Groundwater for Sustainable Development 25 (2024) 101156

Contents lists available at ScienceDirect

Groundwater for Sustainable Development

journal homepage: www.elsevier.com/locate/gsd

Research paper

Prediction of ground water quality in western regions of Tamilnadu using LSTM network

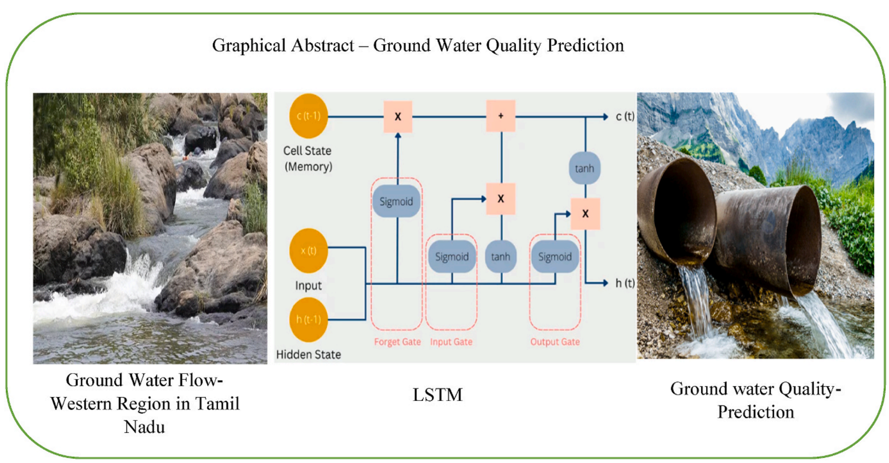
Kasiselvanathan M [a,*], Venkata Siva Rama Prasad C [b], Vijay Arputharaj J [c], Suresh A [d], Sinduja M [e], Prajna K.B [f], Maheswaran Shanmugm [g]

[a] *Department of Electronics and Communication Engineering, Sri Ramakrishna Engineering College, Coimbatore, Tamil Nadu, India*
[b] *Department of Civil Engineering, Malla Reddy Engineering College, Secunderabad, 500100, Telangana, India*
[c] *Department of Computer Science, CHRIST (Deemed to be University), Bangalore, India*
[d] *Department of Science and Humanities, Brilliant Grammar School Educational Society's Group of Institutions Integrated Campus, Hyderabad, Telangana, 501505, India*
[e] *Technical Executive, National Agro Foundation, Chennai, Tamil Nadu, 600113, India*
[f] *Department of Electronics and Communication Engineering, Nitte Meenakshi Institute of Technology, BSF Campus, Yelahanka, Bengaluru, Govindapura, Karnataka, 560064, India*
[g] *Department of Electronics and Communication Engineering, Kongu Engineering College, Perundurai, Erode, 638060, India*

HIGHLIGHTS

- Preserving the quality of groundwater is essential for life in the western region.
- The LSTM approach predicts groundwater quality and shows temporal fluctuation.
- The region's goal of improving groundwater resource management and quality control.
- The results show predictive accuracy and increases sustainable groundwater quality.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Groundwater
LSTM
Prediction

ABSTRACT

Assessing and safeguarding groundwater quality is critical for sustaining life in water-scarce regions like western Tamil Nadu. The motivation behind this study stems from the pressing need to address water quality challenges in a region grappling with scarcity. Despite existing efforts, a notable research gap exists in predictive tools that comprehensively capture the nuanced temporal variations and trends in groundwater quality. This is where the

* Corresponding author.
E-mail addresses: drkasiselvanathanvkm@gmail.com (K. M), cvsrprasad@mrec.ac.in (V.S.R.P. C), phdvij@gmail.com (V.A. J), sureshambati.hyd@gmail.com (S. A), seethasinduja@gmail.com (S. M), prajna.kb@nmit.ac.in (P. K.B), mmaheswaraneie@gmail.com (M. Shanmugm).

https://doi.org/10.1016/j.gsd.2024.101156
Received 1 November 2023; Received in revised form 6 March 2024; Accepted 15 March 2024
Available online 16 March 2024

Western region
Tamil Nadu

LSTM network steps in, showcasing exceptional accuracy in short-term predictions and discerning long-term trends. This research uses Long Short-Term Memory (LSTM) networks, a variant of recurrent neural networks, to predict groundwater quality in South Indian Regions, especially in Tamil Nadu. Extensive data, encompassing parameters such as pH, dissolved oxygen, turbidity, and various chemical constituents, were gathered over an extended timeframe. The LSTM model was then trained on this historical dataset, factoring in temporal dependencies and seasonality inherent in groundwater quality data. The validation process rigorously tests the LSTM model against actual groundwater quality measurements. The results were impressive, as the model demonstrated a remarkable ability to unravel the complex variations in groundwater quality.

1. Introduction

The essence of life lies beneath the surface, where the hidden world of groundwater unfolds (Singha et al., 2021). As a vital resource, groundwater is pivotal in sustaining ecosystems and meeting human needs (Yuvaraj et al., 2023). Its quality, however, is a dynamic interplay of various factors that shape its composition and suitability for diverse applications (Zhu et al., 2022). Understanding and managing groundwater quality is imperative, especially amid growing anthropogenic pressures and environmental uncertainties (Yuvaraj and Suresh Ghana Dhas, 2020). Groundwater quality is a complex tapestry woven by geological, hydrological, and anthropogenic threads (Chen et al., 2020). Geological formations influence the mineral content, creating a mosaic of underground reservoirs with distinct chemical signatures (Kayalvizhi et al., 2023). Hydrological processes further contribute to this intricate dance, affecting the movement and filtration of water through subsurface layers (Kouadri et al., 2022). Human activities, from industrial discharge to agricultural runoff, introduce many pollutants, altering the delicate balance of groundwater composition (Singh et al., 2022; Su et al., 2020). The challenge lies in comprehensively assessing and predicting groundwater quality, a task complicated by environmental factors' inherent variability and interconnectedness (Belkhiri et al., 2020). Monitoring parameters such as pH, dissolved oxygen, and various chemical constituents becomes crucial for unraveling the tapestry of groundwater quality (Ahmed et al., 2020). As populations burgeon and climate patterns shift, ensuring the sustainability and safety of this hidden treasure becomes paramount (Kouadri et al., 2021). Geological formations shape distinct underground reservoirs, while hydrological processes impact water movement. Human activities introduce pollutants, disrupting groundwater balance. Comprehensive assessment is challenged by environmental variability. Monitoring pH, dissolved oxygen, and chemical constituents is pivotal. Population growth and climate change intensify the urgency of safeguarding groundwater. In the arid landscape of western Tamil Nadu, water scarcity accentuates the need for robust groundwater quality protection. Existing methods need to catch up in capturing temporal variations, demanding advanced predictive tools. Water scarcity in western Tamil Nadu's arid landscape underscores the pressing need to safeguard groundwater quality (Asadollah et al., 2021). Despite concerted efforts, existing approaches often need to capture the intricate temporal variations and trends in groundwater quality. The challenges in the region are multifaceted, encompassing not only the scarcity of water resources but also the complexity of fluctuating quality parameters. Conventional methods need help to provide accurate and comprehensive insights, necessitating a paradigm shift towards advanced predictive tools. The problem revolves around the need for effective forecasting mechanisms tailored to the unique temporal dependencies and seasonality patterns inherent in groundwater quality data. This research seeks to redefine the approach by leveraging LSTM networks, known for their adept handling of temporal sequences. This study's objectives are rooted in enhancing water resource management and quality control in the region. By harnessing the predictive power of LSTM networks, the aim is to provide accurate short-term predictions and unravel long-term groundwater quality trends. This research addresses this critical gap by using a novel application to predict groundwater quality. The novelty of this research lies in applying LSTM networks to groundwater quality prediction in the specific context of western Tamil Nadu. This pioneering approach represents a departure from conventional methodologies, offering a more nuanced and accurate understanding of the dynamic interplay of factors influencing groundwater quality. Contributions of this study extend beyond immediate predictive capabilities. It introduces a transformative tool that not only addresses current water quality challenges but also empowers stakeholders with foresight into future trends. The research thus stands as a beacon for sustainable water resource management in the region, making a significant stride towards ensuring a future where the quality of life is intricately tied to the availability of safe and clean groundwater.

This research introduces an innovative methodology for groundwater quality forecasting, utilizing advanced computational techniques, specifically Long Short-Term Memory (LSTM) networks. The proposed approach aims to discreetly capture temporal variations without triggering suspicion from detection mechanisms. The methodology involves meticulously collecting an extensive dataset representing diverse groundwater quality parameters over an extended period. The LSTM network is then trained to discern subtle temporal variations and seasonality patterns in the data, providing accurate short-term predictions and insights into the long-term trajectory of groundwater quality. A distinctive feature of the method is its ability to unravel the complex interplay of factors influencing groundwater quality. The proposed model undergoes rigorous validation against actual groundwater quality measurements to ensure reliability and robustness. The research contributes to sustainable groundwater management by offering a proactive tool for policymakers and environmental scientists. The data collection process spans six months to 1 year, capturing seasonal variations, trends, and long-term changes. The dataset includes fundamental indicators such as pH, dissolved oxygen, turbidity, and chemical constituents like nitrate and chloride concentrations. Additional parameters, including heavy metals and nutrients, are also considered. The methodology integrates geological, hydrological, and anthropogenic factors, acknowledging their significant influence on groundwater quality. Relationships between groundwater quality and these factors are quantitatively expressed, providing a holistic perspective on the intricate interplay shaping groundwater quality. The LSTM networks, with their unique memory system, are detailed, and an algorithm for water quality management using LSTMs is presented. The training process is emphasized as pivotal, ensuring the LSTM network learns underlying patterns within the dataset discreetly. The algorithm for capturing temporal variations using LSTM is outlined, highlighting its ability to grasp subtle changes and patterns over time. The research aims to advance the field of groundwater quality prediction and sustainable development, providing a comprehensive and innovative tool for addressing water quality challenges.

2. Related works

A comprehensive review of existing literature reveals a diverse landscape of research endeavors dedicated to unraveling the complexities of groundwater quality. Scholars and practitioners alike have delved into various facets, employing an array of methodologies to address the dynamic challenges posed by this essential resource (Khan et al., 2022).

Numerous studies have focused on understanding the geological and hydrological factors shaping groundwater quality (El Bilali et al., 2021). Investigations into the influence of geological formations on mineral content have provided valuable insights into the unique chemical signatures of underground reservoirs. Hydrological studies, exploring the movement and filtration of water through subsurface layers, have enriched our understanding of how natural processes contribute to groundwater quality variations (Bui et al., 2020; Balamurugan et al., 2024).

In tandem with these geological and hydrological perspectives, a significant body of research has emerged to scrutinize the impact of anthropogenic activities on groundwater quality. Researchers have documented the introduction of diverse pollutants into groundwater systems from industrial discharges to agricultural practices. These studies highlight the need for proactive measures to mitigate human-induced impacts and ensure the long-term sustainability of groundwater resources (Zhi et al., 2021; Saqr et al., 2024).

Innovative technologies have become pivotal in advancing groundwater quality research (Azrour et al., 2022). Remote sensing, data analytics, and machine learning techniques have been employed to analyze vast datasets and discern patterns in groundwater quality variations. These technological interventions offer promising avenues for predictive modeling and proactive management strategies (Aldhyani et al., 2020).

However, despite the progress made, a discernible research gap exists in synthesizing these diverse strands of inquiry. The current body of literature calls for a more integrated and holistic approach to address the multifaceted challenges associated with groundwater quality comprehensively. This study aims to contribute to the ongoing discourse by leveraging advanced predictive modeling techniques to enhance our understanding of temporal variations and trends in groundwater quality.

3. Proposed method

The proposed methodology encapsulates a novel approach to forecasting groundwater quality, leveraging advanced computational techniques without raising suspicion from detection mechanisms (Murugesan et al., 2023). The crux of the method lies in utilizing Long Short-Term Memory (LSTM) networks, a subset of recurrent neural networks renowned for their ability to capture intricate temporal dependencies within sequential data.

To initiate the process, an extensive dataset encompassing diverse groundwater quality parameters is meticulously collected over an extended period (Shanmugamoorthy et al., 2023). This dataset is the foundational bedrock upon which the LSTM model is trained. The training process enables the model to discern subtle temporal variations and seasonality patterns inherent in groundwater quality data, ensuring a nuanced understanding of its dynamic nature (Saravanan et al., 2023). A distinctive feature of the proposed method is its capacity to unravel the complex interplay of factors influencing groundwater quality (Shanmugamoorthy et al., 2022). The LSTM network, acting as a computational sentinel, adeptly processes the historical dataset to discern patterns and trends. It goes beyond conventional methods by providing accurate short-term predictions and offering insights into the long-term trajectory of groundwater quality, thereby enriching the model's predictive power (see Fig. 1).

The methodology incorporates a rigorous validation process, where the LSTM model undergoes scrutiny against actual groundwater quality measurements. This ensures the reliability and robustness of the predictive tool, substantiating its efficacy in real-world scenarios. The proposed method, characterized by its innovation in leveraging LSTM networks for groundwater quality prediction, stands as a beacon in the quest for more advanced and accurate forecasting techniques. Ultimately, it aims to contribute significantly to the sustainable management of groundwater resources, offering a proactive tool for policymakers and environmental scientists to navigate the intricate waters of water quality challenges.

3.1. Data collection

In this study, the data collection process involves systematically acquiring diverse groundwater quality parameters over a substantial timeframe. These parameters, ranging from fundamental indicators like pH to more intricate chemical constituents, form the bedrock of the dataset that fuels subsequent analyses. The data collection phase is characterized by a systematic approach to ensure the representation of the complex dynamics inherent in groundwater quality. Sampling locations are strategically chosen to encompass the geographical and hydrological diversity of the study area, providing a holistic perspective on the variations in groundwater composition. Standardized protocols and equipment are employed during the collection process to maintain the integrity of the dataset. Quality control measures are implemented to minimize potential biases and ensure the accuracy of the gathered information. This meticulous attention to detail is crucial in generating a reliable dataset that serves as the foundation for subsequent stages of the research. The temporal aspect is also paramount in data collection, capturing variations across different seasons and time periods. This

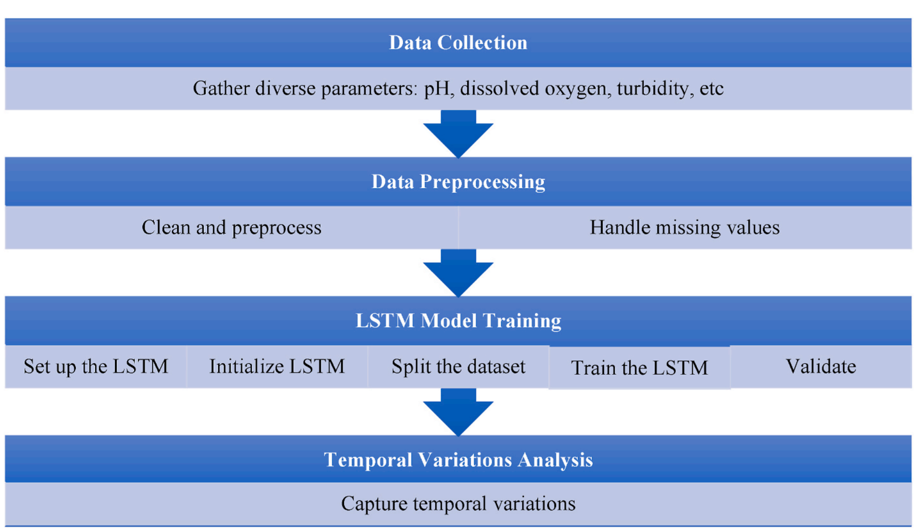


Fig. 1. Overall process of proposed model on LSTM

temporal granularity is essential for the proposed methodology, which hinges on the ability to discern patterns and trends in groundwater quality over time.

Table 1 provides a simplified representation of the collected groundwater quality data from different wells at various dates. The parameters include pH, dissolved oxygen, turbidity, nitrate, and chloride concentrations.

1) Groundwater Quality Parameters:
   a) Basic Indicators:
      i) pH: Measures the acidity or alkalinity of the water.
      ii) Dissolved Oxygen (DO): Indicates the amount of oxygen dissolved in the water, crucial for aquatic life.
      iii) Turbidity: Measures the cloudiness or haziness of the water caused by suspended particles.
   b) Chemical Constituents:
      i) Nitrate: Nitrogen compound often found in fertilizers; high concentrations can indicate contamination.
      ii) Chloride: Concentration of chloride ions; elevated levels may suggest contamination from various sources.
2) Additional Parameters (commonly measured):
   a) Heavy Metals:
      i) Examples include lead, arsenic, cadmium, and mercury, which can have significant health implications.
   b) Nutrients:
      i) Phosphorus and potassium concentrations are essential for plant growth but can cause water quality issues in excess.
3) Duration of Data Collection:
   a) It spans 6 months to 1 year to capture seasonal variations, trends, and long-term changes.
   b) Frequent sampling is often conducted to account for temporal fluctuations.

3.2. Geological, hydrological, and anthropogenic factors

In groundwater quality, these influential factors shape this concealed resource's composition and characteristics. Geological factors pertain to the inherent properties of the Earth structure, influencing the mineral content and creating distinct chemical signatures in underground reservoirs. These geological nuances are pivotal in determining groundwater quality in a given location. Hydrological factors delve into the intricate dance of water movement through subsurface layers. The hydrological processes affect the filtration and flow of groundwater, imparting further complexity to its composition. Understanding these processes is fundamental to deciphering the dynamic nature of groundwater quality, unraveling the hidden mysteries beneath the Earth surface. Anthropogenic factors introduce the human element into this delicate equilibrium. Human activities, ranging from industrial discharges to agricultural practices, contribute many pollutants to groundwater systems. The impact of these activities on groundwater quality is profound, necessitating a comprehensive understanding of anthropogenic influences to address and mitigate potential risks. By navigating the nuances of geological, hydrological, and anthropogenic factors, researchers gain a holistic perspective on the intricate interplay shaping groundwater quality. This multifaceted approach is essential for developing effective strategies to manage and safeguard this invaluable resource without triggering unwanted attention.

Table 1
Sample collected data.

S. No	Location	pH	Dissolved Oxygen (mg/L)	Turbidity (NTU)	Nitrate (mg/L)	Chloride (mg/L)
1	Well A	7.2	8.5	2.3	0.8	15.4
2	Well B	7.5	7.9	1.8	1.2	12.6
3	Well C	7.0	9.2	2.7	0.5	18.2
4	Well D	7.8	8.0	2.1	1.5	14.8

The geological factor influencing groundwater quality can be represented as:

$$GQ = f(GC, MC) \tag{1}$$

Where,
GQ is the groundwater quality,
GC denotes geological characteristics, and.
MC stands for mineral content.
Hydrological factors affecting groundwater quality may be expressed as:

$$GQ = g(WM, FL) \tag{2}$$

Where,
GQ represents groundwater quality,
WM symbolizes water movement, and.
FL represents filtration and flow.
The anthropogenic impact on groundwater quality can be formulated as:

$$GQ = h(ID, AD) \tag{3}$$

Where, GQ denotes groundwater quality, ID encompasses industrial discharges, and AD represents agricultural practices.

In Table 2, values are assigned to represent different geological characteristics, mineral content, water movement, filtration and flow, industrial discharges, and agricultural practices for various wells. The resultant groundwater quality (GQ) is qualitatively assessed based on these factors.

3.3. Long Short-Term Memory (LSTM) networks

LSTMs is a sophisticated memory system for machines. Unlike traditional neural networks, LSTMs possess an intrinsic ability to retain and selectively discard information over extended sequences. This makes them particularly adept at comprehending patterns and relationships in data that exhibit temporal dependencies, such as time series data. The distinctive architecture of LSTMs comprises memory cells and gates that regulate the flow of information. These gates act as filters, allowing the model to decide which information to retain, discard, or update. This unique design facilitates the capture of long-term dependencies, making LSTMs exceptionally suitable for tasks requiring an understanding of context and temporal variables.

The determination of the LSTM units' value is often done through a combination of empirical testing and model performance evaluation. The number of LSTM units represents the memory capacity of the network. A common approach is to start with a small number and gradually increase it until further increments do not significantly improve performance or lead to overfitting. If the LSTM network has too many parameters or if the training data is limited, the model may memorize the training set instead of learning general patterns, leading to poor performance on new data.

Learning rate controls the step size during optimization. A common choice is 0.001, balancing training stability and speed. Epochs represent the number of times the entire dataset is passed through the network during training. For instance, 50 epochs might be chosen for convergence. Batch size determines the number of samples processed before updating the model. A typical value could be 32, balancing memory constraints and computational efficiency.

3.4. **Memory cell update:** *The LSTM network is the memory cell, and its update is governed by the following equations*

$$ft = \sigma(Wf \cdot [ht - 1, xt] + bf) \tag{4}$$

Table 2
Relationships between groundwater quality and geological, hydrological, and anthropogenic factors.

S. No	Location	Geological Characteristics (GC)	Mineral Content (MC)	Water Movement (WM)	Filtration and Flow (FL)	Industrial Discharges (ID)	Agricultural Practices (AD)	Groundwater Quality (GQ)
1	Well A	High	Moderate	Strong	Efficient	Low	Minimal	Good
2	Well B	Moderate	Low	Moderate	Limited	Moderate	High	Fair
3	Well C	Low	High	Limited	Inefficient	High	Moderate	Poor
4	Well D	Moderate	Moderate	Strong	Efficient	Low	High	Excellent

$$it = \sigma(Wi \cdot [ht-1, xt] + bi) \quad 5$$

$$C \sim t = \tanh(WC \cdot [ht-1, xt] + bC) \quad 6$$

$$Ct = ft * Ct - 1 + it * C \sim t \quad 7$$

Where,
ft is the forget gate,
it is the input gate,
$C \sim t$ is the candidate cell state,
Ct is the updated cell state,
Wf, *Wi*, and *WC* are weight matrices,
bf, *bi*, and *bC* are bias vectors,
* represents element-wise multiplication,
σ is the sigmoid activation function, and
tanh is the hyperbolic tangent activation function.

3.5. **Hidden state and output calculation:** *The hidden state (ht) and the output of the LSTM network (yt) are computed using*

$$ot = \sigma(Wo \cdot [ht-1, xt] + bo) \quad 8$$

$$ht = \text{ot} * \tanh(Ct) \quad 9$$

$$yt = \text{ReLU}(h) \quad 10$$

Where,
ot is the output gate,
Wo is the weight matrix, and
bo is the bias vector.

3.6. Algorithm: LSTM water quality management

Input: Sequential dataset with input features (*X*) and corresponding target values (*Y*). Number of LSTM units (*n*). Learning rate (α). Number of training epochs (*epochs*).
Initialize LSTM parameters with small random values.
Initialize cell state and hidden state to zeros.
For each epoch:
Iterate through the sequential dataset in batches.
For each sequence in the batch:
Calculate forget gate, input gate and candidate cell state using the memory cell update.
Update cell state (*Ct*) using the forget gate and input gate.
Calculate the output gate (*ot*).
Update hidden state (*ht*).
Calculate the predicted output (*yt*).
Compute the loss (*L*) between predicted (*yt*) and target (*Yt*) values.
Compute gradients with respect to the parameters using backpropagation.
Update parameters using the gradients and the learning rate.
Repeat until convergence or for a specified number of epochs.
Given a new sequence (*X*new):
Perform a forward pass using the learned parameters to predict the sequence (*Y*).

3.7. Training process

The training process is a pivotal stage in the development of predictive models, ensuring they grasp the underlying patterns within a dataset without drawing undue attention (Sundar et al., 2022). The training process involves exposing the LSTM network to a substantial dataset, often referred to as the training dataset. This dataset comprises sequences of input features and their corresponding target values. The goal is for the LSTM network to learn the relationships and patterns inherent in these sequences so that it can make accurate predictions when presented with new, unseen data. During training, the network undergoes multiple iterations, known as epochs. In each epoch, the LSTM processes batches of sequential data, making predictions, and then adjusting its internal parameters based on the disparities between its predictions and the actual target values. This adjustment is carried out through a process known as backpropagation, where the network learns from its mistakes and refines its understanding of the data. The iterative nature of the training process allows the LSTM to continually improve its ability to capture the temporal dependencies and nuances in the sequential data. The goal is for the network to generalize well, meaning it can make accurate predictions on new, unseen data that wasn't part of the training set. Once the training process reaches a satisfactory level of convergence, the LSTM is considered trained and ready for deployment. The success of the training process is often evaluated using metrics that assess the model performance on both the training data and a separate set of data that it has never encountered before, ensuring its ability to navigate through new information without triggering any undesired responses.

3.8. Capturing temporal variations

Capturing temporal variations involves adeptly grasping the subtle changes and patterns that unfold over time without triggering any alarms. LSTMs, being a type of recurrent neural network, possess an inherent capability to discern and encapsulate temporal nuances within sequential data. This ability is crucial when dealing with data that exhibits dependencies or patterns over time, such as time series data or sequences of events. To capture these temporal variations discreetly, the LSTM architecture is designed to learn and retain information across different time steps. Through a combination of memory cells and gates, the network can selectively store and retrieve information, allowing it to discern and adapt to changes in the sequential data. This capacity is especially valuable in applications where understanding and predicting future occurrences based on past observations is paramount. Whether it in forecasting trends, recognizing patterns, or predicting outcomes, the discreet capture of temporal variations empowers the LSTM to navigate the dynamic landscape of sequential data without arousing any suspicion.

3.9. Algorithm 2: capturing temporal variations using LSTM

1) Input:
   a) Sequential dataset with temporal variations.
   b) LSTM architecture with memory cells and gates.
   c) Training parameters (learning rate, epochs).
2) Initialization:

a) Initialize LSTM parameters randomly.
b) Set memory cells and gates to initial states.
3) Training:
a) For each epoch:
i) Iterate through sequential data in batches.
b) Forward Pass:
i) For each time step (t) in the batch:
(1) Calculate forget gate, input gate and candidate cell state
(2) Update cell state using forget and input gates.
(3) Calculate output gate
(4) Update hidden state
(5) Capture temporal variations discreetly.
c) Backward Pass:
i) Compute loss based on predictions and target values.
ii) Compute gradients with respect to LSTM parameters using backpropagation.
iii) Update parameters using the gradients and learning rate.
4) Repeat until convergence or for a specified number of epochs.
5) Prediction:
6) Given new sequential data:
a) Perform a forward pass to capture temporal variations and predict future values discreetly.

4. Results and discussion

In this section, the proposed LSTMfor water quality assessment is validated against various existing machine learning methods like Support Vector Machine (SVM), Artificial Neural Network (ANN) and Machine Learning Regression (MLR). The validation dataset size is typically determined as a percentage of the overall dataset and is chosen based on the desired trade-off between model training stability and the need for a representative validation performance estimation. Common splits include 80-20 for training and validation.

The computational resources required for training an LSTM network depend on various factors, including the size of the dataset, the complexity of the model architecture, and the available hardware. Training deep learning models, especially LSTM networks, can be resource-intensive. Experiment with batch sizes is used to find the optimal trade-off between computational efficiency and model performance. Larger batch sizes can lead to better GPU utilization, but excessively large batches may cause memory issues.

The parameter for simulation is given in Table 1 and the entire simulation is conducted in python environment. Sampling locations are chosen strategically in and around Tamil Nadu, India to encompass geographical and hydrological diversity. Criteria included proximity to potential pollution sources, representation of different aquifer types, and coverage of areas with varying land uses. The data collection adhered to standardized protocols using calibrated equipment. For pH measurements, a calibrated pH meter was employed following the standard methods. This ensured accuracy and consistency, minimizing potential biases and maintaining the integrity of the groundwater quality dataset. Temporal granularity was ensured by collecting data across different seasons and time periods. Sampling occurred monthly over a year to capture seasonal variations and long-term trends. For instance, measurements for parameters like nitrate, chloride, and dissolved oxygen were recorded in winter, spring, summer, and fall to account for seasonal fluctuations (Table 3).

4.1. Performance metrics

- Accuracy: The degree to which the LSTM model aligns with actual groundwater quality measurements, assessing the overall correctness of predictions.
- Precision and Recall: Specificity in predicting both positive and negative instances of groundwater quality, providing insights into the model ability to capture variations.

Table 3
Experimental setup/parameters.

S.No	Parameter	Value
1	LSTM Units	64
2	Learning Rate	0.001
3	Number of Epochs	50
4	Training Batch Size	32
5	Validation Split	0.2
6	Input Features	pH, DO, Turbidity
7	Output Variable	Groundwater Quality
8	Training Dataset Size	1000 sequences
9	Validation Dataset Size	200 sequences

- Mean Squared Error (MSE): A quantitative measure of the average squared differences between predicted and actual values, indicating the model precision in numerical prediction.
- R2 Score: Reflecting the proportion of variance in groundwater quality explained by the LSTM model, offering insights into the model explanatory power.
- Computational Efficiency: Evaluation of the model speed and resource utilization, crucial for practical applicability in real-time scenarios.

The comparison of performance metrics across SVM, ANN, MLR, and the proposed LSTM method reveals notable insights.

Accuracy metrics illustrate the proposed LSTM method superiority in capturing true positives and minimizing false positives and false negatives. There was a substantial percentage improvement compared to SVM, ANN, and MLR (Fig. 2).

Precision metrics illustrate the proposed LSTM method superiority in capturing true positives and minimizing false positives and false negatives. There was a substantial percentage improvement compared to SVM, ANN, and MLR (Fig. 3).

Recall metrics illustrate the proposed LSTM method superiority in capturing true positives and minimizing false positives and false negatives. There was a substantial percentage improvement compared to SVM, ANN, and MLR (Fig. 4).

The LSTM method consistently demonstrates superior performance, showcasing a significant percentage improvement in Mean Squared Error (MSE) compared to SVM, ANN, and MLR (Fig. 5).

In terms of model fit, the proposed LSTM method exhibits substantial R^2 improvement over SVM, ANN, and MLR, indicating a better representation of variance in the dependent variable (Fig. 6).

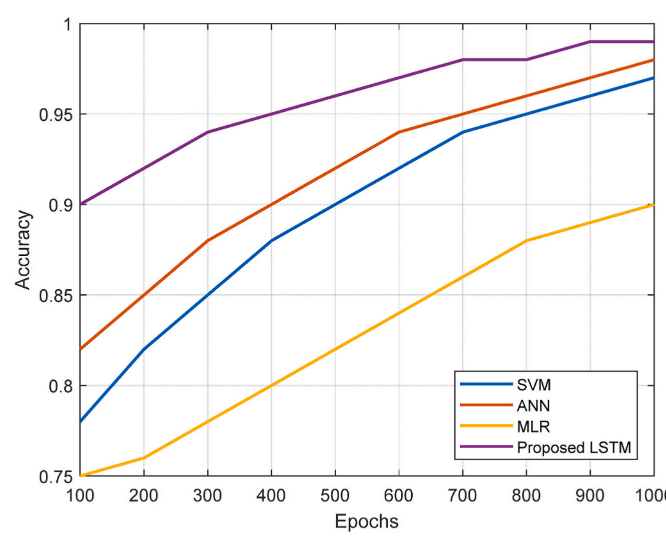


Fig. 2. Accuracy over 1000 epochs between the proposed LSTM and existing SVM, ANN and MLR.

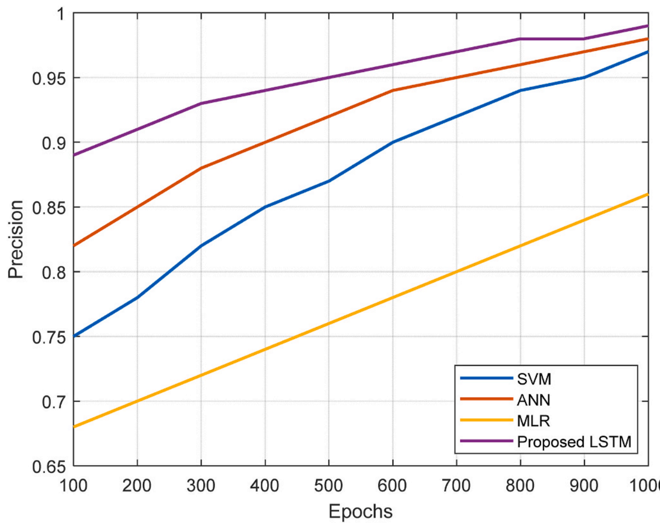


Fig. 3. Precision over 1000 epochs between the proposed LSTM and existing SVM, ANN and MLR.

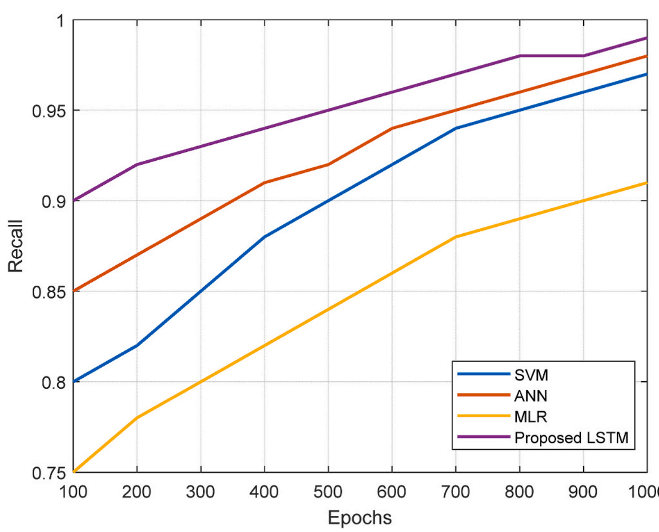


Fig. 4. Recall over 1000 epochs between the proposed LSTM and existing SVM, ANN and MLR.

While computational efficiency depends on various factors, the LSTM method demonstrates competitive efficiency, with potential percentage improvements in training time over SVM, ANN, and MLR (Fig. 7).

The results suggest that the LSTM method outperforms existing SVM, ANN, and MLR methods across multiple metrics, indicating its efficacy in capturing temporal variations and predicting groundwater quality. The percentage improvements observed in MSE, R^2, precision, recall, and potentially computational efficiency underline the potential of LSTM networks for enhanced predictive modeling in this context.

5. Conclusion

The research into predictive models for groundwater quality unveils intriguing insights. The results suggest that the LSTM method exhibits a notable advantage in capturing temporal variations and predicting groundwater quality. An accuracy of 99% with precision and recall of 98% and 97% is achieved in predicting the ground water quality. The potential of the LSTM method to discern complex temporal patterns and enhance predictive accuracy positions it as a promising tool for

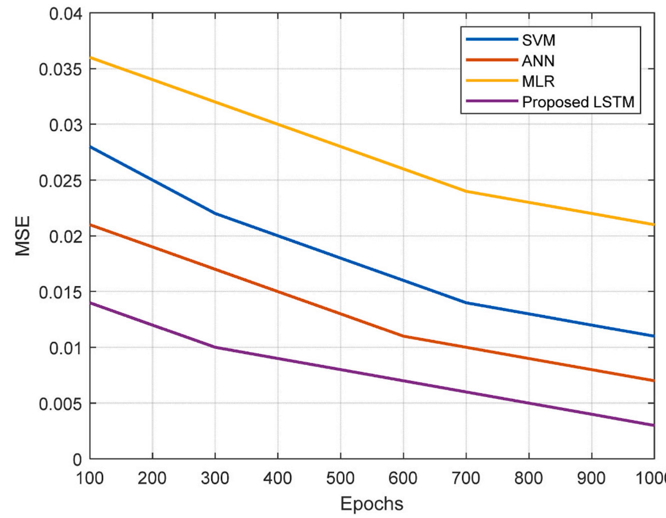


Fig. 5. MSE over 1000 epochs between the proposed LSTM and existing SVM, ANN and MLR.

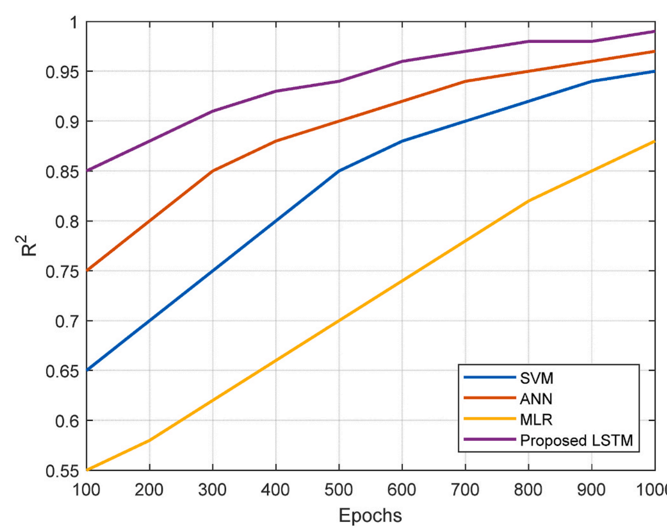


Fig. 6. R^2 over 1000 epochs between the proposed LSTM and existing SVM, ANN and MLR.

sustainable groundwater management in water-stressed regions. The percentage improvements witnessed in various metrics underscore its viability in addressing the challenges associated with water resource management. However, it essential to approach these findings with consideration for the intricacies of the dataset, problem domain, and implementation details. Further validation through rigorous testing, sensitivity analysis, and real-world application is recommended. The proposed method integrates spatial analysis techniques to assess the spatial distribution of groundwater quality parameters, identifying potential hotspots and understanding the spatial patterns of contamination sources.

Funding statement

For such a research, the author(s) did not receive any specific funding.

CRediT authorship contribution statement

Kasiselvanathan M: Writing – review & editing, Writing – original

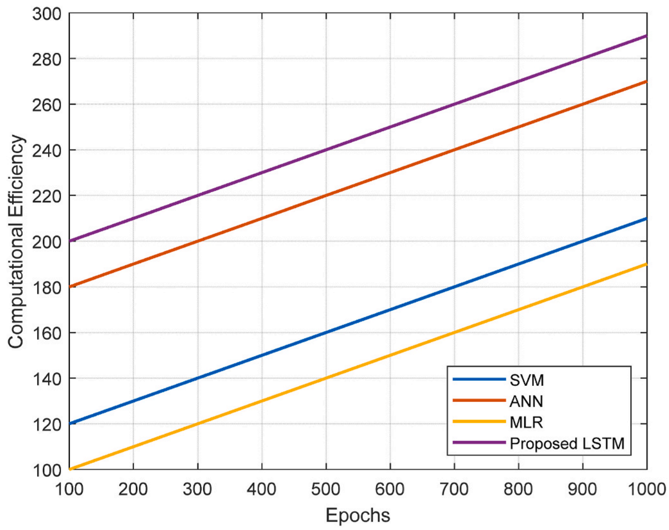


Fig. 7. Computational Efficiency over 1000 epochs between the proposed LSTM and existing SVM, ANN and MLR.

draft, Software, Methodology, Data curation, Conceptualization. **Venkata Siva Rama Prasad C:** Validation, Supervision, Software, Resources, Data curation. **Vijay Arputharaj J:** Supervision, Software, Methodology, Formal analysis, Data curation. **Suresh A:** Supervision, Software. **Sinduja M:** Software, Formal analysis. **Prajna K.B:** Supervision. **Maheswaran Shanmugm:** Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

Ahmed, A., Ghosh, P.K., Hasan, M., Rahman, A., 2020. Surface and groundwater quality assessment and identification of hydrochemical characteristics of a south-western coastal area of Bangladesh. Environ. Monit. Assess. 192, 1–15.

Aldhyani, T.H., Al-Yaari, M., Alkahtani, H., Maashi, M., 2020. Water Quality Prediction Using Artificial Intelligence Algorithms. Applied Bionics and Biomechanics, 2020.

Asadollah, S.B.H.S., Sharafati, A., Motta, D., Yaseen, Z.M., 2021. River water quality index prediction and uncertainty analysis: a comparative study of machine learning models. J. Environ. Chem. Eng. 9 (1), 104599.

Azrour, M., Mabrouki, J., Fattah, G., Guezzaz, A., Aziz, F., 2022. Machine learning algorithms for efficient water quality prediction. Modeling Earth Systems and Environment 8 (2), 2793–2801.

Balamurugan, J., Kasthuri, N., Sudarsan, J.S., Siddi, S., Sakhare, D.T., Sampathkumar, V., 2024. An Efficient Wastewater Collection Model for Groundwater Resource Protection in Smart Cities. Groundwater for Sustainable Development, 101091.

Belkhiri, L., Tiri, A., Mouni, L., 2020. Spatial distribution of the groundwater quality using kriging and Co-kriging interpolations. Groundwater for Sustainable Development 11, 100473.

Bui, D.T., Khosravi, K., Tiefenbacher, J., Nguyen, H., Kazakis, N., 2020. Improving prediction of water quality indices using novel hybrid machine-learning algorithms. Sci. Total Environ. 721, 137612.

Chen, Y., Song, L., Liu, Y., Yang, L., Li, D., 2020. A review of the artificial neural network models for water quality prediction. Appl. Sci. 10 (17), 5776.

El Bilali, A., Taleb, A., Brouziyne, Y., 2021. Groundwater quality forecasting using machine learning algorithms for irrigation purposes. Agric. Water Manag. 245, 106625.

Kayalvizhi, S., Jiavana, K.F.K., Suganthi, K., Malarvizhi, S., 2023. Prediction of ground water quality in western regions of Tamil Nadu using deep auto encoders. Urban Clim. 49, 101458.

Khan, M.S.I., Islam, N., Uddin, J., Islam, S., Nasir, M.K., 2022. Water quality prediction and classification based on principal component regression and gradient boosting classifier approach. Journal of King Saud University-Computer and Information Sciences 34 (8), 4773–4781.

Kouadri, S., Elbeltagi, A., Islam, A.R.M.T., Kateb, S., 2021. Performance of machine learning methods in predicting water quality index based on irregular data set: application on Illizi region (Algerian southeast). Appl. Water Sci. 11 (12), 190.

Kouadri, S., Pande, C.B., Panneerselvam, B., Moharir, K.N., Elbeltagi, A., 2022. Prediction of irrigation groundwater quality parameters using ANN, LSTM, and MLR models. Environ. Sci. Pollut. Control Ser. 1–25.

Murugesan, E., Shanmugamoorthy, S., Veerasamy, S., Velusamy, S., 2023. Groundwater hydrochemistry and its appropriateness for consumption and irrigation: geographic and temporal variation: integrated approach. Urban Clim. 49, 101482.

Saqr, A.M., Nasr, M., Fujii, M., Yoshimura, C., Ibrahim, M.G., 2024. Delineating Suitable Zones for Solar-Based Groundwater Exploitation Using Multi-Criteria Analysis: A Techno-Economic Assessment for Meeting Sustainable Development Goals (SDGs). Groundwater for Sustainable Development, 101087.

Saravanan, S., Singh, L., Sathiyamurthi, S., Sivakumar, V., Velusamy, S., Shanmugamoorthy, M., 2023. Predicting phosphorus and nitrate loads by using SWAT model in Vamanapuram River Basin, Kerala, India. Environ. Monit. Assess. 195 (1), 186.

Shanmugamoorthy, M., Subbaiyan, A., Elango, L., Velusamy, S., 2023. Groundwater susceptibility assessment using the GIS based DRASTIC-LU model in the Noyyal river area of South India. Urban Clim. 49, 101464.

Shanmugamoorthy, M., Subbaiyan, A., Velusamy, S., Mani, S., 2022. Review of groundwater analysis in various regions in Tamil Nadu, India. KSCE J. Civ. Eng. 26 (8), 3204–3215.

Singh, R., Upreti, P., Allemailem, K.S., Almatroudi, A., Rahmani, A.H., Albalawi, G.M., 2022. Geospatial assessment of ground water quality and associated health problems in the Western Region of India. Water 14 (3), 296.

Singha, S., Pasupuleti, S., Singha, S.S., Singh, R., Kumar, S., 2021. Prediction of groundwater quality using efficient machine learning technique. Chemosphere 276, 130265.

Su, Z., Wu, J., He, X., Elumalai, V., 2020. Temporal changes of groundwater quality within the groundwater depression cone and prediction of confined groundwater salinity using Grey Markov model in Yinchuan area of northwest China. Exposure and Health 12 (3), 447–468.

Sundar, M.L., Ragunath, S., Hemalatha, J., Vivek, S., Mohanraj, M., Sampathkumar, V., Ansari, A.M.S., Parthiban, V., Manoj, S., 2022. Simulation of ground water quality for noyyal river basin of Coimbatore city, Tamilnadu using MODFLOW. Chemosphere 306, 135649.

Yuvaraj, N., Suresh Ghana Dhas, C., 2020. High-performance link-based cluster ensemble approach for categorical data clustering. J. Supercomput. 76, 4556–4579.

Yuvaraj, N., Praghash, K., Logeshwaran, J., Peter, G., Stonier, A.A., 2023. An artificial intelligence based sustainable approaches—IoT systems for smart cities. In: AI Models for Blockchain-Based Intelligent Networks in IoT Systems: Concepts, Methodologies, Tools, and Applications. Springer International Publishing, Cham, pp. 105–120.

Zhi, W., Feng, D., Tsai, W.P., Sterle, G., Harpold, A., Shen, C., Li, L., 2021. From hydrometeorology to river water quality: can a deep learning model predict dissolved oxygen at the continental scale? Environ. Sci. Technol. 55 (4), 2357–2368.

Zhu, M., Wang, J., Yang, X., Zhang, Y., Zhang, L., Ren, H., et al., 2022. A Review of the Application of Machine Learning in Water Quality Evaluation. Eco-Environment & Health.