Journal of Industrial Engineering and Applied Science
ISSN 3005-6071 (Print) | ISSN 3005-608X (Online) | Vol. 2, No. 2, 2024

Machine Learning Based Prediction of Water Demand in Megacities: A Case Study of Beijing

FU, Hongpeng [1*] XIANG, Sunan [2] KONG, Xiangji [3]

[1] Northeastern University, USA
[2] University of Chicago, USA
[3] Dalian Nationality University, China

**** FU, Hongpeng is the corresponding author, E-mail: fuhppku@163.com***

Abstract: With the acceleration of urbanization, smart cities are increasingly entering the public consciousness. Beijing, a city facing severe water scarcity, has seen some relief in its water supply pressure through the South-to-North Water Diversion Project. However, as economic development progresses and the population continues to expand, the demand for water in Beijing is still on the rise. Conducting a scientific and rational prediction of water demand is a prerequisite and foundation for planning and constructing future water supply projects. This paper embarks on a study of water demand prediction in Beijing, China, initially identifying 13 explanatory variables related to economics, society, water usage, and resources. Utilizing data from Beijing from 2004 to 2020, a predictive model encompassing both statistical and machine learning models for water demand was established. The findings indicate that among all the models considered, the Random Forest model performed the best, with R2 scores of 97.9% and 97.8%, respectively. A comparative analysis of the model's predictive performance further demonstrates the superiority of machine learning models over statistical models. The results of this study offer valuable insights for the planning and construction of future water supply projects in Beijing. They can serve as a reference for the formulation of water supply management policies in other cities.

Keywords: sustainability, nature-based solution, sustainable development goals.

DOI: https://doi.org/10.5281/zenodo.10791242

1 Introduction

Over the past few decades, rapid population growth and economic development have resulted in a sharp increase in water consumption, leading to water shortages in many countries [1]. The problem of imbalance between water supply and demand, influenced by factors such as population size and economic development, is intensifying [2]. Global annual water demand has reached approximately 4600 cubic kilometers, expected to increase by 30%-50% by 2050 [3]. In response to these challenges, some developed countries have begun predictive analyses on future water demand [4]. However, we are confronted with issues such as continuous population concentration, over-exploitation of water resources, and a severe shortfall in water resource carrying capacity [5]. Water resources, a vital component of natural resources, play a crucial role in societal development, significantly influencing the economic development of nations and regions [6]. Thus, sustainable water resource management is essential for protecting limited water resources and averting economic losses, making it necessary to have accurate demand forecasting for optimizing the planning, and design of water supply systems [7].

Selecting an accurate model for predicting urban water demand remains a significant challenge. Traditional water demand forecasting models often utilize statistical models, including exponential smoothing, moving averages, linear regression, and Seasonal Autoregressive Integrated Moving Averages [8-10]. However, due to the nonlinearity in water demand datasets, the accuracy of these linear models is often reduced [11].

With the emergence of artificial intelligence (AI) and machine learning technologies, algorithmic models have advanced significantly, becoming more efficient, dynamic, and robust [12] Among the nonlinear AI models are various sophisticated methods, including Artificial Neural Networks (ANN), Support Vector Machines (SVM), Deep Neural Networks (DNN), and Extreme Learning Machines (ELM) [13-15] Despite their advanced capabilities, these AI techniques can sometimes produce less than satisfactory outcomes [16]

In response to these limitations, ensemble methods have gained prominence. These techniques enhance predictive accuracy by combining multiple base models to form a superior predictive model. Prominent examples of ensemble methods include Random Forest (RF), Light Gradient Boosting Machine (LGBM), and Extreme Gradient Boosting (XGBoost). These methods represent a significant evolution in the field, offering improved performance by leveraging the strengths of individual models while mitigating their weaknesses [17-19]. Despite producing accurate results, these ensemble models have been sparingly researched in the field of water demand prediction [20]. Our study focuses on Beijing as a subject for water demand prediction research. Utilizing historical data from 2000 to 2021, we aim to identify the best water demand prediction model through machine learning and ensemble methods [21]. This research will assist policymakers in making data-driven decisions, providing a basis for the rational layout, design, and construction of subsequent water supply projects, and ensuring that the region can obtain water resources stably in the long term [22].

2 Literature Review

2.1 Predicting Water Demand Using Traditional Methods

Choosing the best predictive model in the process of urban water demand forecasting is a significant challenge. Linear or traditional models such as univariate time series have been widely used due to their clear structure and ease of implementation and interpretation [23-25]. Arbue & Villanu´a (2006) built a residential water demand model based on a linear model to evaluate the role of water price strategies in water resource management in Zaragoza, Spain. House-Peters et al. (2010) established a model for single-family residential water usage in Oregon, USA, using the least squares method, identifying key factors influencing increased water demand as population growth, climate change, and types of urban development. Kontokosta & Jain (2015) analyzed the impact of socio-economic and population characteristics on water consumption in New York at the building level using geographically weighted regression. Ashoori et al. (2016) used a multivariate linear regression model to examine the impact of residential, commercial, industrial, and government water categories on Los Angeles' water demand. Polebitsk et al. (2010) used a regression model to analyze the impact of population, weather, and economic factors on residential water consumption in Seattle, USA. Bable et al. (2007) constructed a multivariate linear regression model to predict household water use in Kathmandu Valley, Nepal, and validated the model's applicability. Despite the simplicity of linear models, water demand datasets often exhibit varying degrees of non-linearity, reducing these models' accuracy [26].

2.2 Water Demand Prediction Research Based on Machine Learning Models

Non-linear algorithms and machine learning techniques, which utilize historical data and various parameters, excel in detecting intricate non-linear patterns [27]. The adoption of Artificial Intelligence (AI) and machine learning innovations significantly augments the effectiveness, adaptability, and resilience of these algorithmic models [28-30]. The performance of machine learning algorithms is primarily influenced by four pivotal elements: the predictive model employed, its capabilities and constraints, and the selection of parameters for model inputs [31]. The spectrum of non-linear AI models includes, but is not limited to, extensively analyzed Artificial Neural Networks (ANNs), Support Vector Machines, Deep Neural Networks (DNNs), and Extreme Learning Machines (ELMs) [32-35].

Machine learning models are increasingly favored for their minimal application restrictions and pronounced robustness, demonstrating superior predictive accuracy across various domains such as urban infrastructure, credit risk assessment, energy, ecological studies, and water resource management [36].

Despite the acknowledged precision of machine learning algorithms over conventional models, there are instances where AI methodologies fall short of expectations. Ensemble methods, a novel approach, amalgamate different foundational models to forge an optimal predictive framework. These methods employ strategies like bagging, boosting, and stacking. Bagging involves the selection of dataset subsets for replacement with randomly trained models, whereas boosting assigns increased weight to data points misclassified by preceding models, diminishing the weight of accurately classified instances. This targeted focus on problematic data incrementally enhances model performance. Notable implementations of bagging and boosting include Random Forest (RF) and Gradient Boosting Machine (GBM) [37].

In the context of water demand forecasting, Parisouj et al. (2020) utilized Support Vector Regression, Artificial Neural Networks with back-propagation, and Extreme Learning Machines for predicting monthly and daily flow rates across four U.S. river basins [38]. Villarin & Rodrigez-Galiano (2019) applied Classification and Regression Trees alongside RF to develop a multivariate prediction model for water demand in Seville, Spain, demonstrating the RF model's superior predictive capabilities which facilitate a deeper understanding of water demand patterns [39]. Conversely, stacking methods, exemplified by Stacked regressions (STK), consolidate outputs from multiple machine learning models into a singular model [40]. Nevertheless, these innovative models have not yet gained significant traction in water demand forecasting. Emerging ensemble model concepts such as RF, Light Gradient Boosting (LGB), and Extreme Gradient Boosting (XGB) present promising avenues, though their application in water demand forecasting remains relatively unexplored [40].

2.3 Factors Influencing Water Demand

Understanding water demand is a complex challenge due to a wide variety of influencing factors, including climatic, socio-economic, and demographic elements [40]. Climatic factors like temperature, humidity, and precipitation significantly affect water use, while socio-economic aspects like income, water charges, and local economic strength play a crucial role in defining the demand. Similarly, demographic factors such as population characteristics and household composition influence water consumption patterns [41].

Numerous studies have been conducted to better understand and predict urban water demand, incorporating these diverse factors. For instance, models developed by Lu et al. and Li et al. revealed high correlations of water demand with local population and average climate conditions, as well as GDP, respectively. Other research has expanded the model's features, including education level, seasonal changes, and total water supply from utilities, providing more detailed and robust predictions [42-43]. Recent studies started to utilize advanced techniques such as BP neural networks and other machine learning models to predict water demand, considering additional factors including industrial added value, effective irrigation area, and environmental water volume, among others, enhancing their accuracy and comprehensiveness. However, these studies rarely used integrated models, and the selected indicators are not comprehensive enough for predicting water demand.

3. Materials and Methods

3.1 Study Area

Beijing, in 2020, had 61.3 billion cubic meters of water resources, or 0.21% of China's total. The per capita water resource was just 117.8 cubic meters/person, far below the recognized extreme water shortage standard of 500 cubic meters/person. Despite carrying 8% of China's GDP and population, Beijing's water resources account for less than 1% of the country's total, marking it as a severely water-scarce region. Rapid urbanization and population growth, coupled with rising living standards, are straining the existing water infrastructure. The urgent challenge is to accurately forecast water demand and strategically plan water supply infrastructure improvements to alleviate water scarcity.

3.2 Data collection and processing

Table 1. Indicators for machine learning and data resources

Title	Unit
Total Water Supply	Billion Cubic Meters
GDP Growth Rate	%
Per Capita GDP (USD)	USD/Person
Precipitation	Millimeters
Total Water Resources	Billion Cubic Meters
Urban Population	Ten Thousand People
Rural Population	Ten Thousand People
Residents' Per Capita Disposable Income	Yuan/Person
anization Rate of Permanent Residents	%
Primary Industry-Added Value	Billion Yuan
Secondary Industry-Added Value	Billion Yuan
Tertiary Industry-Added Value	Billion Yuan

Before training predictive models, it's critical to preprocess collected data due to variable ranges. Large differences between maximum and minimum values could lead to prediction inaccuracies. This process involves feature scaling to ensure model comparability and accuracy. While some algorithms like SVM and ANN are sensitive to feature scaling, all models are subject to this step for consistency. Among common scaling methods like normalization and standardization, the former is chosen for this study, as it scales data to fall between 0 and 1 and is suitable for various machine learning algorithms.

3.3 Model Selection

Upon summarizing a broad spectrum of both domestic and international literature, we have identified three types of models to carry out our study. These include the traditional statistical analysis method of Linear Regression, the Support Vector Machine (SVM) model, which is a single learning machine model, and an ensemble learning machine model known as the Random Forest (RF).

Linear Regression serves as a tool that portrays the relationship between the independent variable x and the dependent variable y, captured through a linear function. It employs a linear model that aims to minimize the sum of squared residuals between the actual and estimated values of the dependent variable. The relationship between the various features and the predicted object is understood via the ordinary least squares method, thereby determining the linear regression equation.

The Support Vector Machine (SVM) model is marked by its unique ability to create a boundary with the maximum margin. It manages this by using a separating hyperplane to categorize a dataset, ensuring maximum margin. Through a kernel function (φ), SVM transforms the original training set into a high-dimensional feature space, which unveils the relationship between x and y. The nonlinear separable features are substituted by high-dimensional linear discriminant functions [44].

Lastly, the Random Forest model, as proposed by Breiman [45], incorporates the Bagging algorithm for feature selection. It extracts samples from the original dataset using bootstrap sampling, resulting in individual decision trees. The model applies out-of-bag sampling to select a feature subset randomly for training and repeats this process to build multiple decision trees. The final predictive result for each new test sample is achieved by synthesizing the predictions from these multiple decision trees, following the principle of majority rules.

3.4 Model Training and Performance Evaluation

While model training can confirm the model's accuracy on the training set, there is also a risk of overfitting in a well-trained model. This is when the model exhibits high predictive precision for the training samples but fails to predict the test samples accurately. In this study, we allocated 80% of the data for training and reserved the remaining 20% as the test set. However, when we trained the model using the training set and validated the model's predictive capability using the test set, we found the model to be prone to overfitting.

To address the overfitting issue and gauge model performance more effectively, we applied 10-fold cross-validation to the training data. This involved dividing the training set into 10 distinct subsets or "folds". We trained the algorithm model 10 times and evaluated it, each time using nine folds for training and the remaining one fold for assessment. The cross-validation process is akin to a test set simulation, and it aids in selecting a model with superior performance for evaluation. From the ten scores generated by the cross-validation, we calculated an average score and standard deviation for prediction.

Evaluation metrics are crucial to determine the accuracy of each model's prediction results. By juxtaposing the performance metrics of each model, we can identify the best-performing prediction model - in this case, the most effective model for predicting Beijing's water demand. Given the absence of a universally applicable evaluation metric for all models, we decided to employ multiple evaluation metrics to assess the predictive effectiveness of our model. For this purpose, we utilized three popular performance evaluation metrics, namely the Mean Squared Error (MSE), Mean Absolute Error (MAE), and Coefficient of Determination (R-squared, R2).

4. Results

4.1 Descriptive Analysis of Water Demand in Beijing

Table 2. Descriptive Analysis of Water Demand in Beijing

Indicators	Std	Min	Median	Max
Water Demand	3.72E+01	2.44	35.12	39.20
GDP Growth Rate	3.090	1.10	8.900	14.40
Per Capita GDP (USD)	7707.47	3022.00	12460.50	28517.00
Precipitation	122.75	318.00	483.70	733.20
Total Water Resources	9.87	16.90	24.18	61.30
Residents' Per Capita Disposable Income	21209.91	9229.53	31201.83	75002.20
Urbanization Rate of Permanent Residents	3.43	77.50	86.08	87.55
Urban Population	326.51	1057.40	1715.50	1916.40
Rural Population	16.71	250.80	283.85	306.20
Primary Industry-Added Value	25.43	79.30	112.89	159.80
Secondary Industry-Added Value	1774.86	1023.70	3398.20	7389.00
Tertiary Industry-Added Value	10040.31	2174.90	12549.55	33545.20

From the trend of water demand in Beijing from 2004 to 2020, it can be seen that the demand for water in Beijing has been rising with the increase in years. This is due to the difference in water use structure in Beijing. To alleviate the non-capital functions of Beijing, some manufacturing industries in Beijing have moved to other areas, which has led to a gradual reduction in industrial water use in Beijing. However, the population of Beijing is continuously growing, and the increase in water use for daily life and ecology far exceeds the reduction in industrial water use, leading to an upward trend in water demand in Beijing. In addition, the water demand in Beijing in 2020 shows a downward trend compared to 2019, which may be related to the COVID-19 pandemic that occurred in 2020. Affected by the pandemic, some factories stopped production, leading to a decrease in industrial water use, and thus reducing the annual water demand. The overall situation in Beijing, generally presents a trend of first rising and then leveling off, and the rate of increase is relatively small, and the numerical fluctuations are not large, indicating that the water demand in Beijing has remained stable in recent years, which is beneficial for the prediction of water demand in this region.

4.2 The results of machine learning models for water demand in Beijing

Table 3. Machine learning models for water demand in Beijing

Indicators	Linear Regression		SVM		Random Forest	
	Train Set	Test Set	Train Set	Test Set	Train Set	Test Set
R^2	0.806	0.739	0.893	0.842	0.979	0.978

MAE	0.081	0.092	0.082	0.072	0.029	0.030
MSE	0.013	0.031	0.008	0.007	0.002	0.001
RMSE	0.113	0.176	0.092	0.083	0.041	0.031

In both the training and testing phases, all three models achieved high R^2 scores, with the Random Forest model outperforming the other two. The Random Forest model achieved the highest R^2 score and the lowest MAE, MSE, and RMSE scores. This suggests that the Random Forest model provides the most accurate predictions for water demand in Beijing in the given scenarios. Notably, while the Support Vector Machine model had slightly lower scores than the Random Forest model, it still outperformed the Regression Analysis model, suggesting it also provides reasonably accurate predictions. The Regression Analysis model, while still effective, had lower scores than the other two models.

5 Discussion

Beijing faces a future water resource gap between supply and demand. In addition to relying on water transfer projects like the South-North Water Transfer Project to address water shortages, it is also essential to manage water resources internally through engineering planning and construction to enhance the utilization rate of freshwater resources. Currently, China is vigorously promoting the pilot construction of sponge cities to collect and utilize rainwater resources effectively. A sponge city is designed to absorb, store, infiltrate, and purify water during rainy periods and release the stored water for use when needed, demonstrating good "resilience" in adapting to environmental changes and responding to natural disasters. The construction of sponge cities mainly involves engineering techniques such as infiltration, storage, purification, use, and discharge of rainwater, thereby improving the utilization rate of water resources. During the rainy season, the abundance of surface water in cities can be infiltrated and purified through "sponge bodies" such as roads and green spaces, storing the rainwater. In times of drought or water scarcity, the stored rainwater can be released and utilized, thus alleviating the urban water supply and demand contradiction and meeting part of the water demand. Beijing should actively draw on the experience of sponge city pilot construction, integrating sponge city construction with water-saving projects to promote the development of green and water-saving cities. In 2016, Beijing was selected as one of the second batch of sponge city pilot cities. Tongzhou District in Beijing has developed its unique sponge city construction model through five years of practice and exploration, planning that 80% of the built-up area in Tongzhou will meet sponge city construction requirements by 2030. Tianjin has designated areas such as Jiefang South Road and Sino-Singapore Tianjin Eco-City as pilot zones, planning to construct 15 sponge city demonstration areas. These pilot areas have been completed, bringing social and economic benefits to the region.

Facing the challenge of water scarcity in Beijing, relying solely on external water transfer to meet internal demand is not a sustainable solution. It remains necessary to combine internal water-saving measures with national or regional policies. Integrating water-saving projects with sponge city construction can enhance the utilization rate of rainwater resources and reduce unnecessary waste of water resources [46]. Sponge city construction focuses on the management of rainwater, differing from traditional urban construction models. Traditional rainwater management systems in China primarily aim to discharge rainwater through pipelines without considering its recycling, leading to water resource waste. In contrast, sponge city construction uses permeable materials and green spaces to collect rainwater, avoiding flood disasters while collecting water. Through water-saving projects, rainwater resources can be applied to ecological, industrial, or agricultural water use, saving freshwater resources while enhancing the utilization rate of rainwater[47,48]. In its future development, Beijing should combine water-saving projects with sponge city construction to promote the development of green and water-saving cities.

This paper employs machine learning methods to select the optimal water demand prediction model, offering suggestions for future water supply engineering planning and construction in Beijing. These recommendations could serve as valuable references for formulating water supply management policies in Beijing. However, this study has some limitations that need to be addressed in future research. This paper only examines Beijing's total annual water demand without further segmentation into domestic, ecological environment, industrial, and agricultural water use, which warrants separate investigations to identify key factors influencing these variations. Future research could segment water use types, providing scientifically sound suggestions for different water uses. This study focuses on Beijing with a relatively small dataset of fewer than 100 samples. Future studies should consider using larger datasets covering all provinces and cities in China to predict national water demand and verify the model's predictive accuracy. This paper selects 13 factors influencing water demand based on existing research, but the consideration of factors might not be comprehensive, such as wastewater reuse volume, sewage reuse volume, and average annual temperature, which also affect annual water demand. A more comprehensive selection of influencing factors would enhance the model's predictive accuracy.

6 Conclusion

Water scarcity has become a hot issue for major cities worldwide. The accuracy of water demand forecasts is directly related to the planning and construction of water supply projects. Therefore, accurate water demand forecasting helps water departments and suppliers recognize the supply and demand gap in Beijing, facilitating the formulation of reasonable water supply policies and the execution of water supply engineering planning and construction to maintain the balance of urban water resources, preventing waste and shortage. This paper focuses on Beijing, comparing statistical models and machine learning models to identify the best water demand prediction model. The analysis of Beijing's water demand from 2004 to 2020 established three statistical and machine learning models—regression analysis, support vector machine, and random forest—based on economic, social, water use, and resource availability indicators for forecasting water demand. The models were evaluated using MSE, MAE, and R2 scores. The results indicate that among all considered models, the random forest model performed the best, with R2 scores of 97.9% and 97.8%, respectively. The comparison of model performances also demonstrates the superiority of machine learning models over statistical models in predictive accuracy.

Acknowledgments

The authors thank the editor and anonymous reviewers for their helpful comments and valuable suggestions.

Funding

Not applicable.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Author Contributions

Not applicable.

About the Authors

FU, Hongpeng

Student at Northeastern University, U.S.A.

XIANG, Sunan

Student at University of Chicago, U.S.A.

KONG, Xiangji

Student at Dalian Nationality University, China.

References

[1] Kim, J., Lee, H., Lee, M., Han, H., Kim, D., & Kim, H. S. (2022). Development of a Deep Learning-Based Prediction Model for Water Consumption at the Household Level. Water, 14(9), 1512.

[2] Fu, H. (2023). A Comprehensive Review of Nature-Based Solutions: Current status and future research. AIMS Environmental Science, 10(5).

[3] Sidhu, R. K., Kumar, R., & Rana, P. S. (2020). Machine learning based crop water demand forecasting using minimum climatological data. Multimedia Tools and Applications, 79, 13109-13124.

[4] Zang, H., & Dong, X. (2024). Optimizing Soil Health Management in Smart Agriculture: Deep Learning Algorithms for Nutrient Analysis and Fertilizer Recommendation with Precision Agriculture Systems. Journal of Industrial Engineering and Applied Science, 2(1), 1-7.

[5] Rezaali, M., Quilty, J., & Karimi, A. (2021). Probabilistic urban water demand forecasting using wavelet-based machine learning models. Journal of Hydrology, 600, 126358.

[6] Ni, F., Zang, H., & Qiao, Y. (2024, January). Smartfix: Leveraging machine learning for proactive equipment maintenance in industry 4.0. In The 2nd International scientific and practical conference "Innovations in education: prospects and challenges of today"(January 16-19, 2024) Sofia, Bulgaria. International Science Group. 2024. 389 p. (p. 313).

[7] Guo, G., Liu, S., Wu, Y., Li, J., Zhou, R., & Zhu, X. (2018). Short-term water demand forecast based on deep learning method. Journal of Water Resources Planning

and Management, 144(12), 04018076.

[8] Yan, X., Xiao, M., Wang, W., Li, Y., & Zhang, F. (2024). A Self-Guided Deep Learning Technique for MRI Image Noise Reduction. Journal of Theory and Practice of Engineering Science, 4(01), 109-117.

[9] Ahmad, T., & Chen, H. (2018). Utility companies strategy for short-term energy demand forecasting using machine learning based models. Sustainable cities and society, 39, 401-417.

[10] Chen, S., Li, K., Fu, H., Wu, Y. C., & Huang, Y. (2023). Sea Ice Extent Prediction with Machine Learning Methods and Subregional Analysis in the Arctic. Atmosphere, 14(6), 1023.

[11] Weimin, W. A. N. G., Yufeng, L. I., Xu, Y. A. N., Mingxuan, X. I. A. O., & Min, G. A. O. (2024). Enhancing Liver Segmentation: A Deep Learning Approach with EAS Feature Extraction and Multi-Scale Fusion. International Journal of Innovative Research in Computer Science & Technology, 12(1), 26-34.

[12] Nunes Carvalho, T. M., de Souza Filho, F. D. A., & Porto, V. C. (2021). Urban water demand modeling using machine learning techniques: Case study of Fortaleza, Brazil. Journal of Water Resources Planning and Management, 147(1), 05020026.

[13] Hsieh, Y. T., Anjum, K., & Pompili, D. (2022, October). Ultra-low Power Analog Recurrent Neural Network Design Approximation for Wireless Health Monitoring. In 2022 IEEE 19th International Conference on Mobile Ad Hoc and Smart Systems (MASS) (pp. 211-219). IEEE.

[14] Ibrahim, T., Omar, Y., & Maghraby, F. A. (2020, March). Water demand forecasting using machine learning and time series algorithms. In 2020 International Conference on Emerging Smart Computing and Informatics (ESCI) (pp. 325-329). IEEE.

[15] Smolak, K., Kasieczka, B., Fialkiewicz, W., Rohm, W., Siła-Nowicka, K., & Kopańczyk, K. (2020). Applying human mobility and water consumption data for short-term water demand forecasting using classical and machine learning models. Urban Water Journal, 17(1), 32-42.

[16] Wang, Q., & Wang, S. (2020). Machine learning-based water level prediction in Lake Erie. Water, 12(10), 2654.

[17] Hsieh, Y. T., Qi, Z., & Pompili, D. (2022, November). ML-based Joint Doppler Estimation and Compensation in Underwater Acoustic Communications. In Proceedings of the 16th International Conference on Underwater Networks & Systems (pp. 1-8).

[18] Xu, Z., Lv, Z., Li, J., & Shi, A. (2022). A novel approach for predicting water demand with complex patterns based on ensemble learning. Water Resources Management, 36(11), 4293-4312.

[19] Tianqi, Y. (2022). Integrated models for rocking of offshore wind turbine structures. American Journal of Interdisciplinary Research in Engineering and Sciences, 9(1), 13-24.

[20] Truong, V. H., Ly, Q. V., Le, V. C., Vu, T. B., Tran, T. T., & Goethals, P. (2021). Machine learning-based method for forecasting water levels in irrigation and drainage systems. Environmental Technology & Innovation, 23, 101762.

[21] Potočnik, P., Škerl, P., & Govekar, E. (2021). Machine-learning-based multi-step heat demand forecasting in a district heating system. Energy and Buildings, 233, 110673.

[22] Candelieri, A., Giordani, I., Archetti, F., Barkalov, K., Meyerov, I., Polovinkin, A., ... & Zolotykh, N. (2019). Tuning hyperparameters of a SVM-based water demand forecasting system through parallel global optimization. Computers & Operations Research, 106, 202-209.

[23] Guo, Y., Wang, J., Chen, H., Li, G., Liu, J., Xu, C., ... & Huang, Y. (2018). Machine learning-based thermal response time ahead energy demand prediction for building heating systems. Applied energy, 221, 16-27.

[24] Garzón, A., Kapelan, Z., Langeveld, J., & Taormina, R. (2022). Machine Learning‐Based Surrogate Modeling for Urban Water Networks: Review and Future Research Directions. Water Resources Research, 58(5), e2021WR031808.

[25] Kavya, M., Mathew, A., Shekar, P. R., & Sarwesh, P. (2023). Short term water demand forecast modelling using artificial intelligence for smart water management. Sustainable Cities and Society, 95, 104610.

[26] Deng, T., Chau, K. W., & Duan, H. F. (2021). Machine learning based marine water quality prediction for coastal hydro-environment management. Journal of Environmental Management, 284, 112051.

[27] Hanoon, S. K., Abdullah, A. F., Shafri, H. Z., & Wayayok, A. (2022). A Novel Approach Based on Machine Learning and Public Engagement to Predict Water-Scarcity Risk in Urban Areas. ISPRS International Journal of Geo-Information, 11(12), 606.

[28] Wang, L., & El-Gohary, N. M. (2019, June). Machine Learning-Based Prediction of Building Water Consumption for Improving Building Water Efficiency. In ASCE International Conference on Computing in Civil Engineering 2019 (pp. 139-145). Reston, VA: American Society of Civil Engineers.

[29] Li, X., Wang, X., Chen, X., Lu, Y., Fu, H., & Wu, Y. C. (2024). Unlabeled data selection for active learning in image classification. Scientific Reports, 14(1), 424.

[30] Wang, L., & El-Gohary, N. M. (2019, June). Machine Learning-Based Prediction of Building Water Consumption for Improving Building Water Efficiency.

In ASCE International Conference on Computing in Civil Engineering 2019 (pp. 139-145). Reston, VA: American Society of Civil Engineers.

[31] Nasaruddin, N., Zakaria, S. F., Ahmad, A., Ul-Saufie, A. Z., & Mohamaed Noor, N. (2021). Water demand prediction using machine learning: a review.

[32] Candelieri, A., Soldi, D., & Archetti, F. (2015). Layered machine learning for short-term water demand forecasting. Environmental Engineering & Management Journal (EEMJ), 14(9).

[33] Khan, P. W., Byun, Y. C., Lee, S. J., & Park, N. (2020). Machine learning based hybrid system for imputation and efficient energy demand forecasting. Energies, 13(11), 2681.

[34] Kamoona, A., Song, H., Keshavarzian, K., Levy, K., Jalili, M., Wilkinson, R., ... & Meegahapola, L. (2023). Machine learning based energy demand prediction. Energy Reports, 9, 171-176.

[35] Seo, Y., Kwon, S., & Choi, Y. (2018). Short-term water demand forecasting model combining variational mode decomposition and extreme learning machine. Hydrology, 5(4), 54.

[36] Yi, S., Kondolf, G. M., Sandoval-Solis, S., & Dale, L. (2022). Application of machine learning-based energy use forecasting for inter-basin water transfer project. Water Resources Management, 36(14), 5675-5694.

[37] Mokhtar, A., Al-Ansari, N., El-Ssawy, W., Graf, R., Aghelpour, P., He, H., ... & Abuarab, M. (2023). Prediction of Irrigation Water Requirements for Green Beans-Based Machine Learning Algorithm Models in Arid Region. Water resources management, 37(4), 1557-1580.

[38] Liao, Z., Zang, N., Wang, X., Li, C., & Liu, Q. (2021). Machine learning-based prediction of chlorophyll-a variations in receiving reservoir of world's largest water transfer project—a case study in the miyun reservoir, North China. Water, 13(17), 2406.

[39] Ji, G., Wang, J., Ge, Y., & Liu, H. (2014, May). Urban water demand forecasting by LS-SVM with tuning based on elitist teaching-learning-based optimization. In The 26th Chinese Control and Decision Conference (2014 CCDC) (pp. 3997-4002). IEEE.

[40] Ye, G., Wan, J., Deng, Z., Wang, Y., Zhu, B., Yan, Z., & Ji, S. (2024). Machine learning-based prediction of biological oxygen demand and unit electricity consumption in different-scale wastewater treatment plants. Journal of Environmental Chemical Engineering, 12(2), 111849.

[41] Ye, G., Wan, J., Deng, Z., Wang, Y., Zhu, B., Yan, Z., & Ji, S. (2024). Machine learning-based prediction of biological oxygen demand and unit electricity consumption in different-scale wastewater treatment plants. Journal of Environmental Chemical Engineering, 12(2), 111849.

[42] Liu, W., Yu, H., Yang, L., Yin, Z., Zhu, M., & Wen, X. (2021). Deep learning-based predictive framework for groundwater level forecast in arid irrigated areas. Water, 13(18), 2558.

[43] Liu, W., Yu, H., Yang, L., Yin, Z., Zhu, M., & Wen, X. (2021). Deep learning-based predictive framework for groundwater level forecast in arid irrigated areas. Water, 13(18), 2558.

[44] Houchati, M., Beitelmal, A. H., & Khraisheh, M. (2022). Predictive modeling for rooftop solar energy throughput: a machine learning-based optimization for building energy demand scheduling. *Journal of Energy Resources Technology*, *144*(1), 011302.

[45] Groppo, G. D. S., Costa, M. A., & Libânio, M. (2023). Predicting time-series for water demand in the big data environment using statistical methods, machine learning and the novel analog methodology dynamic time scan forecasting. *Water Supply*, *23*(2), 624-644.

[46] Murthy, A., Green, C., Stoleru, R., Bhunia, S., Swanson, C., & Chaspari, T. (2019, November). Machine learning-based irrigation control optimization. In *Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation* (pp. 213-222).

[47] Aggarwal, S., & Sehgal, S. (2021, August). Prediction of Water Consumption for New York city using Machine Learning. In *2021 8th International Conference on Signal Processing and Integrated Networks (SPIN)* (pp. 486-490). IEEE.

[48] Li, W., Finsa, M. M., Laskey, K. B., Houser, P., & Douglas-Bate, R. (2023). Groundwater level prediction with machine learning to support sustainable irrigation in water scarcity regions. Water, 15(19), 3473.