Literature Review: Predicting Access to Clean Water Using Machine Learning

1)Introduction

Access to clean water is a critical global challenge, with 2 billion people lacking safely managed drinking water services. This issue is particularly acute in Sub-Saharan Africa, where water scarcity and inadequate infrastructure exacerbate health risks and hinder socioeconomic development. As highlighted by the World Economic Forum, Sub-Saharan Africa faces a significant water accessibility divide, further complicating sustainable development goals. Machine learning (ML) offers promising solutions for predicting water scarcity by analyzing complex interactions between environmental, socioeconomic, and infrastructure factors. This review of existing literature examines the evolution of water scarcity metrics, evaluates the effectiveness of ML models, and identifies research gaps to guide the development of predictive models.

2)Organization

2.1 Water Scarcity Metrics

Water scarcity metrics have evolved significantly over time:

- Early indicators like the Water Stress Index (WSI) and Withdrawal-to-Availability ratio (WTA) primarily focused on physical water availability but failed to account for temporal variability and groundwater resources [1][2].
- Recent approaches incorporate broader factors such as green water (soil moisture), environmental flows, and virtual water trade. For example, the water footprint-based assessment developed by Hoekstra et al. integrates monthly variability and ecological thresholds to provide a more nuanced understanding of water scarcity [2].

Key research gaps include:

- ✓ Limited integration of climate change impacts.
- ✓ Insufficient consideration of sociocultural and governance factors [1].

2.2 Machine Learning Models

Machine learning models have been applied to predict both water quality and scarcity:

Random Forest: Demonstrated superior performance in predicting water quality with an accuracy of 70.4%, leveraging ensemble learning to handle non-linear relationships and noisy data [3].

Deep Neural Networks (DNNs): Found to be the most effective model for predicting water scarcity based on historical data, outperforming traditional algorithms like Support Vector Machines (SVM) and logistic regression [4].

Gradient Boosting Machines (GBM): Achieved high accuracy in urban water demand forecasting, particularly during disruptions like COVID-19[3].

Hybrid Models: Optimization techniques such as Particle Swarm Optimization (PSO) have been combined with neural networks to enhance prediction accuracy [4].

These models excel in handling large datasets, capturing complex patterns, and providing actionable insights.

2.3 Environmental and Socioeconomic Factors

Water access is influenced by a combination of environmental, demographic, and socioeconomic factors:

Environmental Factors: Climate variability, groundwater depletion, and seasonal changes significantly impact water availability [5].

Socioeconomic Factors: Population growth, urbanization, education levels, and income disparities play critical roles in determining access to clean water. For instance, a study in urban Kenya found that higher education levels significantly increased the likelihood of accessing clean water [6].

2.4 AI and Efficient Water Quality Prediction AI's Role:

The application of Artificial Intelligence (AI) in water quality prediction has gained popularity for its accurate and rapid detection of water contaminants. AI-based models can efficiently analyze a multitude of water quality parameters and provide timely interventions, enhancing water standards across various settings [9].

Classification Algorithms: Machine Learning classification algorithms provide accurate and efficient methods for predicting and analyzing water quality parameters.

Novel ML Model: A novel Machine Learning approach represents a breakthrough in water quality prediction with high accuracy. This enhances the efficiency of water quality prediction, enabling swift and reliable assessments for proactive interventions and resource allocation.

2.5 Policy Integration

Case studies from various countries highlight the importance of integrating ML predictions with policy frameworks:

Egypt’s National Water Resources Plan emphasizes decentralized management systems and participatory approaches to improve urban water access [5].

Investment in educational campaigns and infrastructure development has been recommended as a strategy to enhance access in regions like urban Kenya [6].

3)Summary and Synthesis

Water Scarcity Metrics Early metrics like the Water Stress Index (WSI) and Withdrawal-to-Availability ratio (WTA) provided foundational frameworks for assessing scarcity but relied heavily on mean annual river runoff (MARR), which fails to capture temporal variability [8]. Recent approaches emphasize distinguishing between physical and economic water scarcity. Physical scarcity reflects over-extraction of water resources, while economic scarcity highlights inadequate infrastructure despite sufficient resources [8]. These distinctions are particularly relevant in regions like Sub-Saharan Africa.

Machine Learning Models

Several ML models have been applied to predict water scarcity:

1. Gaussian Process Regression (GPR):
   Key Findings: GPR excels in capturing non-linear trends and uncertainty in agricultural water demand forecasting ($R^2 = 0.98$) [8].
   Contribution: Effective for long-term predictions with limited data availability.

2. Gradient Boosting Machines (GBM):
   Key Findings: GBM achieved high accuracy ($R^2 = 0.88$) in urban water demand forecasting during disruptions like COVID-19[8].
   Contribution: Robust handling of temporal data variability.

3. Multilayer Perceptron (ANN):
   Key Findings: ANN provided superior short-term predictions in tourism-driven water consumption studies [8].
   Contribution: Captures rapid changes in demand patterns.

4. Hybrid Optimization Models:
   Key Findings: PSO-optimized ANN improved accuracy by 12–15% compared to standalone models [8].
   Contribution: Combines optimization techniques with neural networks for enhanced performance.

Environmental and Socioeconomic: Factors Population growth, climate change, and inadequate infrastructure are major drivers of water scarcity [7][8]. Studies highlight the importance of integrating climate indices, population growth rates, sanitation infrastructure quality, and groundwater depletion rates into predictive models6. Vulnerable groups such as women and children disproportionately bear the burden of water scarcity due to time spent fetching water, which hinders education and economic opportunities [7].

Policy Integration: Case studies from Egypt, Peru, Namibia, and South Australia demonstrate effective policies for integrated urban water management6. Recommendations include decentralized management systems, participatory approaches, wastewater treatment innovations, and public awareness campaigns. These strategies can be adapted to other regions facing similar challenges [8].

4)Conclusion

This review highlights significant advancements in using machine learning for predicting access to clean water:

- ML models like Random Forests and Deep Neural Networks are highly effective for predicting water quality and scarcity.
- Incorporating environmental and socioeconomic factors enhances model relevance for policymaking.
- Gaps remain in integrating climate change impacts and addressing data limitations.

Contribution of Research: This project aims to develop an integrated ML framework that combines advanced algorithms with comprehensive datasets on environmental, socioeconomic, and infrastructure factors. Such a framework can guide targeted investments in clean water access, particularly in vulnerable regions like Sub-Saharan Africa.

Resources

[1] Evolution and Trends of Water Scarcity Indicators: Unveiling Gaps, Challenges, and Collaborative Opportunities (2024)
http://ui.adsabs.harvard.edu/abs/2024WCSE....9....8H/abstract

[2] Water Scarcity Assessments in the Past, Present, and Future (2017)
https://pmc.ncbi.nlm.nih.gov/articles/PMC6204262/

[3] Water Quality Prediction Using Machine Learning Models (2024)
https://www.e3sconferences.org/articles/e3sconf/pdf/2024/126/e3sconf_iccmes2024_01025.pdf

[4] Water Scarcity Prediction for Global Regions Using Machine Learning (2021).
https://www.inderscienceonline.com/doi/10.1504/IJW.2020.112746

[5] Factors Affecting Water Availability (2024)
https://www.savemyexams.com/dp/environmental-systems-and-societies-ess/ib/24/sl/revision-notes/4-water/4-2-water-access-use-and-security/factors-affecting-water-availability/

[6] Effect of Socioeconomic Determinants on Access to Clean Water in Urban Kenya (2023).
https://erepository.uonbi.ac.ke/handle/11295/164632

[7] A Drop of Hope: The Vital Role of Clean Water in Developing Communities(2023)
https://idrf.ca/2023/06/14/vital-role-of-clean-water-in-developing-communities/

[8] Lessons learned from water-scarce cities: Proposed policies toward an integrated urban water management in Egypt
https://www.frontiersin.org/journals/water/articles/10.3389/frwa.2022.981261/full

[9] AI for clean water: efficient water quality prediction leveraging machine learning (2024)
https://iwaponline.com/wpt/article/19/5/1986/102177/AI-for-clean-water-efficient-water-quality

[10] Machine learning based groundwater prediction in a data-scarce basin of Ghana
https://www.tandfonline.com/doi/full/10.1080/08839514.2022.2138130#d1e343