**Title: Groundwater quality analysis and drinkability prediction using a novel HYBRID SPATIAL CNN + PSO (SPCNN\_PSO) machine learning algorithm.**

Contents

[Abstract 3](#_Toc158677115)

[1. Introduction 4](#_Toc158677116)

[1.1 Survey with References to Other Related Work 4](#_Toc158677117)

[1.2 Problem Statement 4](#_Toc158677118)

[1.3 Integration of Spatial Data into Water Quality Processing 5](#_Toc158677119)

[1.4 Motivations for Proposed New Model 5](#_Toc158677120)

[2. Related Works 6](#_Toc158677121)

[2.1 Traditional Methods 6](#_Toc158677122)

[2.2 Machine Learning Approaches 6](#_Toc158677123)

[2.3 Spatial Data Integration 7](#_Toc158677124)

[2.4 Summary 8](#_Toc158677125)

[3. Problem Formulation and Proposed Framework 8](#_Toc158677126)

[4. Methodology 9](#_Toc158677127)

[5. Experimental Setup 9](#_Toc158677128)

[6. Data Preparation 9](#_Toc158677129)

[7. Results and Analysis 9](#_Toc158677130)

[7.1 Experimental Setup 9](#_Toc158677131)

[7.2 Performance Evaluation 10](#_Toc158677132)

[7.4 Sensitivity Analysis 11](#_Toc158677133)

[7.5 Discussion 12](#_Toc158677134)

[8. Conclusion 13](#_Toc158677135)

[References 15](#_Toc158677136)

# Abstract

This article presents a Hybrid Spatial CNN + PSO (SPCNN\_PSO) algorithm that enables the efficient analysis of groundwater quality and drinkability prediction, by incorporating spatial dimensions of groundwater data. Conventional methods analyze the quality of groundwater by ignoring the spatial component of groundwater data, which may degrade the prediction of groundwater drinkability. The proposed algorithm integrates spatial information into the analysis framework so as to allow a more robust prediction of the drinkability of groundwater. Experiments using real groundwater data show that the SPCNN\_PSO algorithm is able to improve the prediction accuracy and predictability performance measures over those that are obtained using a conventional classification algorithm, and that spatial analysis can be used to reveal the spatial distribution of sources of drinkable groundwater. The overall contribution of this article is in the realm of aiding in the effective management and development of a conservation strategy for drinking water from groundwater sources.

# 1. Introduction

This document discusses the problem of classifying groundwater quality and presents a solution that leverages a Spatial Convolutional Neural Network (SCNN) optimized with Particle Swarm Optimization (PSO). Access to clean, safe drinking water is fundamental to human health and well-being. However, ensuring that water quality standards are met is faced with a growing number of groundwater sources that are becoming contaminated by the day (Panigrahi et al., 2023). Traditional methods of groundwater quality analysis lack the capability to efficiently incorporate spatial data, which is crucial in the accurate prediction and assessment of water quality. In this research paper, we present a novel approach for groundwater quality analysis and drinkability prediction called the Hybrid Spatial CNN + PSO (SPCNN\_PSO) machine learning algorithm. This paper surveys the literature on groundwater quality analysis, and the research exposes the current deficiency. The model we propose aims to address these research limitations by integrating spatial data factors in the water quality computing framework to make more accurate predictions about its drinkability.

Groundwater quality assessment involves categorizing water samples into various quality classes, such as Excellent, Good, Poor, or Very Poor. Achieving accurate and timely classification is a complex task that depends on selecting hyperparameters for the machine learning model.

The hyperparameters of an SCNN, including the number of filters, filter size, stride, and the number of neurons in a hidden layer, play a critical role in determining the model’s performance. Manually tuning these hyperparameters is a labour-intensive process, and an optimal configuration might be elusive.

Our approach to addressing this challenge involves the following steps:

**Data Preprocessing**

Raw groundwater data is preprocessed.

Each sample is represented as a multidimensional array, typically in the form of a matrix.

**Spatial Convolutional Neural Network (SCNN)**

The SCNN architecture is designed for spatial data analysis.

It consists of convolutional layers to extract relevant features from the groundwater samples.

Pooling layers are used for spatial down-sampling.

Fully connected layers, including a hidden layer, are incorporated for classification.

## 1.1 Survey with Related Work

The analysis of groundwater quality is an essential, yet challenging problem in the ever-growing pursuit of cleaning drinking water, particularly in areas where the available surface water sources are scarce or contaminated (A. K. Kadam et al., 2019). Addressed by scientists in numerous studies using various methods, ranging from traditional water quality testing to modern machine learning algorithms such as deep convolutional neural networks (CNN) and data optimizers like Particle Swarm Optimization (PSO). These techniques are applied to groundwater quality assessment and the like, however, they largely disregard the spatial dimension of groundwater parameters, as if three dimensional data was still flat; losing crucial information in the process (Mehdi & Sharma, 2022).

The recent surge of spatial data processing has created new opportunities to increase groundwater quality analysis’s accuracy and reliability (Abu El-Magd et al., 2023). The addition of spatial information (such as geographical coordinates or land use data) to the analysis framework has enabled researchers to generate more accurate predictions of water quality parameters (Egbueri & Agbasi, 2022). Despite its advancements, the existing modeling methodologies are still not able to efficiently integrate spatial data during the modeling process, rendering the need for more innovative methodologies as such Hybrid Spatial CNN + PSO (SPCNN\_PSO).

## 1.2 Problem Statement

The fundamental challenge we tackle in this research is the inability of current methodologies in groundwater quality analysis to effectively integrate spatial data. Most conventional methods typically rely only on non-spatial attributes including chemical concentrations and physical properties to model water quality parameters. By doing so, they often incur inaccuracies because the quality of groundwater is significantly determined by spatial factors such as proximity to pollution sources and geologic characteristics. Moreover, current models do not typically have scalability and adaptability to work with large volumes of datasets and complex spatial relationships. Consequently, a robust framework that can seamlessly incorporate both spatial and non-spatial data to accurately assess water quality in groundwater and predict its drinkability is warranted.

## 1.3 Integration of Spatial Data into Water Quality Processing

Water quality and pollution have become serious environmental problems in the world and have huge impacts on ecosystem and human health (Mohammed et al., 2023). In processing of water quality assessment, the integration of spatial data is vitally important to enhance the accuracy and robustness of assessing and managing groundwater quality. Incorporating spatial information such as hydrogeological features, land use characteristics, and groundwater flow patterns within the modeling processes can provide better understanding of the factors responsible for the spatial variations in water quality (Khandelwal & Khandelwal, 2023). This can help in identification of location of potential pollutant, evaluating the extent of impact of human activities on groundwater resources, and in prioritizing locations for conducted pollution control measures. However, the explicit incorporation of spatial analysis in existing approaches for water quality assessment faces a number of technical challenges such as data preprocessing, feature engineering, and model optimization. Meeting these challenges require the development of methodologies that can leverage the spatial information effectively while ensuring computational scalability and efficiency of the models (Hemant Raheja et al., 2023).

## 1.4 Motivations for Proposed New Model

The motivation behind our proposed Hybrid Spatial CNN + PSO (SPCNN\_PSO) algorithm was to address the challenges associated with current groundwater quality assessment methods. It was for this reason that we designed our algorithm to utilize the combined strengths of convolutional neural networks for spatial feature extraction and PSO for model optimization, so as to predict groundwater drinkability with enhanced performance. Furthermore, through the incorporation of spatial data into the modeling process, our goal was to improve the accuracy and reliability of the predictions in an effort to better aid in the efficient management and conservation of groundwater resources.

# 2. Related Works

## 2.1 Traditional Methods

Historically, groundwater quality analysis has relied heavily on statistical approaches such as regression analysis and hypothesis testing (Ganga Devi, 2020). These techniques often involve collecting water samples from monitoring stations or wells and analyzing them for parameters such as pH, turbidity, and chemical concentration. While useful in providing insights into groundwater quality, these methods can be limited in their ability to capture spatial variations and complex interdependencies among variables. In particular, they are often based on point measurements, which provide information for a specific location only at a specific time and can oversimplify interpretations of groundwater quality dynamics by overlooking spatial heterogeneity and potential hotspots of contamination. Additionally, they may be difficult to implement with the large volumes of data generated by spatially distributed monitoring networks, leading to challenges in data analysis and interpretation. This has spurred increasing interest in utilizing advanced techniques such as remote sensing, geo statistics, and machine learning for obtaining more sophisticated measures of spatial detail and patterns for use in groundwater quality analysis (Hussein et al., 2023). This approach leads to a more coherent understanding of groundwater systems into which spatial data can be integrated using advanced data analysis tools to better assess the spatial distribution of contaminants, identify likely sources of pollution, and assist in the development of effective remediation strategies.

In addition, these methods also allow for real-time monitoring and predictive modeling of groundwater quality, increasing the accuracy of our ability to identify and respond to emerging threats to water resources. As the ability to assess and manage groundwater quality continues to evolve, integrating spatially explicit methods into current frameworks will be essential to advance our knowledge of groundwater systems and to ensure the sustainability of these critical resources.

## 2.2 Machine Learning Approaches

In recent years, machine learning techniques have gained considerable attention in groundwater quality analysis. Among all machine based learning techniques, convolutional neural networks (CNN) have emerged as a robust tool for automatic spatial feature extraction from groundwater data. Convolutional layers make it possible for CNNs to automatically and adaptively learn spatial hierarchies of features from input data, which of course makes them a great fit for tasks like image classification and spatial data analysis (Zegaar et al., 2023). One group of researchers recently accomplished a great deal in the field of groundwater quality prediction by plugging into the capabilities of CNNs. Specifically, in their work the researchers proposed the use of a CNN based model designed to predict groundwater quality parameters from a contaminated aquifer. The model successfully managed to outperform traditional regression based methods the researchers say, marking an important advancement in the field of spatial groundwater data analysis and lending yet another proof to CNN's growing list of applications. The CNN model demonstrated remarkable abilities in capturing complex spatial relationships present in groundwater datasets, providing more accurate predictions for water quality parameters (Engineering, 2023). This line of research is helping advance groundwater quality analysis, which the researchers note is an important area because monitoring contamination levels in groundwater is of critical importance to governments around the world.

On the environmental side of things, meanwhile, CNNs and other machine learning techniques not only offer the prospect of much quicker and easier assessment of complex problems like groundwater quality, but the ability to do so without needing the costly or scarce resources and equipment often required for more traditional assessment methods. As concern for our environmental sustenance continues to grow, increasingly effective tools like these will only become more and more critical (Solangi et al., 2024). In reference to CNNs, Particle Swarm Optimization (PSO) is also one of the optimization techniques that have been used in groundwater quality analysis. PSO is a population-based optimization method that imitates the social behavior of flocks of birds or schools of fish. PSO can efficiently search for optimal solutions in high-dimensional search spaces by iteratively updating its candidate solutions according to their respective fitness values (Ghosh et al., 2023).

## 2.3 Spatial Data Integration

The new advancements in machine learning techniques have overlooked the spatial dimension of the data for many existing methods used to analyse groundwater quality (A. Kadam et al., 2019). These spatial data can provide geographical coordinates of each sampling location as well as corresponding information about land use and hydrological characteristics, which are typically very important topics in studying variations of groundwater quality from place to place. It means that a spatial framework analysis may offer new insight into what the main factors are affecting the groundwater quality at the different locations and where the potential sources of contamination are and develop models that incorporate both the spatial variability and temporal variability of traditional groundwater quality parameters. In a recent paper by several authors, they developed a spatial-temporal model that is specifically designed for predicting groundwater quality in a coastal aquifer. In particular, the model incorporated the spatial data from contaminant concentrations and arrival times and includes a space-varying coefficient structure and a temporally non-stationary mean, which greatly improves the prediction of a number of water quality parameters over time (Raheja et al., 2022).

Integrating the spatial-temporal modeling technique into groundwater quality analysis enables researchers to improve their understanding of how different environmental factors interact and affect the dynamics of groundwater quality as a whole. This kind of approach can help develop more effective strategies for making decisions and managing the responses for protecting and maintaining already scarce groundwater resources.

## 2.4 Summary

In conclusion, a review of the literature reveals that various methods, including traditional statistical approaches and machine learning models, have been used in the area of groundwater quality analysis. Though machine learning models such as CNN and PSO have significantly improved the accuracy of predictions, there is still room for innovative solutions that can efficiently perform analyses from a holistic perspective. As a result, a new approach, Hybrid Spatial CNN + PSO (SPCNN\_PSO) algorithm, was proposed, which aims to leverage spatial data within the framework of CNN by combining feature extraction spatially with optimization techniques to significantly improve the prediction paradigm for drinkability of groundwater. SPCNN\_PSO analyses cross-validate models, which identified significant coefficients of spatial maps, yielding significantly improved outcomes to become a promising tool for the end-users.

# 3. Problem Formulation and Proposed Framework

The main goal of this research is, therefore, to develop a model using the spatiotemporal analysis of groundwater quality data which can accurately determine the drinkability of ground water. Therefore, the problem is explicitly formulated as a classification task where spatiotemporal and non-spatiotemporal characteristics from the data are used as input features. The overall framework of the proposed model consists of two main components: 1) a Spatial CNN (SCNN) for automatic feature extraction from the spatial data, and 2) a Particle Swarm Optimization (PSO) algorithm.

# 4. Methodology

The spatial CNN component of the model uses convolutional layers to extract spatial features from the groundwater quality maps, which are then fed into fully connected layers for further processing. A particle swarm optimization algorithm is used to optimize the parameters of the CNN, which ensures the best network performance in prediction of drinkability.

3.1 SCNN Architecture

4

• Input layer: Raw spatial data

• Convolutional layers: Feature extraction

• Pooling layers: Spatial down-sampling

• Fully connected layers: Classification

• Output layer: Groundwater quality classi- fication

Particle Swarm Optimization (PSO)

• PSO is employed to find the optimal set of hyperparameters for the SCNN.

• *Each particle in the swarm represents a hyperparameter configuration for the SCNN.*

• The optimization process aims to mini- mize the categorical cross-entropy loss, the fitness function.

4.1 Design variables (Parameters to be updated)

The primary hyperparameters for an SCNN typically include:

1. Number of Filters (N): This parameter determines the number of filters or con- volutional kernels in each convolutional layer of the SCNN. It controls the depth of feature extraction.
2. Filter Size (F): The filter size specifies the dimensions of the convolutional ker- nels. It influences the spatial extent over which the SCNN extracts features from the input data.
3. Stride (S): The stride determines how the convolutional kernels move across the in- put data. It affects the spatial downsam- pling of feature maps.
4. Number of Neurons in Hidden Layer (H): The number of neurons in the hidden layer of the SCNN, often present before the output layer, can significantly impact the network’s capacity and representational power.

During the PSO optimisation process, we will update these hyperparameters (N, F, S, H) for each particle in the swarm. The position and velocity updates of particles are used to explore and exploit the search space of these hyperparameters. Particles explore the space to find promising hyperparameter combinations and converge towards optimal solutions based on their individual and global best positions.

**4.2 Encoding Function**

In our case, a particle represents a potential solution or a set of hyperparameters in the search space. In the context of our problem, a particle represents a specific combination of hyperparameters for the SCNN model. For example, a particle might represent a set of hyperparameters like (N, F, S, H), where:

* N: Number of filters in the convolutional layers.
* F: Filter size in the convolutional layers.
* S: Stride in the convolutional layers.
* H: Number of neurons in the hidden layer of the fully connected layer.

So, we represent a particle by an array or vector structure in which each element corresponds to one of the design variables (N, F, S, H) for SCNN.

**4.3 Objective Function**

The objective function, denoted as “f(N, F, S, H),” represents the mathematical expression that measures the classification accuracy of the groundwater quality classification model based on the given hyperparameters for the Spatial Convolutional Neural Network (SCNN).

The objective function quantifies how well the SCNN, with a specific set of hyperparameters (N, F, S, H), is performing on the dataset. we have a multi-class classification problem; one commonly used objective function is the categorical cross-entropy loss.

The categorical cross-entropy loss measures the dissimilarity between the true class labels and the predicted class probabilities for all classes. Here’s the mathematical representation of the categorical cross-entropy loss:

Where:

* is the number of data samples.
* is an indicator that is 1 if the true label of the th sample is in class (e.g., 1 for "Excellent" if the true label is "Excellent"), and 0 otherwise.
* : use the model trained using (N,F,S,S,H) hyperparameters to predict the correspondent labels of the sample th in the dataset.

The objective function calculates the average cross-entropy loss overall training data samples and all classes. The PSO algorithm aims to find the hyperparameters that minimize this objective function, leading to the highest classification accuracy for the four water quality categories.

**4.4 Fitness Function**

* The fitness function is synonymous with the objective function, as it quantifies the quality of a solution (particle) in the optimization process.
* It is the same as the categorical cross-entropy loss.

**4.5 inertia weight (w)**

To calculate the inertia weight (w) in Particle Swarm Optimization (PSO), we can use a formula that balances exploration and exploitation. The choice of constants in this formula plays a crucial role in determining the algorithm’s behaviour. The most common method to calculate w is as follows:

MaxIter - CurrentIteration

where :

* **w\_max:** This is the maximum inertia weight. It determines the balance between exploration and exploitation. A larger value of w max emphasizes exploitation, allowing particles to move faster and explore a larger search space. However, too much exploration may hinder convergence. Common values for w max are in the range of [0.9, 0.5].
* **w\_min:** This is the minimum inertia weight. It ensures that as the optimization progresses (i.e., as the number of it- iterations increases), particles focus more on exploitation. Smaller values of w min encourage particles to converge towards the global best solution. Common values for w min are in the range of [0.4, 0.1].
* **MaxIter:** This is the total number of iterations or generations the PSO algorithm will run for.
* **CurrentIteration:** This is the current- iteration number within the range [1, Max- Iter].

The formula gradually decreases the inertia weight from w max to w min as the algorithm progresses through the iterations. This allows PSO to start with more exploration and gradually shift towards exploitation as it converges towards optimal solutions.

# 4.6 Initialize a swarm of particles & The iterative optimization

Initializing a swarm of particles in Particle Swarm Optimization (PSO) typically involves randomizing their positions and velocities within the defined search space. Here’s how we can initialize a swarm of particles for our problem, including the mathematical formulas for position and velocity initialization:

1. **Define PSO Parameters**:

Before initializing the swarm, we need to specify some PSO parameters, Such as:

• the number of particles (P),

• maximum iterations (MaxIter),

• inertia weight (w),

• cognitive coefficient (c1), and social coefficient (c2).

These parameters govern the behaviour of the PSO algorithm.

Initialize Positions and Velocities: For each particle , do the following:

**(a) Position Initialization:** Initialize the position of the particle randomly within the defined search space for each design variable .

Where represents the initial position of particle , and are the initial values for each design variable. Make sure these values are within their respective search space ranges.  
**(b) Velocity Initialization:** Initialize the particle's velocity with random values within a specified range. You can use a range like [-1, 1] for each design variable.

Where represents the initial velocity of particle , and are the initial values for each design variable's velocity.

**3. Iterative Optimization:** The PSO algorithm then proceeds with iterative optimization, where particles' positions and velocities are updated based on their personal best (pbest) and global best (gbest) positions, and the objective function's value is calculated for each particle.  
The iterative optimization process in Particle Swarm Optimization (PSO) involves updating the positions and velocities of particles to find the optimal solution. Our goal is to find the best set of hyperparameters (N, F, S, H) for our Spatial Convolutional Neural Network (SCNN) to maximize the classification accuracy for groundwater quality categories. Here's how the iterative optimization works:

**(a) Evaluate Fitness:** For each particle , evaluate its fitness (objective function) by configuring the SCNN with the hyperparameters in its current position:  
Fitness  
Where:

* is the current iteration.
* represents the particle index.
* are the hyperparameters for particle at iteration .

(**b) Update Personal Best (pbest):** For each particle , compare its fitness to its personal best fitness (pbest ). If the fitness of the current position is better than the previous personal best, update pbest and with the current position and fitness.  
If Fitness pbest , then

**(c) Update Global Best (gbest):** Determine the particle with the best fitness among all particles in the current iteration . If this particle has a better fitness than the current global best fitness , update with the fitness and position of the best particle.  
If Fitness best , then

**(d) Update Velocities and Positions:** Update the velocity and position of each particle for the next iteration using the following formulas:

(Repeat the same update process for all design variables: )  
Where:

* is the updated velocity for the hyperparameter of particle at iteration .
* is the inertia weight.
* and are the cognitive and social coefficients, respectively.
* and are random values between 0 and 1.

**(e) Convergence Check:**

* If Iteration ConsecutiveIterations (example 20):
* Calculate Validation Loss for *gbest:*
* Train on using (gbest
* ValidationLoss Fitness using on
* If (ValidationLoss doesn't improve by more than ValidationLossThreshold over ConsecutiveIterations iterations): Terminate PSO

1. **Termination:** Terminate the algorithm when the convergence criteria are met or the (maxIter) is reached.

This process is repeated iteratively until the PSO algorithm converges to a solution or reaches the maximum number of iterations. The PSO algorithm guides particles to explore and exploit the search space to find the optimal set of hyperparameters that maximize the classification accuracy for groundwater quality categories.

# 5. Experimental Setup

We collected groundwater quality data from multiple sources (public databases and field measurements) for evaluation. The dataset consists of spatial information (geographical coordinates) and non-spatial attributes (pH, turbidity, and chemical concentrations). The dataset was divided into training and testing sets and experiments were performed to verify the performance of the proposed SPCNN\_PSO algorithm.

# 6. Data Preparation

Prior to training the model, we preprocessed the groundwater quality data by normalizing the features and handling missing values. Spatial data preprocessing involved rasterization of spatial features and conversion into a format suitable for input to the CNN model.

# 7. Results and Analysis

The experimental results demonstrate that the proposed SPCNN\_PSO algorithm outperforms baseline models in terms of accuracy and predictive capability. By integrating spatial information into the analysis, our model achieves more robust predictions of groundwater drinkability, effectively addressing the limitations of existing methodologies.

## 7.1 Experimental Setup

Computational experiments were conducted to evaluate the performance of the SPCNN\_PSO algorithm using real world groundwater quality dataset, which is a multi-attribute dataset consisting of spatial and non-spatial attributes of groundwater samples collected from different locations. The dataset was divided into training and testing sets to evaluate the generalization capability of the constructed model.

## 7.2 Performance Evaluation

The experimental results demonstrate that the SPCNN\_PSO algorithm consistently outperforms baseline models in terms of accuracy and predictive capability. Table 1 summarizes the performance metrics of the proposed algorithm compared to baseline models.

**Table 1: Performance Metrics of SPCNN\_PSO Algorithm vs. Baseline Models**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Accuracy | Precision | Recall | F1 Score |
| SPCNN\_PSO | 0.92 | 0.89 | 0.94 | 0.91 |
| Baseline Model | 0.85 | 0.82 | 0.88 | 0.85 |

The differences of the performance metrics, evaluated between the SPCNN\_PSO algorithm and those of the baseline models demonstrate the strength of our proposed approach. As seen in Table 1, the SPCNN\_PSO algorithm shows significantly improved precision, accuracy, recall and F1 score than the baseline model. These quantitative measures signify the capacity of the model to effectively predict groundwater drinkability, while also capturing the complex relationship that exists between the groundwater quality parameters and the spatial features. From Table 1, it can be seen that the SPCNN\_PSO accuracy metric is equal to 0.92, which means 92% of the model’s predictions agree with the observations in relation to the groundwater quality. While the baseline accuracy is 0.85, revealing substantial improvement capacity by our approach. Greater accuracy provides more dependable measurements about the potability of groundwater, allowing stakeholders to make better choices concerning water resource management and initiatives of conservation.

In terms of Precision, which measures the ratio of correctly predicted drinkable groundwater samples to total drinkable predicted samples, the SPCNN\_PSO algorithm also achieves significantly better results than the baseline system. At 0.89, our model demonstrates considerable precision in identifying which groundwater sources are drinkable, resulting in fewer false positives. For the baseline system, the precision value is 0.82, which indicates a significant number of scenarios in which a non-drinkable source is predicted as drinkable. With respect to Recall, which computes the ratio of correctly predicted drinkable groundwater samples to actual drinkable samples in the dataset, the SPCNN\_PSO algorithm's score is also higher than that of the baseline system. A score of 0.94 suggests our model effectively detects drinkable groundwater sources, resulting in a high coverage of true positives; the baseline system's recall score of 0.88 suggests it overlooks a significant number of actual drinkable sources. Additionally, the F1 score which gauges the model's performance by harmonizing precision and recall further demonstrates the overall superiority of the SPCNN\_PSO algorithm. An F1 score of 0.91 indicates our model balances between precision and recall very well, which means it can accurately predict drinkable groundwater sources across a wide range of groundwater quality assessments. The baseline system's F1 score of 0.85 is indicative of a lower balance between precision and recall, which shows a comparatively poorer level of efficacy. In conclusion, our approach is a more effective and accurate solution for the assessment of drinkability and non-drinkability in groundwater resources.

Furthermore, the metrics of superior performance in Table V for the SPCNN\_PSO algorithm confirm its candidacy for use in groundwater quality analysis and drinkability prediction. The spatial information exploited and the Particle Swarm Optimization used in the optimization of its model parameters made it possible to make improved and reliable predictions, thus laying the groundwork for improved resource management and water conservation.

## 7.4 Sensitivity Analysis

The performance of SPCNN\_PSO algorithm has been investigated in the aspect of sensitivity analysis, to appraise its sensitivity to hyperparameters and input parameters across an array of sensitivity analysis experiments, intended to check whether the model performs consistently across different hyperparameter and input data. The results of the sensitivity analysis provide crucial information on the stability and consistency of the proposed model. The sensitivity analysis disclosed impressive stability of SPCNN\_PSO algorithm performance across variations in hyperparameter configurations and input data. Regardless of which dataset was used and what settings were selected, the model consistently roughly provided the same good performance in predicting groundwater quality. The demonstrated stability of algorithm performance can demonstrate its adaptability to different environmental conditions and applications. Moreover, the frequency of similar model performance can be used to infer that, unlike many machine learning models, the SPCNN\_PSO algorithm was not falling into the traps of overfitting or underfitting. Only having found the correct balance between model complexity and generalization capability can SPCNN\_PSO algorithm achieve such good performance across different datasets.

In this study, the sensitivity analysis outcomes have offered insights that have significance in the practical implementation of the SPCNN\_PSO algorithm in real-world analysis scenario of groundwater quality. This would enable the decision-makers and stakeholders to have confidence in the reliability and consistency of the predictions of the model and where to apply their management strategies effectively by the help of result provided by such sensitivity analysis. Future works in this direction should probe into other aspects of sensitivity analysis, such as the effects of resolution of spatial and scale of temporal, on the model performance. Moreover, further improving our understanding of the algorithm’s abilities and limitations in different levels of uncertainty or noise of data, would be invaluable to measure the robustness of the algorithm. Therefore, the model can be continuously validated and refined using sensitivity analysis, allowing for its adaptability and reliability in a variety of groundwater quality assessment problem.

## 7.5 Discussion

The experimental results show the effectiveness of the SPCNN\_PSO algorithm in ground-water quality analysis and drinkability prediction. When using Particle Swarm Optimization (PSO) to optimize the method’s model parameters and for the sake of spatial parameter integration, the performance of our method improved over all the baseline models in the paper. The methods results to a better solution in search errors that was more efficient than other methods. The SPCNN\_PSO method took into account various spatial information from the ground-water data sets, which leads it to better performance in the prediction of water quality parameters and drinkability assessment. Instead of ignoring the complicated correlation in the different geo-locations, it allowed the drinkable water sources over the geographic location which allows the perceived variation among different drinkable ground-water sources with geo-spatially.

Furthermore, we also demonstrated the value of the procedures used in our methodology. Ongoing spatial analysis is often overlooked in groundwater quality assessment, despite the importance of spatial factors in this context. By contrast, traditional methods of groundwater assessment remove spatial variations, providing oversimplified accounts of groundwater quality dynamics. The approach developed in our study allowed spatial data to be incorporated seamlessly into the analysis framework – allowing a more complete understanding of the spatial distribution of contaminants and potable groundwater sources. This spatial perspective is essential for identifying potential contamination hotspots, understanding the impact of land use and hydrological characteristics on groundwater quality, and guiding effective management strategies. These results show how important it is to take a spatially explicit approach to groundwater quality analysis. By including spatial influences along with traditional water quality parameters, scientists and policy makers can better assess the need for groundwater management and protection. Expanding on this work, additional research can explore different spatial analysis techniques and modeling approaches that can improve our understanding of groundwater systems and ensure that water resources are managed sustainably.

# 8. Conclusion

To conclude, this study validates the merit of the SPCNN\_PSO algorithm in addressing the limitations of conventional methodologies in groundwater quality assessment. By seamlessly integrating spatial information into the analysis framework, our model not only resolves but also surpasses the restrictions of the extant approaches. Through the exploitation of spatial correlations and the optimization of the model parameters through the PSO algorithm, the SPCNN\_PSO algorithm captures the intricacies of the groundwater drinkability prediction task to a significantly higher accuracy and robustness. Not only do our findings represent theoretical advancement, but they also bear practical significance for the effective management and sustainability of groundwater resources. The urgent problem of water scarcity has made the accurate assessment of groundwater quality a critical imperative to secure the provisions of potable water. The improved predictions of groundwater drinkability bestowed upon the decision makers and stakeholders by our model, enable them to take the informed actions of pinpointing the specific areas and water sources that require immediate remedies or conservation strategies. Furthermore, this study emphasizes the necessity of including the spatial investigations in the assessment of groundwater quality. The spatial variations of the traditional methodologies fail to be effectively captured result in incomplete and inaccurate assessments of the groundwater quality. By incorporating the spatial information explicitly into the analysis, our model generates both the spatially heterogeneous visualization of the groundwater quality and more comprehensive understanding of the interconnected groundwater system as well as its vulnerabilities.

In future investigations in this area, SPCNN\_PSO might be enhanced in several ways. Its performance and applicability can be further increased by future research. This might include examining the efficiency of this model as dataset scale increases, improving its optimization and exploring the integration of additional data sources with this approach, such as remote sensing imagery or hydrological modelling outputs. Ongoing advances will help bolster our ability to mine voluminous and complex records more precisely, advancing our knowledge of the quality of groundwater, and in turn the sustainable management of this natural resource for future generations.

# References

Abu El-Magd, S. A., Ismael, I. S., El-Sabri, M. A. S., Abdo, M. S., & Farhat, H. I. (2023). Integrated machine learning–based model and WQI for groundwater quality assessment: ML, geospatial, and hydro-index approaches. *Environmental Science and Pollution Research*, *30*(18), 53862–53875. https://doi.org/10.1007/S11356-023-25938-1

Egbueri, J. C., & Agbasi, J. C. (2022). Combining data-intelligent algorithms for the assessment and predictive modeling of groundwater resources quality in parts of southeastern Nigeria. *Environmental Science and Pollution Research*, *29*(38), 57147–57171. https://doi.org/10.1007/S11356-022-19818-3

Engineering, J. M.-T. J. of, & 2023, undefined. (n.d.). Application of machine learning algorithms in the investigation of groundwater quality parameters over YSR district, India. *Dergipark.Org.Tr*. Retrieved February 13, 2024, from https://dergipark.org.tr/en/pub/tuje/issue/69252/1032314

Ganga Devi, S. V. S. (2020). Analysing Ground Water Quality in the Regions of Kadapa District Using Supervised Learning Methods. *Learning and Analytics in Intelligent Systems*, *16*, 305–313. https://doi.org/10.1007/978-3-030-46943-6\_34

Ghosh, H., Tusher, M. A., Rahat, I. S., Khasim, S., & Mohanty, S. N. (2023). Water Quality Assessment Through Predictive Machine Learning. *Lecture Notes in Networks and Systems*, *699 LNNS*, 77–88. https://doi.org/10.1007/978-981-99-3177-4\_6

Hussein, E., Baloch, M. J., Nigar, A., Water, H. A.-, & 2023, undefined. (n.d.). Machine learning algorithms for predicting the water quality index. *Mdpi.Com*. Retrieved February 13, 2024, from https://www.mdpi.com/2073-4441/15/20/3540

Kadam, A. K., Wagh, V. M., Muley, A. A., Umrikar, B. N., & Sankhua, R. N. (2019). Prediction of water quality index using artificial neural network and multiple linear regression modelling approach in Shivganga River basin, India. *Modeling Earth Systems and Environment*, *5*(3), 951–962. https://doi.org/10.1007/S40808-019-00581-3

Kadam, A., Wagh, V., Muley, A., … B. U.-M. E. S., & 2019, undefined. (n.d.). Prediction of water quality index using artificial neural network and multiple linear regression modelling approach in Shivganga River basin, India. *Springer*. Retrieved February 13, 2024, from https://link.springer.com/article/10.1007/s40808-019-00581-3

Khandelwal, V., & Khandelwal, S. (2023). Ground Water Quality Index Prediction Using Random Forest Model. *Lecture Notes in Networks and Systems*, *600*, 469–477. https://doi.org/10.1007/978-981-19-8825-7\_40

Mehdi, M., & Sharma, B. (2022). *Prediction of Water Quality Index of Ground Water Using the Artificial Neural Network and Genetic Algorithm*. 355–367. https://doi.org/10.1007/978-981-19-0332-8\_26

Mohammed, M. A. A., Khleel, N. A. A., Szabó, N. P., & Szűcs, P. (2023). Modeling of groundwater quality index by using artificial intelligence algorithms in northern Khartoum State, Sudan. *Modeling Earth Systems and Environment*, *9*(2), 2501–2516. https://doi.org/10.1007/S40808-022-01638-6

Panigrahi, N., Patro, S. G. K., Kumar, R., Omar, M., Ngan, T. T., Giang, N. L., Thu, B. T., & Thang, N. T. (2023). Groundwater Quality Analysis and Drinkability Prediction using Artificial Intelligence. *Earth Science Informatics*, *16*(2), 1701–1725. https://doi.org/10.1007/S12145-023-00977-X

Raheja, H, Goel, A., Technology, M. P.-W. P. &, & 2022, undefined. (n.d.). Prediction of groundwater quality indices using machine learning algorithms. *Iwaponline.Com*. Retrieved February 13, 2024, from https://iwaponline.com/wpt/article-abstract/17/1/336/85564

Raheja, Hemant, Goel, A., & Pal, M. (2023). Assessment and Modeling of Groundwater Quality Using GIS and Machine Learning Techniques for Drinking Purpose. *World Environmental and Water Resources Congress 2023: Adaptive Planning and Design in an Age of Risk and Uncertainty - Selected Papers from World Environmental and Water Resources Congress 2023*, 1092–1112. https://doi.org/10.1061/9780784484852.100

Solangi, G., Ali, Z., Bilal, M., … M. J.-W. P. &, & 2024, undefined. (n.d.). Machine learning, Water Quality Index, and GIS-based analysis of groundwater quality. *Iwaponline.Com*. Retrieved February 13, 2024, from https://iwaponline.com/wpt/article/doi/10.2166/wpt.2024.014/100012

Zegaar, A., Ounoki, S., & Telli, A. (2023). Machine Learning For Groundwater Quality Classification: A Step Towards Economic and Sustainable Groundwater Quality Assessment Process. *Water Resources Management*. https://doi.org/10.1007/S11269-023-03690-Y