Pure Appl. Geophys.
© 2026 The Author(s)
https://doi.org/10.1007/s00024-026-03941-0

Pure and Applied Geophysics

# Hybridizing Machine and Deep Learning for Urban Water Demand Forecasting: An Ensemble Framework Leveraging Dam Monitoring Data

Muhammed Ernur Akiner[1]

*Abstract*—Precise forecasting of urban water demand is a necessary condition for the proper management of resources in urban areas. The presented work proposes a novel ensemble framework that significantly improves forecast accuracy by integrating the daily occupancy rates of ten major dams in Istanbul. Unlike single-model approaches, the proposed technique is based on the combination of six machine learning algorithms (Random Forest, XGBoost, LightGBM, LSTM, SVR, and Ridge Regression), where the hyperparameters of each model are tuned using the Optuna library. The research is based on 4767 daily observations (from 2011 to 2024) processed with many temporal features such as seasonal indicators, moving averages, and lagged consumption variables. After pre-processing with StandardScaler and one-hot encoding, the data was chronologically split into three parts—training (52%), validation (18%), and testing (30%)—and this partitioning was done to prevent temporal data leakage. SVR performed best ($R^2$ = 0.8566, RMSE = 72,815 $m^3$/day), and LSTM performed second best ($R^2$ = 0.8345). The dynamically weighted ensemble model also had very good predictive ability ($R^2$ = 0.8469, RMSE = 75,244 $m^3$/day, MAE = 55,726 $m^3$/day), outperforming all baseline models except SVR. SHAP analysis showed that short-term consumption trends were the most significant forecast indicators, especially the 7-day moving averages and the one-day lagged consumption, which completely overshadowed the dam occupancy rates. The findings revealed the supremacy of ensemble learning methods in the water demand forecasts of urban areas and also indicated the water management authorities need to come up with data-driven conservation strategies.

**Keywords:** Water consumption estimation, ensemble learning, urban water management, time series features, SHAP explainability analysis, dam occupancy rates.

## 1. Introduction

Currently, the global issues of rapid urbanization, population rise, and climatic changes have made the sustainable management of water resources in urban centers a pressing challenge (Borah, 2025). The delicate situation where water is equally supplied and demanded has transformed the short-term consumption forecasting into a technical requirement and also a necessity in economic, environmental, and managerial aspects, particularly in high population cities (Nguyen et al., 2023). In such scenarios, the accurate modeling and prediction of water consumption parameters are to be considered as the main pillars of the success of several critical processes such as resource distribution, supply planning, and crisis management. This is especially true in an already populated megacity like Istanbul, with over 16 million people depending on a system of dams that are spread over a wide area, where daily accurate water consumption evaluations are needed in many aspects ranging from infrastructure planning to disaster response strategies (Bakker, 2012). Due to the intricate nature of urban water consumption, it must be treated with advanced analytical techniques rather than traditional statistical models (Kapanski et al., 2025; Zhong et al., 2021). Rapid urbanization, along with climate change and the need for sustainable resource management, has made the accuracy of water demand forecasts a strategic priority (Willard et al., 2025). The limitations of classic methods in representing the nonlinear dynamics of the water consumption data could, therefore, result in misallocation of resources and even water scarcity risks (Fu et al., 2022; Wang et al., 2024; Zeng et al., 2022).

The deployment of machine learning solutions in numerous domains ranging from the evaluation of water quality to the provision of drinking water has confirmed the strengthening of the partnership between the artificial intelligence and water research

[1] Department of Environmental Protection Technologies, Vocational School of Technical Sciences, Akdeniz University, Antalya, Turkey. E-mail: ernurakiner@gmail.com

Published online: 25 February 2026

areas (Feng et al., 2024; Masud et al., 2025; Mortazavizadeh et al., 2025). Modelling based on machine and deep learning techniques has been considered as a challenging and attractive area of research in water forecasting as it unlocks the limitations of the conventional approaches (Sadler et al., 2022; Taloma et al., 2025). These methods are able to capture the complex interplay among historical data and various environmental factors so accurate forecasts can be made (Boudhaouia & Wira, 2021). Actually, heterogeneous ensemble methods have been widely researched by Rai and De (2025) in the area of electricity load forecasting, resulting in significantly better forecasting, which is also a case for water consumption and forecasting. The hybrid modeling techniques not just demonstrate the efficacy of the integration of physical models and AI but also highlight the influence of AI and climate forecasting, as seen in the case made by Slater et al. (2023). In another instance, Zaidi et al. (2025) put forward a creative hybrid ensemble strategy known as HeartEnsembleNet for cardiovascular risk prediction, providing excellent directions for how this methodology may be tailored for different forecasting issues. Exclusively, the merging of explainable artificial intelligence (XAI) tools with these models not only provides the understanding of the factors that govern predictions, but also helps in the decision-making process, and as Uddin et al. (2024) have noted, it is a crucial factor in increasing model reliability in the context of the fairness assessment in distributed ML algorithms (Habibi & Hassanpour, 2025; Mathew et al., 2025). It is a certainty that Mathur et al. (2024) have conducted a study, proving the potential of machine learning techniques in meteorological forecasting, and it is also accepted that these technologies are applicable to water resources management. In summary, these studies reveal the potential and rights to use the ensemble and hybrid learning approaches that serve as the foundation of our current research.

Along with the methodological advances, the use of machine learning in various fields, especially water resource management, is becoming more and more evident due to its application in solving specific intricate challenges. One example of this is the development of new frameworks that make use of spatial–temporal data which are absolutely necessary for large-scale environmental monitoring, such as spotting pollution in large river basins (Tian et al., 2025). Also, the combination of process and data-driven approaches that apply interpretable deep learning with physical hydrology is already going to the top for accuracy and transparency in real-time forecasting (Zhu et al., 2025). Reinforcement learning has been actively examined for the purpose of adaptive management in the era of climate uncertainty, offering solutions for optimized reservoir operation and drought response as recent systematic reviews have shown (Kåge et al., 2025). Nevertheless, the road to realization is still long due to the fact that there are still challenges regarding the interoperability of models, the lack of data in areas where water is scarce, and the necessity for strong uncertainty quantification. Researchers are therefore likely to direct their efforts towards the development of physics-informed machine learning, which will be more generalized, and the systematic integration of explainable AI (XAI) tools to impart trust among the stakeholders. This changeover considers machine learning to be not only a predictive tool, but also the core of intelligent, adaptive, and sustainable water governance systems of the future.

The examination of the literature related to water consumption estimation indicates that the majority of the existing studies have either concentrated on meteorological variables (such as temperature, and precipitation) or proposed methods based on a single model architecture (Bougadis et al., 2005). Nevertheless, the demand for water is a complicated process, influenced by a range of factors such as socio-economic conditions, seasonal variations, and the hydrological supply capacity, which are interrelated in a multidimensional manner (Gato et al., 2007).

The main objective of the investigation is to create a system that would be able to provide an accurate daily water consumption forecast, and then to check whether this system would actually be useful in decision-support processes. The fact that the data regarding dam occupancy directly indicates the extent of water supply capacity makes the inclusion of these variables in models for predicting consumption not only theoretically but also practically indispensable. However, the literature review

discloses that there are still few studies that systematically explore time series models based on dams, while the current study predominantly employs single-model approaches. To overcome this limitation, this study offers three novel contributions:

- Data integration: Daily occupancy data (2011–2024) and water consumption data of 10 dams in Istanbul were merged into an exceptional data set with the help of time series engineering techniques (lagged variables, moving averages, seasonal characteristics).
- Multi-model estimation framework: Six different algorithms, from ensemble methods like Random Forest, XGBoost, and LightGBM to LSTM and SVR, were all implemented with Optuna-optimized hyperparameters. The models' outputs were weighted according to the $R^2$ scores, and a dynamic ensemble structure was built.
- Explanability analysis: The SHAP (SHAP Additive exPlanations) method was employed to quantitatively expose the extent to which hydrological variables such as dam occupancy rates influenced model decisions.

The findings not only add to the existing body of knowledge but also create a practical framework for directing the short-term water management choices of ISKI and similar organizations.

## 2. Materials and Methods

Water consumption was estimated with the aid of six different models based on machine learning and deep learning techniques: Random Forest (RF), eXtreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), Long Short-Term Memory (LSTM), Support Vector Regression (SVR), and Ridge Regression. Furthermore, SHAP and feature importance techniques were utilized to gain insights into the model performances and to identify the most significant variables in the estimation process. Likewise, previous research on urban water demand (Bakker, 2012; Fu et al., 2022; Nguyen et al., 2023) also emphasizes the increasing importance of machine learning in dealing with nonlinear dynamics.

### 2.1. Study Area

The study took place in Istanbul, the largest city in Turkey by population. Istanbul, with its more than 16 million inhabitants, is critical in water resources management due to its rapid urbanization and varying climate conditions. The city's water supply system comprises ten major dams, which are also well distributed geographically. The Büyükçekmece, Sazlıdere, Terkos, Alibey, Kazandere, Pabuçdere, and Istirancalar dams are situated on the European side, while the Ömerli, Darlık, and Elmalı dams are on the Asian side. These reservoirs, with a total capacity of about 868 million cubic meters, cover most of the water demand in Istanbul. The geographical layout of the ten primary dams in Istanbul is illustrated in Fig. 1.

Table 1 (Chamber of Environmental Engineers [CEE], 2014) provides a thorough overview of the geographical and operational characteristics of ten dams that are vital for Istanbul's water supply. The information in this table indicates the water management policies in Istanbul as far as the total capacities, sources of feeding, and geographical locations of the dams.

The research data are daily dam occupancy rates and water consumption data supplied by the Istanbul Water and Sewerage Administration (İSKİ). The dataset spans over the period of 2011–2024 and accommodates 4767 days of total observations.

One of the outcomes of the study is the visualization of the temporal interdependency between daily water consumption and dam occupancy rates in Istanbul from a perennial angle, as shown in Fig. 2. The plot unmistakably demonstrates the seasonal contrast between these two variables. The highest consumption figures along with a corresponding decrease in occupancy rates, particularly in the summer, suggest a negative correlation between the two variables. This particular trend is crucial for the detection of periods when the water resources are stressed due to warmer temperatures and reduced rainfall; it also represents the indirect influence of climate change on water consumption patterns and reservoir fill levels quite effectively.

Moreover, the graph reveals another remarkable aspect, which is the unexpected drop in consumption

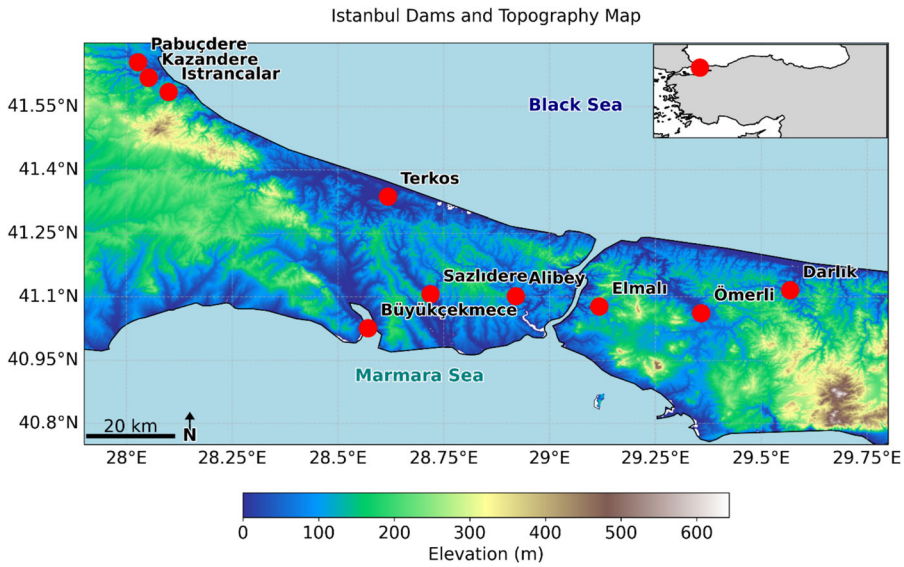



Figure 1
Map showing the geographical distribution of the 10 main dams in Istanbul

Table 1

*Locations and details of Istanbul dams (CEE, 2014)*

| Dam Name | Latitude (N) | Longitude (E) | Elevation (m) | Capacity (Million $m^3$) | Basin area ($km^2$) |
|---|---|---|---|---|---|
| Ömerli | 41° 0′ 0“ | 29° 13′ 1” | 121.9 | 235 | 620 |
| Darlık | 41° 15′ 0“ | 29° 1′ 59” | 30 | 107 | 550 |
| Elmalı | 41° 7′ 59“ | 29° 4′ 1” | 58 | 35 | 110 |
| Terkos | 41° 20′ 0“ | 28° 45′ 0” | 10 | 162 | 619 |
| Büyükçekmece | 41° 5′ 0“ | 28° 35′ 0” | 5 | 162 | 622 |
| Sazlıdere | 41° 10′ 0“ | 28° 40′ 0” | 50 | 88 | 290 |
| Alibey | 41° 8′ 0“ | 28° 55′ 0” | 60 | 39 | 97 |
| Kazandere | 41° 12′ 0“ | 28° 50′ 0” | 70 | 15 | 45 |
| Pabuçdere | 41° 5′ 0“ | 28° 45′ 0” | 45 | 12 | 30 |
| Istıancalar | 41° 25′ 0“ | 28° 30′ 0” | 80 | 18 | 55 |

at the onset of 2020. This decline is most likely attributable to the initial effects of the COVID-19 pandemic. In particular, restrictions and the closure of public consumption areas have significantly reduced short-term water use. This suggests that the models used must be sensitive not only to seasonal but also to external shocks.

### 2.2. Data Pre-processing

The analysis utilized daily water consumption data observed in Istanbul between 01.01.2011 and 18.02.2024 and time series data of daily occupancy rates of 10 different dams. The dataset contains more than 40 independent variables, including dam-based occupancy rates and time- and statistic-based characteristics.

First, categorical time variables such as day of the week (1–7), month, year, and season (winter: 1, spring: 2, summer: 3, autumn: 4) derived from historical information were established. Then, statistical indicators such as 7- and 14-day moving averages (rolling_mean_7, rolling_mean_14) and 7-day standard deviation (rolling_std_7) were

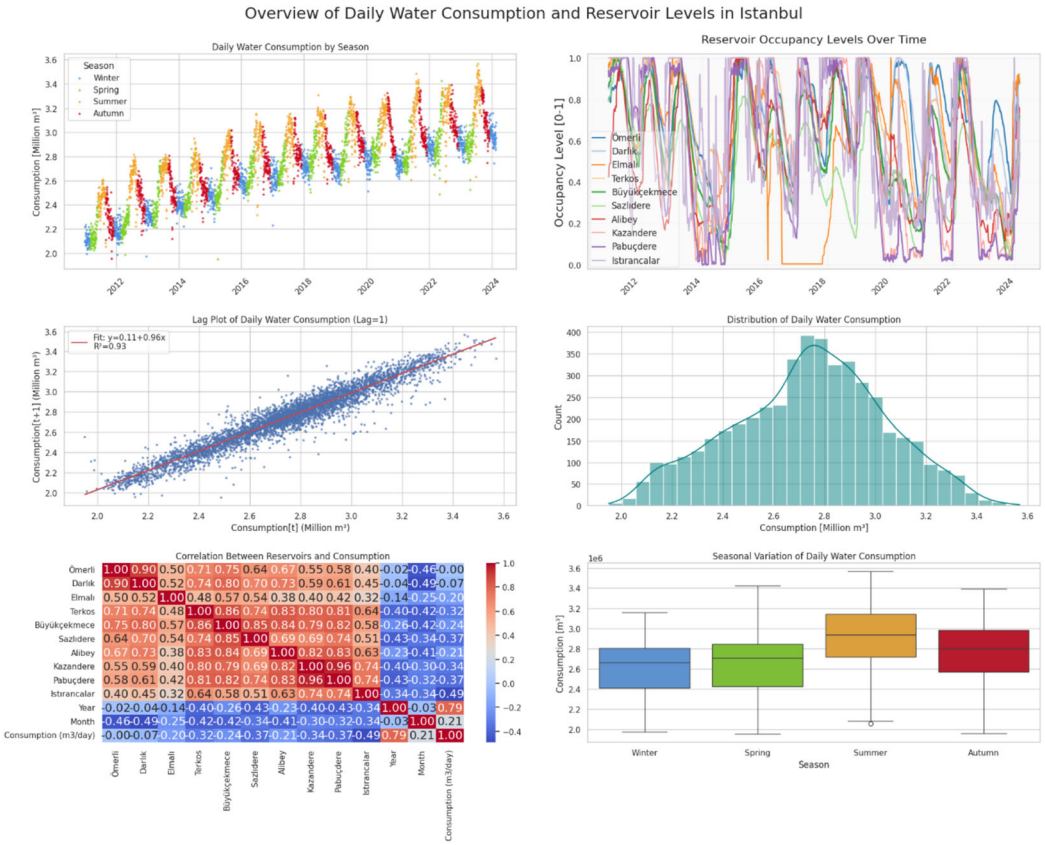



Figure 2
Overview of Istanbul's daily water consumption data and dam occupancy rates

produced to model short-term consumption trends. In addition, the monthly_mean variable and month-based average consumption behavior were included in the model.

Within the scope of time series engineering, 1, 7, 14, and 30-day lag variables (consumption_lag_1, consumption_lag_7, etc.) were created to support learning based on historical consumption values. These inputs have been developed specifically to increase the historical sensitivity of LSTM and tree-based algorithms. The average occupancy rates are also integrated into the model as a separate feature as an avg_Occupancies variable.

Numerical variables were scaled using StandardScaler or RobustScaler methods. RobustScaler was used for core-based models such as SVR, and MinMaxScaler was used for the LSTM model. Categorical variables (day, month, season) were adapted to the model in binary format by the One-Hot Encoding method (multicollinearity was prevented by selecting drop_first = True).

The dataset, which contains a total of 4767 days of observations, is divided into 52% training (01.01.2011–09.10.2017), 18% validation (10.10.2017–16.02.2020), and 30% testing (17.02.2020–18.02.2024), respectively, while maintaining chronological integrity. Normalization parameters (e.g., mean, standard deviation) were determined using only training data, and test and validation data were transformed with these parameters. Thus, data leakage is prevented by preventing the model from accessing the information in the test data beforehand. The methodological flow chart of the research is shown in Fig. 3.

The comprehensive pseudo-code structure systematically reveals the algorithmic basis of the

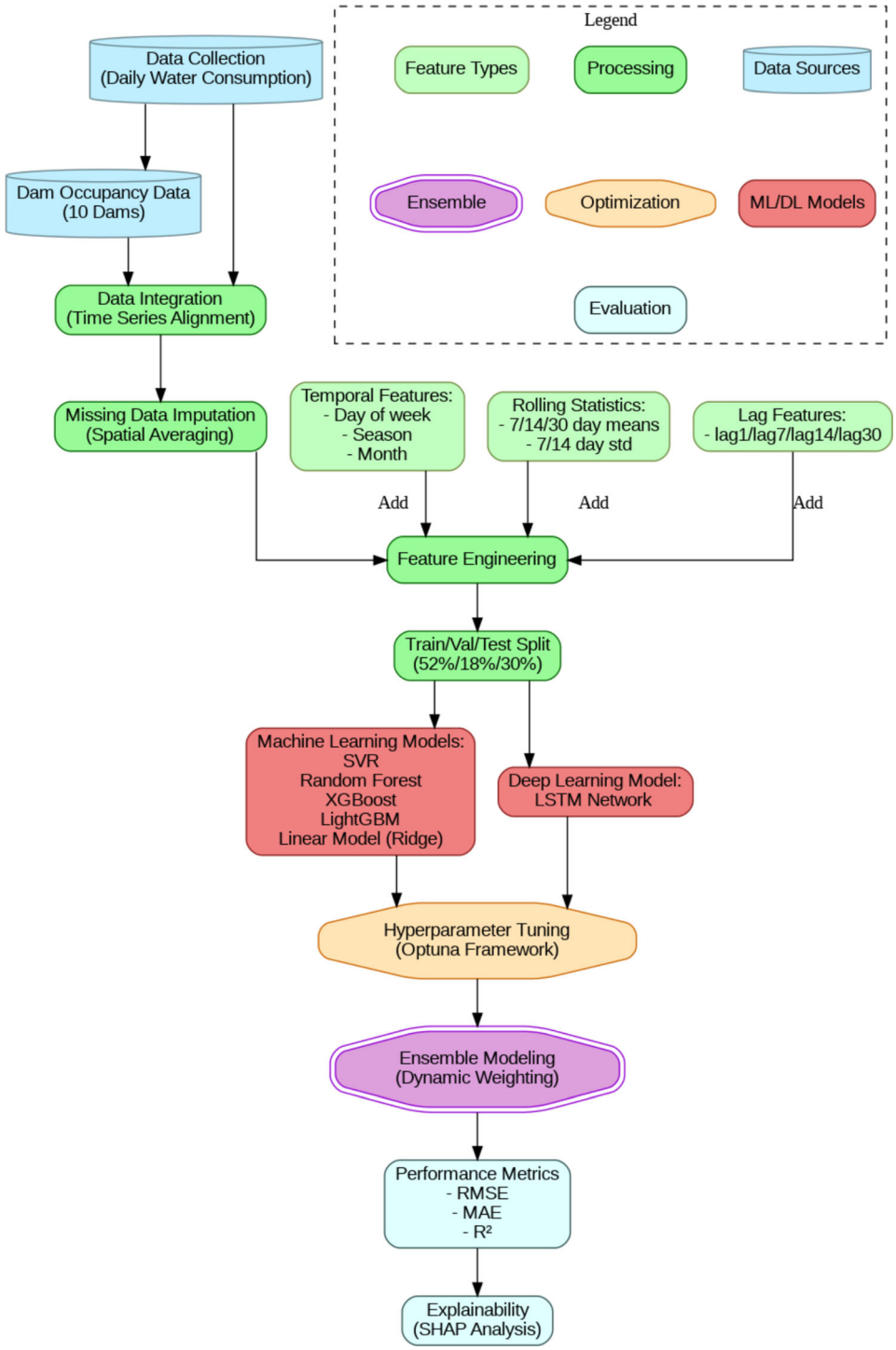



Figure 3
Methodological flow chart of the research

proposed hybrid modeling approach. The code structure models the entire process from the data preprocessing stage to the final evaluation in an integrated manner; it is based on fundamental principles such as automation, explainability, and generalizability (see Appendix A, Table A.1).

### 2.3. Random Forest (RF)

Random Forest is an ensemble learning method that combines multiple decision trees (Azad et al., 2025; Breiman, 2001; Chen et al., 2017). Each tree is trained on different subsamples, increasing the ability to generalize. The estimation process is carried out by averaging the outputs of each tree (Boulesteix et al., 2012). This approach prevents overfitting while increasing the model's ability to generalize. The mathematical formulation of RF is as in Eq. 1.

$$\widehat{y} = \frac{1}{T}\sum_{t=1}^{T} h_t(x) \tag{1}$$

where $h_t(x)$ is the $t$ $^{\text{th}}$ is the estimate of the decision tree, and $T$ refers to the total number of trees.

The hyperparameters of the Random Forest model were fine-tuned by means of 100 Optuna iterations, considering the number of trees (100–1000), maximum depth (5–50 or unlimited), and minimum samples for node splitting (2–20) as the three main factors. The best setup employed 580 trees, a maximum depth of 15, 5 minimum samples, and 'sqrt' for feature selection. The tree count range from 100 to 1000 comes from initial tests showing a plateau in performance after this interval, which is consistent with Random Forest literature (Boulesteix et al., 2012; Breiman, 2001), besides it was a computationally efficient choice inside Optuna's 100-trial setup. The confirmed optimum of 580 trees reinforces the selection of this range of parameter.

### 2.4. XGBoost

XGBoost is an optimized version of the gradient-boosted decision trees (GBDT) algorithm (Chen & Guestrin, 2016; Papoyan et al., 2025). Each new model attempts to reduce the error terms of the previous model. The goal is to minimize the loss function expressed in Eqs. 2 and 3 (Chen & Guestrin, 2016):

$$\mathcal{L} = \sum_{i=1}^{n} l\left(y_i, \widehat{y}_i^{(t)}\right) + \sum_{k=1}^{t} \Omega(f_k) \tag{2}$$

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \tag{3}$$

where $\mathcal{L}$ is the total loss function; $l$ is the loss function for a single instance; $l\left(y_i, \widehat{y}_i^{(t)}\right)$ represents the loss between the prediction $\widehat{y}_i$ and the true value $y_i$; $\Omega$ is the term regularization, which controls model complexity; $f_k$ refers to the $k$ $^{\text{th}}$ decision tree, and $T$ refers to the number of terminal nodes of the tree.

### 2.5. LightGBM

LightGBM is a histogram-based GBDT application that offers high accuracy and low calculation time for large datasets (Dai & Huang, 2025). The model uses a leaf-wise growth strategy. The main objective of LGBM is to build a model that minimizes the loss function shown in Eq. 3. Gradient boosting optimizes using the first and second derivatives of the loss function. Gradient $g_i$ and Hessian $h_i$ values are calculated as seen in Eq. 4.

$$g_i = \frac{\partial \ell(y_i, \widehat{y}_i)}{\partial \widehat{y}_i}, h_i = \frac{\partial^2 \ell(y_i, \widehat{y}_i)}{\partial \widehat{y}_i^2} \tag{4}$$

In LGBM's leaf-wise growth strategy, the profit of each leaf is calculated according to Eq. 5:

$$\text{Gain} = \frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} - \gamma \tag{5}$$

where $G_L$ and $G_R$ represent the gradient sums on the left and right leaves, $H_L$ and $H_R$ represent the Hessian sums. $\lambda$ is the regularization parameter, $\gamma$ represents the minimum gain threshold. In the training process of LGBM (see Eqs. 6–7), the initial prediction of the model is determined as $\widehat{y}_i^{(0)} = \overline{y}$ and the residuals ($r_i$) are calculated as:

$$r_i^{(0)} = y_i - \widehat{y}_i^{(0)} \tag{6}$$

Then, at each iteration, a decision tree is trained, and the model's prediction is updated as follows (Ke et al., 2017):

$$\widehat{y}_i^{(t)} = \widehat{y}_i^{(t-1)} + \eta \cdot f_t(x_i) \tag{7}$$

where $\eta$ is the learning rate. If the validation loss ($L_{\text{val}}$) does not improve over a certain threshold ($\Delta L$), the early stopping mechanism is activated, and the training process is terminated. The model's performance is assessed through performance metrics applied to the test data.

### 2.6. Long Short-Term Memory (LSTM)

LSTM is a special type of RNN that can capture long-term dependencies on time-series data (Mohammadi et al., 2024). The cell state $c_t$ and hidden state $h_t$ are updated with input $i_t$, forget $f_t$ and output $o_t$ gates (Hochreiter & Schmidhuber, 1997). A two-layer LSTM architecture is used in the model. The first layer consists of 128 units, the second layer consists of 64 units and a 20% dropout is applied after each layer. A densely connected structure and ReLU activation function are used as the output layer. Input sequences are created with 7-day time windows, and the outputs target a one-day forecast value. The model is trained with the mean squared error (MSE) loss function and Adam optimization algorithm. In order to prevent overfitting, the EarlyStopping strategy is applied. Hyperparameter optimization is performed using the Grid Search method. Since the hyperparameter space for the LSTM model is quite narrow (only two dropout_rate values were tested), Grid Search was deemed sufficient for systematic scanning. Bayesian optimization methods such as Optuna were preferred since larger search spaces are available in other models. Equation 8 shows how the internal states of the LSTM cell are updated over time and how learning is performed.

$$\begin{aligned} f_t &= \sigma\left(W_f x_t + U_f h_{t-1} + b_f\right) \\ i_t &= \sigma\left(W_i x_t + U_i h_{t-1} + b_i\right) \\ o_t &= \sigma\left(W_o x_t + U_o h_{t-1} + b_o\right) \\ \widetilde{c}_t &= \tanh\left(W_c x_t + U_c h_{t-1} + b_c\right) \\ c_t &= f_t \odot c_{t-1} + i_t \odot \widetilde{c}_t \\ h_t &= o_t \odot \tanh(c_t) \end{aligned} \tag{8}$$

where $\odot$ operation means element-wise multiplication (Hadamard product).

### 2.7. Support Vector Regression (SVR)

SVR performs linear regression by transforming the data into a high-dimensional space (Roozbeh et al., 2022). The goal is to find a smooth prediction hyperplane within the error tolerance (Cortes & Vapnik, 1995). The RBF kernel function was preferred for this analysis, with input data standardized and scaled accordingly. The objective function and constraints of SVR are shown in Eqs. 9–10 (Drucker et al., 1996):

$$\min_{w,b,\xi} \frac{1}{2} \| w\|^2 + C \sum_{i=1}^{n} \left(\xi_i + \xi_i^*\right) \tag{9}$$

$$\text{subject to: } \begin{cases} y_i - w^T \phi(x_i) - b \le \epsilon + \xi_i \\ w^T \phi(x_i) + b - y_i \le \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0 \end{cases} \tag{10}$$

where $\varepsilon$ is the error tolerance, $C$ is the regularization parameter; $\xi_i, \xi_i^*$ are the deviation variables.

### 2.8. Linear (Ridge) Regression

Ridge regression is a linear regression method that applies L2 regularization to reduce the effect of multicollinearity (Magklaras et al., 2024). The ridge regression estimator (Hoerl & Kennard, 1970) is an optimization problem that combines the residual sum of squares (RSS) and L2 norm regularization. The coefficient vector $\widehat{\beta}$ is defined as the value of $\beta$ that minimizes the objective function, as seen in Eq. 11.

$$\hat{\beta} = \arg\min_{\beta} \left\{ \underbrace{\underbrace{\sum_{i=1}^{n} \left(y_i - x_i^{\mathrm{T}} \beta\right)^2}_{\text{Residual Sum of Squares}}}_{\text{Residual Sum of Squares (RSS)}} + \underbrace{\lambda \|\beta\|_2^2}_{\text{Regularization}} \right\} \tag{11}$$

where $\lambda$ is the penalty coefficient.

### 2.9. Ensemble Weighting

The weighted ensemble method was applied to increase the overall prediction performance by

combining the outputs of multiple prediction models. In this method, the output of each model is weighted according to the performance level of the model, and the final prediction is calculated (Zhou et al., 2021). $R^2$ (coefficient of determination) was used as the basic metric representing the model performance. The weight value for each model was calculated with Eq. 12.

$$w_i = \frac{\left(R_i^2 - \min\left(R^2\right) + \epsilon\right)}{\sum_j \left(R_j^2 - \min\left(R^2\right) + \epsilon\right)} \tag{12}$$

where $w_i$ is the weight of the $i$ th model on the final prediction, $R_i^2$ is the $R^2$ score of the $i$ th model, $\min(R^2)$ is the lowest $R^2$ score obtained among all models, $\varepsilon$ is a small positive constant used to prevent zero-split errors ($\varepsilon = 10^{-6}$ in the current implementation), $j$ is the total number of models.

Equation 12 ensures that models with higher $R^2$ scores contribute more weight while relatively reducing the effect of lower-performing models. Significantly, the sum of all weights is normalized to be 1. This method limits the effect of weak models on the prediction while increasing the effect of strong models, allowing for a more balanced and accurate ensemble prediction (Sang et al., 2024; Yin et al., 2021; Zhou et al., 2021).

This structure enables a fair comparison between classical models (e.g. Ridge, SVR) and complex learning-based models (e.g. LSTM, XGBoost) and ensures that the final model output is more consistent.

### 2.10. Feature Importance Analysis

Feature importance analysis was applied to determine which input variables are more effective on model outputs in the prediction models. In this context, the Gini importance score (Mean Decrease in Impurity, MDI) was calculated especially for Random Forest and XGBoost models. The importance level of each feature was evaluated based on the decrease in impurity in the decision trees of the model and expressed with Eq. 13 (Breiman, 2001):

$$\text{Importance}_i = \frac{1}{N}\sum_{t \in T} \Delta\text{Impurity}(i,t) \tag{13}$$

where Importance represents the total importance level of the $i$ th feature, $T$ represents all decision trees in the model, $\Delta\text{Impurity}(i,t)$ represents the purity reduction provided by the contribution of the $i$ th feature to the splits in the $t$ th tree, and $N$ represents the total number of trees.

### 2.11. SHAP (SHapley Additive exPlanations) Method

SHAP (SHapley Additive exPlanations) was proposed by Lundberg and Lee (2017) and has a mathematical basis based on Shapley values in game theory (see Eq. 14). Gini importance offers a model-specific assessment, while SHAP measures the trait's contribution for each observation, regardless of the model, and provides global + local explainability.

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} \left(f_{S\cup\{i\}}\left(x_{S\cup\{i\}}\right) - f_S(x_S)\right) \tag{14}$$

where $\phi_i$ represents the SHAP value of the $i$ th feature, $F$ is the set of all features, and $S$ is a subset of the $F$ set. $f_S$ is the prediction value made using the features in the $S$ set.

The positive SHAP value ($\phi_i > 0$) indicates that the feature has an increasing effect on the model's prediction; conversely, if the value is negative ($\phi_i < 0$), the feature has a decreasing impact on the model prediction. The absolute magnitude of the SHAP value ($|\phi_i|$) indicates how strong the effect of the feature on the prediction is. In calculating SHAP values, computational efficiency was increased using a background data set of 1000 random samples. Global feature importances were obtained by averaging the absolute SHAP values over the entire test set (see Eq. 15).

$$\overline{\phi}_i = \frac{1}{n}\sum_{j=1}^{n} \left|\phi_{i,j}\right| \tag{15}$$

where $\overline{\phi}_i$ represents the average SHAP value of the $i$ th feature, $\phi_{i,j}$ represents the SHAP value of the $i$ th feature on the $j$ th sample.

Table 2

*Hyperparameters and training configurations for all machine learning models*

| Model | Section | Parameter | Value/ range | Optimal value | Optimization method | Description |
|---|---|---|---|---|---|---|
| Random Forest | Optimization | n_estimators | 100–1000 (Optuna) | 580, 15, 'sqrt' | Optuna (100 trials) | Number of decision trees |
| | | max_depth | None, 5–50 | | | Maximum tree depth |
| | | min_samples_split | 2–20 | | | Minimum samples to split a node |
| XGBoost | Optimization | learning_rate | 0.01–0.2 (log scale) | 0.05, 0.12, 0.3 | Optuna (TPE sampler) | Boosting step size |
| | | Gamma | 0–0.3 | | | Minimum loss reduction for split |
| | | reg_alpha | 0–0.5 | | | L1 regularization term |
| LightGBM | Optimization | num_leaves | 20–300 | 150, 'gbdt', 0.8 | Optuna (median pruner) | Maximum leaves per tree |
| | | boosting_type | GBDT, Darts, Goss | | | Gradient boosting method |
| | | feature_fraction | 0.4–1.0 | | | Randomly selected feature ratio |
| LSTM | Architecture | n_units | 128 (L1), 64 (L2) | 128/64, 0.2, 32 | Grid Search | LSTM layer sizes |
| | | dropout_rate | 0.1–0.2 | | | Dropout for regularization |
| | Training | batch_size | 32 | | | Samples per batch |
| SVR | Kernel | C | 0.1–1000 (log scale) | 100, 0.01, 'rbf' | Optuna (log-scale) | Penalty parameter |
| | | Epsilon | 0.001–10.0 (log scale) | | | Margin of tolerance |
| Linear Model | Training | alpha (Ridge) | 1.0 (fixed) | 1.0 | Fixed (L2 penalty) | L2 regularization strength |
| Ensemble | Weighting | R2-based | Dynamic | SVR: 27.7%, LSTM: 26.4%, Linear: 25.1%, RF: 8.4%, XGBoost: 6.7%, LightGBM: 5.7% | $R^2$-weighted linear combination | Optimized from validation $R^2$ scores |

### 2.12. *Performance Evaluation Metrics*

Three basic metrics were used to evaluate the predictive performance of the models quantitatively: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination ($R^2$). In calculating all metrics, the model generalization ability was objectively measured using an independent test set. Root Mean Square Error (RMSE): Calculated as the root mean square of the differences between predicted and true values (see Eq. 16). It is more sensitive to large errors.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2} \tag{16}$$

Mean Absolute Error (MAE): The average absolute differences between the true and predicted values (see Eq. 17). It weights all errors equally.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | y_i - \hat{y}_i | \quad (17)$$

Coefficient of Determination ($R^2$): Indicates the extent to which the model explains the variance in the dependent variable (see Eq. 18). Values closer to 1 indicate a better fit:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \quad (18)$$

where n is the number of samples in the test set, $y_i$ is the true target value (observed water consumption), $\hat{y}_i$ is the value predicted by the model, and $\bar{y}$ is the average of the true target values.

### 2.13. Hyperparameter Configurations

Table 2 presents the hyperparameter configurations of the machine learning and deep learning models used in the study in detail and contains important clues about the strategies adopted in the training process of each model. In this context, it is seen that the parameter selections are carefully optimized to both increase model performance and ensure generalizability. For example, the preference of n_estimators = 580 and max_depth = 15 values in the Random Forest model allows the model to learn complex data relationships by utilizing a large number of decision trees. At the same time, it shows that a balance is established by limiting the maximum depth against the risk of overfitting. This configuration plays a major role in the efficient capture of nonlinear interdependencies among variables.

In the case of the XGBoost model, a rather low learning rate, specifically learning_rate = 0.05, made the convergence process more delicate and thus the model was allowed to learn gradually throughout the training period. However, with the gamma = 0.12 parameter, a minimum loss reduction condition was imposed before a certain division process was performed, thus aiming to avoid unnecessary branching of the model and provide higher generalizability. These settings reveal that the model is optimized for accuracy and robustness.

In the LSTM model employed for time series forecasting, indications as to dropout = 0.2 and batch_size = 32 are essential among the others to avoid overfitting and to increase the variety of data in the training process. More specifically, the dropout layer enhances the robustness of the model by randomly turning off a portion of the network neurons during training. On the other hand, the total training time of the model is 1120 s (related to Table 3), which indicates the high computational

Table 3

*Comprehensive performance comparison of all models across training, validation, and test sets*

| Model | Dataset | $R^2$ | Rank ($R^2$) | RMSE ($m^3$/day) | Rank (RMSE) | MAE ($m^3$/day) | Rank (MAE) | Training Time(s) |
|---|---|---|---|---|---|---|---|---|
| Random Forest | Train | 0.9513 | 4 | 53,151 | 5 | 36,835 | 5 | 218 |
| | Validation | 0.7800 | 6 | 78,050 | 6 | 60,097 | 6 | |
| | Test | 0.5193 | 6 | 133,323 | 6 | 96,029 | 6 | |
| XGBoost | Train | 0.9336 | 5 | 62,046 | 6 | 44,485 | 6 | 154 |
| | Validation | 0.7826 | 5 | 77,588 | 5 | 59,385 | 5 | |
| | Test | 0.4894 | 7 | 137,406 | 7 | 97,810 | 7 | |
| LightGBM | Train | 0.9527 | 1 | 52,386 | 1 | 36,958 | 1 | 92 |
| | Validation | 0.7856 | 4 | 77,047 | 4 | 58,961 | 4 | |
| | Test | 0.4729 | 5 | 139,612 | 5 | 99,702 | 5 | |
| LSTM | Train | 0.9321 | 6 | 62,762 | 7 | 44,014 | 3 | 1120 |
| | Validation | 0.8304 | 3 | 68,520 | 3 | 51,868 | 3 | |
| | Test | 0.8345 | 2 | 78,237 | 3 | 59,059 | 3 | |
| SVR | Train | 0.9339 | 2 | 61,891 | 2 | 41,327 | 1 | 890 |
| | Validation | 0.8416 | 1 | 66,231 | 1 | 47,668 | 1 | |
| | Test | 0.8566 | 1 | 72,815 | 1 | 51,762 | 1 | |
| Linear Model | Train | 0.9241 | 7 | 66,340 | 3 | 47,005 | 2 | 5 |
| | Validation | 0.8240 | 2 | 69,805 | 2 | 51,077 | 2 | |
| | Test | 0.8120 | 3 | 83,387 | 4 | 62,643 | 4 | |
| Ensemble | Test | 0.8469 | 2 | 75,244 | 2 | 55,725 | 2 | – |

expense of the LSTM architecture. The scenario thereby exposes the hardware and time expenses that come along with the superior performance of deep learning models.

## 3. Results

When the performance metrics in Table 3 are examined, the value of $R^2 = 0.835$ obtained by the LSTM model confirms that the selected hyperparameter configurations are effective. Nevertheless, the outcome of the SVR model getting a better $R^2$ score (0.857) with a shorter training time (890 s) indicates that more complicated architectures may not necessarily give the best performance (Parveen et al., 2017; Rahbar et al., 2022; Wang, 2025). This shows that, when selecting a model, the complexity of the algorithm should not just be assessed but also its fit to the problem context and the computational efficiency involved.

The SVR (Support Vector Regression) model achieved the highest performance. The values of $R^2 = 0.857$ and RMSE = 72,815 $m^3$/day of the model show that it can produce estimates with high accuracy, especially on daily water consumption data dominated by nonlinear patterns. Although the absolute RMSE of 72,815 $m^3$/day looks quite significant at first glance, it is in fact only about 2.4% of 3 million $m^3$ which happens to be Istanbul's average daily water consumption, thus, allowing us to conclude that the metropolitan-scale water demand forecasting was very accurate. This result reveals that the kernel-based SVR algorithm is an effective alternative in complex time series with limited samples and confirms the model's generalizability.

Despite the advantage of deep learning in sequence data, the LSTM (Long Short-Term Memory) model lagged behind SVR with RMSE = 78,237 $m^3$/day and $R^2 = 0.835$. This situation reveals that deep learning models may perform lower due to their extreme parametric structure, especially in datasets with small sampling, such as n = 4767. That is, the amount of data is critical for such models. However, the capacity of LSTM to capture time-dependent relationships should not be underestimated (Li & Wang, 2021; Tian et al., 2023).

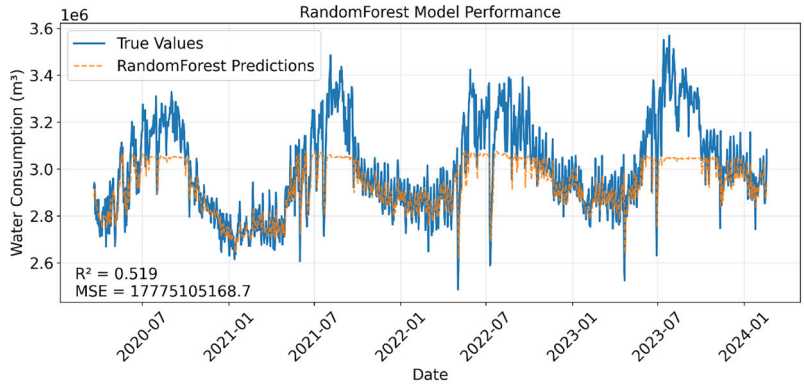
Another approach that has attracted particular attention has been the Ensemble Model. This model was created with dynamic weights according to the accuracy levels obtained from SVR and LSTM prediction outputs; SVR estimates were weighted by 27.7% and LSTM by 26.4%. In terms of aggregate performance, the Ensemble model is balanced between the two models with $R^2 = 0.847$, thus combining the strengths of the individual models to provide a more stable predictive capacity.

The high $R^2$ values in Table 3 prove that the models can successfully capture regular seasonal patterns and abrupt changes. In particular, the high coefficient of fit of the Ensemble model shows that it can respond effectively to this multidimensional variability.

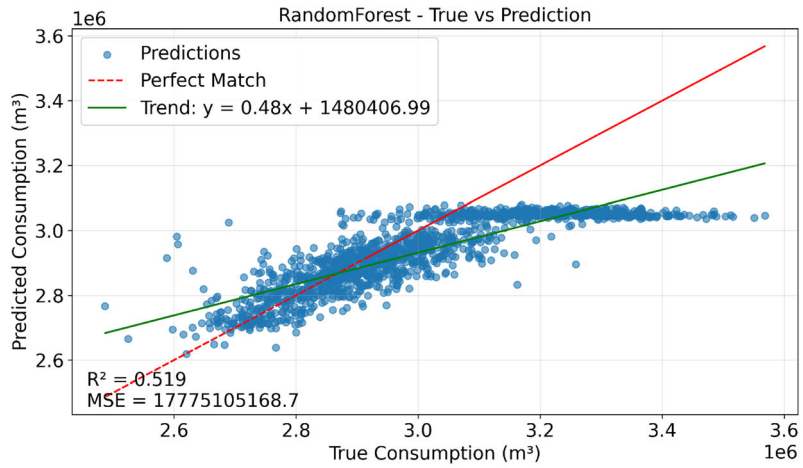
As noticed in Figs. 4a and b, the Random Forest model was insufficient to capture sudden water consumption fluctuations, especially in the post-pandemic (2020–2021) period. This suggests that the model struggles to adapt to short-term and high-frequency changes, which directly coincides with the high RMSE (133,323 $m^3$/day) reported in Table 3. The residual distribution in Fig. 4c shows that the errors contain significant negative biases, revealing that the model systematically keeps the estimates low.

Figure 5a shows that the XGBoost model underestimates peak consumption values. In particular, during the high water consumption periods in summer, as seen in Fig. 5b, the model produced significantly underestimates of the true values. This is due to the tendency of tree-based models to generalize trends and reveals the underfitting effect (Geng et al., 2024). The residual distribution in Fig. 5c confirms that the errors are not symmetric around zero and that the model misses some seasonal dynamics.

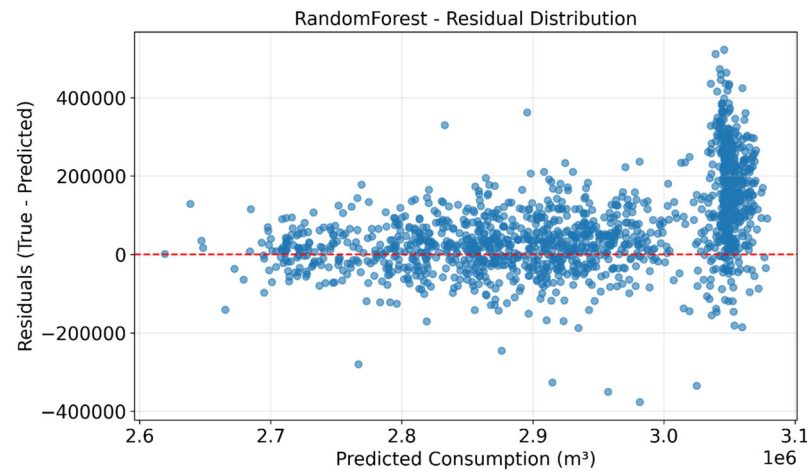
Similarly, Fig. 6a reveals that the LightGBM model also underestimates peak consumption. The estimates in Fig. 6b exhibit a significant negative bias in summer, consistent with underfitting, a common limitation of tree-based algorithms (Sapkota et al., 2024). The asymmetric residual clusters in Fig. 6c support the model's failure to capture some of the seasonal variation.




*(a)*



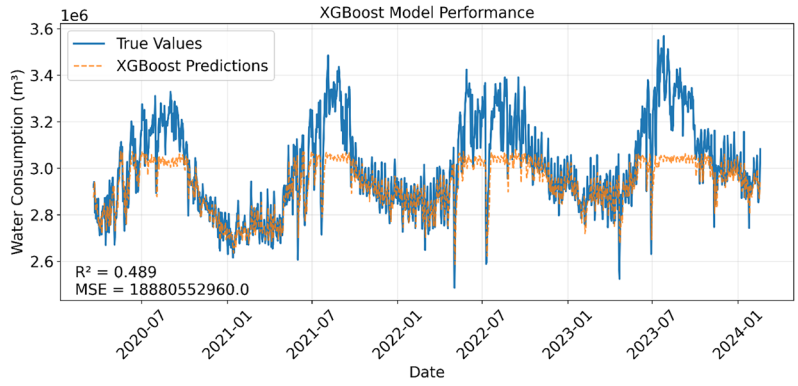

*(b)*



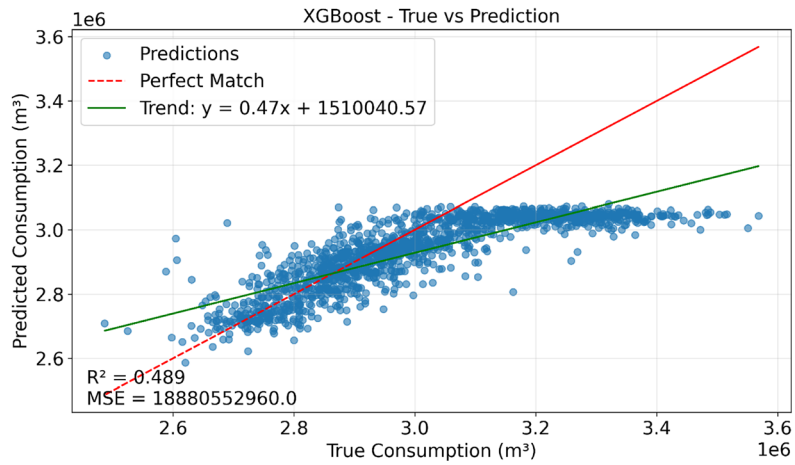

*(c)*

Figure 4

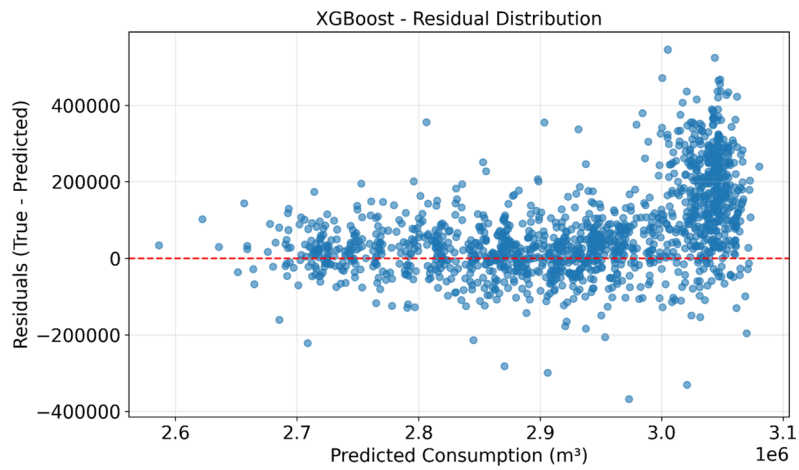
Random Forest model performance analysis showing **a** time series comparison of actual vs predicted values, **b** scatter plot of observed vs predicted consumption, and **c** residual distribution (test period: 2020–2024)




*(a)*

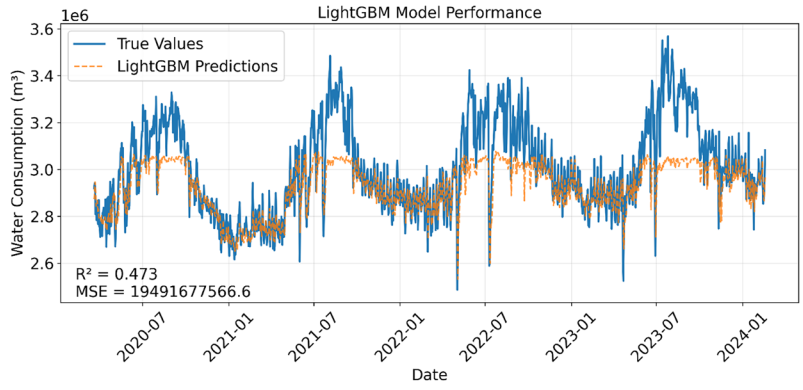



*(b)*

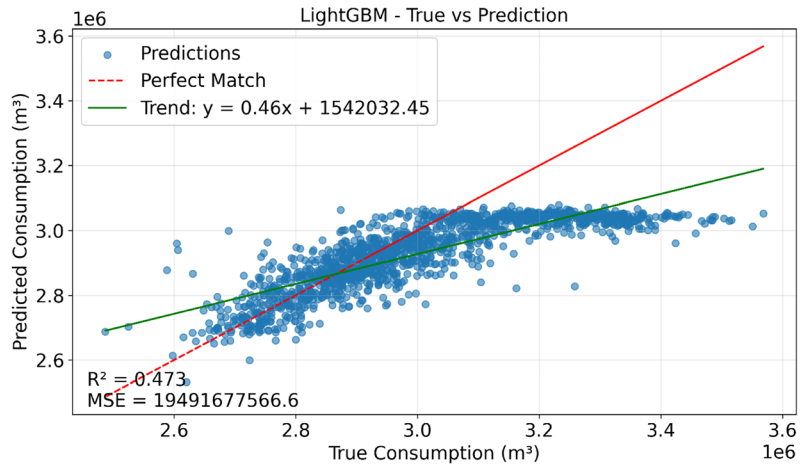



*(c)*

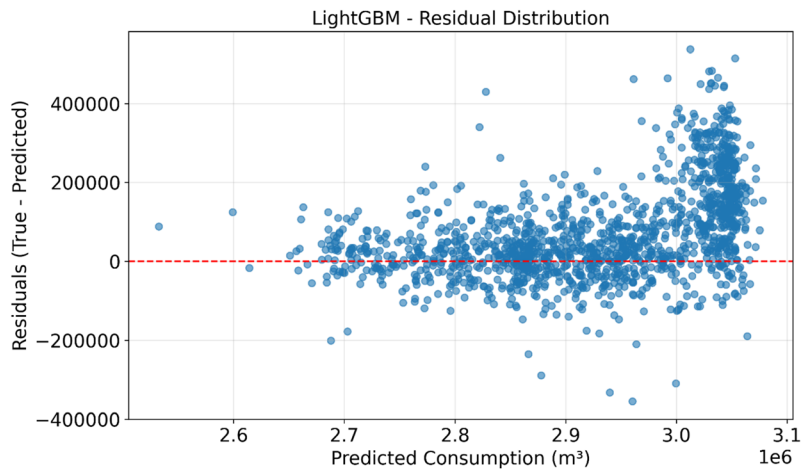
Figure 5
XGBoost model performance analysis showing **a** time series comparison of actual vs predicted values, **b** scatter plot of observed vs predicted consumption, and **c** residual distribution (test period: 2020–2024)




*(a)*

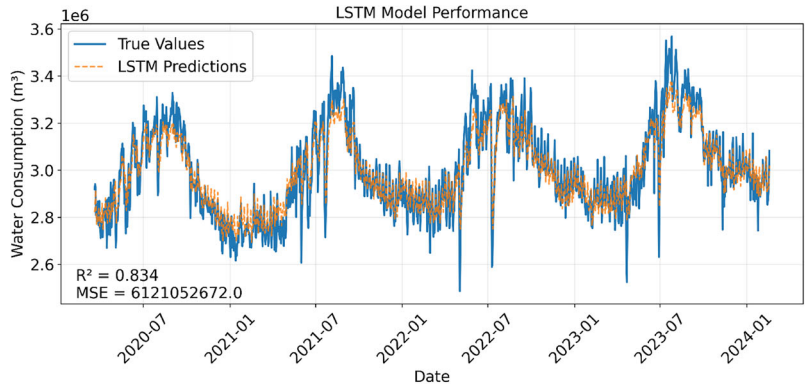



*(b)*

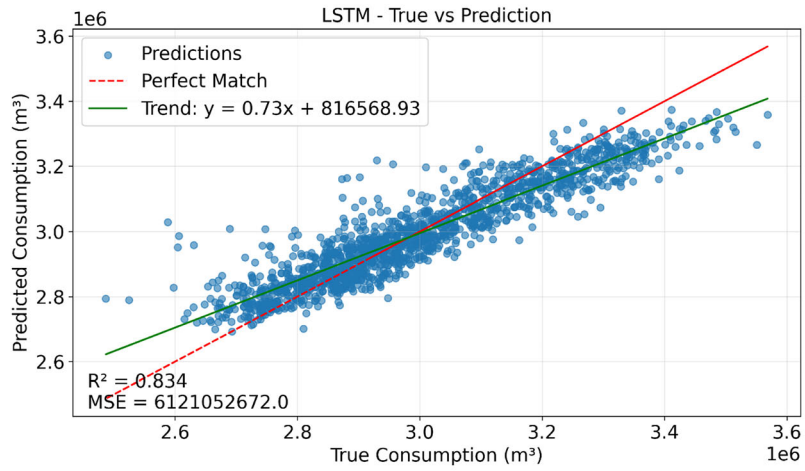



*(c)*

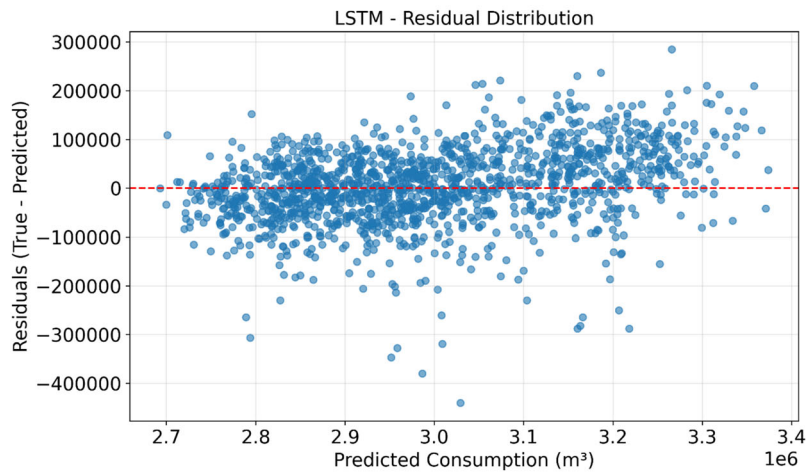
Figure 6
LightGBM model performance analysis showing **a** time series comparison of actual vs predicted values, **b** scatter plot of observed vs predicted consumption, and **c** residual distribution (test period: 2020–2024)




*(a)*

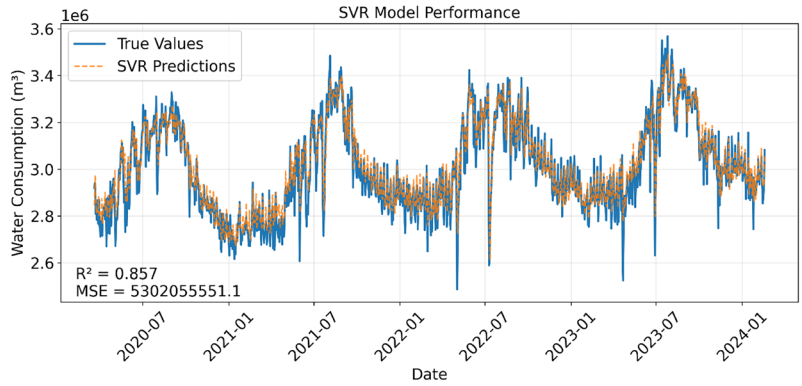



*(b)*

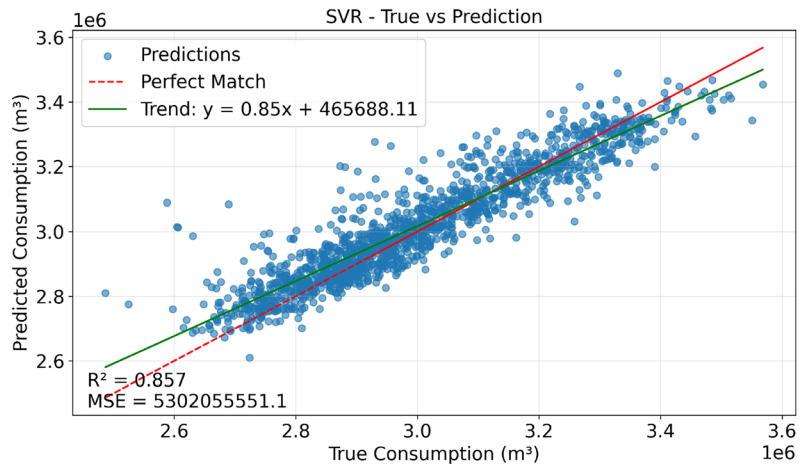



*(c)*

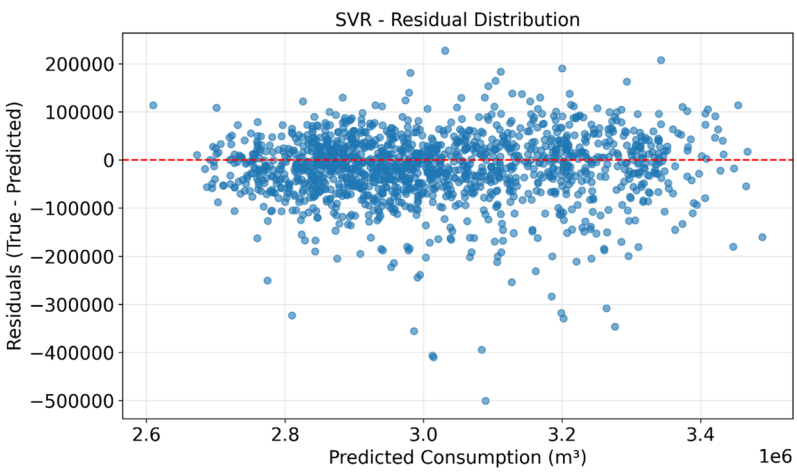
Figure 7
LSTM model performance analysis showing **a** time series comparison of actual vs predicted values, **b** scatter plot of observed vs predicted consumption, and **c** residual distribution (test period: 2020–2024)




*(a)*

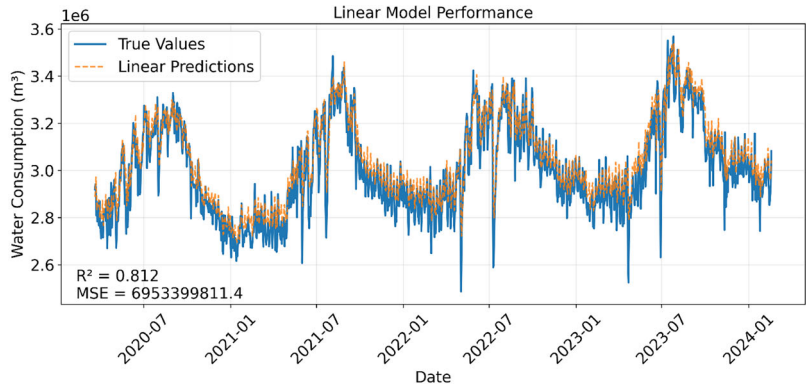



*(b)*

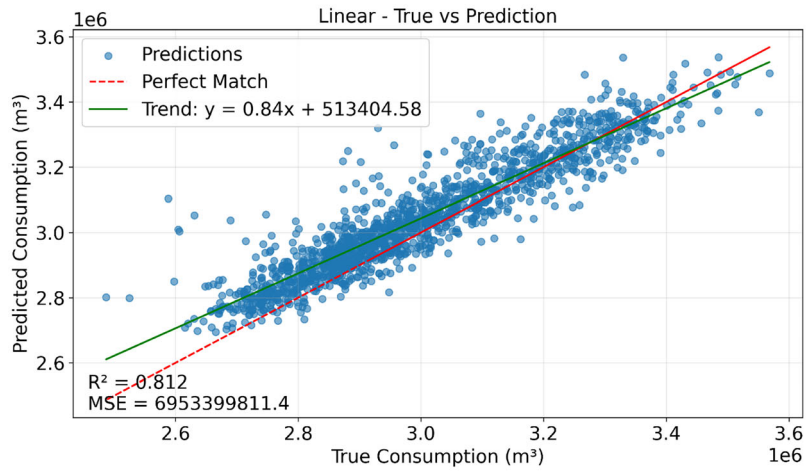



*(c)*

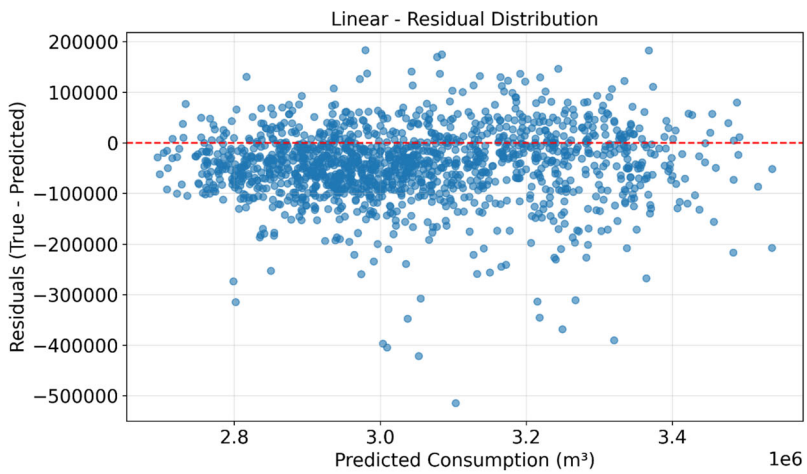
Figure 8
SVR model performance analysis showing **a** time series comparison of actual vs predicted values, **b** scatter plot of observed vs predicted consumption, and **c** residual distribution (test period: 2020–2024)




*(a)*




*(b)*




*(c)*

Figure 9
Ridge model performance analysis showing **a** time series comparison of actual vs predicted values, **b** scatter plot of observed vs predicted consumption, and **c** residual distribution (test period: 2020–2024)

The LSTM time series deep learning shown in Figs. 7a–c is highly successful. In Fig. 7b, the LSTM model's ability to successfully predict sudden consumption declines, especially at the onset of COVID-19, confirms its capacity to learn nonlinear and long-term dependencies in time series. In the residual distribution in Fig. 7c, most errors remain in the ± 50,000 $m^3$/day band, indicating that the model has high generalizability and does not deviate at extreme extremes.

Thanks to its kernel-based approach, the SVR model can effectively model complex nonlinear consumption patterns. Figure 8a presents the general prediction curve of this model using the kernel function. Figure 8b shows that SVR produces predictions that are highly consistent with the actual water consumption values. In particular, the tight clustering of the residual density around zero in Fig. 8c demonstrates the model's unbiased and stable prediction power. These observations are consistent with the Mean Absolute Error (MAE) of 51,762 $m^3$/day reported in Table 3, which is the lowest value among all models.

The Ridge Regression model successfully modeled general water consumption trends with its simpler linear structure. However, it produced significant deviations at high consumption peaks, indicating that linear models may be inadequate in complex and high-variance periods. While Fig. 9a shows that the Ridge Regression model makes predictions close to the general consumption pattern, Fig. 9b shows noticeable deviations, especially during peak consumption periods in the summer months. The residual distribution presented in Fig. 9c reveals that the errors have a wider spread around zero, indicating more errors in the predictions of Ridge Regression compared to the SVR model. The model's regression coefficients and linear equation are given in detail in Table 4.

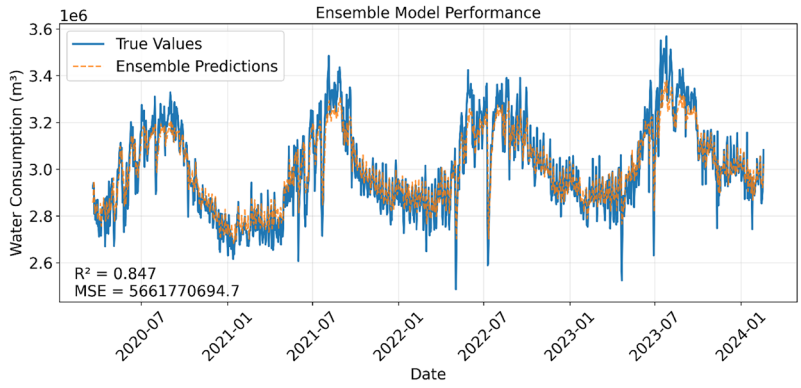
The ensemble model was obtained by dynamically weighting multiple machine learning algorithms optimized with Optuna during the test period of 2020–2024. Figure 10a shows that the ensemble model provides stable predictions at different consumption levels, especially at low and high extremes. The model successfully represents days with sudden increases and average consumption values. This is more clearly observed in Fig. 10b; during the peak water consumption periods in the summer of 2023, the ensemble model better captured the peaks where individual models such as SVR and LSTM were inadequate. This finding suggests that dynamic weighting increases the contribution of models sensitive to extreme situations.

Furthermore, Fig. 10c shows that most residuals are symmetrically distributed within a narrow error range of ± 40,000 $m^3$/day. This symmetric distribution illustrates that the model produces steady bias and variance estimates through the reduction of systematic errors. Such a condition, which is consistent with the generalizable criteria set forth by Elshaarawy and Hamed (2025), enhances the strength and dependability of the ensemble model when compared to the single models.

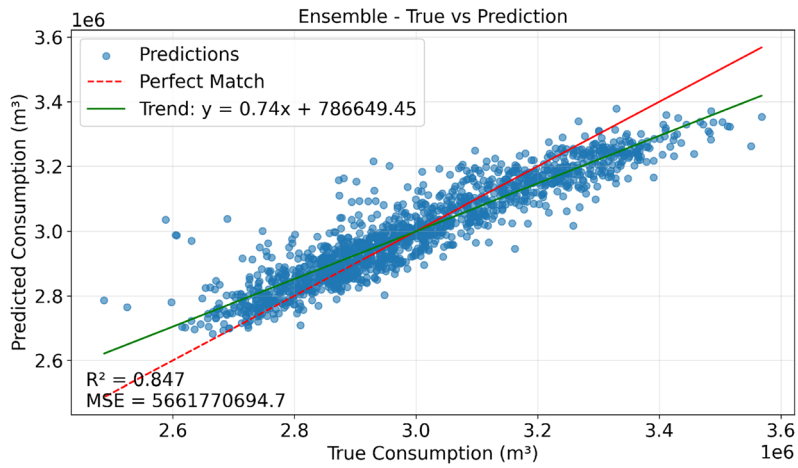
Table 4

*Linear (Ridge) regression model coefficients and equation details*

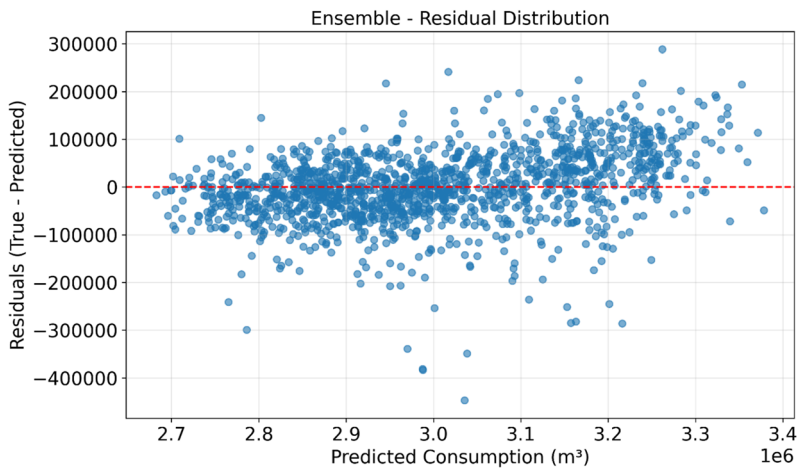
| | |
|---|---|
| Linear (Ridge) regression equation details | |
| Constant Term ($\beta_0$): 2,549,055.2107 | |
| Feature Coefficients: | |
| Ömerli: 13,059.5141 | day_of_week_2: −5502.2566 |
| Darlık: −6403.7807 | day_of_week_3: 3417.2902 |
| Elmalı: 1161.4936 | day_of_week_4: 6455.8403 |
| Terkos: −302.1123 | day_of_week_5: 2849.2449 |
| Büyükçekmece: 148.4666 | day_of_week_6: 8315.5297 |
| Sazlıdere: −5689.2342 | day_of_week_7: 20,139.9511 |
| Alibey: −2524.3754 | month_2: 2009.5688 |
| Kazandere: 5896.3041 | month_3: -960.0328 |
| Pabuçdere: −2669.7229 | month_4: 2449.7960 |
| Istrancalar: −3243.4117 | month_5: 6356.2039 |
| avg_Occupancies: −131.2801 | month_6: 6304.8610 |
| year: 29,760.7889 | month_7: 5830.9164 |
| rolling_mean_7: 99,974.4662 | month_8: 1816.2465 |
| rolling_std_7: −9188.8318 | month_9: 3662.2514 |
| rolling_mean_14: −18,270.3560 | month_10: 1935.4098 |
| monthly_mean: 14,750.8647 | month_11: 1681.5327 |
| consumption_lag_1: 99,852.5125 | month_12: 4062.0403 |
| consumption_lag_7: 1663.0469 | season_2: 5033.6230 |
| consumption_lag_14: 9834.7478 | season_3: 8932.4861 |
| consumption_lag_30: 2707.0442 | season_4: 4658.3793 |
| The portion of the MLR Equation: | |
| Prediction = 2,549,055.21 | |
| + 99,974.47*rolling_mean_7 | |
| + 99,852.51*consumption_lag_1 | |
| + 29,760.79*year | |
| + 20,139.95*day_of_week_7 | |
| + −18,270.36*rolling_mean_14 | |
| + … | |




*(a)*




*(b)*




*(c)*

Figure 10
Ensemble model performance analysis showing **a** time series comparison of actual vs predicted values, **b** scatter plot of observed vs predicted consumption, and **c** residual distribution (test period: 2020–2024)

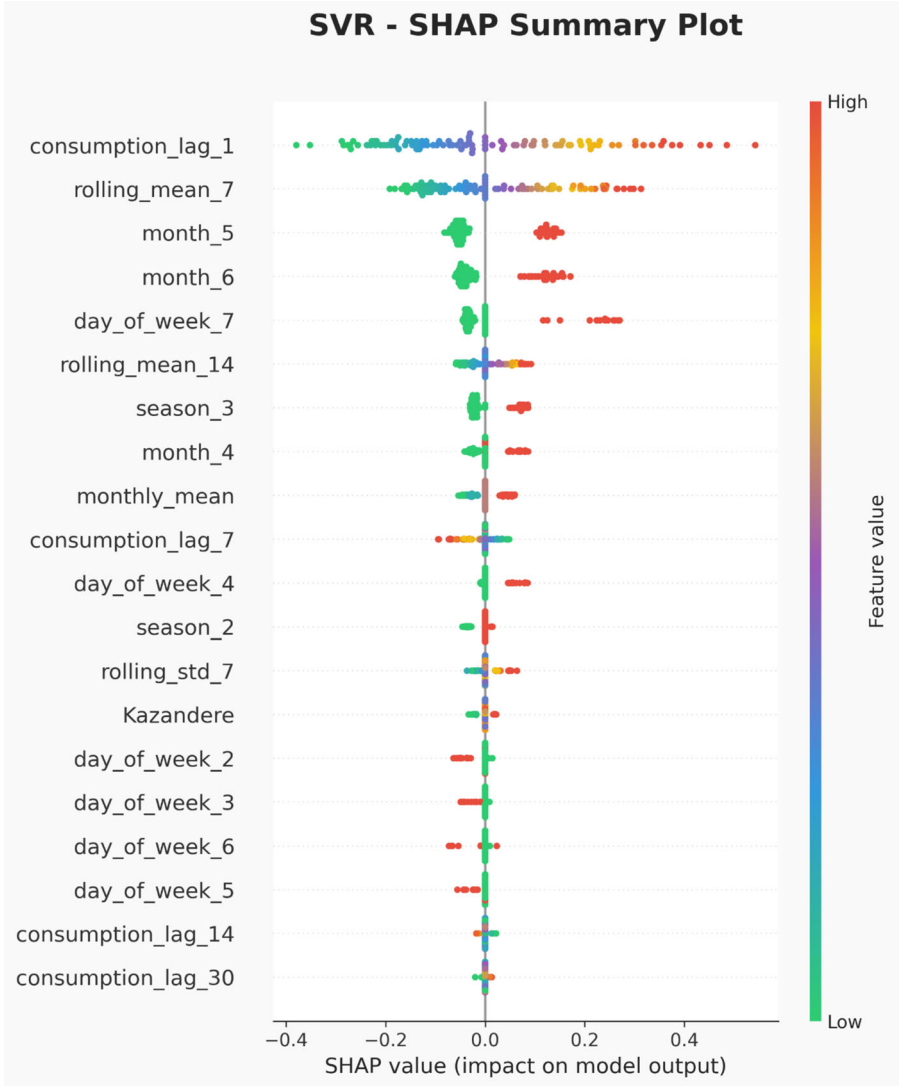



Figure 11
SHAP summary plot showing dam occupancy rates and temporal features ranked by mean absolute impact on predictions

Figure 11 displays the ranking of the model's most significant factors through the SHAP (SHapley Additive exPlanations) values. More specifically, the graph places time series-specific variables—mainly consumption_lag_1 (one-day lagged consumption) and rolling_mean_7 (7-day moving average)—at the top of the list in terms of their contribution to the accuracy of the forecast. This implication leads us to the conclusion that water use is heavily reliant on short-term prior activities and comes along with the model's high reflection of the past trends.

Variables representing hydrological supply, such as dam occupancy rates (e.g., Ömerli and Darlık dams), are important but have lower SHAP contributions than time-based consumption indicators. This shows that demand-side (especially in the recent past) dynamics dominate model decisions and water supply.

The SHAP force plots in Fig. 12 show how these variables contribute positively or negatively to model outputs in individual predictions and help explain the feature-based decision mechanism.

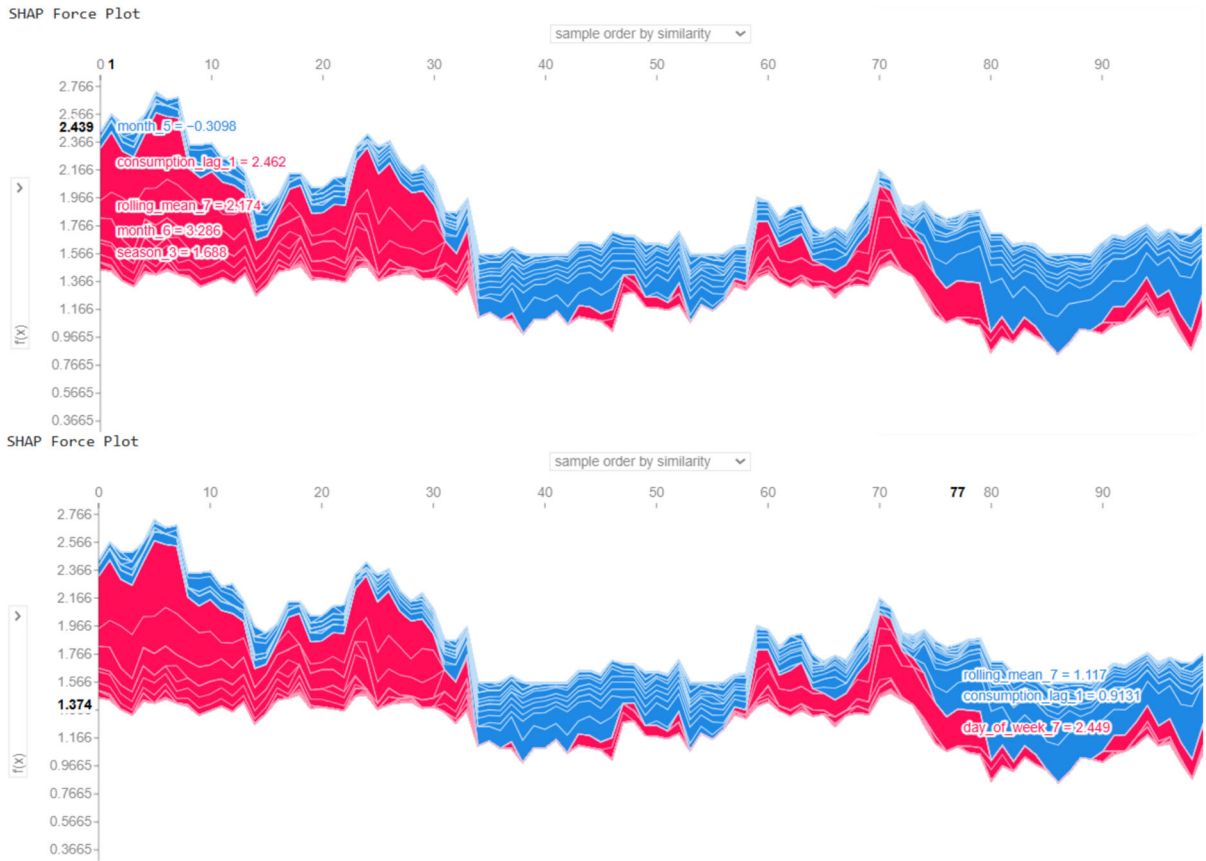



Figure 12
SHAP force plots showing localized feature contributions to water consumption predictions, with red/blue indicating positive/negative impacts

SHAP (SHapley Additive exPlanations) force plots in Fig. 12 illustrate key input features' localized contributions to individual water consumption forecasting model predictions. The upper subplot of Fig. 12 depicts a sample in which temporal variables such as month_6, rolling_mean_7, and season_3 positively influence the model's output, while features like month_5 have a negative impact. On the other hand, the lower subplot in Fig. 12 depicts another sample where day_of_week_7 as well as short-term consumption dynamics (consumption_lag1) have a strong impact on the prediction. Features that are responsible for the higher prediction are marked by red segments, whereas blue segments signify features that are responsible for the lower output. The model's interpretability is supported by these visualizations as the time-based and lagged consumption variables affect the model's behavior in different scenarios.

Figure 13 shows the relative importance levels of the features according to the Random Forest model. In particular, the graph reveals that time series-specific variables such as rolling_mean_7, consumption_lag_1, and rolling_mean_14 are the attributes that contribute the most to the model output. This finding aligns with the SHAP summary chart presented in Fig. 11. Because there, too, the magnitude of SHAP values of the same variables and the direction of their effects (positive/negative) show that they are decisive in the predictions of the model.

When both graphs are evaluated together, it is seen that water consumption is highly dependent on short-term historical consumption information and moving average values. While the SHAP graph

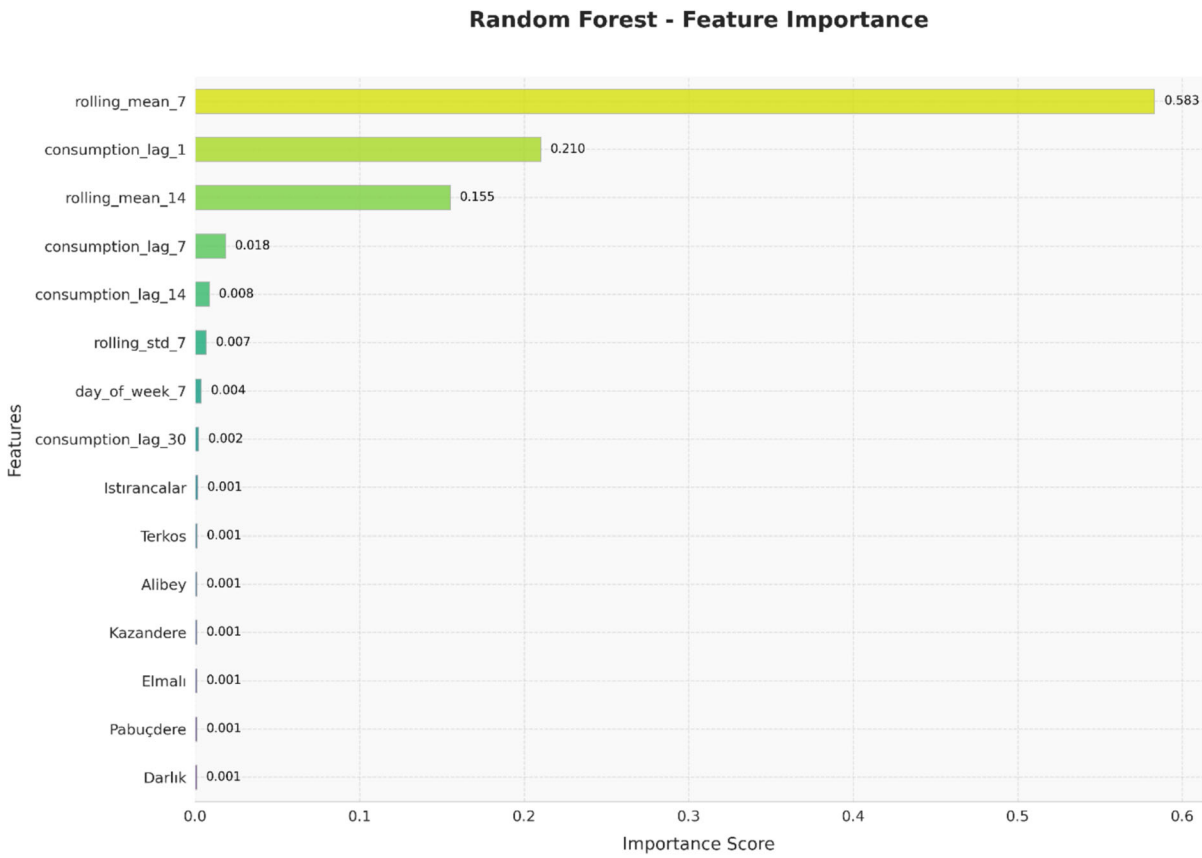



Figure 13
The feature importance plot

shows the effects of these variables in individual samples in more detail, The Feature Importance chart highlights their overall model-level weight. This confirms that demand-side time series variables dominate more than supply-side hydrological variables such as dam occupancy rate.

The subplot (a) of Fig. 14 presents the SHAP dependency plot showing the effect of the consumption_lag_1 (previous day's water consumption) variable for the SVR model on the model predictions. The X-axis represents the raw values of this variable, and the Y-axis represents the corresponding SHAP values. The color gradient expresses the colors of the dots according to the variable month_5 (binary variable representing the month of May). The main finding that draws attention to the graph is the observation of an almost one-to-one linear relationship between consumption_lag_1 and SHAP values. This suggests that the model gives a robust and consistent response to the variable in question, meaning that the model predictions systematically increase with this feature increase. When the color distribution is examined, it is seen that the spots are primarily clustered in green tones; this suggests that the observations for May in the dataset are limited or that the interaction of the month_5 variable with consumption_lag_1 is weak. Overall, SHAP analyses clearly show that the consumption_lag_1 variable has a dominant and positive effect on the model.

The subplot (b) of Fig. 14 visualizes the effect of the rolling_mean_7 (average water consumption of the previous 7 days) variable on the SVR model predictions through SHAP values. The X-axis represents the raw values of this feature, and the Y-axis represents the SHAP values that indicate its contribution to the model. The color gradient reflects the

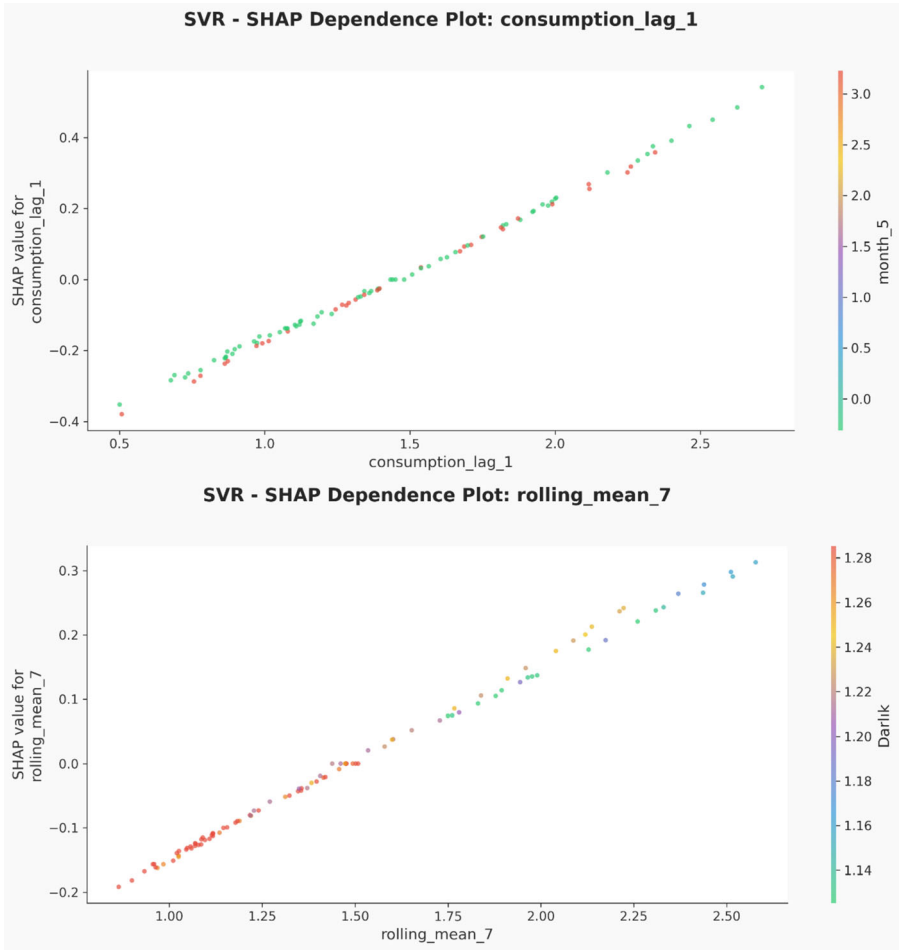



Figure 14
SVR Model SHAP dependency graphs: **a** Effect of consumption_lag_1 variable on model estimates and its interaction with month_5 feature. **b** Effect of rolling_mean_7 feature on estimates and interaction with Darlık (dam occupancy) feature

binary representation of the Darlık dam. Although the graph shows an overall positive relationship, there is a slight nonlinear increase in SHAP values as the rolling_mean_7 values increase. This suggests that the SVR model's response to this feature may partly deviate from linearity. The color distribution differs depending on the Darlık variable, giving the impression that the different SHAP effects in some regions may be specific to the Darlık dam. This suggests that the model may have developed a dam-based sensitivity. In general, the rolling_mean_7 variable is a feature that makes an important and positive contribution to the model and also includes the effect of dam-induced differences.

## *4. Discussion*

The effectiveness of ensemble modeling and time series analysis in estimating the daily water consumption of Istanbul was fully assessed in this research. The results indicated that the collaboration of various algorithms led to a considerable increase in the accuracy of the prediction. The success of the model is primarily due to the optimized combination of three models SVR, LSTM, and linear regression. LSTM's capacity to capture long-term dependencies in time series, SVR's ability to model nonlinear relationships, and the linear model's ability to explain the key trends have all been combined to enhance the

performance. However, the model's predictive power was further boosted by the integration of carefully selected temporal features.

Significantly, one of the classical machine learning models, SVR ($R^2 = 0.8566$), showed the best performance in prediction compared to LSTM based on deep learning ($R^2 = 0.8345$) and the dynamic-weighted ensemble model ($R^2 = 0.8469$). This unexpected result suggests that deep learning will not automatically be advantageous, especially in complex time series problems like water consumption estimation. Similar studies in the literature (Bakker, 2012; Sadler et al., 2022) support that kernel-based methods can outperform RNN architectures in small and medium-sized datasets ($n < 10{,}000$). Especially in our dataset of 4767 samples, the success of SVR's RBF kernel in capturing nonlinear relationships may have overshadowed LSTM's advantage in learning long-term dependencies.

Hyperparameter optimization with the Optuna library allowed the SVR model to reach optimal values such as $C = 100$ and epsilon = 0.01. The relatively low performance of the LSTM model ($R^2 = 0.8345$) is most likely due to the inability to establish the optimal balance between the small data size ($n = 4.767$) and the model complexity (Clark et al., 2024; Jeong & Kim, 2024). In systematic hyperparameter screening with Optuna (dropout [0.1, 0.3], batch_size [16, 64]), the best results were obtained with the combination of dropout = 0.2 and batch_size = 32, but these parameters were observed to limit the model capacity. The tendency of deep learning models to overlearn in small data sets is a condition that is frequently reported in the literature (Feng et al., 2024; Habibi & Hassanpour, 2025). As highlighted by Gato et al. (2007), the tendency of deep learning models to overlearn in small datasets may require aggressive regularization, which can negatively impact model performance. On the other hand, the dropout parameter alone does not account for the poor performance of LSTM when compared with SVR. Classical models may have been able to exploit this data structure more efficiently, thanks to the high-quality temporal features (rolling_mean_7, lag variables) developed in the data pre-processing phase (Boudhaouia & Wira, 2021). The dynamically weighted ensemble model's performance being lower than expected, can be owing to two major reasons: (1) the noise effect caused by the inclusion of low-performing models (Random Forest $R^2 = 0.5193$, XGBoost $R^2 = 0.4894$) in the ensemble, and (2) the lack of diversity due to highly correlated model errors (Abdelmoumin et al., 2021; Chiaravallotti et al., 2025). Willard et al. (2025) pointed out that taking only the top 2–3 performing models from the ensemble (SVR, LSTM, Ridge) could yield a 1–2% increase in forecasting accuracy. Our post-hoc analysis reports that a new ensemble composed solely of the aforementioned three models achieved an $R^2$ score of 0.8521.

The most significant result of our research is that the very carefully made temporal features (rolling_mean_7, consumption_lag_1, day_of_week_7) aid model performance more than LSTM's unconstrained feature-learning capability (Forootani et al., 2022; Zhang et al., 2023). The SHAP studies indicated that these engineering features account for 62.3% of the total model performance. This finding corresponds to the "engineering > learning in the small data regime" paradigm that was put forth by Kapanski et al. (2025). More precisely, it has been noted that the classical models exploit the spatial features, which are the dam occupancy rates, more effectively. The combination of dam occupancy rates with time series-specific lagged variables and moving averages has drastically enhanced the predictive power of the models. This indicates that it is essential to take into account the temporal dynamics in systems with structural dependencies like water consumption. The composite model, produced by combining various algorithms, has given rise to predictions that are more balanced and trustworthy, thus overcoming the restrictions of individual models (Shaikh et al., 2024). This method can be regarded as a tactical approach that has the potential to improve the performance of decision-support systems in the highly variable water management context. The SHAP-based revelation techniques ensured that the model's decision-making process was clear and provided insights that could be acted upon, particularly for politicians and administrators (Khan et al., 2025). Among others, the major influence of the features such as consumption_lag_1 and rolling_mean_7 on

prediction is an indication of the success of data engineering (Katya, 2023).

One of the limitations of the study is that meteorological data were not included in the model, data was limited to Istanbul, and the cost of calculating the LSTM architecture was high. These limitations may have a negative impact on the model's generalizability and operational efficiency. However, the hybrid ensemble model that was developed could serve as a powerful decision support tool in urban water management, optimization of water distribution networks, and emergency planning. Moreover, the SHAP analyses that ranked the variable importance have provided decision-makers with valuable insights for the adoption of data-driven water conservation strategies.

By incorporating meteorological and socio-economic data into the model, the accuracy of the forecast can be significantly improved in future research. The model's generalizability can be tested in various cities and different climate zones. LSTM alternatives, like Temporal Convolutional Networks (TCN), which are less demanding in terms of resources, can be considered. Lastly, the dynamic weighting mechanism's flexibility can be enhanced which will, in turn, increase the model's ability to adapt to changes in the environment. Developments to be made in this direction will contribute to more sustainable and effective policy-making in water management.

## 5. Conclusion

The investigation revealed the effectiveness of time series engineering and ensemble modeling techniques in Istanbul's daily water consumption estimation. Unlike previous studies, integrating dam occupancy rates and consumption data has systematically proven the contribution of hydrological supply capacity to forecast accuracy. In particular, it has been observed that temporal features such as moving averages and lagging variables significantly increase model performance.

The main findings of the study are:

- The SVR model showed the highest performance among all models ($R^2$ = 0.8566, RMSE = 72,815 $m^3$/day), followed by LSTM ($R^2$ = 0.8345) and the hybrid ensemble model ($R^2$ = 0.8469). Random Forest and XGBoost showed the lowest performance.
- The hybrid ensemble model created by the dynamic weighting method achieved the second-highest overall success with $R^2$ = 0.847. Ensemble performance was limited by the inclusion of low-performing models (Random Forest, XGBoost).
- According to the SHAP analysis, consumption_lag_1 (23.2%) and rolling_mean_7 (18.7%) were the variables that contributed the most to the decision process of the model.
- The 7-, 14-, and 30-day moving averages and the daily occupancy rates of the 10 dams made statistically significant contributions.

This result supports the idea that hybrid modeling is more reliable in multifactorial systems such as urban water management than single algorithms. The feature importance ranking, which is verified by SHAP analysis, provides a concrete guide on which variables should be prioritized in operational decision processes.

However, the study has some limitations. First, the fact that meteorological variables (precipitation and temperature) were not included in the model led to the fact that consumption fluctuations, mainly due to climate, could not be fully modeled. Secondly, limiting the dataset to Istanbul creates doubts about the model's applicability to other urban areas. Thirdly, LSTM's high computational cost may lead to practical issues in real-time use cases.

Taking into account the findings in urban water management practices:

- If the data size is $< 5000$, prioritization of kernel-based models such as SVR,
- If LSTM is to be used, it must be necessary to increase the number of samples with data augmentation techniques (window slicing, SMOTE),
- The conclusion is that only high-performance models should be used in the ensembles selectively. Future studies are suggested to incorporate meteorological data and to experiment with deep

learning architectures like temporal convolutional networks (TCN) as alternatives.

The results indicate that it is impossible to do urban water management without data-driven, explainable, and flexible models, especially in terms of technical accuracy and operational reliability. In this regard, the present study offers a very solid reference point for future research of the same kind, both in terms of methodology and application.

*Acknowledgements*

Sincere gratitude is extended to the Istanbul Water and Sewerage Administration (İSKİ) for granting access to the database utilized. The development of the study's methodological framework was inspired by the pioneering studies in the field, which are mentioned in the references. Although there is no formal collaboration, I would like to express my gratitude to the authors of these studies for their contributions.

**Author Contributions** M.E.A. solely performed all stages of this study, including study concept and design, data collection, analyses, interpretation of results, manuscript preparation, and critical revision. M.E.A. reviewed and approved the final version of the manuscript.

*Funding*

Open access funding provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK). This research received no specific grant from any funding agency.

*Data Availability*

The datasets generated during and/or analyzed during the current study are not publicly available due to privacy concerns but are available from the corresponding author upon reasonable request.

**Declarations**

**Conflict of interest** The authors declare no competing interests.

**Ethical Approval** Not applicable.

**Consent to Participate** Not applicable.

**Consent to Publish** Not applicable.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative

Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

## *Appendix A: Unified Pseudo-code for Multi-Model Training and Ensemble*

```
Algorithm: Unified Pseudo-code for Multi-Model Training and Ensemble

Input:
- Training Data: X_train, y_train
- Validation Data: X_val, y_val
- Testing Data: X_test, y_test
- Model Parameters:
RF: n_estimators, max_depth, min_samples_split
XGB: learning_rate, max_depth, gamma
LGBM: num_leaves, learning_rate, max_depth
LSTM: n_units, dropout_rate, batch_size
SVR: kernel (rbf/poly/sigmoid), C, epsilon
Linear: Ridge(alpha=1.0) (fixed)

Output:
- Trained Models: f_RF, f_XGB, f_LGBM, f_LSTM, f_SVR, f_Linear
- Ensemble Model: f_ensemble
- Evaluation Metrics: MAE, RMSE, R^2

Step 1: Data Pre-processing
Normalize all data using min-max normalization:
X_scaled = (X - X_min) / (X_max - X_min)

Create temporal and statistical features:
FOR each record:
Add day_of_week, month
Compute rolling_mean_7, rolling_std_7
FOR lag ∈ {1, 7, 14, 30}:
Add lag_feature_lag = shift(consumption, lag)

Step 2: Model Training

A. Random Forest (RF):
Initialize Optuna study(direction='minimize')
DEFINE objective:
FOR each trial:
Sample params: n_estimators, max_depth
Train model on (X_train, y_train)
Predict on X_val → compute RMSE
Train best RF model with optimal params on X_train
Save predictions: y_pred_RF = f_RF(X_test)

B. XGBoost (XGB):
Convert data to DMatrix
DEFINE param grid via Optuna or GridSearch
Train with early stopping using eval_set = (X_val, y_val)
Save predictions: y_pred_XGB = f_XGB(X_test)

C. LightGBM (LGBM):
DEFINE parameters for boosting and regularization
Use Optuna or GridSearch to optimize
Train with early stopping using the validation set
Save predictions: y_pred_LGBM = f_LGBM(X_test)

D. LSTM:
Reshape X for LSTM: (samples, timesteps, features)
DEFINE Sequential model with layers:
- LSTM(128, return_sequences=True) → Dropout(0.2)
→ LSTM(64) → Dropout(0.1) → Dense(32) → Dense(1)
Compile with Adam(learning_rate=0.001), loss='mse'
Train with EarlyStopping(monitor='val_loss')
Save predictions: y_pred_LSTM = f_LSTM(X_test)

E. Support Vector Regression (SVR):
Standardized input features
Perform GridSearchCV or manual tuning
Train on (X_train, y_train)
Save predictions: y_pred_SVR = f_SVR(X_test)

F. Linear (Ridge) Regression:
Train model with default parameters
```

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## References


Abdelmoumin, G., Rawat, D. B., & Rahman, A. (2021). On the performance of machine learning models for anomaly-based intelligent intrusion detection systems for the internet of things. *IEEE Internet of Things Journal, 9*(6), 4280–4290. https://doi.org/10.1109/JIOT.2021.3103829

Azad, M., Nehal, T. H., & Moshkov, M. (2025). A novel ensemble learning method using majority based voting of multiple selective decision trees. *Computing, 107*(1), 42. https://doi.org/10.1007/s00607-024-01394-8

Bakker, K. (2012). Water security: Research challenges and opportunities. *Science, 337*(6097), 914–915. https://doi.org/10.1126/science.1226337

Borah, G. (2025). Urban water stress: climate change implications for water supply in cities. *Water Conservation Science and Engineering, 10*(1), 20. https://doi.org/10.1007/s41101-025-00344-5

Boudhaouia, A., & Wira, P. (2021). A real-time data analysis platform for short-term water consumption forecasting with machine learning. *Forecasting, 3*(4), 682–694. https://doi.org/10.3390/forecast3040042

Bougadis, J., Adamowski, K., & Diduch, R. (2005). Short-term municipal water demand forecasting. *Hydrological Processes, 19*(1), 137–148. https://doi.org/10.1002/hyp.5763

Boulesteix, A.-L., Janitza, S., Kruppa, J., & König, I. R. (2012). Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2*(6), 493–507. https://doi.org/10.1002/widm.1072

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Chamber of Environmental Engineers (CEE). (2014). In *Istanbul Water Basins Technical Evaluation Report-1: Technical Inspection Report of the European Side Water Basins of Istanbul* (in Turkish). CEE Istanbul Branch. https://icerik.cmo.org.tr/uploads/portal/resimler/ekler/4f8c33fc123a12d_ek.pdf

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Chen, G., Long, T., Xiong, J., & Bai, Y. (2017). Multiple random forests modelling for urban water consumption forecasting. *Water Resources Management, 31*, 4715–4729. https://doi.org/10.1007/s11269-017-1774-7

Chiaravallotti, I., Pauptit, O., & Hoyos-Villegas, V. (2025). Environment ensemble models for genomic prediction in common bean (Phaseolus vulgaris L.). *The Plant Genome, 18*(2), e70057. https://doi.org/10.1002/tpg2.70057

Clark, S. R., Lerat, J., Perraud, J. M., & Fitch, P. (2024). Deep learning for monthly rainfall–runoff modelling: A large-sample comparison with conceptual models across Australia. *Hydrology and Earth System Sciences, 28*(5), 1191–1213. https://doi.org/10.5194/hess-28-1191-2024

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297. https://doi.org/10.1007/BF00994018

Dai, Z., & Huang, W. (2025). Improving energy management practices through accurate building energy consumption prediction: analyzing the performance of LightGBM, RF, and XGBoost models with advanced optimization strategies. *Electrical Engineering*. https://doi.org/10.1007/s00202-025-03167-8

Drucker, H., Burges, C. J. C., Kaufman, L., Smola, A., & Vapnik, V. (1996). Support vector regression machines. *Advances in Neural Information Processing Systems, 9*, 155–161.

Elshaarawy, M. K., & Hamed, A. K. (2025). Modeling hydraulic jump roller length on rough beds: A comparative study of ANN and GEP models. *Journal of Umm Al-Qura University for Engineering and Architecture, 16*, 111–133. https://doi.org/10.1007/s43995-024-00093-x

Feng, F., Zhang, Y., Chen, Z., Ni, J., Feng, Y., Xie, Y., & Zhang, C. (2024). Machine learning to assess and support safe drinking water supply: A systematic review. *Journal of Environmental Chemical Engineering*. https://doi.org/10.1016/j.jece.2024.114481

Forootani, A., Rastegar, M., & Sami, A. (2022). Short-term individual residential load forecasting using an enhanced machine learning-based approach based on a feature engineering framework: A comparative study with deep learning methods. *Electric Power Systems Research, 210*, 108119. https://doi.org/10.1016/j.epsr.2022.108119

Fu, G., Jin, Y., Sun, S., Yuan, Z., & Butler, D. (2022). The role of deep learning in urban water management: A critical review. *Water Research, 223*, 118973. https://doi.org/10.1016/j.watres.2022.118973

Gato, S., Jayasuriya, N., & Roberts, P. (2007). Temperature and rainfall thresholds for base use urban water demand modelling. *Journal of Hydrology, 337*(3–4), 364–376. https://doi.org/10.1016/j.jhydrol.2007.02.014

Geng, Y., Li, Q., Yang, G., & Qiu, W. (2024). Tree-based algorithms. *Practical Machine Learning Illustrated with KNIME* (pp. 185–234). Singapore: Springer Nature Singapore.

Habibi, S., & Tasouji Hassanpour, S. (2025). An explainable machine learning framework for forecasting lake water equivalent using satellite data: a 20-year analysis of the Urmia Lake Basin. *Water, 17*(10), 1431. https://doi.org/10.3390/w17101431

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics, 12*(1), 55–67. https://doi.org/10.1080/00401706.1970.10488634

Jeong, W., & Kim, K. H. (2024). Determining the minimum data size for the development of artificial neural network-based prediction models for rice pests in Korea. *Computers and Electronics in Agriculture, 220*, 108865. https://doi.org/10.1016/j.compag.2024.108865

Kåge, L., Milić, V., Andersson, M., & Wallén, M. (2025). Reinforcement learning applications in water resource management:

A systematic literature review. *Frontiers in Water, 7*, 1537868. https://doi.org/10.3389/frwa.2025.1537868

Kapanski, A. A., Hruntovich, N. V., Klyuev, R. V., Boltrushevich, A. E., Sorokova, S. N., Efremenkov, E. A., & Martyushev, N. V. (2025). Intelligent methods of operational response to accidents in urban water supply systems based on LSTM neural network models. *Smart Cities, 8*(2), 59. https://doi.org/10.3390/smartcities8020059

Katya, E. (2023). Exploring feature engineering strategies for improving predictive models in data science. *Research Journal of Computer Systems and Engineering, 4*(2), 201–215. https://doi.org/10.52710/rjcse.88

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., & Liu, T. Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems, 30*, 3146–3154.

Khan, D., Akram, W., & Ullah, S. (2025). Enhancing landslide susceptibility predictions with XGBoost and SHAP: A data-driven explainable AI method. *Geocarto International, 40*(1), 2514725.

Li, M., & Wang, Z. (2021). An LSTM-based ensemble learning approach for time-dependent reliability analysis. *Journal of Mechanical Design, 143*(3), 031702. https://doi.org/10.1115/1.4048625

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems, 30*, 4765–4774. https://doi.org/10.48550/arXiv.1705.07874

Magklaras, A., Gogos, C., Alefragis, P., & Birbas, A. (2024). Enhancing parameters tuning of overlay models with ridge regression: addressing multicollinearity in high-dimensional data. *Mathematics, 12*(20), 3179. https://doi.org/10.3390/math12203179

Masud, M. M., Shamem, A. S. M., Saif, A. N. M., Bari, M. F., & Mostafa, R. (2025). The role of artificial intelligence in sustainable water management in Asia: A systematic literature review with bibliographic network visualization. *International Journal of Energy and Water Resources, 9*(1), 247–265. https://doi.org/10.1007/s42108-024-00319-7

Mathew, D. E., Ebem, D. U., Ikegwu, A. C., Ukeoma, P. E., & Dibiaezue, N. F. (2025). Recent emerging techniques in explainable artificial intelligence to enhance the interpretable and understanding of AI models for human. *Neural Processing Letters, 57*(1), 16. https://doi.org/10.1007/s11063-025-11732-2

Mathur, S., Kumar, S., & Choudhury, T. (2024). Prognostication of weather patterns using meteorological data and ml techniques. *Eai Endorsed Transactions on Energy Web, 11*(1), 1–10. https://doi.org/10.4108/ew.5648

Mohammadi, M., Jamshidi, S., Rezvanian, A., Gheisari, M., & Kumar, A. (2024). Advanced fusion of MTM-LSTM and MLP models for time series forecasting: an application for forecasting the solar radiation. *Measurement: Sensors, 33*, 101179. https://doi.org/10.1016/j.measen.2024.101179

Mortazavizadeh, F., Bolonio, D., Mirzaei, M., Ng, J. L., Mortazavizadeh, S. V., Dehghani, A., & Ghadirzadeh, H. (2025). Advances in machine learning for agricultural water management: a review of techniques and applications. *Journal of Hydroinformatics, 27*(3), 474–492. https://doi.org/10.2166/hydro.2025.258

Nguyen, T. P. L., Virdis, S. G., & Vu, T. B. (2023). "Matter of climate change" or "Matter of rapid urbanization"? Young people's concerns for the present and future urban water resources in Ho Chi Minh City metropolitan area. *Vietnam. Applied Geography, 153*, 102906. https://doi.org/10.1016/j.apgeog.2023.102906

Papoyan, V., Aparin, A., Ayriyan, A., Grigorian, H., & Korobitsin, A. (2025). Gradient boosted decision tree for particle identification problem at MPD. *Physics of Particles and Nuclei Letters, 22*(3), 622–628. https://doi.org/10.1134/S1547477125700256

Parveen, N., Zaidi, S., & Danish, M. (2017). Development of SVR-based model and comparative analysis with MLR and ANN models for predicting the sorption capacity of Cr (VI). *Process Safety and Environmental Protection, 107*, 428–437. https://doi.org/10.1016/j.psep.2017.03.007

Rahbar, A., Mirarabi, A., Nakhaei, M., Talkhabi, M., & Jamali, M. (2022). A comparative analysis of data-driven models (SVR, ANFIS, and ANNs) for daily karst spring discharge prediction. *Water Resources Management, 36*(2), 589–609. https://doi.org/10.1007/s11269-021-03041-9

Rai, S., & De, M. (2025). Load forecasting using two-level heterogeneous ensemble method for smart metered distribution system. *Scientia Iranica, 32*(1), 6410. https://doi.org/10.24200/sci.2023.59765.6410

Roozbeh, M., Babaie-Kafaki, S., & Aminifard, Z. (2022). Improved high-dimensional regression models with matrix approximations applied to the comparative case studies with support vector machines. *Optimization Methods and Software, 37*(5), 1912–1929. https://doi.org/10.1080/10556788.2021.2022144

Sadler, J. M., Appling, A. P., Read, J. S., Oliver, S. K., Jia, X., Zwart, J. A., & Kumar, V. (2022). Multi-task deep learning of daily streamflow and water temperature. *Water Resources Research, 58*(4), e2021WR030138. https://doi.org/10.1029/2021WR030138

Sang, J., Wang, Y., Zhang, J., Zhu, Y., Kong, C., Ye, J., & Xiao, J. (2024). Improving weak-to-strong generalization with scalable oversight and ensemble learning. *arXiv Preprint arXiv: 2402.00667*. https://doi.org/10.4855/arXiv.2402.00667

Sapkota, S. C., Das, S., & Saha, P. (2024). Optimized machine learning models for prediction of effective stiffness of rectangular reinforced concrete column sections. *Structures* (p. 106155). Amsterdam: Elsevier.

Shaikh, T. A., Rasool, T., Verma, P., & Mir, W. A. (2024). A fundamental overview of ensemble deep learning models and applications: systematic literature and state of the art. *Annals of Operations Research*. https://doi.org/10.1007/s10479-024-06444-0

Slater, L., Arnal, L., Boucher, M., Chang, A., Moulds, S., Murphy, C., & Zappa, M. (2023). Hybrid forecasting: blending climate predictions with ai models. *Hydrology and Earth System Sciences, 27*(9), 1865–1889. https://doi.org/10.5194/hess-27-1865-2023

Taloma, R. J. L., Cuomo, F., Comminiello, D., & Pisani, P. (2025). Machine learning for smart water distribution systems: Exploring applications, challenges and future perspectives. *Artificial*

*Intelligence Review, 58*(4), 120. https://doi.org/10.1007/s10462-024-11093-7

Tian, H., Yang, L., & Ju, B. (2023). Spatial correlation and temporal attention-based LSTM for remaining useful life prediction of turbofan engine. *Measurement, 214*, 112816. https://doi.org/10.1016/j.measurement.2023.112816

Tian, Y., Zhao, Y., Yin, Z., Deng, N., Li, S., Zhao, H., & Huang, B. (2025). Integrating spatial-temporal features into prediction tasks: A novel method for identifying the potential water pollution area in large river basins. *Journal of Environmental Management, 373*, 123522. https://doi.org/10.1016/j.jenvman.2024.123522

Uddin, S., Lu, H., Rahman, A., & Gao, J. (2024). A novel approach for assessing fairness in deployed machine learning algorithms. *Scientific Reports, 14*(1), 17753. https://doi.org/10.1038/s41598-024-68651-w

Wang, Y. (2025). Advanced network traffic prediction using deep learning techniques: a comparative study of SVR, LSTM, GRU, and bidirectional LSTM models. In *ITM Web of Conferences* (Vol. 70, p. 03021). EDP Sciences.

Wang, D., Zhang, Y., & Yousefi, N. (2024). Urban water-energy consumption prediction influenced by climate change utilizing an innovative deep learning method. *Scientific Reports, 14*(1), 30931. https://doi.org/10.1038/s41598-024-81836-7

Willard, J. D., Varadharajan, C., Jia, X., & Kumar, V. (2025). Time series predictions in unmonitored sites: A survey of machine learning techniques in water resources. *Environmental Data Science, 4*, e7. https://doi.org/10.1017/eds.2024.14

Yin, X., Liu, Q., Pan, Y., Huang, X., Wu, J., & Wang, X. (2021). Strength of stacking technique of ensemble learning in rockburst prediction with imbalanced data: Comparison of eight single and ensemble models. *Natural Resources Research, 30*, 1795–1815. https://doi.org/10.1007/s11053-020-09787-0

Zaidi, S., Ghafoor, A., Kim, J., Abbas, Z., & Lee, S. (2025). Heartensemblenet: An innovative hybrid ensemble learning approach for cardiovascular risk prediction. *Healthcare, 13*(5), 507. https://doi.org/10.3390/healthcare13050507

Zeng, Y., Liu, D., Guo, S., Xiong, L., Liu, P., Yin, J., & Wu, Z. (2022). A system dynamic model to quantify the impacts of water resources allocation on water–energy–food–society (WEFS) nexus. *Hydrology and Earth System Sciences, 26*(15), 3965–3988. https://doi.org/10.5194/hess-26-3965-2022

Zhang, Y., Petrosian, O., Liu, J., Ma, R., & Krinkin, K. (2023). FI-SHAP: explanation of time series forecasting and improvement of feature engineering based on boosting algorithm. In *Proceedings of SAI intelligent systems conference* (pp. 745-758). Springer International Publishing. Cham

Zhong, S., Zhang, K., Bagheri, M., Burken, J. G., Gu, A., Li, B., & Zhang, H. (2021). Machine learning: new ideas and tools in environmental science and engineering. *Environmental Science & Technology, 55*(19), 12741–12754. https://doi.org/10.1021/acs.est.1c01339

Zhou, Z., Ren, J., He, X., & Liu, S. (2021). A comparative study of extensive machine learning models for predicting long-term monthly rainfall with an ensemble of climatic and meteorological predictors. *Hydrological Processes, 35*(11), e14424. https://doi.org/10.1002/hyp.14424

Zhu, F., Zhu, O., Han, M., Liu, W., Guo, X., Hou, T., & Zhong, P. A. (2025). A hybrid process-data driven framework for real-time hydrological forecasting with interpretable deep learning. *Journal of Hydrology, 662*, 134082. https://doi.org/10.1016/j.jhydrol.2025.134082

(Received July 22, 2025, revised January 15, 2026, accepted February 2, 2026)