ISSN 2959-6157

Water-use Efficiency with Machine Learning: A Global Perspective

Xizhe Shen[1, *], Muchen Wu[2] and Yanbang Zhou[3]

[1] St. Mary's Wuhan, Wuhan, China
[2]Wuhan Gangcheng No. 4 Middle School, Wuhan, China
[3]Jiangsu Jiangyin No.1 Middle School, Jiangsu, China
*Corresponding author: shenxizhe12138@gmail.com

Abstract:

Water scarcity is a pressing social, economic, and environmental issue with significant corporate sustainability and planetary health implications. Therefore, improving the efficiency of water resources use is an important research issue. With the rapid development of machine learning technology, relying on the learning ability of learning algorithms and the support of computer computing power, it is possible to optimize the global allocation of water resources through machine learning technology. This paper investigates the progress made by machine learning methods in water use efficiency. Specifically, this paper introduces the basic principle of the machine learning method for evaluating water resources efficiency, including classification and regression tasks. Then, several mainstream evaluation indexes of water resource utilization efficiency are introduced. Finally, this paper analyzes the main challenges and future development direction of machine learning to improve the efficiency of water resources utilization. This paper aims to provide a complete research view for researchers using machine learning methods to improve water efficiency.

Keywords: Water use efficiency; Machine learning; Deep learning; Artificial Intelligence.

1. Introduction

Water is an indispensable resource for all life forms and underpins societal and economic development. Despite covering two-thirds of the Earth's surface, freshwater is not uniformly distributed, and human activities and climate change threaten its availability. Water scarcity affects approximately 80% of the global population and is expected to worsen, potentially displacing 700 million people by 2030. Corporate facilities, particularly those in the agriculture, energy, and manufacturing sectors, are significant water consumers. Their water management practices can either exacerbate or alleviate water scarcity issues. As such, corporations must assess and improve their water-use efficiency (WUE), especially in areas where water is scarce [1].

The United Nations has recognized the importance of water and sanitation, dedicating one of the 17 SDGs to "Clean Water and Sanitation". This goal emphasizes the need for sustainable water management and reducing water scarcity. Several metrics have been proposed to monitor and manage water demand and supply. These include water withdrawal, discharge, and consumption, essential for understanding a facility's water usage. However, these metrics do not consider the geographical context of water stress. Water stress is the ratio of freshwater withdrawal to the available freshwater resources. It is a critical indicator of the balance between water supply and demand. High water stress levels can lead to environmental degradation and social issues. Figure 1 shows that the Regional and global changes in water stress from 2008 to 2018 (The data from United Nations World Water Development Report 2023) [2].

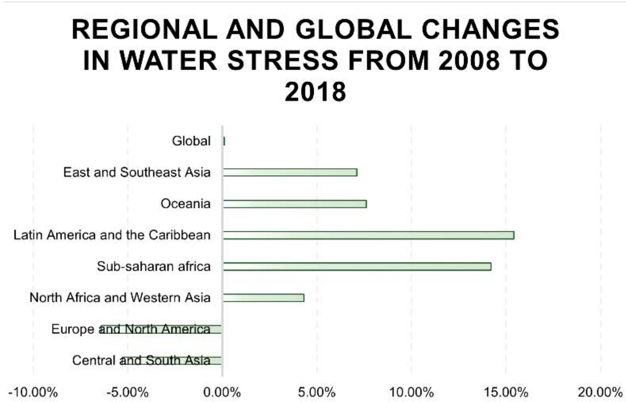

Fig. 1 Regional and global changes in water stress from 2008 to 2018

Pervious works suggested that context-oriented metrics can provide insights that may not be apparent from traditional metrics. These metrics consider water usage's output, outcome, and impact, offering a more holistic view of water management. The World Economic Forum (WEF) proposes combining water quantity and geographical metrics to monitor water consumption and withdrawal in water-stressed areas. This approach allows evaluating water use that may negatively impact other users or industries [3].

With the development of artificial intelligence technology, it plays a vital role in improving the efficiency of water resources utilization. Artificial intelligence (AI) technology can comprehensively analyze climate, hydrology, and usage data and accurately predict the supply and demand of water resources. By collecting a large number of historical data and using a machine learning algorithm for in-depth analysis, the AI system can find the relationship between water consumption and various factors (such as weather, holidays, population movement, etc.) to predict water consumption in the future and help water departments allocate resources and make plans more reasonably. Based on the supply and demand forecast results, AI can optimize the allocation strategy of water resources to ensure the maximum utilization of water resources while meeting different water demands.

In practical application, machine learning algorithms must make clear the index of water resource utilization efficiency. Firstly, this paper introduces the basic principle of the machine learning method for water resources efficiency evaluation, including classification tasks and regression tasks. Then, several mainstream evaluation indexes of water resource utilization efficiency are introduced. Finally, this paper analyzes the main challenges and future development direction of machine learning to improve the efficiency of water resources utilization.

2. The Water Use Efficiency with Machine Learning

2.1 Classification Model

In water resources prediction, machine learning classification algorithms commonly include 1) decision tree. Classify by building a tree-like structure. Each node represents a feature, each branch represents a feature value, and leaf nodes represent a classification result. Decision trees are easy to understand and explain but easy to overfit. 2) Random forest. An ensemble learning method classifies by constructing multiple decision trees and voting. Random forest can reduce the risk of over-fitting and is highly efficient when dealing with large data sets. 3) Support Vector Machine (SVM). By mapping the samples into a high-dimensional space, we can find a hyperplane to maximize the interval between classes for classification. SVM performs well in dealing with high-dimensional data and nonlinear problems. 4) Gradient lifting tree (such as XGBoost). An ensemble learning method gradually improves the classification performance by iteratively training multiple weak classifiers and adjusting the sample weights according to the prediction results of the previous round. Gradient Boosted Decision Tree(GBDT), such as XGBoost, is favoured for its high efficiency and accuracy in water resources prediction. Machine learning classification plays a vital role in water resources prediction. By choosing suitable machine learning algorithms and tuning parameters, accurate prediction and effective management of water resources can be realized: 1) water quality prediction. Using machine learning algorithms to classify and predict water quality data can evaluate the future changing trends and pollution risk of water quality. This helps find water quality problems in time and take corresponding control measures. 2) Water quantity forecast. The future range of water quantity change can be predicted by classifying and forecasting hydrological data through a machine learning algorithm. This is significant to the rational allocation and scheduling of water resources. 3) Prediction of groundwater level. Using machine learning algorithms to classify and predict groundwater level data can evaluate the future fluctuation trend of groundwater levels. This is helpful to make a reasonable groundwater exploitation plan and protect groundwater resources [4].

2.2 Regression Model

The regression model is a statistical method used to study the quantitative relationship between independent variables (explanatory variables) and dependent variables (response variables). In water resources prediction, independent variables may include rainfall, temperature, population, economic development level, and other factors, while dependent variables may include water quantity, water quality index, or groundwater level. The mathematical expression between independent and dependent variables, the regression equation, can be established to predict the future water resources situation through regression analysis [5].

Whether a classification model or a regression model, accurate indicators (independent and dependent variables) are essential factors affecting the model's performance in machine learning. In addition, a noteworthy fact is that the rapid development of machine learning has shown good potential in improving water resource utilisation efficiency. However, the energy consumption of large-scale neural network models has attracted wide attention. Scope-1 water consumption refers to AI servers' heat dissipation, requiring vast amounts of clean water for cooling towers and outside air cooling. Liquid cooling within servers may also transfer heat to facility cooling systems, which ultimately consume water. Scope 2 covers water used offsite for electricity generation, including cooling in thermal, nuclear, and hydropower plants. Both scopes are operational water consumption. Additionally, Scope-3 encompasses embodied water in AI supply chains, such as microchip production and significant language model training/inference, with GPT-4 likely consuming more water than GPT-3 due to its larger size, as shown in Figure 2.

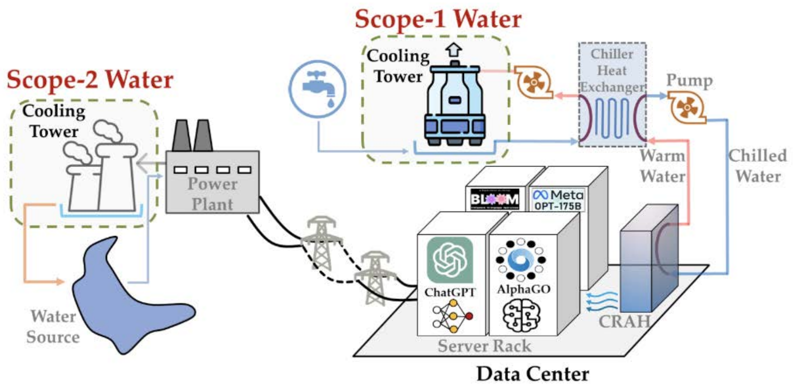

Fig.2 An example of a data center's operational water usage [6]

3. Indicators of Water Use Efficiency

Physiological WUE indicators. The researchers calculated the water use efficiency at the plant physiological level. They linked agro-ecosystem's water budget and water use efficiency with plants' fundamental physiological characteristics related to carbon assimilation and water loss. Understanding the mechanisms of photosynthesis and transpiration parameters on the whole plant WUE can be enhanced by expanding from the leaves to the entire plant. For instance, a deeper comprehension of the relationship between leaf physiology and whole plant attributes would be beneficial for crop management and plant breeding, as it would aid in cross-scale variety testing. Each of the several components that make up physiological WUE is a complex trait with minimal effect on WUE. It is vital to clarify each component's properties and how they affect WUE in order to enhance physiological WUE. The second difficulty is that it is harder to recognize, characterize, and quantify the properties of WUE components due to measurement and instrument constraints.

Biomass/grain yield WUE indicators. In agricultural production, WUE is often determined by unit water input, water consumption, or aboveground biomass or grain output. Plant responses are integrated over a longer time scale in biomass/grain yield WUE, in contrast to leaf-level plant physiological WUE, which displays a moment in time (seconds to minutes). These WUE indicators are essential variables to assess agricultural productivity from the standpoints of grain production and economy. They can cover a variety of spatial (such as plots, fields, and small watersheds) and temporal (such as seasons to years) scales, and they are particularly helpful in comparing similar variables (such as grain yield and grain yield). It can be difficult to compare WUE between various agricultural environments and plant measurement factors, even though metrics like grain yield or ANPP aid in the comparison of WUE between comparable agricultural production variables. Including non-cash crops, like cover crops, in the calculation of biomass/grain yield WUE presents another difficulty because they have an impact on consumptive water usage and storage. Furthermore, the fallow season needs to be taken into account in regions that get rain, as it provides soil moisture for income crops.

Ecosystem WUE indicators. The amount of water and carbon transported between the land surface and atmosphere is reflected in WUE at the patch or ecosystem size, which is often determined using the micrometeorology method. The ability to measure WUE almost instantly every half hour and characterize its diurnal cycle is one of the key benefits of utilizing the eddy covariance technique. The measurements from the continuous eddy covariance technique can also be combined to find the ecosystem WUE over a range of time periods (from hourly to annual). Measurements using the long-term, continuous eddy covariance approach yield useful, directly comparable data sets that may be used to assess ecosystem WUE across global areas, production systems, soil, climate, land use, management, and disturbance gradients. Only the land area contributing to the measured flux—which can range from less than 100 meters to several kilometers, depending on a number of factors—is evaluated by the horizontal carbon absorption (NEP or GPP) and water loss (ET or t) measured or obtained by eddy covariance (for example, tower height, weather conditions and canopy characteristics). The coverage of eddy covariance is limited by the expensive cost of the equipment, intricate logistics needs, and challenge of understanding the flux in complicated diverse terrain. Rather of being the directly measured values of the canopy functional flux of GPP and T, the measured values of NEP and ET of the eddy covariance system indicate the comprehensive flux above the canopy. Conversely, ER and evaporation (e) are confounded with NEP and ET, respectively.

Landscape-to-global scale WUE indicators. To make informed decisions, land managers, researchers, and others need accurate estimates of WUE at the local, regional, national, and international levels. On the other hand, a lot of WUE measures are made on a much smaller scale, including physiology, grain yield, and eddy covariance. On a larger scale, they frequently fail to convey the underlying diversity. To mechanically relate the smaller-scale WUE indicators to a wider spatial range, satellite remote sensing (such Landsat, Sentinel, and MODIS) can be utilized at different spatial and temporal resolutions on a global scale. When estimating WUE using remote sensing, denominators such as soil moisture, irrigation, ET or T, grid precipitation, and vapour pressure difference can be used. WUE, which is dependent on remote sensing, is limited in a few ways. Many aspects impact the performance, usability, and application of all remote sensing data products (e.g., source, technique, ground resolution distance or cell granularity, revisit frequency). The calculation of remote sensing biomass, GPP, and ET is imprecise due to the simplified algorithm and parameter estimation. Hydrological variables (denominator) are typically obtained by interpolation or process-based models, hence it might be difficult to generate high-resolution estimates with minimal uncertainty. Furthermore, the widely used spectral reflectance measurement, or “greenness,” highlights the need to improve the direct measurement of physiological activities like solar-induced fluorescence, as it represents the photosynthetic potential rather than actual photosynthesis.

Water use indicators across supply chains. The WUE indices in the agricultural ecosystem that have been discussed thus far are computed at small scales (like leaves and fields) or large scales (like basins and the entire world). They typically concentrate on plants and take into account factors related to plant production (like photosynthesis, grain yield, and biomass) or carbon absorption. However, there is a long and intricate supply chain involved in getting agricultural products from the field to the table, which requires a variety of water resources and raises the overall water consumption of particular items. Water footprint, water productivity, and life cycle assessment are the three methods available for doing a thorough assessment of the water consumption of the entire supply chain. The WUE indices in the agricultural environment that have been covered so far can be computed at large or tiny sizes, such as fields and leaves, or even the entire world. Usually focusing on plants, they investigate aspects of carbon absorption or plant production (such as photosynthesis, biomass, and grain yield). But transporting agricultural products from the field to the table involves a convoluted and lengthy supply chain that increases the total water consumption of specific products and calls for a range of water resources. Three approaches are available to undertake a comprehensive assessment of the water consumption of the complete supply chain: water productivity, water footprint, and life cycle assessment.

4. Discussion

Data Quality and Missing Data. The quality of data in machine learning for agricultural water management directly determines the model’s accuracy. Especially for real-time monitoring, sensor failures, errors in data transmission, and other environmental interference may lead to missing or poor data, which further affects the model training and prediction performance. This remains a challenge. The ELM model performs highly in data with missing values. It could fill the gap between all the meteorological data and field observations to improve the model forecasting. Future research is further conducted to optimize the development of new data pre-processing techniques and imputation algorithms using a more comprehensive set of environmental variables to enhance the robustness of models.

Model Complexity and Computational Resources. Although more complex models, such as deep learning models, could give higher accuracy in prediction, they demand a considerable computational budget, creating challenges for resource-constrained agricultural setups like small farms. One possible way is the use of a layered model structure. Much more straightforward and less computationally expensive models like k-nearest neighbours (kNN) can be used for the primary filtering. Later on, more complex models, such as Extreme Learning Machine (ELM), can be used in further processing. This will help us move closer to high prediction accuracy with less computational need. Another advantage is the exploration of platforms for edge computing or cloud computing, thus offloading computational burdens.

Environmental Heterogeneity Across Different Regions. Climate, soil, and crop types vary from region to region, making it harder to have a generalistic approach to one machine-learning model across areas. This environmental heterogeneity leads to low generalization capabilities in the model. Enhancing the input data’s diversity and sources helps provide a model with more excellent generalization capability, including microclimate data, soil characteristics, and growth stage information regarding the crop. It would be possible to design adaptive structures of the models where the parameters are altered dynamically to the given regional conditions.

Models Practical Application and User Adoption: Complex machine learning models are too complicated for practical application by a farmer or agricultural manager without a technical background. Low user adoption and lack of intuitive application interfaces may lead to poor adoption. Easy-to-use Decision Support Systems or mobile applications to translate the technical, model-generated outputs into plain recommendations for operations. The farmers can easily access irrigation recommendations and decision support through intuitive interfaces without necessarily knowing complex algorithms.

Although machine learning techniques promise to be beneficial in increasing the efficiency of water resources in agriculture, many challenges still need to be addressed to improve data quality, meeting the high demand for computational resources, environmental heterogeneity, and lack of user adoption. Advanced data imputation techniques layered model structures, diversity in the input variable pool, and user-friendly application systems help the intelligent and efficient management of agricultural water resources. Careful planning of these applications can help prevent trouble in service.

5. Conclusion

Water scarcity is a formidable challenge that transcends traditional boundaries, impacting societies, economies, and the delicate balance of our natural environment, with far-reaching consequences for corporate sustainability efforts and the overall health of our planet. As populations continue to grow and climate change exacerbates water availability issues, enhancing the efficient utilization of this precious resource has emerged as a paramount research priority. The advent of machine learning technology, fueled by its remarkable learning capabilities and the exponential growth in computing power, offers an innovative avenue for addressing this multifaceted problem. By harnessing the power of machine learning algorithms, researchers can embark on a journey to optimize the global allocation and management of water resources. These algorithms, trained on vast amounts of data, can identify patterns and trends otherwise imperceptible to humans, enabling us to make data-driven decisions that maximize water efficiency. This paper delves into the exciting advancements made by machine learning in this realm, showcasing how it is revolutionizing our approach to water use efficiency. At the core of this exploration, this paper presents the fundamental principles of machine learning as applied to evaluating water resource efficiency. This encompasses techniques such as classification, where algorithms categorize water usage patterns into distinct groups, and regression, which predicts future water demand or consumption based on historical trends and other variables. By understanding these foundational concepts, readers gain insight into the methodological backbone of this research. Furthermore, this paper introduces several key performance indicators (KPIs) and evaluation indexes commonly employed to measure the effectiveness of water resource utilization. These metrics, including but not limited to water productivity, water use efficiency ratios, and environmental impact assessments, provide a comprehensive framework for assessing the success of machine learning-driven interventions.

Authors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

[1] Francisco Suárez, Claudio Latorre, Magdalena Mendoza, et al. Using foliar δ13C from high-Andean plants (Silala River basin) as a measure of potential evapotranspiration through water use efficiency. Wiley Interdisciplinary Reviews: Water, 2024, 11(1): e1707.

[2] Van Dijk, A. I. J. M., Beck, H. E., et al. Global water monitor 2023, summary report. 2024.

[3] Jie Liu, Hui Li, Zhenyu Yuan, et al. Effects of microbial fertilizer and irrigation amount on growth, physiology and water use efficiency of tomato in greenhouse. Scientia Horticulturae, 2024, 323: 112553.

[4] Haiyang Shi, Yu Zhang, Geping Luo, et al. Machine learning-based investigation of forest evapotranspiration, net ecosystem productivity, water use efficiency and their climate controls at meteorological station level. Journal of Hydrology, 2024: 131811.

[5] Agus Kristian, Thomas Sumarsan Goh S, Ahmad Ramadan, et al. Application of ai in optimizing energy and resource management: Effectiveness of deep learning models. International Transactions on Artificial Intelligence, 2024, 2(2): 99-105.

[6] Pengfei Li, Jianyi Yang, Mohammad A. Islam, et al. Making ai less“ thirsty“: Uncovering and addressing the secret water footprint of ai models. arXiv preprint arXiv:2304.03271, 2023.