MACHINE LEARNING APPROACHES TO IDENTIFY HYDROCHEMICAL PROCESSES AND PREDICT DRINKING WATER QUALITY FOR GROUNDWATER ENVIRONMENT IN A METROPOLIS

AUTHORS:
ZAKIRUL MOBIN ZISAN [ID: 241-50-006]
TASNIHA TASNIM TONI [ID: 241-50-007]
SUPERVISOR:
DIPTO BISWAS, SENIOR LECTURER
DEPARTMENT OF ICE, DAFFODIL INTERNATIONAL UNIVERSITY

ARTICLEINFO

Keywords:
Groundwater environment
Machine learning approaches
Water quality index

ABSTRACT

This study evaluates groundwater quality in Chongqing using geochemical analysis and machine learning.
Two hydrochemical clusters were identified: ridge areas impacted by agriculture and valleys by sewage.
The entropy-weighted water quality index (EWQI) rated most samples as excellent or good. Linear regression achieved the best prediction accuracy ($R^2 = 0.9999$).
The results support sustainable groundwater management in urban environments.

INTRODUCTION

Groundwater is critical for urban water supply, especially in megacities like Chongqing. population growth and rapid economic development, substantial volumes of domestic and industrial wastewater are discharged, which causes different levels of groundwater pollution .ML offers a modern approach to assess and predict groundwater suitability.

PROBLEM STATEMENT

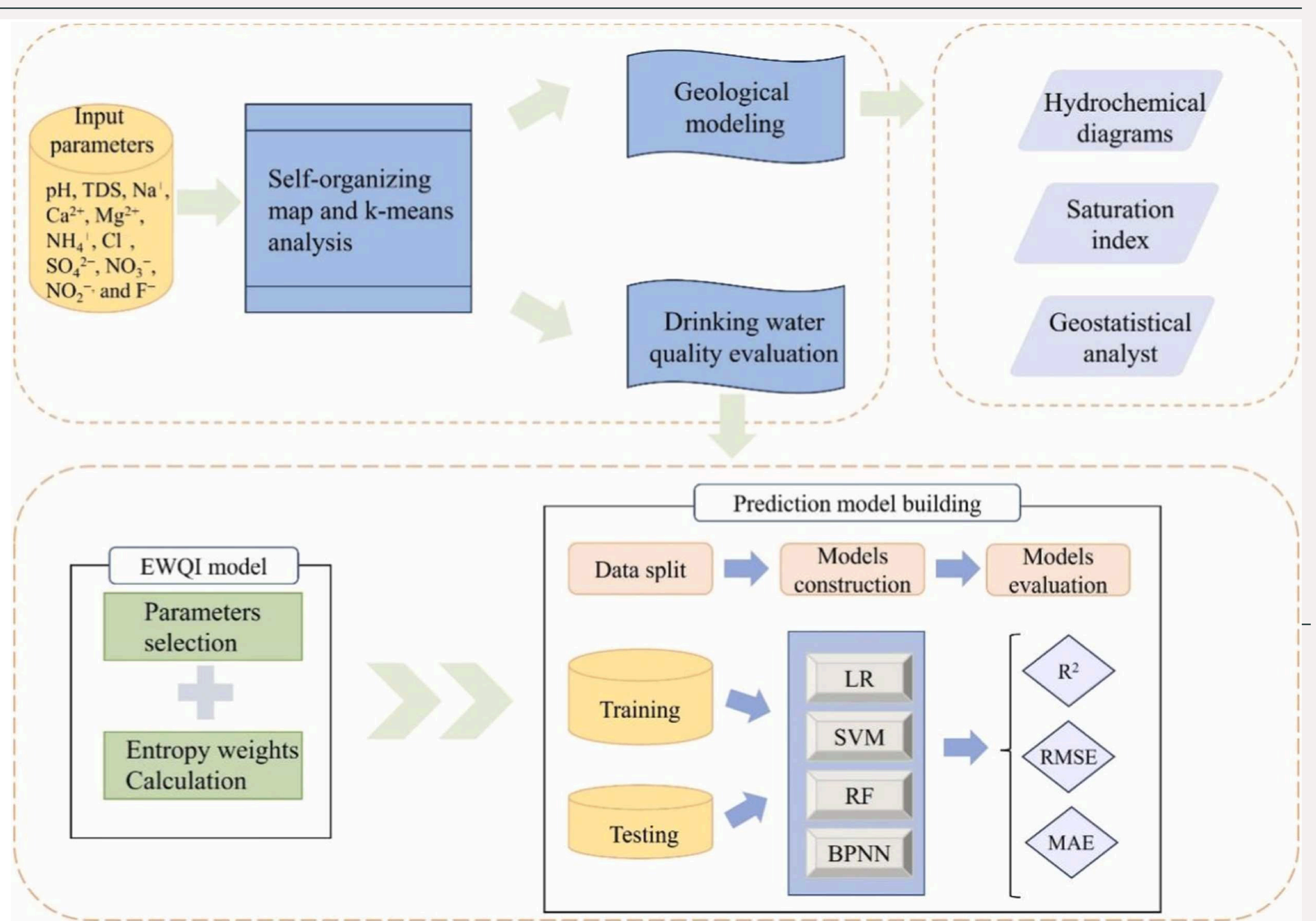
Urbanization and other human activities have impacted the quality of water in the groundwater in Chongqing, China. There are no good predictive models to utilize in assessing the potability of the groundwater. Lack of understanding of hydrochemical processes and groundwater quality prediction limits sustainable water management

OBJECTIVE

1. Identify hydrochemical processes driving groundwater composition.
2. Assess drinking water quality using Entropy-Weighted Water Quality Index (EWQI).
3. Develop machine learning (ML) models to predict water quality.

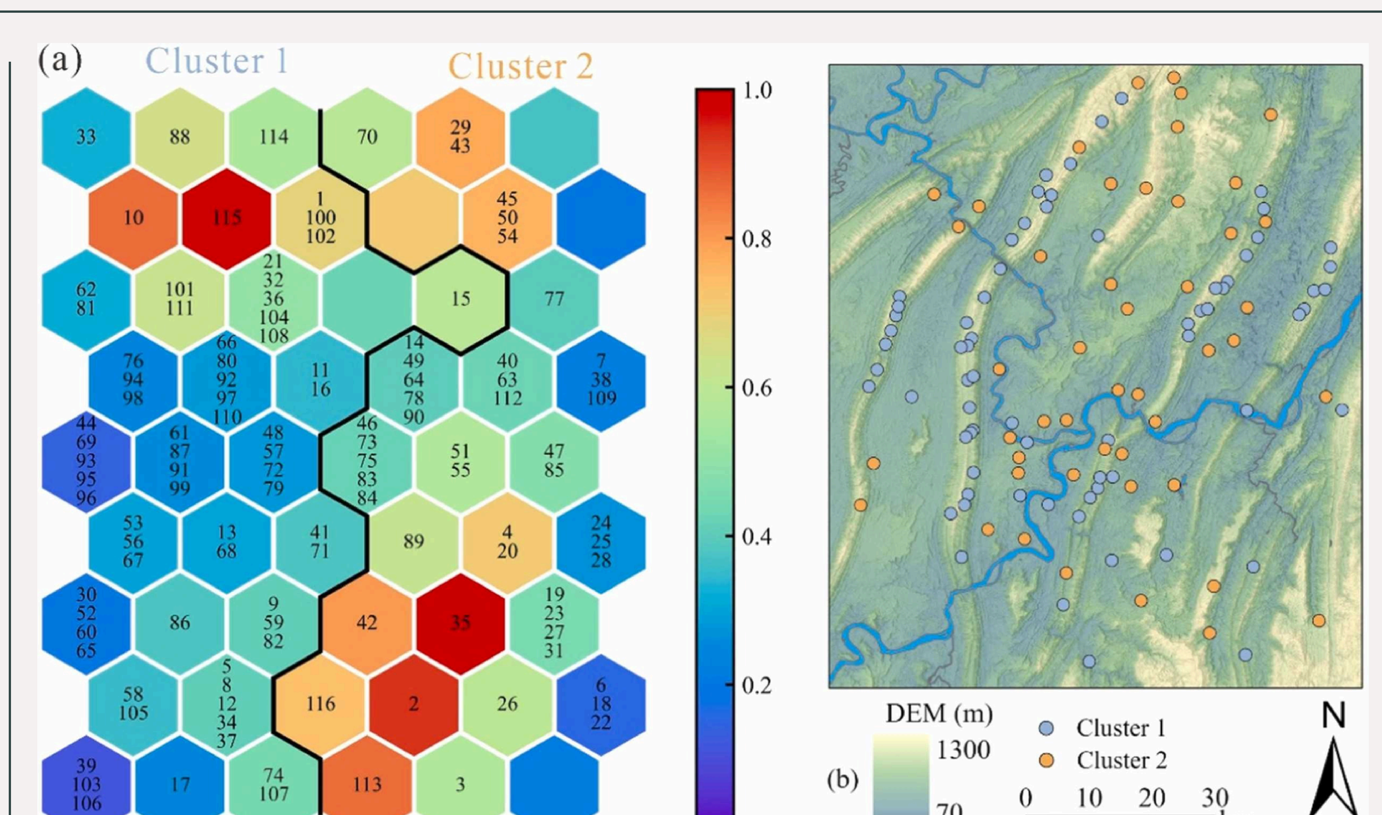
METHODOLOGY

Researcher's collected 116 groundwater samples from Chongqing's urban areas. Sampling locations were chosen to represent different geological conditions and land use types. The models used standardized water quality data like pH, TDS, major ions (Na^+, Ca^{2+}, SO_4^{2-}, etc.). Four machine learning models were tested: Linear Regression, SVM, Random Forest, and BPNN. The study used k-means clustering to explore groundwater chemistry patterns. Three statistical metrics, e.g., coefficient of determination (R2), root mean square error (RMSE), and mean absolute error (MAE), are used for the model performance evaluation



RESULTS

Cluster 1(ridge areas), the mean pH value stood at 7.5 and exhibited a neutral character of the water. HCO3- emerged as the dominant anion in Cluster 1 with an average content of 292.15 mg/L. Thus, the HCO3–Ca type was identified as the primary hydrochemical facies for Cluster 1.
For cluster 2(valley areas), The average pH value, at 8.1, underscores the slightly alkaline nature of the water. The predominant cation in this cluster was Na+, while the anions were composed of HCO3- and Cl.
a notable distribution of water quality for domestic drinking purposes emerged, with 59.48 % of samples classified as excellent, 31.90 % as good and even 8.62 % as medium. The calculated EWQI and predicted EWQI achieved a near-perfect correlation of 0.9999, coupled with low RMSE and MAE values of 0.49 and 0.26, respectively.



The cluster results of SOM and k-means: (a) the u-matrix diagram, where the numbers in the neurons are the samples' ID, and (b) the spatial distribution of each cluster

PRACTICAL APPLICATION

Urban Relevance: Cities like Dhaka, Delhi, or Jakarta, facing similar groundwater pressures, could apply these techniques.
Data-Driven Decision Making: Application of machine learning models (like Linear Regression and BPNN) allows for accurate prediction of water quality, supporting real-time monitoring systems for groundwater safety.
Sustainable Development Planning: Supports SDG Goal 6 ("Clean Water and Sanitation") by promoting sustainable groundwater use in megacities and can be adapted for other urban areas with similar hydrogeological challenges.
Public Health Protection: The entropy-weighted water quality index (EWQI) and sensitivity analysis pinpoint the most harmful pollutants (NO_2^-, NO_3^-, NH_4^+), helping reduce health risks associated with contaminated drinking water.

REFERENCE

1. Xie et al. (2025). Journal of Hydrology: Regional Studies.
2. Yang et al. (2023). Hydrochemical classification criteria.
3. Uddin et al. (2022). EWQI advancements.
4. Subba Rao, N. et al. (2022). Groundwater quality assessment using entropy-weighted water quality index (EWQI) in agricultural areas.
5. Haggerty, R. et al. (2023). Machine learning in groundwater quality modeling.
6. Singha, S. et al. (2021). Prediction of groundwater quality using machine learning.
7. Gaillardet, J. et al. (1999). Global silicate weathering and CO_2 consumption rates deduced from river chemistry.
8. Jia, X. et al. (2019). Groundwater depletion and contamination in China.