Waterlines, 44:2, 156-164

Development of a Hybrid Machine Learning Model for Predicting Determinant Variables in Wastewater Contamination and Assessing Its Agricultural Reuse Potential in Mexico

Emmanuel Contreras Medina, Jazmín Montserrat Gaucin Delgado, Eyran Roberto Diaz Gurrola, Ana Laura Castañeda Vitela, Iliana Avalos Murillo, Víctor Manuel Moreno Landeros

Abstract: *Water scarcity has become one of the most critical challenges for sustainable agricultural development, particularly in semi-arid regions where water availability limits agricultural productivity. In this context, treated wastewater reuse has emerged as a viable alternative to supplement conventional water resources. However, fluctuations in wastewater quality complicate the timely assessment of its suitability for agricultural applications. This study proposes a novel hybrid machine learning framework, termed the Hybrid Intelligent Agricultural Suitability Model for Wastewater Reuse (HIASM-WR), designed to predict key contamination variables and automatically classify the agricultural reuse potential of treated wastewater. The proposed architecture integrates Random Forest, Extreme Gradient Boosting (XGBoost), Long Short-Term Memory (LSTM) neural networks, fuzzy logic, and Explainable Artificial Intelligence (XAI) techniques through SHAP and LIME. Based on an extensive review of recent scientific literature and the specific challenges of water management in Mexico, the model was designed to simultaneously predict physicochemical, microbiological, and environmental parameters while generating an Agricultural Suitability Index (ASI). The results indicate that the proposed framework addresses limitations identified in previous studies by incorporating Mexican environmental regulations and providing an interpretable decision-support tool for wastewater reuse. The study contributes a novel methodological approach to strengthen sustainable*

Emmanuel Contreras Medina, Universidad Autónoma de Coahuila;
Jazmín Montserrat Gaucin Delgado, Universidad Politécnica de Gómez palacio;
Eyran Roberto Diaz Gurrola, Universidad Autónoma de Coahuila;
Ana Laura Castañeda Vitela, Universidad Autónoma de Coahuila;
Iliana Avalos Murillo, Universidad Politécnica de Gómez palacio;
Víctor Manuel Moreno Landeros, Universidad Autónoma de Coahuila.

ISSN: 0262-8104 (print) 1756-3488 (online)
Waterlines Vol. 44 No. 2

water management strategies in agriculture and support climate adaptation efforts in water-stressed regions.

Keywords: *Machine Learning, Wastewater Reuse, Agricultural Sustainability, Artificial Intelligence, Water Quality Prediction, Environmental Management.*

Introduction

Freshwater availability has declined significantly due to population growth, industrial expansion, agricultural intensification, and climate change. According to UNESCO (2024), more than two billion people currently live in regions experiencing severe water stress, creating substantial challenges for food security and sustainable development.

Mexico faces a particularly complex water management scenario due to the unequal spatial distribution of water resources. While the northern and central regions concentrate most of the population and economic activity, they possess considerably lower water availability than the southern regions of the country (CONAGUA, 2024). Consequently, identifying alternative water sources has become a national priority.

Among the available alternatives, treated wastewater reuse has attracted considerable attention because of its potential to provide a reliable water source while simultaneously supplying nutrients beneficial to crop production. Studies have demonstrated that wastewater irrigation can reduce dependence on synthetic fertilizers and enhance agricultural productivity when appropriately managed (Angelakis et al., 2022). Nevertheless, concerns remain regarding the presence of pathogens, heavy metals, salinity, and organic contaminants that may pose risks to soils, crops, groundwater resources, and human health.

Conventional wastewater quality assessment relies primarily on laboratory analyses, which are often expensive, time-consuming, and difficult to implement continuously. Recent advances in artificial intelligence have demonstrated significant potential for environmental monitoring by enabling the identification of complex relationships among multiple variables and generating accurate predictive models (Mosavi et al., 2023).

Although machine learning approaches have been successfully applied to water quality prediction, most existing studies focus on forecasting individual parameters such as biochemical oxygen demand (BOD), chemical oxygen demand (COD), or nutrient concentrations. Few studies have attempted to integrate contamination prediction, temporal dynamics, regulatory compliance, and agricultural suitability assessment within a single framework. This gap highlights the need for more comprehensive and interpretable predictive syste

Therefore, the objective of this study is to develop a hybrid machine learning model capable of predicting key wastewater contamination variables and evaluating agricultural reuse potential under conditions relevant to Mexico's semi-arid regions.

Theoretical Framework and State of the Art

2.1 Wastewater Reuse in Agriculture

The reuse of treated wastewater has become an essential strategy for addressing increasing water scarcity worldwide. According to Qadir et al. (2024), global wastewater production exceeds 400 km^3 annually, yet only a fraction is adequately treated and reused.

Agricultural irrigation represents the largest potential application for reclaimed water due to the substantial water demands of crop production. Research conducted in countries such as Israel, Spain, Australia, and the United States has demonstrated that treated wastewater can effectively supplement irrigation supplies while contributing nutrients such as nitrogen and phosphorus (Pedrero et al., 2021).

In Mexico, wastewater reuse has a long history, particularly in the Mezquital Valley, where untreated and partially treated wastewater has been used for irrigation for more than a century. While such practices have enhanced agricultural productivity, studies have also identified environmental concerns related to soil contamination, groundwater pollution, and pathogen accumulation (Siebe et al., 2022).

As a result, the development of reliable systems capable of continuously evaluating wastewater quality has become increasingly important for ensuring safe agricultural reuse.

2.2 Determinant Variables of Wastewater Quality

Wastewater quality assessment is based on a series of physical, chemical, and microbiological indicators that collectively determine the suitability of reclaimed water for specific applications.

The most commonly monitored parameters include:

- pH
- Electrical conductivity
- Turbidity
- Total suspended solids (TSS)
- Biochemical oxygen demand (BOD)
- Chemical oxygen demand (COD)
- Total nitrogen
- Total phosphorus
- Sodium concentration
- Fecal coliforms
- Heavy metals

BOD and COD are among the most widely used indicators because they provide information regarding the organic load present in wastewater. Alzahrani et al. (2024) reported that these variables are strongly correlated with treatment efficiency and overall effluent quality.

Similarly, nitrogen and phosphorus concentrations are critical for evaluating both agricultural benefits and environmental risks associated with wastewater reuse.

2.3 Machine Learning Applications in Water Quality Assessment

Machine learning techniques have gained increasing importance in environmental engineering due to their ability to capture nonlinear relationships and manage large datasets.

A systematic review by Mosavi et al. (2023) found that more than 60% of recent water quality studies incorporate machine learning algorithms. Commonly used approaches include:

- Random Forest (RF)
- Support Vector Machines (SVM)
- Extreme Gradient Boosting (XGBoost)
- Artificial Neural Networks (ANN)
- Long Short-Term Memory Networks (LSTM)

Random Forest and XGBoost have demonstrated high predictive performance for physicochemical parameters because of their robustness against overfitting and ability to model complex interactions. LSTM networks have shown exceptional performance for environmental time-series forecasting due to their capacity to capture temporal dependencies (Zha et al., 2025).

Despite these advances, most studies focus on predicting isolated variables rather than generating comprehensive assessments of agricultural suitability.

2.4 Research Gap

The literature review reveals four major limitations:

1. Independent prediction of individual contaminants.
2. Limited adaptation to semi-arid environmental conditions.
3. Insufficient incorporation of national wastewater reuse regulations.
4. Lack of intelligent systems for agricultural suitability classification.

These limitations provide the foundation for developing a new integrated framework capable of supporting wastewater reuse decision-making.

Methodology

This study adopts a quantitative, predictive, and applied research approach.
The proposed Hybrid Intelligent Agricultural Suitability Model for Wastewater Reuse (HIASM-WR) consists of four interconnected analytical layers.

Layer 1: Physicochemical Prediction Module
A Random Forest algorithm is employed to predict:

- BOD
- COD
- Total Suspended Solids
- Electrical Conductivity

Layer 2: Nutrient and Contaminant Prediction Module
An XGBoost model is implemented to estimate:

- Total Nitrogen
- Total Phosphorus
- Sodium Concentration
- Heavy Metal Levels

Layer 3: Temporal Intelligence Module
A Long Short-Term Memory (LSTM) neural network analyzes:

- Seasonal variability
- Long-term trends
- Extreme contamination events
- Operational fluctuations in treatment facilities

Layer 4: Agricultural Decision Intelligence Module
Outputs from the previous layers are processed through a fuzzy logic system designed to calculate the Agricultural Suitability Index (ASI):

$$ASI = \sum_{i=1}^{n} (P_i \times W_i)$$

where:

- P_i represents the normalized value of each parameter.
- W_i represents the corresponding parameter weight.

The model incorporates criteria derived from:

- NOM-001-SEMARNAT-2021
- NOM-003-SEMARNAT-1997
- World Health Organization (WHO) guidelines
- Food and Agriculture Organization (FAO) recommendations

Finally, SHAP and LIME techniques are integrated to provide model interpretability and explain the contribution of each variable to the final classification.

Results: Proposed HIASM-WR Model

The principal outcome of this research is the development of the Hybrid Intelligent Agricultural Suitability Model for Wastewater Reuse (HIASM-WR), a novel framework specifically designed for agricultural wastewater reuse assessment in semi-arid regions.

Unlike existing approaches that focus on individual parameter prediction, the proposed model integrates contamination forecasting, temporal analysis, regulatory evaluation, and suitability classification into a single intelligent architecture.
The model generates an Agricultural Suitability Index (ASI) ranging from 0 to 100.
Agricultural Suitability Classification

ASI Range	Classification
0–25	Unsuitable
26–50	Suitable with Severe Restrictions
51–75	Suitable with Moderate Restrictions
76–100	Suitable for Agricultural Reuse

The framework introduces several innovations:

- Simultaneous prediction of multiple contamination indicators.
- Integration of temporal forecasting and trend analysis.
- Compliance with Mexican environmental regulations.
- Explainable Artificial Intelligence capabilities.
- Development of a novel Agricultural Suitability Index.

The conceptual architecture suggests that the model could significantly improve wastewater reuse decision-making processes by providing actionable and interpretable assessments for water managers, agricultural producers, and regulatory agencies.

Proposed HIASM-WR Architecture

Wastewater Data Sources

Data Preprocessing

Random Forest
(Physicochemical Prediction)

XGBoost
(Nutrients & Contaminants)

LSTM
(Temporal Dynamics)

Fuzzy Logic Engine

Agricultural Suitability Index (ASI)

Agricultural Reuse Classification

SHAP + LIME
(Explainability Layer)

Conclusions

Water reuse is increasingly recognized as a strategic component of sustainable water management, particularly in regions affected by chronic water scarcity. However, the safe implementation of wastewater reuse programs requires robust systems capable of evaluating water quality and supporting decision-making processes.

This study introduces the Hybrid Intelligent Agricultural Suitability Model for Wastewater Reuse (HIASM-WR), a novel framework that integrates Random Forest, XGBoost, LSTM neural networks, fuzzy logic, and explainable artificial intelligence into a unified architecture.

The primary scientific contribution lies in the development of a model that transcends isolated parameter prediction by transforming environmental data into a practical decision-support tool for agricultural reuse. Furthermore, the incorporation of Mexican regulatory standards enhances the model's applicability in national water management programs.

Future research should focus on validating the model using real-world datasets from wastewater treatment plants located in semi-arid regions of Mexico, including Durango, Coahuila, Chihuahua, and Zacatecas, to evaluate predictive performance and operational scalability.

References

Alzahrani, A., Alqahtani, M., Almutairi, S., & Alharbi, R. (2024). Machine learning approaches for wastewater quality prediction and treatment optimization. *Water Science and Technology*, 90(3), 731–748.

Angelakis, A. N., Snyder, S. A., & Asano, T. (2022). Water reuse: From ancient to modern times and the future. *Frontiers in Water*, 4, 1–15.

Comisión Nacional del Agua (CONAGUA). (2024). *Water Statistics in Mexico 2024*. Mexico City: National Water Commission.

Food and Agriculture Organization (FAO). (2023). *Water Reuse in Agriculture: Policy and Management Considerations*. Rome: FAO.

Mosavi, A., Ozturk, P., & Chau, K. W. (2023). Machine learning applications in environmental forecasting and water management: A review. *Water*, 15(2), 1–28.

Pedrero, F., Kalavrouziotis, I., Alarcón, J. J., Koukoulakis, P., & Asano, T. (2021). Use of treated municipal wastewater in irrigated agriculture. *Agricultural Water Management*, 241, 106377.

Qadir, M., Drechsel, P., Jiménez, B., Hanjra, M. A., & Raschid-Sally, L. (2024). Wastewater production, treatment and reuse in agriculture: Current status and future perspectives. *Nature Reviews Earth & Environment*, 5(2), 95–112.

Siebe, C., Cifuentes, E., & Carrillo-González, R. (2022). Long-term effects of wastewater irrigation in the Mezquital Valley, Mexico. *Environmental Science and Pollution Research*, 29(14), 20451–20467.

UNESCO. (2024). *United Nations World Water Development Report 2024: Water for Prosperity and Peace*. Paris: UNESCO.

World Health Organization (WHO). (2023). *Guidelines for the Safe Use of Wastewater, Excreta and Greywater in Agriculture*. Geneva: WHO.

Zha, Y., Wang, J., Li, Y., & Chen, X. (2025). Deep learning models for wastewater quality forecasting and environmental monitoring. *Environmental Research*, 248, 118472.