J. Ind. Geophys. Union, 30(1) (2026), 70-77
https://doi.org/10.71122/JIGU.30(1)2026.007

Deep learning-based model for groundwater quality prediction in Kanyakumari District, Tamil Nadu, India.

Annie Jose* and Srinivas Y
Centre for Geotechnology, Manonmaniam Sundaranar University, Tirunelveli – 627 012, Tamil Nadu (India)
*Corresponding Author: anniejose1996@gmail.com

ABSTRACT

In this study, a Deep learning-based model, Stochastic Neural Network (SNN), which excels in handling uncertain and complex data, is customized to predict groundwater quality in the Kanyakumari District of Tamil Nadu (India), where groundwater is considered a critical source for drinking and agricultural purposes. The SNN model captures the stochastic nature of the data and provides reliable predictions by simulating multiple possible outcomes, making the model ideal for groundwater quality prediction. A range of groundwater quality indicators, such as pH, EC, TDS, etc. were used to train the model. Rainfall patterns were found to have a considerable impact on water quality over a ten-year period, emphasizing the importance to include seasonal data into prediction models. Hence, rainfall data was also included in order to evaluate its impact on groundwater quality. The deep learning model demonstrated its effectiveness with 95% prediction accuracy. The model's capacity to distinguish between classes was evaluated by the Classification Report, Receiver Operating Characteristic (ROC) curves, Area Under the Curve (AUC) values and the confusion matrix. In addition to this, Cross-validation (CV) was employed to confirm the model's performance and also to test the reliability of the results. This study provides an efficient method that can assist in sustainable use of groundwater resources.

Key Words: Deep learning, Groundwater quality prediction, Rainfall, Stochastic Neural Network, Kanyakumari District, Public Health.

INTRODUCTION

Water is a necessity for survival, economic development and human well-being, hence access to clean water is a fundamental right (Kirsch, 2006). With rising living standards, population demands, and expanding domestic, industrial, agricultural, and urban activities, water consumption has increased, resulting in overuse and deterioration of both surface and groundwater resources (Gilli et al., 2012). Groundwater in particular, is difficult to manage and more susceptible to contamination from both geogenic and anthropogenic sources (Karamouz et al., 2011). As water easily dissolves substances, monitoring groundwater quality changes is critical for its protection and conservation. Exceeding permissible limits of physicochemical parameters, may pose serious health risks (Karnena and Saritha, 2019). Groundwater is vital for drinking as well irrigation, especially in the study area of Kanyakumari, in Tamil Nadu and hence proper quality monitoring is critical for sustainable resource management. Given the complex and variable nature of groundwater, which is influenced by both natural and human influences (Giao et al., 2023), traditional evaluation methods based on laboratory testing and statistical analysis can be time and resource draining, especially for large datasets. To address these limitations, this study utilizes data-driven modelling to improve the accuracy and reliability of groundwater quality predictions.

Over the last decade, Machine Learning (ML) has grown rapidly in groundwater quality modeling, as ML algorithms learn patterns directly from data rather than depending on pre-set equations, allowing for precise predictions of water quality parameters (Haggerty et al., 2023). Deep Learning (DL), a subset of machine learning, inspired by neural processing in the human brain, employs multilayer neural networks to evaluate complicated, structured data (Chollet, 2017). These developments indicate great potential in groundwater research, as neural networks often deliver accurate and dependable predictions. They have been extensively used to predict groundwater levels, evaluate water quality near shale gas reserves, improve MODFLOW simulations, predict nitrate contamination and model groundwater drawdown (Lohani and Krishnan, 2015; Kulisz et al., 2021; Gholami and Sahour, 2022; Stylianoudaki et al., 2022; Kishor et al., 2025). All studies indicate that the neural networks may effectively assess groundwater quality and quantity, but most studies utilize conventional Artificial Neural Networks (ANN), which do not capture or forecast uncertainty. Stochastic Neural Networks (SNNs), on the other hand, can reduce overfitting and capture uncertainty. Despite being widely used in domains such as, finance (Kalariya et al., 2022), civil engineering (Emig et al., 2023), microbiology (Sarmadi et al., 2022), and physics (Schneider et al., 2017), their use in groundwater research remains limited. SNN's strength lies in making probabilistic predictions and adapting to changing environmental conditions. In this study, SNN with Monte Carlo Dropout, was used to simulate numerous possibilities per input, resulting in uncertainty-aware groundwater quality prediction. Rainfall was also included to better express climate-related effects on groundwater chemistry. The main objective of the study is to create a sophisticated deep learning model that is specifically suited to capturing the diversity and complexity present in groundwater quality data, to study the prediction reliability using stochastic inference and assess how rainfall affects the accuracy of water quality classification in the Kanyakumari District of Tamil Nadu, India.

STUDY AREA

The study area, Kanyakumari District (Figure 1), is one of those areas that heavily relies on groundwater as a freshwater