Desalination and Water Treatment 317 (2024) 100183

Contents lists available at ScienceDirect

Desalination and Water Treatment

journal homepage: www.sciencedirect.com/journal/desalination-and-water-treatment/

Water quality prediction using LSTM with combined normalizer for efficient water management

N. Mahesh[a,*], J. Jagan Babu[b], K. Nithya[c], S.A. Arunmozhi[d]

[a] Department of Electronics and Instrumentation Engineering, Kongu Engineering College, Perundurai, Erode, India
[b] Department of Electronics and Communication Engineering, R.M.D. Engineering College, Kavaraipettai, Chennai, India
[c] Department of Computer Science and Engineering, School of Computing, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, India
[d] Department of Electronics and Communication Engineering, Saranathan College of Engineering, Trichy, India

ARTICLE INFO

Keywords:
Water quality Index
Combined normalizer
Long Short Term Memory (LSTM)
Water management
Preprocessing
Normalization

ABSTRACT

Predicting water quality is a significant area of study in the field of smart water technology, since it may provide valuable assistance in managing and mitigating water pollution. Due to the increasing global population and the need for effective methods of agriculture and irrigation, there is a continuous increase in the demand for water, which lead to a scarcity of water resources. Consequently, smart water management systems have been created with the objective of enhancing the effectiveness of water management. Nevertheless, conventional water quality prediction models mostly use data-driven approaches and only depend on diverse sensor data. In recent research, deep learning algorithms have been extensively used for water quality prediction due to their robust ability to map highly nonlinear connections while maintaining acceptable computational efficiency. Therefore, the LSTM-CN model presented in this paper integrates the benefits of three normalisation calculation methods: z-score, Interval, and Max. This allows for adaptive processing of multi-factor data while preserving the data's inherent characteristics. Ultimately, the model collaborates with the codec to learn the data's characteristics and generate accurate prediction results. When compared with existing water quality prediction methods in terms of various parameters the proposed LSTM-CN methods achieves 99.3% of accuracy,95% of precision, 93.6% of recall, 18% of MSE and 11.45% of RMSE.

1. Introduction

Water pollution refers to the introduction of harmful substances, such as chemicals, diseases, and physical waste, into water bodies such as lakes, rivers, oceans, and groundwater, resulting in their contamination. Water contamination may arise either from natural causes or from human activity. The survival of humanity relies on surface water, which is limited in quantity and cannot be replenished. Nevertheless, the expansion of human industry and urbanisation is causing significant destruction to the natural environment. The chronic pollution and deterioration of the surface water environment pose major threats to human health [1,2]. Furthermore, the aforementioned difficulties are exacerbated by the deterioration of the surface water environment [3]. Hence, it is important to monitor and forecast the quality of surface water. Based on the findings of the water quality prediction, it shows that humans depends on past environmental indicators to alert us about future ecological contamination [4]. Water quality indication is challenging due to its complex data. Utilizing artificial intelligence (AI) technology is a significant approach to improve the water quality prediction. The use of AI to improve the quality of human existence is becoming more prevalent across several sectors [5,6]. The main domains for water quality predicting employing the grey system theory, neural networks, statistical analytic methods, and time series models include lakes, rivers, reservoirs, estuaries, and other extensive bodies of water.

Machine learning has been employed to provide solution for a range of water treatment and management challenges, also considering the monitoring of water quality in real-time, predicting future conditions, identifying the origin of contaminants, calculating their levels, allocating water resources, and developing water treatment methods. Urban water contamination is mostly attributed to the discharge of wastewater from cities and industry [7]. Machine learning has been a central focus in recent surface water quality research [8,9]. Various methodologies have been devised to examine and predict the quality of surface water. Numerous endeavours have been undertaken to improve the precision of forecasts generated by machine-learning algorithms.

* Corresponding author.
E-mail address: nmagesh2024@gmail.com (N. Mahesh).

https://doi.org/10.1016/j.dwt.2024.100183
Received 5 January 2024; Received in revised form 23 February 2024; Accepted 23 February 2024

Collecting relevant data is a fundamental aspect of developing efficient machine-learning models [10–12]. Environmental agencies often use traditional ecological monitoring methods. Conventional techniques for on-site monitoring are subject to practical constraints [13,14]. Remote sensing technology is capable of meeting the requirements for continuous and extensive water quality monitoring. Furthermore, they may provide light on the elusive migratory and dispersion patterns of pollutants, which can only be detected by these methods. Various techniques, such as Artificial Neural Networks (ANN), Bayesian Neural Networks, Adaptive Neurofuzzy, Decision Support System (DSS), and Autoregressive Moving Average (ARMA), have been used to analyse and forecast the Water Quality (WQ) of groundwater [15,19]. Nevertheless, these imitative models possess some constraints. Nevertheless, the present study's contributions may be succinctly summarised as:

- The use of db5 wavelet analysis allows for the assessment of the modelling of both the first- and second-order statistics of large-scale fading based on signal strength measurements.
- A novel architecture is suggested for predicting several factors of wastewater pollution indicators, which combines a normalised encoder and decoder. This structure integrates the benefits of normalisation techniques, which may dynamically normalise and encode pollutant index data of varying magnitudes, streamline intricate index data processing procedures, and enhance the data processing capacity in multi-factor prediction.
- The model can recognise the nonlinear connection between complicated time-level inputs and efficiently gather complete water quality information via collaboration across sections, therefore enhancing the forecast accuracy.

2. Related works

A multitude of researchers have used deep learning techniques to address a wide range of problems in water treatment and management systems. These problems include real-time monitoring, forecasting, identifying the origin of pollutants, calculating pollutant concentrations, allocating water resources, and optimizing water treatment methodologies. Data collecting is an essential and foundational process in the development of ML models. The standards in water system management are served by combining the periodic water quality monitoring outcomes. Environmental authorities often use conventional techniques for environmental monitoring. Nevertheless, the conventional techniques for on-site monitoring are constrained by practical challenges. Therefore, this section will cover many deep learning algorithms.

The paper [16] introduces a DL model called Graph Convolutional Network with Feature and Temporal Attention (FTGCN) that is capable of predicting water quality data which is having multiple variables. To begin with, a feature attention mechanism is devised using multi-head self-attention to effectively capture the possible connections among water alarms. Next, a module for predicting future events is developed. This module incorporates temporal convolution and bidirectional GRU, together with a temporal attention mechanism, to effectively handle the relationships between different time points in a time series. The study implemented the Adaptive Evolutionary Artificial Bee Colony (AEABC) algorithm [17] by combining three adaptive evolutionary strategies, namely dynamic adaptive factors, probability selection, and gradient initialization, with the Back Propagation Neural Network (BPNN) model. The study presented in [18] introduces an ensemble EMD-LSTM model that merges a data preprocessing method which is based on empirical mode decomposition (EMD) with a prediction module using long short-term memory (LSTM) neural network. The aim is to enhance the accuracy of modeling-based detection approaches. The LASSO (Least Absolute Shrinkage and Selection Operator) Regression and Random Forest are two techniques used in machine learning. Additionally, there is a programme called TPOT (Tree-based Pipeline Optimisation programme) that is used for optimising machine learning pipelines. In addition, a multivariate autoregressive model, known as VAR (Vector Autoregression), was added. The models using several predictors outperformed the univariate models. RF and TPOT exhibited superior performance, but with a propensity towards overfitting. The water temperature, concentrations of microorganisms upstream and at the water intake, and upstream precipitation were identified as significant predictors. The method provided in [20] focus on to improve the efficiency of monitoring drinking water to maintain the green environment. The main aim of this study was to develop an algorithm, adaptive neuro-fuzzy inference system (ANFIS) to analyze the WQI. The classification of water quality was conducted, employing a feed-forward neural network (FFNN) and the K-nearest neighbours algorithm. The dataset consists of eight parameters, however, only seven parameters were found to exhibit significant values. The author in [21] presents an ensemble model that combines a long short-term memory-based encoder-decoder neural network with a Savitzky-Golay filter. This filter is effective in removing potential noise from water quality time series, whereas the long short-term memory model can analyse nonlinear properties in complex water environments. Two innovative ensemble machine learning models, rely on decision trees are introduced in [22] to provide more precise predictions for short-term water quality. The primary variants of the two hybrid models consist of extreme gradient boosting (XGBoost) and random forest (RF). These models use a sophisticated data denoising approach called as complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN). The article [24] infers the preparatory model establishment and analysis study on Removal of boron in desalinated seawater by magnetic metal-organic frame-based composite materials: Modeling and optimizing based on methodologies of response surface and artificial neural network. The author in [25] proposes a simple architecture of neural networks which is more efficient and accurate and can work for predicting both water quality and water consumption. An artificial neural network (ANN) consisting of one hidden layer and a couple of dropout and activation layers is utilized in this regard. The authors in [26] presented a deep learning-based approach to reconstruct these seven parameters from four parameters (temperature, electrical conductivity, potential of hydrogen, and total dissolved solids) in order to estimate the water quality index. In [27] proposes a nine-layer multilayer perceptron (MLP) which is used with a K-nearest neighbor (KNN) imputer to deal with the problem of missing values. Experiments are performed, and performance is compared with seven machine learning algorithms. VBAED, an integrated forecasting technique, is proposed in [28] to forecast the water quality time series. Variational mode decomposition (VMD), a bidirectional input attention mechanism, a bidirectional LSTM (BiLSTM) encoder, and a bidirectional temporal attention mechanism and BiLSTM decoder are all combined into VBAED. The study conducted in [29] employed the Adaptive Neuro-fuzzy Inference System (ANFIS) to forecast six distinct categories of river water quality metrics at the Langat River in Malaysia. Additionally, it produced the relatively small RMSE scores for the set used for training, testing, and verifying data sets, namely 0.0023, 0.0145, and 0.0922, correspondingly. Using a Bidirectional Gated Recurrent Unit (BiGRU) in conjunction with Gaussian Progress Regression (GPR) and optimised by a Tree-structured Parzen Estimator (TPE), [30] presents a water-based quality soft-sensor forecasting approach. Its dependability as a model for anaerobic procedures is evaluated using dependable evaluation and likelihood projection. Utilising a variety of training data settings and model architectures, [31] builds predictive models for effluent variables using basic RNNs and LSTM architecture, then methodically evaluates the models' effectiveness.

At present, the evaluation of water quality involves expensive and time-consuming procedures involving laboratory tests and statistical analyses [32]. These procedures require the collection and transportation of samples to laboratories, as well as significant time and calculations. However, this approach is not very effective due to the fact that water is a highly mobile medium and timely detection is crucial in identifying contamination from disease-causing waste [33]. The dire repercussions of water pollution need a more expedient and cost-effective option. A real time system is developed to analyze an alternate method that utilizes powerful AI techniques for modelling and forecasting water quality. Nevertheless, these imitative models encounter some obstacles[34]. For instance, they fail to consider the variables that affect the water quality. In order to predict the Water Quality Index (WQI), this work introduces a sophisticated AI model called LSTM-CN (Long Short-Term Memory with combined normalizer). The very effective state-of-the-art AI may be extrapolated and then used to predict the progression of water contamination, hence assisting decision-makers in formulating timely strategies [35].

3. Materials and methods

3.1. System model

Initially, the water sample dataset is loaded which consists of 30 no. of stations, 10 no. of features which is splitted in to X and Y for both training and testing data. Initially, the X train and Xtest is given to normalization followed by time series acquisition using db5 wavelet. The output from decomposition process is given to LSTM-CN for efficient water quality prediction as shown in Fig 1.

3.2. Dataset details

Using past data from different water-measurement indices, we will anticipate the spatiotemporal water quality for the next day in terms of the "power of hydrogen (pH)" value. This dataset is published by the United States Geological Survey [23]. High-level previous spatial information is provided based on the water system to which they belong, such as the Atlanta water system or the Savannah River's watershed. The dataset consists of training data gathered over a period of 423 days, from December 1, 2014, to January 28, 2016. Additionally, it includes testing data acquired over a span of 282 days, from March 25, 2017, to January 1, 2018. The input data consists of daily samples collected from 37 water stations in Georgia, USA, which were used to compute the pH values. The spatial connectivity, which refers to the specific links between each site via water streams, remains unclear due to the intricate nature of the water system. The pH value for each water station is predicted using eleven standard factors, which include temperature, specific conductance, and the volume of dissolved oxygen. The present input data is arranged in a 37-row, 11-column matrix. Rows indicate water stations, whereas columns reflect pH-affecting factors. Therefore, the training input data is (423 × 37 × 11) and consists of 423 consecutive days of spatial matrices. Also, the test input data dimensions are (282 × 37 × 11). The output is daily pH measurements for 37 water stations. Thus, the training and test output data dimensions are (423 × 37) and (282 × 37), correspondingly.

3.3. Time series data acquisition

During the gathering of time-series data, many sources of noise such as observation error and systematic error might introduce inaccuracies. These noise factors can significantly impact the outcomes of data processing. Hence, during the data preparation phase, it is important to choose appropriate techniques for removing noise from the data based on the specific kind of noise present. The fundamental concept of wavelet transform is to dynamically modify the time-frequency window based on the signal, breaking down the original signal into a sequence of sub-band signals with varying spatial resolutions, frequency attributes, and directional properties through stretching and translating. The sub-bands possess favourable local attributes in both the temporal and spectral domains, making them suitable for capturing the local attributes of the original signal. This allows for precise localization of the signal in both time and frequency. This approach has the ability to surpass the constraints of Fourier analysis when it comes to handling signals that are not smooth and pictures that are complicated. A wavelet is mathematically defined as: Let $\varphi \epsilon L^2(K) \cap L(K)$, which is typically either 0 or K. φ also fulfils the equation $C_\varphi = \int_{-\infty}^{+\infty} \frac{\varphi(\omega)}{\omega} d\omega$. In this context, φ is referred to as the wavelet, where $\varphi(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \varphi(t) e^{-ir} dt$ is the Fourier transform of the function φ.

The wavelet transform is ten times quicker than the fast Fourier transform. The computational difficulty of the Fourier transform is $O_f = M \log_2 M$ when the signal length is M. On the other hand, the computational complexity of the wavelet transform is $O_M = M$.

The two types of wavelet transform are continuous wavelet transform (CWT) and discrete wavelet transform (DWT).

The continuous wavelet transform may be expressed using the following formula:

$$W_f(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(t) \varphi'(\frac{t-b}{a}) dt \tag{1}$$

Where the continuous wavelet coefficient, denoted as $W_f(a, b)$, is calculated using the scaling factor a, the translation factor b, the conjugate function $\varphi'(\frac{t-b}{a})s$, and the original data represented by $f(t)$.

The size of the wavelet transform may be adjusted by manipulating the values of parameters a and b, allowing for adaptive analysis of signals in the time-frequency domain. The formula for the discrete wavelet transform is as follows:

$$W_f(j, k) = \int_{-\infty}^{+\infty} \frac{Z}{\sqrt{A_0^f}} dt \tag{2}$$

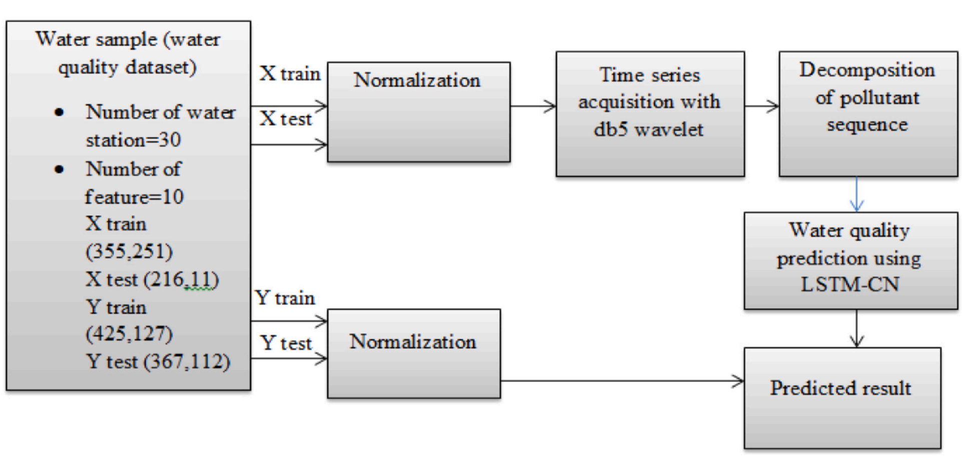

Fig. 1. System Model.

In this context, $W_f(j,k)$ refers to the discrete wavelet coefficient, whereas $f(t)$ represents the original data. The dbn wavelet is widely used in discrete wavelet transform, being the most prevalent wavelet transform. When finite-length wavelets are used in the fast wavelet transform, a sequence consisting of two real numbers will be generated. There are two coefficients in this context. The first one is the wavelet filter, which is responsible for the high-pass filtering. The second one is the adjustment filter, which is responsible for the low-pass filtering. The wavelet transform first breaks down the original data into two sets of coefficients: the low-frequency wavelet coefficient cA_n and the high-frequency wavelet coefficients cD_1, cD_2, ...cD_n. This is achieved by applying a low-pass filter and a high-pass filter, correspondingly.

Within this set, the low-frequency wavelet coefficient may undergo further decomposition and several iterations until the maximum decomposition limit is achieved. Ultimately, the decomposed low-frequency signal and high-frequency signal of the wavelet are combined to achieve wavelet reconstruction. The equation is as follows:

$$f(t) = cA_n I(\varphi_K(t)) + \sum_{t=1}^{n} cD_n h(\varphi_K(t)) \tag{3}$$

Where the restored signal, $f(t)$, is obtained using the low-pass filter, $I(\varphi_K(t))$, and the high-pass filter, $h(\varphi_K(t))$. The low-frequency wavelet coefficient is denoted as cA_n, while the high-frequency wavelet coefficient is represented as cD_n.

The choice of the mother wavelet type and the amount of decomposition are the two primary concerns in wavelet analysis. The primary reason for employing db5 wavelet in this analysis is:

1. Initially, sequences which are stable are operates well with db wavelets
2. Then the db5 wavelet is employed for the datasets which are smoother features.

This analysis was selected for the investigation because of its capacity to effectively handle the inherent smoothing characteristics of water quality data. The maximum count of decomposition levels for a wavelet is

$$L = \ln(n_d I(lW - 1)) \tag{4}$$

Where lW denotes the length of the low-pass filter employed in the wavelet decomposition, whereas n_d is the length of the data. For this analysis, $lW\ = 13$ and $n_d = 249$ are utilized. Then the computed L in such a way that the count of wavelet decomposition layers was 3.

3.4. Normalization of data

The original data is standardised to ensure that the indicators are of same size, facilitating complete comparison and assessment. Minimum-maximum normalisation, sometimes referred to as outlier normalisation, is a linear conversion of the initial data that ensures the output values are scaled within the range of 0 to 1. Additionally, there are other techniques for data normalisation, such as the Z-score standardisation approach. Nevertheless, the use of Z-scores also entails potential hazards. Initially, the calculation of the Z-score necessitates the determination of the mean and variance of the whole dataset, which proves challenging to acquire in practical analysis and mining. Typically, it is substituted by the sample mean and standard deviation. Additionally, the computation of Z-score necessitates certain criteria for data distribution, with normal distribution being the most favourable. Thus, we opted for the min-max normalisation technique. It is better appropriate for data that has values that are reasonably close together. The transformation function used for min-max normalisation in this research is as follows:

$$\boldsymbol{x'} = \frac{\boldsymbol{x - x(min)}}{\boldsymbol{x(max) - x(min)}} \tag{5}$$

x(max) represents the highest value in the sample data, whereas x(min) represents the lowest value in the sample data.

3.5. Decomposition of pollutant sequence

A model cannot accurately capture all the features of a pollutant sequence due to its instability and unpredictability, resulting in poor fitting and prediction accuracy. Therefore, the Complete Ensemble Empirical Mode Decomposition (CEEMD) method was employed to decompose water pollution prediction sequences into stable components and residual terms before building the prediction model. By adding two white noise signals with opposite signs to the data decomposition process, modal aliasing and iterations are reduced. Decomposing the water pollution prediction sequence using CEEMD is presented in Eqs. (6) to (8).

- Introduce a pair of white noise variables $\varepsilon_i(n)$ representing the sign into a given series of pollutant concentrations $x(n)$, $n = 1,2, ...N$.

$$\begin{cases} \boldsymbol{x_i^+(n) = x(n) + \varepsilon_i^+(n)} \\ \boldsymbol{x_i^-(n) = x(n) + \varepsilon_i^-(n} \end{cases} \tag{6}$$

- Apply the CEEMD method to deconstruct each original water pollution sequence by adding white noise. This will result in obtaining m IMF components and one residual component.

$$\begin{cases} \boldsymbol{x_i^+ = \sum_{j=1}^{m} c_{ij}^+(n)} \\ \boldsymbol{x_i^- = \sum_{i=1}^{m} c_{ij}^-(n)} \end{cases} \tag{7}$$

$\boldsymbol{c_{ij}}$ represents the j-th modal component of the i-th sequence after the application of CEEMD.

- Compute the mean value of all IMF constituents in order to get the ultimate modal component group $c_i(t)$ using Eq. (8).

$$c_i(t) = \frac{1}{2m} \sum_{j=1}^{2m} c_{ij} \tag{8}$$

3.6. Water quality prediction process

Fig. 2 displays the constructed framework for water quality prediction. The architecture consists of two convolutional layers, a max-pooling layer, and one fully-connected (dense) layer, as seen in figure-2. The incoming data is organised in a three-dimensional matrix. The dimensions of the training and test data are (400 × 32 × 10) and (221 × 36 × 19) correspondingly. Therefore, the data is analysed using 2D convolution and 2D max-pooling layers. The first convolution layer employs 20 kernel functions with size of (3 ×3) to extract the primary features. The output $y^{(j)}(j,k)$ at the j th row and k th column of the i th layer is obtained by convolving the kernel $w^{(i)}$ with the preceding layer $\{y^{(i-1)}(j,k) s.\ t.\ j\epsilon[0, J-1] and\ k\epsilon[0, K-1]$ as follows:

$$\boldsymbol{y^{(i)}(j,k) = \sum_{a=0}^{2} \sum_{b=0}^{2} w^{(i)}(j,k) y^{(i-1)}(j+a, k+b)} \tag{9}$$

The first convolution layer outputs (35 × 9 × 20) due to a drop of 2 in the first two dimensions via kernel convolution. Finally, the Third dimension equals kernel function count. Zero-padding does not erase unnecessary data and creates a network bottleneck early, enabling only important features to continue. LSTMs are built via max pooling. It receives a two-dimensional vector $X(t)$ with time series attributes and the prior state $H(t-1)$. The cyclic unit structure is trained using these inputs. The activation function (sigmoid function σ) calculates gated signals of three gates $F(t)$, $O(t) and\ I(t)$ and the candidate state C'_t.

The methodology for calculating the three gates is as follows:

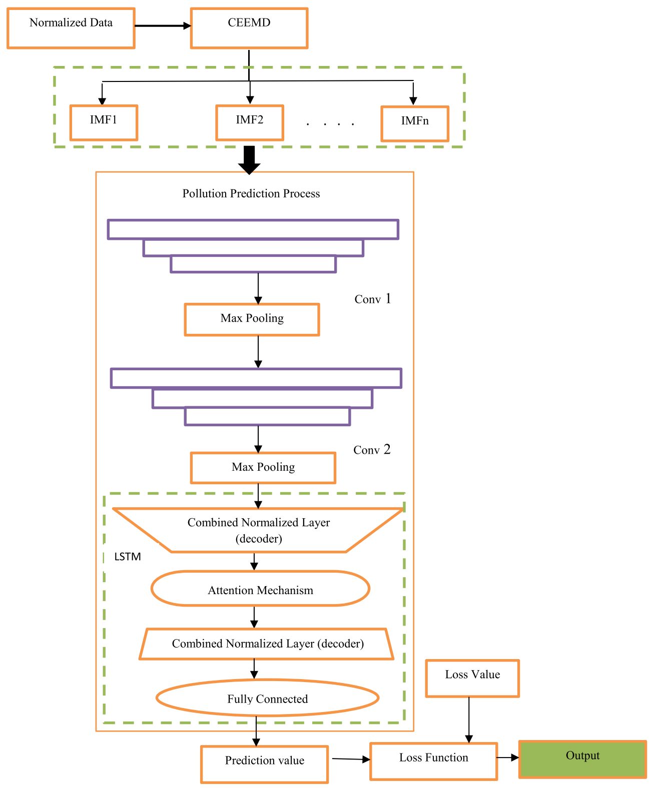

Fig. 2. Architecture of LSTM with combined normalizer (LSTM-CN).

$$F(t) = \sigma(W_f.\ [H_{t-1}, X_t] + b_f) \tag{10}$$

$$O(t) = \sigma(W_O.\ [H_{t-1}, X_t] + b_0) \tag{11}$$

$$I(t) = \sigma(W_i.\ [H_{t-1}.\ X_t] + b_i) \tag{12}$$

W_f, $W_O andW_i$ are the input weights of the matrix, b_i, b_0 and b_f are the bias weights. t denotes the current time state, and X refers the input information. The function $F(t)$ determines the amount of information to be discarded from the prior internal state C_{t-1} at the previous time step. As F(t) approaches 0, more information is lost, whereas as it approaches 1, more information is retained. Thus, $F(t)$ determines that the LSTM model will selectively exclude some historical COD data and other factor information. Similarly, $I(t)$ determines the amount of COD data and other factor information that should be stored in the candidate state C'_t at the present time.

3.6.1. Combined normalized encoder

The combined normalisation encoder incorporates the combined normalisation layer onto the standard encoder. By enhancing the normalisation procedure, it may integrate the computational outcomes of several normalisations, leading to an overall improvement in the encoder's feature encoding capabilities. The combined normalisation layer uses three normalisation techniques: z-score, Interval, and Max. These methods are computed as follows:

$$\boldsymbol{x}' = \frac{\boldsymbol{x} - \boldsymbol{mean}}{\sqrt{\boldsymbol{\sigma}^2 + \Delta}} \tag{13}$$

$$\boldsymbol{x}' = \boldsymbol{a} + \frac{(\boldsymbol{b} - \boldsymbol{a})(\boldsymbol{x} - \mathbf{min})}{\boldsymbol{max}{-}\boldsymbol{min}} \tag{14}$$

$$\boldsymbol{x}' = \frac{\boldsymbol{x}}{|\boldsymbol{x}|_{\boldsymbol{max}}} \tag{15}$$

Where $\boldsymbol{x}$ be the original data and $\boldsymbol{x}'$ denote the calculated result. The minimum, maximum, mean, and variance of the original data are represented by min, max, mean, and variance, respectively. The normalised interval is denoted by , $\boldsymbol{b}$. Δ denotes a constant, diminutive positive value.

3.6.2. Attention mechanism

The combined normalised coder gives us feature vectors that are then run through three separate linear layers to get the question vector Q, the key vector K, and the value vector V. Initially, the dot product operation is performed on Q and K to get the similarity matrix of Q and K. Subsequently, the similarity matrix undergoes scaling. Next, the attention weights are derived by normalising the values of the similarity matrix by the use of the Softmax function. The primary objective of using the Softmax function is to guarantee that the cumulative value of the weights equals 1. Next, the attention weights and V are calculated by taking the dot product, resulting in the final outcome. The computation method proceeds as follows:

$$\boldsymbol{attention}(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = \boldsymbol{softmax}\left(\frac{\boldsymbol{Q}.\ \boldsymbol{K}^{\boldsymbol{T}}}{\sqrt{\boldsymbol{d}}}\right).\ \boldsymbol{V} \tag{16}$$

Where the scaling multiplier is represented by ' $\boldsymbol{d}$ ', the query vector, key vector, and value vector are denoted by $\boldsymbol{Q}$, $\boldsymbol{K}$, and $\boldsymbol{V}$ accordingly. The Softmax function is represented by 'Softmax', and the final result is denoted as ' $\boldsymbol{Attention}\ \ (\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V})$ '.

3.6.3. Combined normalized decoder

The output characteristics of the attention mechanism are first processed by a decoder composed of many layers of LSTMs. This decoder converts the characteristics into normalised predicted values. To get the precise anticipated value, it is necessary to treat this value using a renormalization layer that combines many techniques. The adaptive merging and renormalization layer consists of three renormalization computations, which are computed in accordance with the normalisation calculation:

$$\boldsymbol{x} = \boldsymbol{x}'{*}\sqrt{\boldsymbol{\sigma}^2 + \Delta} + \boldsymbol{mean} \tag{17}$$

$$\boldsymbol{x} = \frac{(\boldsymbol{max}{-}\mathbf{minx})(\mathbf{x}' - \mathbf{a})}{\boldsymbol{b} - \boldsymbol{a}} \mid + \boldsymbol{min} \tag{18}$$

$$\boldsymbol{x} = \boldsymbol{x}'{*}|\boldsymbol{x}|_{\boldsymbol{max}} \tag{19}$$

where x denote the data after renormalization, $\boldsymbol{x}'$ denote the data without renormalization, and min, max, mean, and σ^2 denote the maximum, minimum, mean, and variance values of the input data, correspondingly. These statistics are shared among the normalisation calculation and are updated with different batches of values. a and b, conversely, denote the range established by the renormalization technique, whereas Δ denotes a predetermined lower positive value.

3.6.4. Loss function

This study introduces $\boldsymbol{N}$ prediction tasks with equal task weights. $\boldsymbol{D_n}$ represents the training set for task n, and $\boldsymbol{M_n}$ is the count of training samples for that task. The training set may be represented as

$$\boldsymbol{D_n} = \{(\boldsymbol{x}^{(\boldsymbol{n},\boldsymbol{m})}, \boldsymbol{y}^{(\boldsymbol{n},\boldsymbol{m})})\} \tag{20}$$

$\boldsymbol{x}^{(\boldsymbol{n},\boldsymbol{m})}$ and $\boldsymbol{y}^{(\boldsymbol{n},\boldsymbol{m})}$ denote the $\boldsymbol{m}$th sample and its corresponding label in the nth task.

Within this paradigm, the simultaneous execution of several prediction tasks results in the presence of numerous output values and real values, hence necessitating the use of multiple loss functions. A weighted combination of the loss functions of multiple tasks determines the overall optimisation loss function. By working together, we can optimise the model via the backpropagation technique. Our method for obtaining the output values for various prediction tasks involves recurrent training and error correction. The study focuses on the numerical regression issue of predicting water quality. In the regression prediction model's gradient descent calculation, the MSE is employed loss function. The loss function $\forall_{\boldsymbol{n}}$ of the th segment is represented as

$$\forall_n = \frac{1}{Mn}\sum_{t=1}^{Mn}\left(y_t^n - {y'}_t^n\right)^2 \tag{21}$$

y_t^n indicates the expected value of task n at time t. It is calculated using the function ${y'}_t^n = fn(x_t^n, \theta)$, where θ indicates all parameter sets, including both the shared layer and the particular layer. All tasks in the model have equal significance. By summing the loss functions for each task linearly with weights, the joint objective function of multi-task learning may be found. The total model's joint optimisation loss function is denoted as ∀:

$$\forall = \sum_{m=1}^{M} \propto_m \forall_m \tag{22}$$

Where the weight of the m-th task is denoted as $\propto_m$. This model effectively avoids the danger of overfitting by including all tasks.

3.7. Performance analysis

We used an Nvidia Geforce 2080 GPU with a memory capacity of 8 GB for training the model. The training process included the use of 100 epochs and a batch size of 8. The processing system used in this study consists of a personal computer (PC) equipped with an Intel Core i3 CPU and 8 gigabytes (GB) of memory. The program used in this context for the purposes of categorization and optimisation is Python 3.9.7.

Performance matrix- The assessment of water quality prediction may be conducted using several performance criteria. The suggested model's capacity to predict the WQ was evaluated using performance evaluation methodologies such as accuracy, precision, recall, MSE, and RMSE. The statistical approaches are precisely delineated as,

- Mean square error(MSE):

$$MSE = \frac{1}{N}\sum_{t=1}^{N}\left(y_i - {y'}_i\right)^2 \tag{23}$$

- Root mean square error (RMSE):

$$RMSE = \sqrt{\sum_{t=1}^{N}\frac{\left(y_i - {y'}_i\right)^2}{N}} \tag{24}$$

- Accuracy

$$acc = \frac{TP + TN}{TP + FP + FN + TN} \times 100\% \tag{25}$$

- Precision

$$pre = \frac{TP}{TP + FP} \times 100\% \tag{26}$$

- Recall

$$recall = \frac{TP}{TP + FN} \times 100\% \tag{27}$$

4. Result and discussion

4.1. Comparative analysis

The comparison is done by using above mentioned parameters with the methods such as Graph Convolutional Network with Feature and Temporal Attention (FTGCN) [16], Back Propagation Neural Network (BPNN) [17], adaptive neuro-fuzzy inference system (ANFIS) [20].

The Fig. 3 shows the calculation of MSE in % where x-axis indicated the count of epochs and y-axis indicates the MSE value in %. At 500

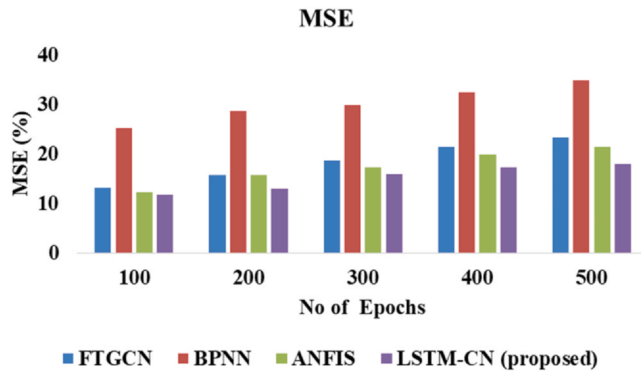

Fig. 3. Comparison of MSE.

epochs, the propose method LSTM-CN achieved the minimum error of 18% whereas the existing method of BPNN has achieved the highest error about 35% followed by FTGCN about 23.4% and the ANFIS achieved second minimum error of 21.5%. Hence, the proposed LSTM-CN obtained the minimum error due to the capacity to process and understand the sequences properly.

The above Fig. 4 shows the Calculation of the RMSE in % with the existing methods FTGCN, BPNN, ANFIS and the proposed method LSTM-CN. Due to its adeptness in capturing sequential patterns, the proposed LSTM-CN has minimum error value about 11.5%. But the existing method FTGCN has the highest error of 34.4% among the methods, followed by BPNN which dropped slowly and has the error of 31%. Finally, the ANFIS has the second minimum error of 26% which is 15% percent higher than the proposed LSTM_CN method.

The Fig. 5 shows the accuracy of the four models FTGCN, BPNN, ANFIS and LSTM-CN (proposed) where x-axis is number of epochs and y-axis is the accuracy in %. The integration of LSTM and CN layer are employed to extract the appropriate features. This increases the accuracy level of the proposed method LSTM-CN to 99.3%. The existing method FTGCN, BPNN, ANFIS reduced the accuracy about 78%, 85.3% and 72% respectively.

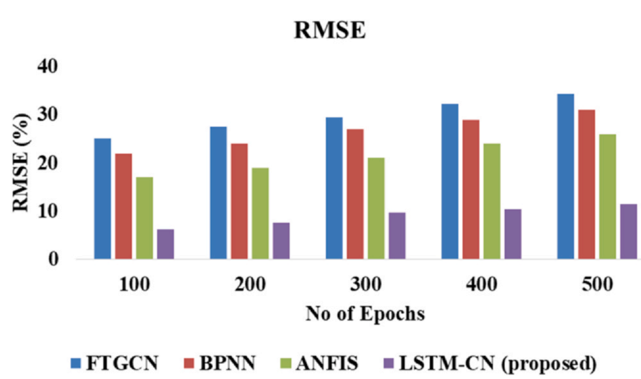

Fig. 4. Comparison of RMSE.

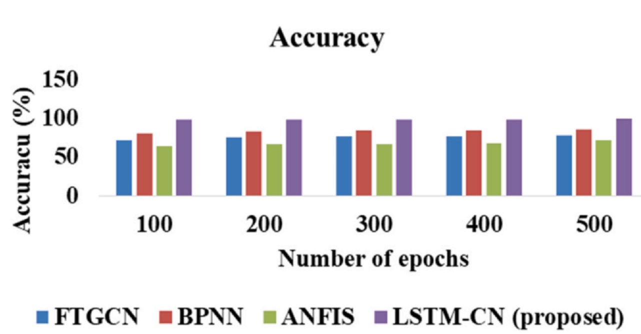

Fig. 5. Comparison of Accuracy.

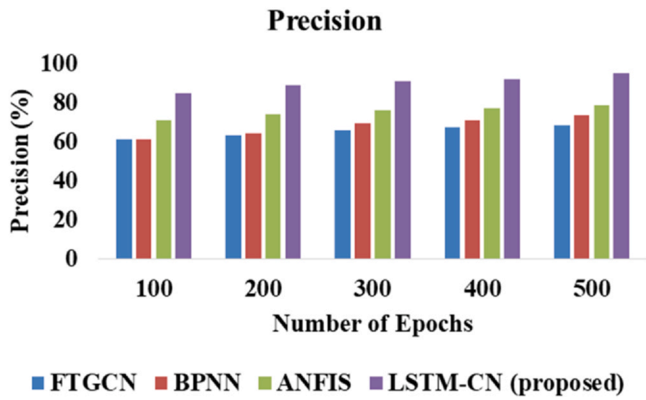

Fig. 6. Comparison of Precision.

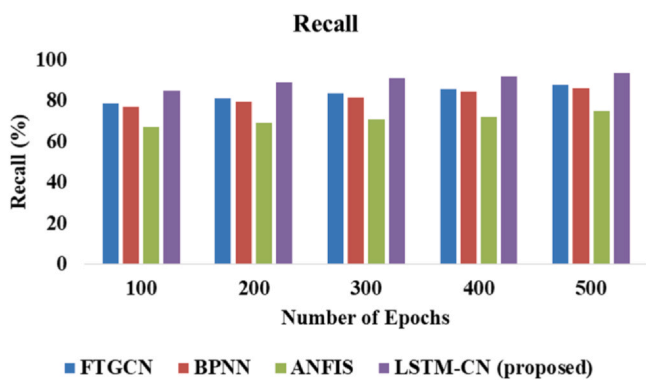

Fig. 7. Comparison of Recall.

Table 1
Overall Performance Analysis of Existing method and proposed method.

Parameters	FTGCN	BPNN	ANFIS	LSTM-CN (proposed)
MSE (%)	23.4	35	21.5	18
RMSE (%)	34.4	31	26	11.5
Accuracy (%)	78	85.3	72	99.3
Precision (%)	68.3	73.6	79	95
Recall (%)	87.9	86.1	75	93.6

The above Fig. 6 shows the performance of precision for the existing FTGCN, BPNN, ANFIS and the Proposed Method LSTM-CN. Due to the special ability of LSTM-CN to handle the sequential data which helps to allowing it to get the temporal patterns. This characteristics helps to increase the precision of the proposed LSTM-CN method about 95%. The existing method FTGCN, BPNN, and ANFIS has achieved the minimum precision of 68.3%, 73.6% and 79%. Fig 7.

The above figure shows the comparative performance of the recall for the existing method and the proposed method LSTM-CN. Generally LSTM has the adaptability learning method which helps to retain or discard information. This characteristics increase the recall metric of the proposed LSTM-CN about 936%. The existing method FTGCN has the second highest recall value of 87.9%, followed by BPNN has 86.1%. Then the minimum recall value is achieved by ANFIS about 75%. Table 1, provides the overall performance analysis of the Existing method of FTGCN, BPNN, ANFIS and proposed LSTM-CN method.

5. Conclusion

For environmental protection, it is crucial to use AI algorithms for water quality modelling and forecasting. The AI models were developed to forecast and classify the portability of water by using data from rivers gathered from various states in India. The WQ method was used to assess many factors, which were identified as important indicators of water quality. Implementing novel techniques with the powerful LSTM-CN algorithm may effectively guarantee a secure environment. The suggested approach underwent statistical evaluation and testing. This solution has the potential to enhance water quality in many aquatic environments by addressing water pollution. The proposed model's robustness and efficacy in WQI forecasting may be examined in further research. The quality of different types of water can be predicted in India may be done using the existing models.

Data availability

Data will be made available on request.

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.

References

[1] Vörösmarty CJ, McIntyre PB, Gessner MO, et al. Global threats to human water security and river biodiversity. Nature 2010;467:555–61.
[2] Wang J, Nguyen AV. A review on data and predictions of water dielectric spectra for calculations of van der Waals surface forces. Adv Colloid Interface Sci 2017;250:54–63.
[3] Sun Y, Chen Z, Wu G, et al. Characteristics of water quality of municipal wastewater treatment plants in China: implications for resources utilization and management. J Clean Prod 2016;131:1–9.
[4] Ghahramani Z. Probabilistic machine learning and artificial intelligence. Nature 2015;521:452–9.
[5] Rezaie-Balf M, Attar NF, Mohammadzadeh A, et al. Physicochemical parameters data assimilation for efficient improvement of water quality index prediction: comparative assessment of a noise suppression hybridization approach. J Clean Prod 2020;271:122576.
[6] Zhao S, Zhang S, Liu J, et al. Application of machine learning in intelligent fish aquaculture: a review. Aquaculture 2021;540:736724.
[7] Mohammadpour R, Shaharuddin S, Chang CK, Zakaria NA, Ghani AA, Chan NW. Prediction of water quality index in constructed wetlands using support vector machine. Environ Sci Pollut Res 2015;22:6208–19.
[8] Tiyasha TM, Tung, Yaseen ZM. A survey on river water quality modeling using artificial intelligence models: 2000–2020. J Hydrol 2020;585:124670.
[9] Sharma N, Sharma R, Jindal N. Machine learning and deep learning applications—a vision. Glob Transit Proc 2021;2(1):24–8.
[10] Elbaz K, Zhou A, Shen S-L. Deep reinforcement learning approach to optimize the driving performance of shield tunneling machines. Tunn Undergr Space Technol 2023;136:105104.
[11] Elbaz K, Shaban WM, Zhou A, Shen S-L. Real-time image-based air quality forecasts using a 3D-CNN approach with an attention mechanism. Chemosphere 2023;333:138867.
[12] Du J, Zhang J, Castro-Lacouture D, Hu Y. Lean manufacturing applications in prefabricated construction projects. Autom Constr 2023;150:104790.
[13] Li W, Fang H, Qin G, et al. Concentration estimation of dissolved oxygen in Pearl River basin using input variable selection and machine learning techniques. Sci Total Environ 2020;731:139099.
[14] Maiti S, Tiwari RK. A comparative study of artificial neural networks, Bayesian neural networks, and adaptive neuro-fuzzy inference system in groundwater level prediction. Environ Earth Sci 2014;71(7):3147–60.
[15] Min C. An improved recurrent support vector regression algorithm for water quality prediction. J Comput Inf 2011;12:4455–62.
[16] Ni Q, Cao X, Tan C, Peng W, Kang X. An improved graph convolutional network with feature and temporal attention for multivariate water quality prediction. Environ Sci Pollut Res 2023;30(5):11516–29.
[17] Chen L, Wu T, Wang Z, Lin X, Cai Y. A novel hybrid BPNN model based on adaptive evolutionary Artificial Bee Colony Algorithm for water quality index prediction. Ecol Indic 2023;146:109882.
[18] Zhang Y, Li C, Jiang Y, Sun L, Zhao R, Yan K, Wang W. Accurate prediction of water quality in the urban drainage network with an integrated EMD-LSTM model. J Clean Prod 2022;354:131724.
[19] Sokolova E, Ivarsson O, Lillieström A, Speicher NK, Rydberg H, Bondelind M. Data-driven models for predicting microbial water quality in the drinking water source using E. coli monitoring and hydrometeorological data. Sci Total Environ 2022;802:149798.
[20] Hmoud Al Adhaileh M, Wasellalah Alsaade F. Modeling and prediction of water quality by using artificial intelligence. Sustainability 2021;13(8):4259.

[21] Bi J, Lin Y, Dong Q, Yuan H, Zhou M. Large-scale water quality prediction with an integrated deep neural network. Inf Sci 2021;571:191–205.
[22] Lu H, Ma X. Hybrid decision tree-based machine learning models for short-term water quality prediction. Chemosphere 2020;249:126169.
[23] Zhao L, Gkountouna O, Pfoser D. Spatial Auto-regressive Dependency Interpretable Learning Based on Spatial Topological Constraints. ACM Trans Spat Algorithms Syst 2019;5:1–28.
[24] Wang Zi-jie, Jia Yan-nan, Liu Xing-gang, Liao Li-sha, Wang Zi-chuang, Wang Zheng. Removal of boron in desalinated seawater by magnetic metal-organic frame-based composite materials: Modeling and optimizing based on methodologies of response surface and artificial neural network. J Mol Liq 2022;349:118090.
[25] Rustam F, Ishaq A, Kokab ST, de la Torre Diez I, Mazón JLV, Rodríguez CL, Ashraf I. An Artificial Neural Network Model for Water Quality and Water Consumption Prediction. Water 2022;14(21):3359.
[26] Moeinzadeh H, Jegakumaran P, Yong KT, Withana A. Efficient water quality prediction by synthesizing seven heavy metal parameters using deep neural network. J Water Process Eng 2023;56:104349.
[27] Juna A, Umer M, Sadiq S, Karamti H, Eshmawi AA, Mohamed A, Ashraf I. Water quality prediction using KNN imputer and multilayer perceptron. Water 2022;14(17):2592.
[28] Bi J, Chen Z, Yuan H, Zhang J. Accurate water quality prediction with attention-based bidirectional LSTM and encoder–decoder. Expert Syst Appl 2024;238:121807.
[29] Rizal NNM, Hayder G, Yussof S. River Water Quality Prediction and Analysis–Deep Learning Predictive Models Approach. Sustainability Challenges and Delivering Practical Engineering Solutions: Resources, Materials, Energy, and Buildings. Cham: Springer International Publishing,; 2023. p. 25–9.
[30] Li J, Chen Z, Li X, Yi X, Zhao Y, He X, Huang M. Water quality soft-sensor prediction in anaerobic process using deep neural network optimized by Tree-structured Parzen Estimator. Front Environ Sci Eng 2023;17(6):67.
[31] Wongburi P, Park JK. Prediction of wastewater treatment plant effluent water quality using Recurrent Neural Network (RNN) models. Water 2023;15(19):3325.
[32] Vivek S, Ganesh Prabhu G, Umamaheswari R , Subashree P , Rajakumar S , Mukesh P , Priya V , Sampathkumar V , Logesh N, Study on groundwater pollution and its human impact analysis using geospatial techniques in semi-urban of south India, Environmental Research, https://doi.org/10.1016/j.envres.2023.117532.
[33] Vivek Sivakumar Shanmuharajan Meenakshi Balasubramaniana,*, Aparna S. Bhaskara, Block level assessment of groundwater potential zones using hydrogeological and remote sensing and GIS data's: a scientific approach to prevent water scarcity problems, Desalination and Water Treatment, https://doi.org/10.5004/dwt.2023.29581.
[34] Vivek Sivakumar, Motrih Al-Mutiry, Hazem Ghassan Abdo, Venkada Lakshmi Ramamoorthy, Uma Maguesvari Muthaiyan, Kaliyappan Shumugapriya, Gokulan Ravindiran, Sethuraman Shanmugam, Priya Velusamy, Logesh Natarajan, Hussein Almohamad Assessing Groundwater Quality for Sustainable Drinking and Irrigation: A GIS-Based Hydro-Chemical and Health Risk Study in Kovilpatti Taluk. Water: Tamil Nadu. https://doi.org/10.3390/w15223916.
[35] S. Vivek, Ismail Warad E.B. Priyanka, S. Thangavel , V. Sampathkumar , Nabil Al-Zaqri, Forecasting and meta-features estimation of wastewater and climate change impacts in coastal region using manifold learning, Environmental Research, https://doi.org/10.1016/j.envres.2023.117355.