters of the SVR algorithm.

Table 6: Support Vector Regression (SVR) Specifications

Category	Description
Kernel	• Radial Basis Function (RBF): • $\gamma = 1/(n\_features \times Var(X))$ • Regularization parameter $C = 10$
Loss Function	• ϵ-insensitive ($\epsilon = 0.1$) with soft margin
Implementation	• scikit-learn (SVR kernel='rbf') • Memory cache limited to 4GB

The Multilayer Perceptron (MLP): The MLP model (Table 7) uses a three-layer feedforward neural network (256-128-64 neurons) with ReLU activation in the hidden layers and linear output for

regression. The Adam optimizer is used for training, with an initial learning rate of 0.001 and dynamic adjustments based on plateau detection (reduction factor 0.5 after 5 stationary epochs). Regularization uses L2 weight decay (λ=0.0001), 20% dropout before the output layer, and batch normalization between layers. The model processes data in 64-sample batches with an early halt triggered after 15 epochs.

Table 7: Multilayer Perceptron (MLP) Specifications

Category	Description
Architecture	• 3 fully-connected layers (256-128-64 neurons) • ReLU activation for hidden layers • No activation function for output (regression)
Hyperparameters	• Initial learning rate: 0.001 (adaptive) • L2 regularization ($\lambda = 0.0001$) • Dropout rate: 20% before output layer
Optimization	• Adam optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.999$) • Batch size: 64 • Early stopping after 15 epochs without improvement

3.2 Deep Learning Models

Figure 2 represents the developed two deep learning architectures: the standard Long Short-Term Memory (LSTM) network and the Bidirectional LSTM (Bi-LSTM) model.

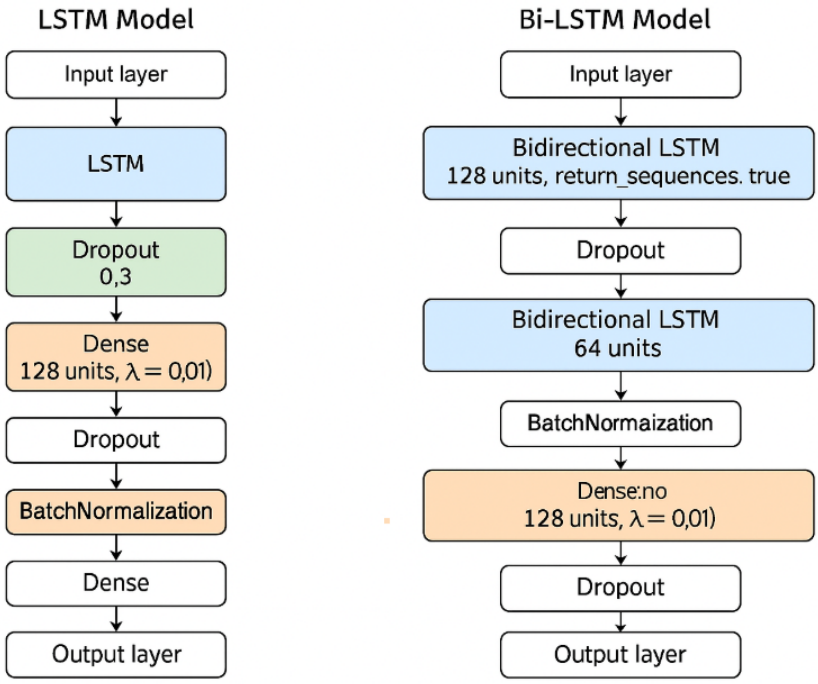

Figure 2: Architecture of models LSTM and BiLSTM.

Long short-term memory: The standard LSTM contains 256 memory units followed by dropout

and batch normalization layers, capturing temporal dependencies in treatment progression. The bidirectional variant stacks two LSTM layers (128→64 units) processing sequences in both directions, proving particularly effective for detecting subtle physiological trends that may manifest differently in forward versus backward analysis. Both architectures use tanh/sigmoid activation gates and Adam optimization for stable gradient flow during backpropagation through time.

Bi-directional Long short-term memory: The Bi-LSTM model, in contrast, is designed to leverage information from both past and future time steps by using bidirectional LSTM layers. It begins with a Bidirectional wrapper around an LSTM layer with 128 units and 'tanh' activation, configured to return sequences. A second Bidirectional LSTM layer with 64 units follows to extract deeper contextual features. Similar to the LSTM model, it includes a 128-unit dense layer with 'ReLU' activation and L2 regularization, batch normalization, dropout, and an additional 64-unit dense layer before the final regression output.

Both models are optimized using the Adam optimizer with a learning rate of 0.0005 to minimize error. The key difference lies in the use of bidirectionality in the Bi-LSTM model, which enables it to model sequences more comprehensively, albeit at a higher computational cost.

The comparative evaluation uses four metrics which are calculated on an independent test set. This quantitative analysis is complemented by clinically relevant visualizations: learning curves to diagnose model convergence, correlation matrices to identify variable dependencies, and prediction-versus-actuality plots to validate physiological coherence.

4 Results and Discussions

To evaluate the performance of the models, several criteria were computed. This section provides a detailed presentation of the simulation results with analyses and discussion. Performance is assessed through standard regression metrics (MSE, MAE, R^2) and comparative error analysis, with additional insights provided by learning curve visualizations and prediction plots.

4.1 Machine Learning Results

The results of simulations using different machine learning models are described in depth in the section that follows.

4.1.1 Machine Learning Prediction

A comparison of prediction models for post-hemodialysis weight estimate is shown in the three graphs. With the lowest Mean Absolute Error (0.2743 kg) and the greatest correlation coefficient (0.9995), the MLP (Figure 3) model performs exceptionally well, showing a close clustering of data points along the optimal prediction line.

The SVR (Figure 4) model, which retains strong accuracy but exhibits somewhat greater dispersion at higher weight ranges (90-110 kg), comes next with a correlation of 0.9984 and an MAE of 0.3256 kg.

The GradientBoosting (Figure 5) model shows the highest prediction variation across all weight ranges, yet maintaining a high level of accuracy (correlation 0.9979, MAE 0.5749 kg). Although all three models exhibit good performance throughout the whole range of patient weights (40–110 kg), the MLP model appears to have the most accurate clinical predictions for post-dialysis weight due to its much reduced error rate, which might lead to better fluid management choices.

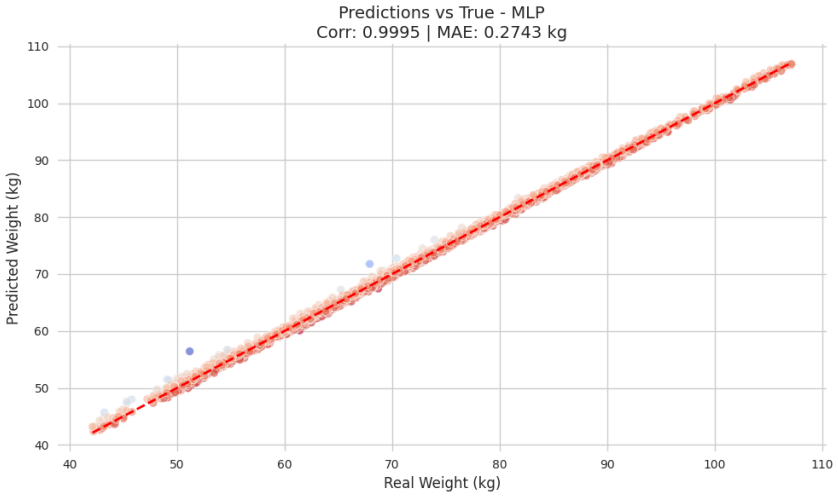

Figure 3: MLP Correlation

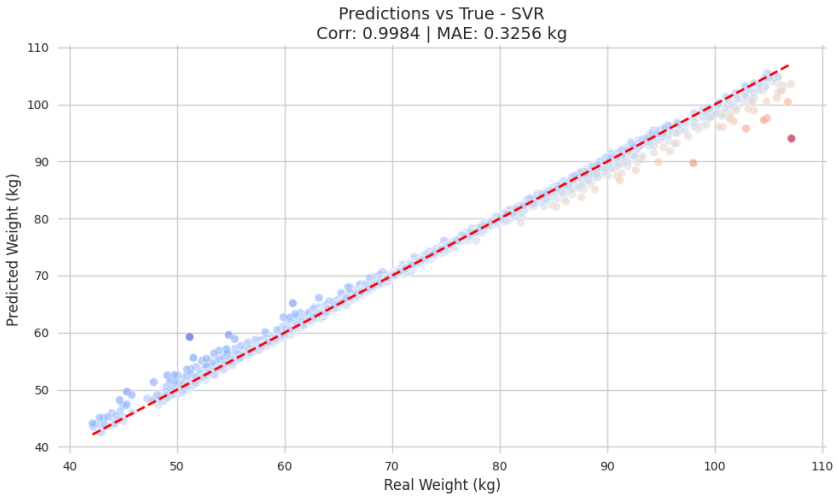

Figure 4: SVR Correlation

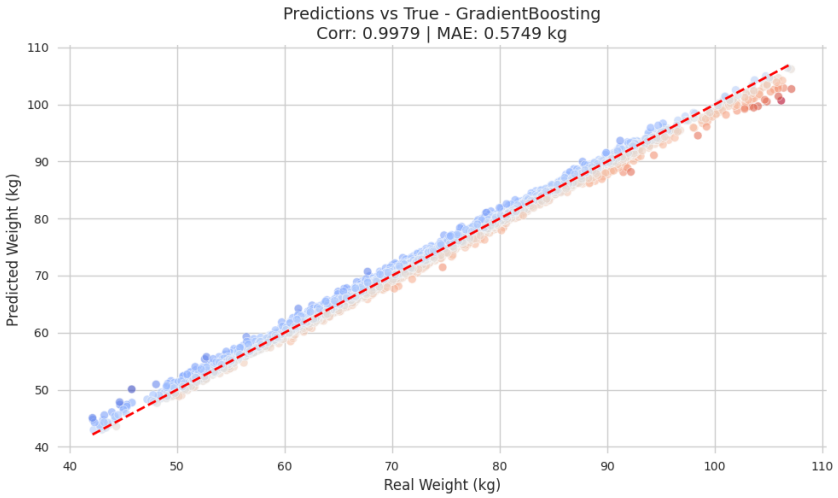

Figure 5: GBR Correlation

4.1.2 Error Curves Index

A time-series display of actual vs predicted post-hemodialysis weights demonstrates significant variations in model performance when tracking patient weight chronologically. The MLP model (Figure 6) has excellent temporal prediction accuracy, with projected values (blue dashed line) closely matching the actual weight trajectory (continuous red line) across the full sample range. This temporal alignment is most visible in the model's capacity to properly represent both slow weight transitions and rapid weight swings, with very little variance around sample point 30.

In contrast, the GradientBoosting model (Figure 7)exhibits more frequent oscillations around the real values, especially in the lower weight ranges (samples 0-20) and upper ranges (samples 80-100), where it tends to slightly overestimate and then underestimate the actual weights.

The SVR model (Figure 8) performs well in the middle of the samples but makes significant prediction errors at the extremes, with significant overestimation around samples 1-15 and a significant underestimation of the sharp weight increase at the end of the series (samples 95-100), where it fails to capture the full magnitude of the weight change.

Each of the models can effectively track the general weight pattern, but the SVR's inability to accurately predict sudden weight increases, as well as the Gradient Boosting's frequent oscillations, highlight the MLP's superior capability for precise weight prediction across various patient scenarios, confirming the lower MAE observed in the previous plots.

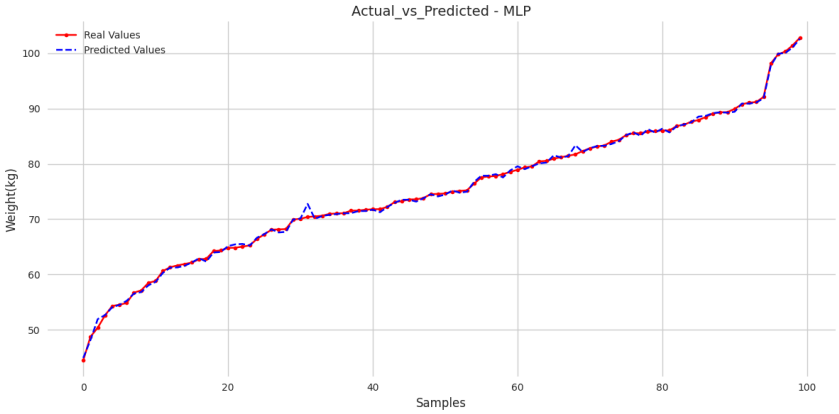

Figure 6: MLP Error

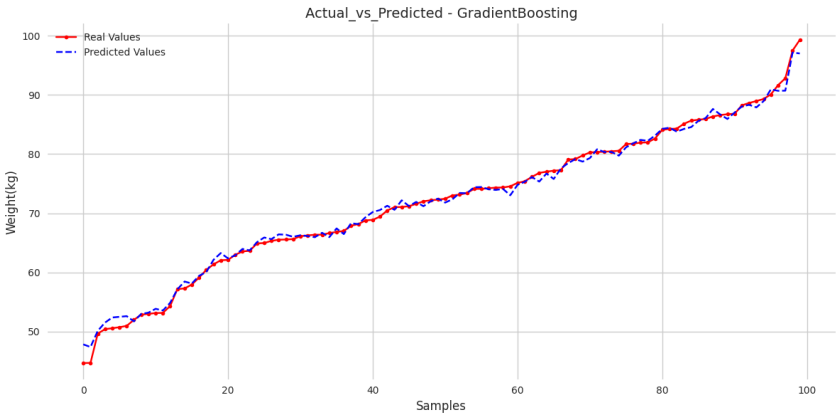

Figure 7: GR Error

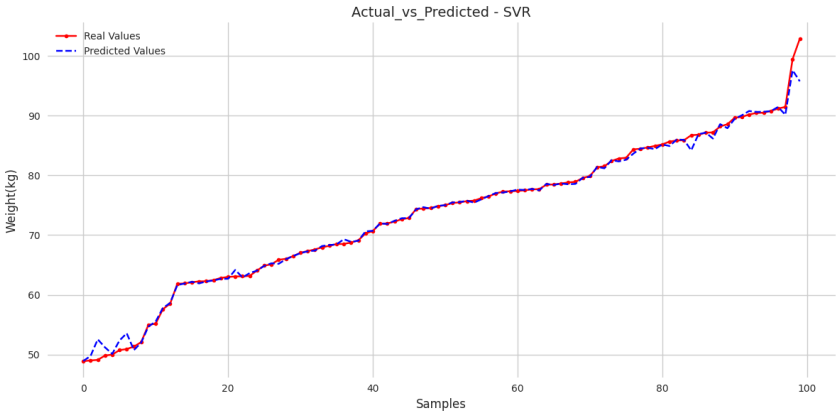

Figure 8: SVR Error

4.2 DL Performance

The results of simulations using different deep learning models are described in depth in the section that follows.

4.2.1 Deep Learning Correlation

The LSTM and Bi-LSTM models have high predictive performance for post-hemodialysis weight estimates, but with differing error profiles.
The LSTM model has a high correlation (0.9952) but a greater Mean Absolute Error (0.8962 kg), indicating steady but less exact predictions across the weight spectrum. Plot as it it is depicted in Figure 9 reveal a distinct pattern in which the LSTM overestimates weights in some areas (as shown in the blue data points above the reference line), with distributed prediction deviation .
In comparison, the Bi-LSTM model has a slightly higher correlation coefficient (0.9970) and a significantly lower MAE (0.6879 kg), indicating better prediction accuracy. The Bi-LSTM's data points are more consistent with the ideal prediction line, but there is still some systematic underestimate of weights, as shown in Figure 10, where the curve fall slightly below the reference line in the mid-range weights.
These temporal models perform well, but have higher error margins than the previously analyzed MLP model (MAE: 0.2743 kg), implying that, despite their ability to capture sequential patterns in dialysis data, simpler architectures may provide superior precision for this specific weight prediction task.

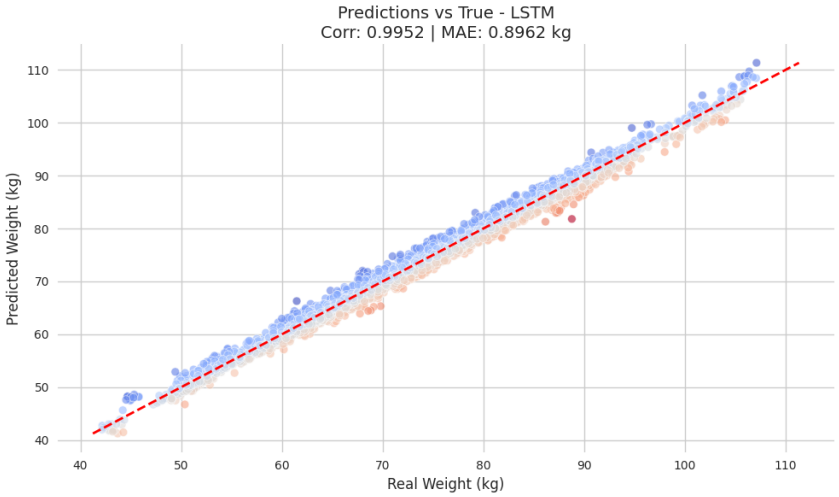

Figure 9: LSTM Correlation

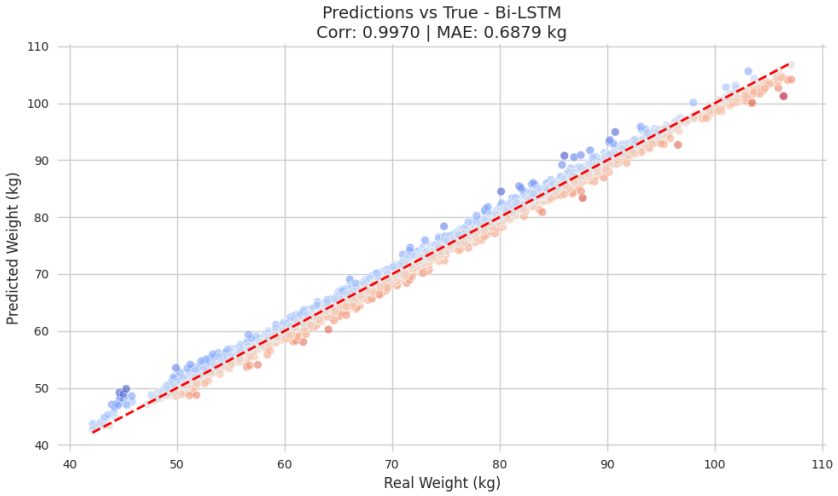

Figure 10: Bi LSTM Correlation

4.2.2 Deep Learning Error Curves

The time-series comparison of the LSTM and Bi-LSTM models reveals distinct prediction behaviors for post-hemodialysis weight estimation. The LSTM model (Figure 11) exhibits notable oscillatory patterns throughout the sample sequence, with frequent fluctuations around the actual weight values. Particularly visible deviations occur around samples 40-45, where the LSTM significantly overshoots the actual weight, and in the 60-85 range, where the predictions oscillate more substantially around the true values. Despite these fluctuations, the LSTM generally tracks the overall weight trend but consistently fails to smoothly follow gradual transitions.

The Bi-LSTM model (Figure 12) demonstrates markedly improved performance, with predictions adhering more closely to the actual weight trajectory throughout most of the sample range. The bidirectional architecture appears to reduce the prediction variance, though some notable deviations still occur, particularly a prominent spike around sample 80 and slight oscillations in the 85-95 range.

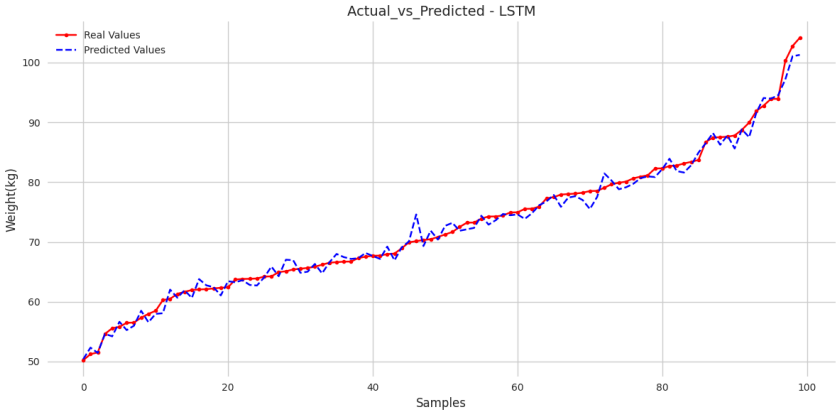

Figure 11: LSTM Error

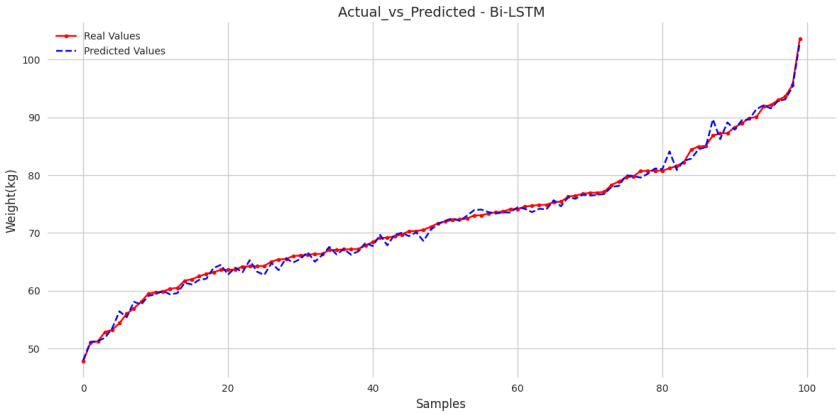

Figure 12: Bi LSTM Error

Both models adequately capture the final sharp weight increase at the end of the sequence, though neither achieves the precision demonstrated by the MLP model in previous visualizations. The Bi-LSTM's improved temporal stability compared to the standard LSTM suggests that incorporating both forward and backward temporal information enhances prediction accuracy, though the sequential models still underperform compared to the MLP's consistent tracking behavior across the entire patient weight range.

A Long Short-Term Memory (LSTM) model's learning curves are depicted in Figure 13, which show Mean Absolute Error (MAE) and Mean Squared Error (MSE) throughout a sequence of training epochs. Early in the training phase, the MAE graph on the left exhibits a sharp fall, suggesting that as the model gains knowledge from the training data, error will rapidly decrease. After around 20 epochs, the training (solid line) and validation (dashed line) curves settle, indicating that the model is generalizing effectively without experiencing substantial overfitting. This is supported by the validation error, which is comparatively flat.
On the other hand, the MSE chart on the right shows a similar pattern, starting with a steep decline and then stabilizing. Because the MSE measure is squared and penalizes greater mistakes more severely, the MSE values are consistently higher than the MAE values. Both indicators show that with the LSTM model learning is effective in reducing error.

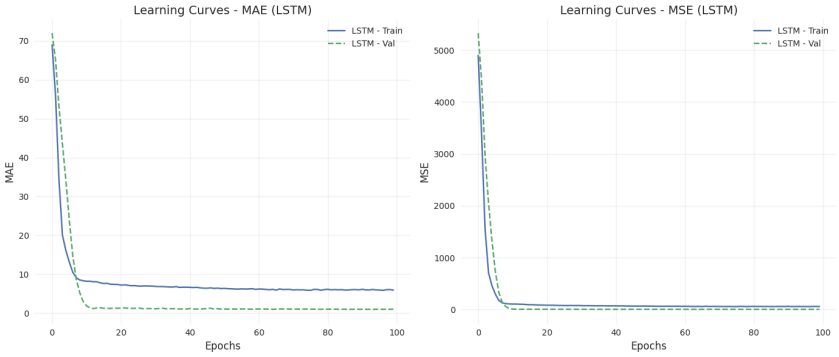

Figure 13: LSTM Train Curves

The Bi-LSTM model's learning curves (Figure 14) show the Mean Absolute Error (MAE) and Mean Squared Error (MSE) performance metrics over training and validation at different epochs. Effective learning and convergence are demonstrated by the significant decrease in error for both the training (solid blue line) and validation (dashed green line) datasets in the left graph illustrating MAE, especially in the early epochs. But at a given period, the validation error stabilizes at a larger value than the training error, indicating possible overfitting.
Similar to the MAE pattern, the MSE trends on the right graph show a considerable decrease in both training and validation errors during the early training stages. Although the model has successfully reduced error rates during the training process, the MSE values, which are more sensitive to big errors,

also raise concerns about overfitting as training goes on. These learning curves highlight the significance of tracking validation measures to guarantee reliable performance while offering insightful information about the model's training dynamics.

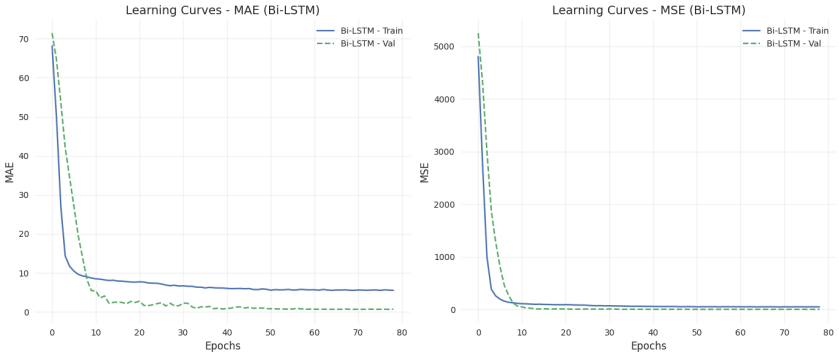

Figure 14: Bi LSTM Train Curves

4.3 Comparative Analysis

With R^2 scores near of 1, the machine learning models (MLP, SVR, and Gradient Boosting) show good overall performance and a good fit to the data. With the lowest MSE (0.146) and the greatest R (0.9989), the Multilayer Perceptron (MLP) is clearly the best-performing model. SVR (Support Vector Regression), which exhibits a fair balance between accuracy (MSE = 0.480) and robustness, comes in second. Although Gradient Boosting shows a higher MAE (0.574), indicating somewhat bigger mistakes in some predictions, it performs marginally worse (MSE = 0.607) but is still competitive. These models are very effective at the task, with MLP being the best option in terms of accuracy and stability (Table 8).

The Deep Learning models (Bi-LSTM and LSTM) presented in same Table 8 demonstrate decreased performance compared to the Machine Learning techniques, while they still produce acceptable results ($R^2 > 0.99$). The Bi-LSTM outperforms the normal LSTM, with an MSE of 0.818 compared to 1.357 for LSTM, demonstrating that bidirectional processing increases the model's capacity to capture temporal relationships. But compared to the Machine Learning models, their mean absolute errors (MAE) are significantly higher (0.688 and 0.896, respectively), which may indicate underfitting or the need for architectural adjustments. These findings indicate that, despite their complexity, Deep Learning techniques do not outperform conventional methods for this application.

With the lowest error rates across all measures and the best correlation coefficient, the MLP model performs noticeably better than any other design. Compared to Gradient Boosting, which exhibits much greater error values, the SVR model performs significantly better and comes second in efficacy with moderate error metrics. Although both sequential models, the Bi-LSTM architecture performs better than the regular LSTM, with reduced error metrics for all algorithms.

The MLP architecture offers significantly higher precision for clinical weight prediction following hemodialysis, with error rates less than half those of the sequential models and significantly lower than other conventional machine learning techniques.

The Root Mean Square Error (RMSE) values for ML and DL models employed in a comparison are displayed in Figure 15.

With an RMSE of around 0.4 kg, MLP performs the best, and SVR, at about 0.5 kg, follows. While Bi-LSTM and LSTM exhibit greater error rates with RMSE values approaching 0.8 kg and 1.0 kg, respectively, Gradient Boosting performs decently with an RMSE of around 0.6 kg. While LSTM and Bi-LSTM need more tweaking to increase their prediction accuracy, this investigation also shows how successful MLP and SVR are.

Error distribution, in kilos, for models, such as Gradient Boosting, Support Vector Regression (SVR), Multi-layer Perceptron (MLP), Long Short-Term Memory (LSTM), and Bidirectional LSTM, is shown in Figure 16 . Although there are notable outliers, the median error for the majority of models is close to zero, as seen by the y-axis, which displays the error values. While the MLP, LSTM, and Bi-LSTM models indicate possibly more stable performance with less extreme values, Gradient Boosting and SVR exhibit significant volatility in their error distributions.

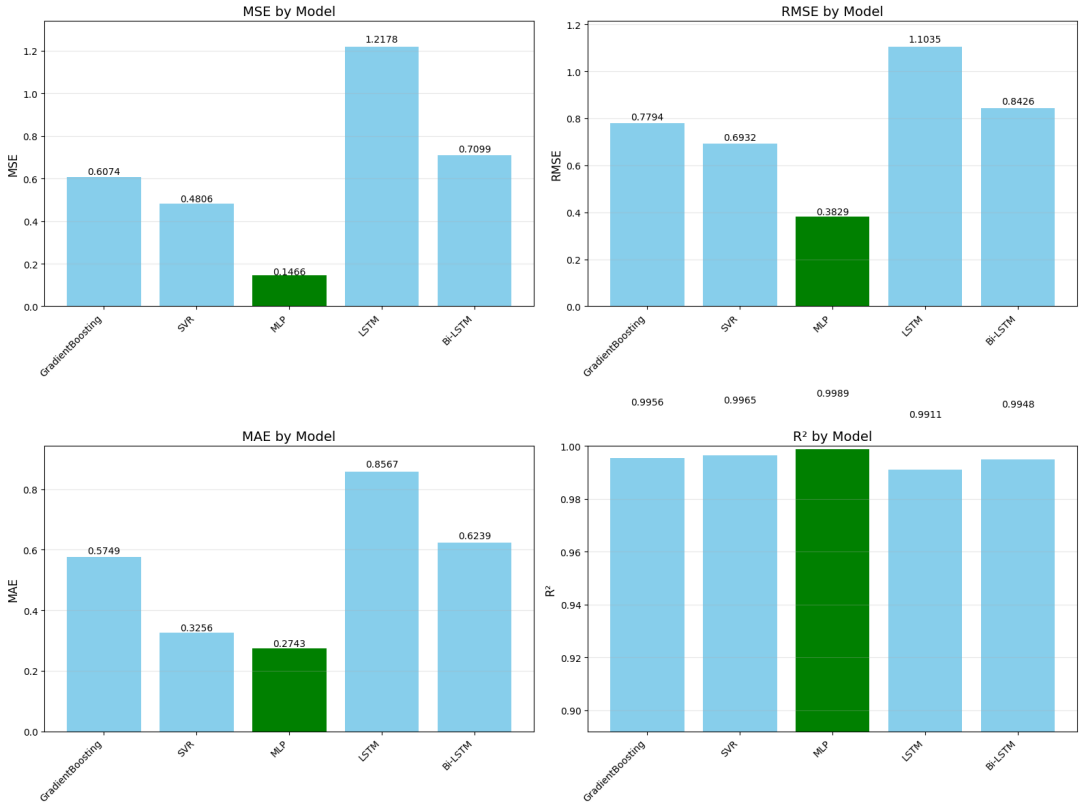

Figure 15: Model Comparison

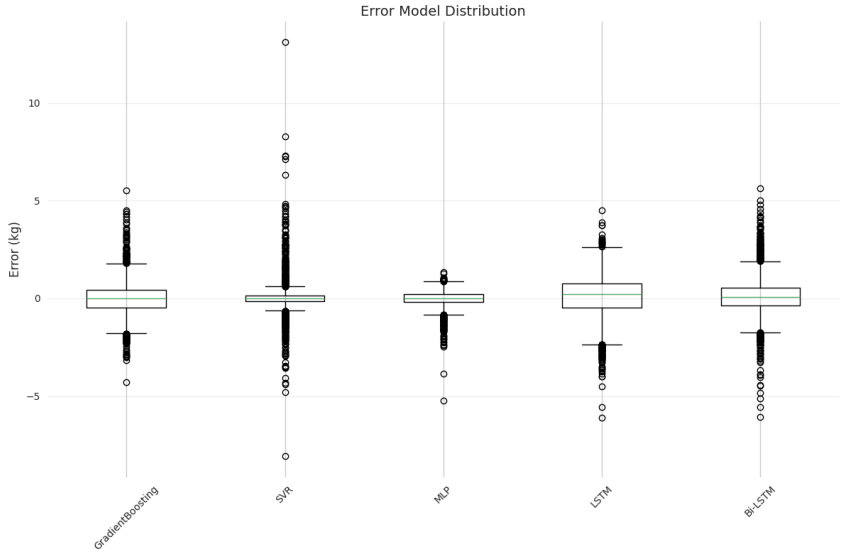

Figure 16: Error Distribution

Table 8 presents a comparative evaluation of different machine learning (ML) and deep learning (DL) models applied to dry weight (DW) prediction, including both the proposed methods and those from related studies. The performance is assessed using standard regression metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2).

In the study by Bi et al. (2019), traditional time-series models (ARIMA) and ensemble methods (Random Forest) achieved low MSE and MAE values, with ARIMA reaching an MSE of 0.1693 and a notably low MAE of 0.0178. However, their LSTM model reported a slightly higher MAE of 0.0314, indicating

limitations in capturing temporal dependencies compared to statistical and tree-based methods in their dataset.
Kim et al. (2021) focused exclusively on MAE as the evaluation metric, where LightGBM outperformed XGBoost and Random Forest, achieving an MAE of 0.2358. These values reflect solid predictive capability but are still dependent on careful feature engineering and parameter tuning.
In comparison, the proposed models demonstrate competitive performance, especially the MLP model, which achieved the best overall performance across multiple metrics: the lowest MSE (0.1466), a relatively low RMSE (0.3829), and an MAE of 0.2743, with a high R^2 of 0.9989, indicating strong generalization and accurate predictions. The SVR and Gradient Boosting models also performed well, though their MAE values were slightly higher.

Reference	Models	MSE	RMSE	MAE	R^2
Bi, et al. [6] (2019)	ARIMA	0.1693	-	0.0178	-
	RF	0.2044	-	0.0167	-
	LSTM	0.168	-	0.0314	-
Kim, et al. [11] (2021)	LightGBM	-	-	0.2358	-
	XGBoost	-	-	0.2845	-
	Random Forest	-	-	0.3082	-
Proposed	MLP	0.146597	0.382879	0.274269	0.998932
	SVR	0.480570	0.693232	0.325633	0.996499
	Gradient Boosting	0.607439	0.779384	0.574930	0.995575
	Bi-LSTM	0.817975	0.904420	0.687907	0.994041
	LSTM	1.357143	1.164965	0.896207	0.990113

Table 8: Comparative Table of the Results of ML and DL Techniques, with Reference to Related Works

While deep learning models such as Bi-LSTM and LSTM were less effective in this set up with higher MAEs (0.6879 and 0.8962, respectively) and MSEs exceeding 0.8—their inclusion demonstrates the potential challenges of training complex temporal models on smaller or noisy datasets without sufficient regularization or data preprocessing.
The proposed MLP model outperforms not only other traditional ML and DL models within the same experiment, but also most models from previous studies in terms of MSE and R^2. Although some of the methods referenced (e.g., ARIMA and RF of Bi et al.) achieved exceptionally low MAEs, these could reflect different experimental conditions, feature sets, or evaluation frameworks. Overall, this comparison highlights the effectiveness and robustness of the proposed approach in DW prediction tasks, and it underlines the importance of choosing appropriate models tailored to the characteristics of the dataset.

5 Conclusion

A comparative analysis involving several popular architectures is performed in order to assess the efficacy of different machine learning models for post-hemodialysis weight prediction, including multi-layer perceptron (MLP), support vector regression (SVR), gradient boost, long-short-term memory (LSTM) and bidirectional LSTM (Bi-LSTM). A data set of time series clinical parameters collected during hemodialysis sessions was used to train and evaluate the models.
This thorough analysis has shown that the MLP model performs noticeably better than other designs. When compared to other models, the MLP's significantly reduced error rates (MAE: 0.274 kg) imply that sophisticated sequential models may not always offer benefits for this particular clinical prediction job. Although the Bi-LSTM model outperformed the conventional LSTM architecture, neither sequential model was able to achieve the MLP approach's easier prediction accuracy. This conclusion highlights the significance of choosing models based on actual performance rather than theoretical benefits and calls into question the notion that temporal models are fundamentally superior for time-series health data.
The application of this predictive approach in clinical settings could offer nephrologists useful decision support and improve outcomes for hemodialysis patients. Future research should concentrate on prospective validation of the MLP model in a variety of patient populations and investigation of feature importance to further optimize prediction accuracy.

6 Declaration

Ethics Declaration: the study was conducted in accordance with institutional and national ethical standards. The data were collected and organized by the first author, a biomedical technician in the Hemodialysis Center of the hospital, as part of routine clinical practice. The analysis was based solely on retrospective clinical data obtained from routine hemodialysis sessions.
Consent to Participate, and Consent to Publish declarations: not applicable.
Funding: No Funding.
The Approval Committee: the study was reviewed and approved by the Head of the Hemodialysis Department and the Hospital Administration, in accordance with institutional policies and national regulations governing the use of anonymized clinical data for research.
Acknowledgment: a special thank for all the medical hemodialysis staff in the Middle west Tunisia Hospital (Doctors, nurses and technical service).

References

[1] Barbieri, C., et al. (2019). Development of an artificial intelligence model... *Kidney Diseases*, **5**(1), 28–33.

[2] Basile, C., et al. (2007). Development and validation of bioimpedance... *Clinical Journal of the American Society of Nephrology*, **2**(4), 675–680.

[3] Chamney, P. W., et al. (2002). A new technique for establishing dry weight... *Kidney International*, **61**(6), 2250–2258.

[4] Boonvisuth, N. (2022). Development of an artificial intelligence model.

[5] Guo, X., et al. (2021). An efficient multiple kernel support vector... *Current Bioinformatics*, **16**(2), 284–293.

[6] Bi, Z., et al. (2019). A practical electronic health record-based... *IEEE Journal of Translational Engineering in Health and Medicine*, **7**, 1–9.

[7] Mussnig, S., et al. (2024). Volume and Body Composition... *Kidney Medicine*, **6**(7), 100837.

[8] Inoue, H., et al. (2023). Predicting dry weight change... *BMC Nephrology*, **24**(1), 196.

[9] Guo, X., et al. (2021). Assessing dry weight... *BioMed Research International*, **2021**(1), 6627650.

[10] de Miranda Guimarães, A. C. A., et al. (2024). Equations for Prediction... *Journal of Renal Nutrition*, **34**(4), 343–349.

[11] Kim, H. R., et al. (2021). A novel approach... *PLOS ONE*, **16**(4), e0250467.

[12] Yang, Z., et al. (2022). Optimization of dry weight assessment... *IEEE Journal of Biomedical and Health Informatics*, **26**(10), 4880–4891.

[13] Germain, M. J., et al. (2021). Assessing accuracy of estimated dry weight... *Journal of Nephrology*, **34**(6), 2093–2097.

[14] Kim, J. Y., et al. (2024). The Prediction of Dry Weight... *Journal of Sport Psychology*, **33**(1).

[15] Jian, Y., et al. (2014). Comparison of bioimpedance... *Blood Purification*, **37**(3), 214–220.