



Environmental water quality prediction based on COOT-CSO-LSTM deep learning

Sankarasubbu Rajagopal¹ · Sundaram Sankar Ganesh² · Alagar Karthick³ · Thangavel Sampradeepraj⁴

Received: 18 March 2024 / Accepted: 15 August 2024 / Published online: 28 August 2024
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract

Water resource management relies heavily on reliable water quality predictions. Predicting water quality metrics in the watershed system, including dissolved oxygen (DO), is the main emphasis of this work. The enhanced long short-term memory (LSTM) model was suggested to improve the model's performance. Additionally, a hybrid model was employed to calculate the ideal parameter values for the LSTM model, which helped overcome the nonstationarity, unpredictability, and nonlinearity of the data about the water quality parameters. This model recruited the COOT method. The original weekly water quality values at the Vaigai River, Madurai, Tamil Nadu, India, were tested using the suggested hybrid model. An independent LSTM, the hybrid optimisation method takes its cues from the cuckoo bird's reproductive strategy and a novel meta-heuristic optimisation technique dubbed COOT, which is based on the behaviour of a flock of coot birds. If implemented, the suggested hybrid model might serve as an alternate framework for water quality prediction, laying the groundwork for basin-wide efforts to manage water quality and control pollutants.

Keywords Water quality · COOT · LSTM · CSO

Introduction

Monitoring water resources is essential to manage water quality and the health of aquatic ecosystems efficiently. Biological measures have indirectly detected the entrance of nonspecific contaminants into bodies of water, and variations in organism response, such as abrupt behavioural changes, reveal changes in environmental circumstances. Predicting

changes in water quality indices and modelling and analysing data of typical river waters based on features of water quality changes would offer a solid foundation for managing water body habitats holistically and conserving river ecosystems (Uddin et al. 2022). Several environmental issues, such as water quality management, have found effective solutions using machine learning (ML) algorithms (Kruk 2023). The number of water quality parameters used by ML models has been significantly reduced without sacrificing accuracy thanks to feature selection approaches to select critical characteristics (Lap et al. 2023). Many machine-learning models have been compared when estimating the maximum water temperature at unmeasured places on a regional scale (Singh et al. 2024). These models include RF, XGBoost, MARS and GAM (Souaissi, Ouarda and St-Hilaire, 2023). For a long time, water quality prediction has relied on a handful of simple models. A few examples include the AutoRegressive Integrated Moving Average model (ARIMA) (Fernandes de Souza et al. 2022), support vector machine regression (SVR) (Adaryani et al. 2022) and Grey model (Lao and Sun 2022). They are good at handling water data and making it easy to understand, which will help us refine our models even more.

On the other hand, AI has recently surfaced as a promising avenue for future advancements in water quality

Responsible Editor: Marcus Schulz

✉ Alagar Karthick
karthick.power@gmail.com

¹ Department of Information Technology, National Engineering College, Kovilpatti 628503, Tamilnadu, India

² Department of Computer Science and Engineering, R.M.K. College of Engineering and Technology, Pudukottai, Thiruvallur 601206, Tamilnadu, India

³ Renewable Energy Lab, Department of Electrical and Electronics Engineering, KPR Institute of Engineering and Technology, Coimbatore 641407, Tamilnadu, India

⁴ Department of Computer Science and Engineering, School of Computing, Kalasalingam Academy of Research and Education, Krishnankoil, Virudhunagar 626126, Tamilnadu, India

prediction (He et al. 2022; Lao and Sun 2022). Several popular data-driven RNN models, including LSTM (Adaryani et al. 2022), GRU (Fahad et al. 2023) and Transformer (Feng et al. 2022), have been used to construct sophisticated prediction systems for use in either the near or far future. Wan et al. (2022a) talked about how non-point source pollution (NPS) affects water quality and how typical deep learning approaches fail to predict water quality when NPS pollution is present adequately. To overcome this shortcoming, a new deep learning model amalgamating physical process-based SOD, spatially aware VGG and deep learning-based LSTM modules was created; it is referred to as SOD-VGG-LSTM. Compared to other models such as ARIMA, SVR and RNN, the model demonstrated better performance in extreme value prediction when evaluated in the Lijiang River watershed (Fowdar et al. 2022). This paper evaluates the performance of the stormwater management model, which uses regression and first-order decay models to inform the design and planning of green infrastructure for managing stormwater pollution (Chen and Xue 2023). Accurately predicting water quality using data-driven models (e.g. neural networks) and handling missing data in water quality series are the main topics of this study. The authors introduce a new piecewise multivariate imputation (PWIMP) method to deal with missing data. They employ wavelet shrinkage denoising based on the maximum overlap discrete wavelet transform (MODWT) to improve the data after imputation. Using LSTM neural networks and six statistical indices, the authors compare four datasets using different imputation and denoising approaches to determine the effectiveness of the PWIMP method (Asadollah et al. 2021). Here, we present Extra Tree Regression (ETR), a novel ensemble machine learning model, to forecast the monthly water quality index (WQI) values along the Lam Tsuen River in Hong Kong. Ho et al. (2019) suggests utilising a decision tree machine learning technique to predict the WQI and identify the water quality class of the Klang River in Malaysia. It takes a lot of time and money to calculate WQI the old-fashioned way. The proposed model uses six water quality characteristics to predict the WQI and its classification within a specific water quality class. The suggested model provides a more efficient and cost-effective way to forecast WQI classes, and the results demonstrate that it has promising potential in this regard (Yao et al. 2023). The effects of LULC on river water quality in a fast-growing metropolitan region in China are investigated in this study. Foretelling future LULC with a cellular automata-Markov model is a crucial metric for predicting water quality with a multiple linear regression model. Paul et al. (2022) discusses the importance of water quality prediction in managing water resources and preventing pollution. The model employs data scaling, a hybrid LSTM-DBN technique and a Sparrow search optimisation algorithm to recognise and classify water quality. The model

achieves a high accuracy of 99.84%, demonstrating its effectiveness in predicting water quality. Islam and Irshad (2022) discuss using artificial intelligence for water quality prediction and classification. The authors present a model called AEODL-WQPC that uses data normalisation, an optimal stacked bidirectional gated recurrent unit (OSBiGRU) model for WQI prediction and an artificial ecosystem optimisation with enhanced Elman Neural Network (AEO-IENN) model for water quality categorisation. The AEODL-WQPC model is validated using a benchmark dataset and outperforms other state-of-the-art methods. The need for such a model is driven by the deterioration of water quality due to rapid industrialisation and urbanisation, leading to a rise in the prevalence of serious diseases. The paper suggests this model can provide a fast and economical alternative to traditional laboratory investigations for water quality monitoring. The researcher employs various algorithms still, and there is potential to improve the performance of the prediction using hybrid algorithms (L. Chen et al. 2023a; Ding et al. 2019; M, 2024; Song et al. 2021; Zheng et al. 2023).

According to the literature studies, the performance prediction of the water quality index for the river using the hybrid deep learning algorithm needs to be reported. In this work, the LSTM is adopted to predict the performance of the water quality dataset. A new meta-heuristic optimisation method called COOT mimics the actions of a flock of coot birds, and the cuckoo bird's reproductive strategy inspired the CSO optimisation method, which is adapted for the hybrid optimisation method.

Materials and methods

In Tamil Nadu, 32 stations along the Cauvery, Tamiraparani, Palar and Vaigai rivers and Ooty, Kodaikanal and Yercaud lakes are monitored for water quality inland. Afterwards, in November 2010, the CPCB approved 23 more water quality monitoring stations for the 11th plan period, and monitoring began in December 2010. While this is happening, TNPCB is keeping an eye on 55 stations. The CPCB, Delhi, provides funding for the programmes. All the lakes and rivers being monitored under the Monitoring of Indian Notional Aquatic Resources Series (MINARS) initiative. This research developed a novel time-series processing framework for water quality prediction. Using past patterns as a guide, this case study will examine how to foretell how water quality will change. Data sourcing, pre-processing, model creation and model assessment are the four essential parts of the suggested framework, as seen in Fig. 1.

Profile of study area

South India's Tamil Nadu state is home to the 7393 square kilometres occupied by the Vaigai River, which flows

Fig. 1 Site location Vaigai River



northeast and passes through the districts of Theni, Madurai, Dindigul, Sivagangai and Ramanathapuram. Near Gandamanaiken Zamin on the eastern slope of the Western Ghats, the river begins its journey at 1524 m above mean sea level. From its headwaters to its eventual merging with the ocean, the river travels about 258 km. You can see the basin map in Fig. 1. Figure 2 shows the methodology adopted in the study.

Implementation to the CSOA and COOA and LSTM-CSOA-COOA

COOT optimisation algorithm

A new meta-heuristic optimisation method called COOT mimics the actions of a flock of coot birds (Naruei and Keynia 2021). The COOT method outperforms other optimisation algorithms in terms of convergence accuracy and speed when compared to them. These algorithms include differential evolution and particle swarm optimisation. Along with pressure vessel design, welded beam design, stepped cone pulley and rolling bearing difficulties, the COOT algorithm has also been verified for these applications (Naruei and Keynia 2021).

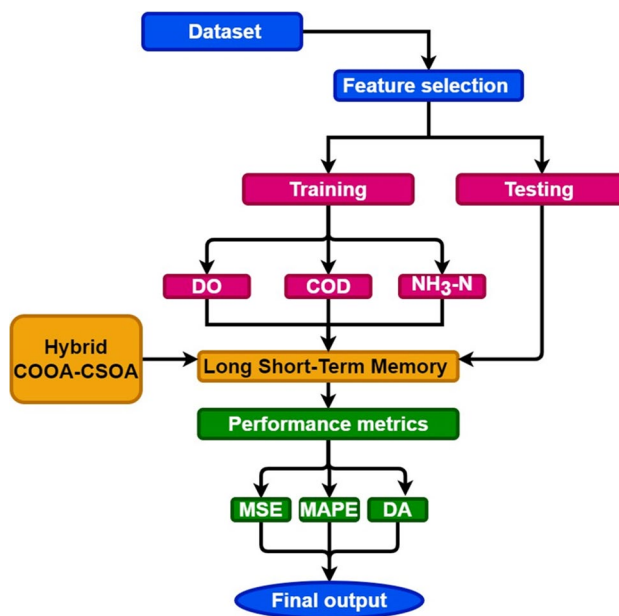


Fig. 2 Methodology adopted in this study

This study employs the Coot optimization algorithm, inspired by the movement behaviours of coot birds on the water, to determine unknown parameters. The algorithm mimics four types of coot movements: chain movement, leader movement, random movement and position adjustment according to the leader, reflecting their group behaviours aimed at reaching food or specific locations. Equations (1) and (2) implement the specific procedure of the Coot algorithm (Naruei and Keynia 2021). Equation (1) builds the random initialization of the population:

1. Determine the starting point for the coot populations' fitness levels and location:

$$R_i = M_{nb} + \text{rand}(1, d) * (M_{xb} - M_{nb}) \quad (1)$$

The random initialisation of individuals produces the variables R_i , where d is the problem's dimension and M_{xb} and M_{nb} are the upper and lower limits of the population's placements, respectively. The coots' whereabouts are shown in this matrix:

$$C_{ps}T = [A_{x,1}, A_{x,d}, A_{x,d}^t] \quad (2)$$

C_{ps} is the total number of coot locations, and N is the population scale. The following is the population-level fitness matrix for coots, obtained by feeding their locations into the fitness function:

$$FT^t = [FT_1^t, FT_2^t, \dots, FT_x^t] \quad (3)$$

Specifically, $fitl$ l is the value of the individual's fitness at the time of the l th iteration.

Cuckoo search optimisation algorithm

The cuckoo bird's reproductive strategy inspired the CSO optimisation method. For a better chance of hatching, these birds deposit their eggs in the nests of other birds and sometimes even steal eggs from the host species. The name for this kind of reproduction is brood parasitism. Below are three examples of the three behavioural patterns used to develop the CSO algorithm, which aims to achieve the global optimal value by mimicking brood parasitism (Yogambal Jayalakshmi et al. 2021):

1. A cuckoo will lay an egg and then, entirely at random, drop it into the nest of another kind of bird. An egg represents a solution or a component of an array of solutions. A solution array or a cluster of eggs forms a nest.
2. The following generation will get the nests that have the best eggs or solutions.

3. The nest count is constant. The host bird will either abandon the nest or build a new one close by if it discovers that all of the eggs are not its own. It will also discard the alien eggs.

Result and discussion

Numerous sources contribute to the water supply that the Vaigai River receives. Notable rivers that flow into the Vaigai include Varahanadhi, Manjalar, Theniar, Marudhandhi, Varatar, Sirumalayar, Nagalar, Sathiar, Uppar and Suriliar. Suriliar is a significant tributary of the Vaigai as it flows from Thoovanam to the last river, Surulipatti. Midway through June marks the beginning of the basin's monsoon season, which lasts into December. January begins the non-monsoon season, which concludes in the second week of June. With the monsoon accounting for 58% and the non-monsoon for 42%, the total annual rainfall is estimated to be 693.30 mm. Wind speeds average 1.86 km/h, sunshine hours are 7.5 per day, and relative humidity is 57.55%. Temperatures range from 23.62 °C during the low season to 34.38 °C in the high season.

The model suggested specific enhancements to the LSTM-CSOA, LSTM-COOA and LSTM-CSOA-COOA to accomplish the dual goals of increasing the rate of convergence and the precision of predictions. To assess the performance of the modified algorithm, we use the specific water quality monitoring data from the Vaigai River, given in Fig. 3. This data includes six indicators: pH, DO, COD, NH₃-N, TP and WQG, all collected from weekly measurements. The performance of the approach is greatly affected by four parameters of the LSTM model used in this paper: the learning rate Lr , the number of training iterations K , the number of neurons in the LSTM hidden layer $L1$, and the number of LSTM units in the first and second layers, respectively. With the help of the LSTM CSOA, LSTM-COOA and LSTM-CSOA-COOA algorithms, the LSTM model is fine-tuned and optimised with these four critical characteristics serving as features for particle search. Both the test and training datasets for the Vaigai River are displayed in Fig. 3. Maximising LSTM accuracy is the primary objective of LSTM CSOA, LSTM-COOA and LSTM-CSOA-COOA, which aim to optimise LSTM hyperparameters. The best mean square error (MSE) obtained by The LSTM-COOA, LSTM-CSOA-COOA and LSTM-LSTM CSOA fitness functions are utilised during training. Figure 4 shows the DO results of LSTM-COOA, LSTM-CSOA and LSTM-COOA combined for DO prediction in the Vaigai River. The errors of LSTM-COOA, LSTM-CSOA and LSTM-CSOA-COOA converged rapidly as the number of iterations increased. In just five cycles, the fitness evolution

Fig. 3 Test and training dataset of the Vaigai River

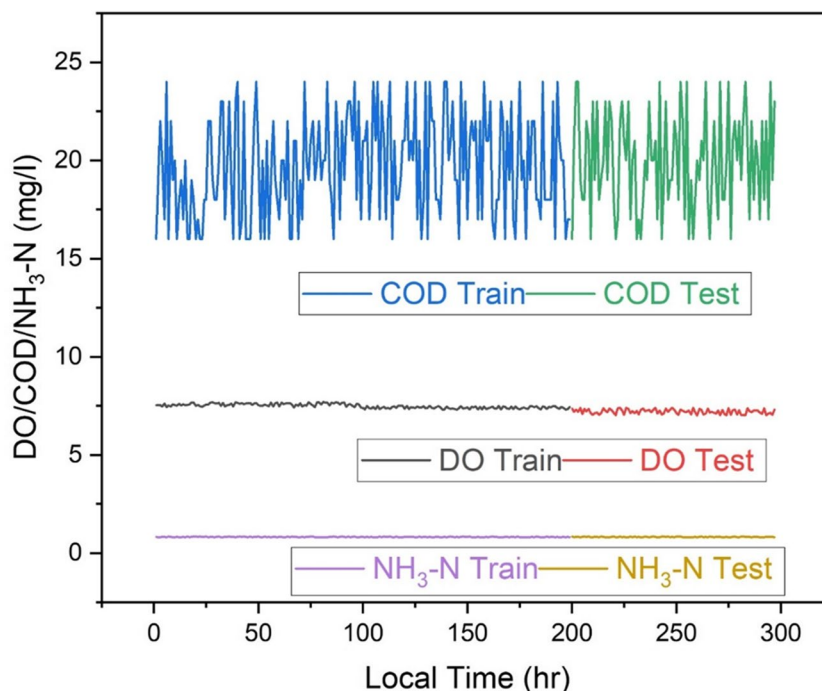
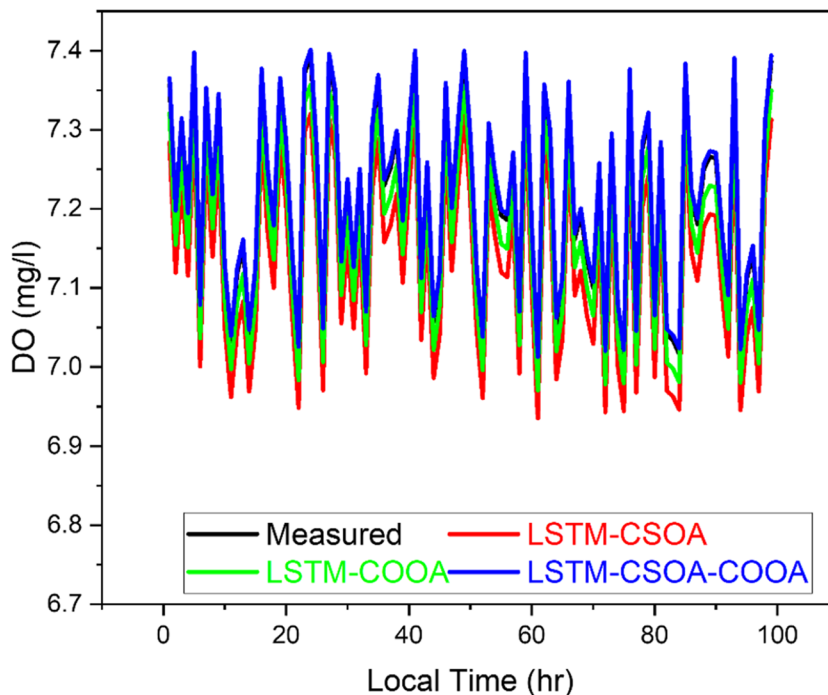


Fig. 4 Prediction of DO in the Vaigai River with LSTM CSOA, LSTM-COOA and LSTM-CSOA-COOA



curves of LSTM-CSOA, LSTM-COOA and LSTM-CSOA-COOA achieved the needed accuracy and maintained the appropriate fitness value, demonstrating excellent performance and high learning capacity. Coots update their positions based on certain rules inspired by their natural behaviours: foraging behaviour: Coots search for food by moving towards promising areas. This behaviour helps

in exploring the search space. Nesting behaviour: Coots build nests in areas they deem safe and resource-rich. This behaviour helps in exploiting good solutions. Avoidance of predators: coots move away from threats, helping them to avoid poor solutions.

Following the smoothing of the time series, the in-sample data is fitted using the LSTM CSOA, LSTM-COOA and

LSTM-CSOA-COOA model in the first step. To find the optimum model for predicting the out-of-sample data, the parameters of the prediction model are further optimised depending on the fit of the multivariate data. To obtain the final prediction results, the residuals from data fitting are fed into a residual feedback network, which in turn produces the out-of-sample data prediction residuals. Applying a moving average to the output predictions can help reduce noise and improve the smoothness of the predicted sequence. Proper tuning of hyperparameters such as learning rate, batch size and the number of epochs can significantly affect the smoothness and stability of the model's predictions.

Forecasts of COD in the Vaigai River using LSTM CSOA, LSTM-COOA and LSTM-CSOA-COOA have their data inaccuracies remedied due to non-strictly stable and nonlinear variables in the multidimensional series. Figure 5 compares the hybrid model's COD prediction results to those of the other models.

The best performing model is determined using three assessment metrics: RMSE, MAPE and NSE. LSTM-COOA, LSTM-CSOA and LSTM-COOA. The study's three techniques for predicting water quality measurements are displayed in Table 1, along with their respective anticipated inaccuracies. The network model is unable to produce more accurate predictions of multivariate water quality indicators and has poor prediction accuracy when working with limited data sets. For the majority of water quality data, the LSTM-COOA, LSTM-CSOA and LSTM-COOA models provide good forecast accuracy. Nevertheless, a notable inaccuracy in parameter prediction exists for $\text{NH}_3\text{-N}$. When it comes to

Table 1 Performance metrics of the variable methods

Features	Method	RMSE	MAE	MAPE
DO	LSTM	0.2223	0.1791	0.1575
	LSTM CSOA	0.1611	0.1251	0.1071
	LSTM-COOA	0.1215	0.0828	0.0711
	LSTM-CSOA-COOA	0.0747	0.0495	0.0396
COD	LSTM	0.0531	0.0342	0.2619
	LSTM CSOA	0.0441	0.0261	0.1719
	LSTM-COOA	0.0351	0.0189	0.1278
	LSTM-CSOA-COOA	0.0144	0.0099	0.072
$\text{NH}_3\text{-N}$	LSTM	0.0135	0.0126	0.5382
	LSTM CSOA	0.0117	0.0099	0.1971
	LSTM-COOA	0.0072	0.0063	0.1341
	LSTM-CSOA-COOA	0.00495	0.00405	0.1152

multiparameter prediction jobs, the SVR model falls short since it is not sensitive to certain characteristic factors and does not capture all the aspects of the water quality data. All four data points show that the LSTM-CSOA, LSTM-COOA and LSTM-CSOA-COOA models produce identical prediction results to the hybrid model. Nonetheless, the hybrid model's generalizability is diminished due to the increased difficulty in applying and deploying it in practice caused by its more substantial input data needs. As shown in Table 1, they achieved MAPE and RMSE for every water quality indicator and practically flawless data matching rates. Figure 6 displays the $\text{NH}_3\text{-N}$ data comparison results, which involve comparing the prediction outcomes of the hybrid

Fig. 5 Prediction of COD in the Vaigai River with LSTM CSOA, LSTM-COOA and LSTM-CSOA-COOA

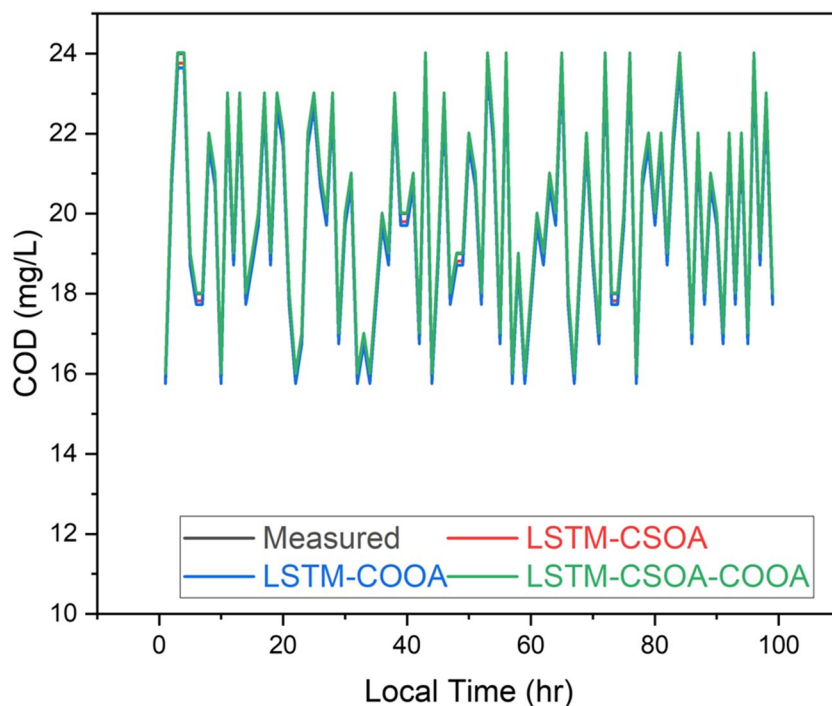
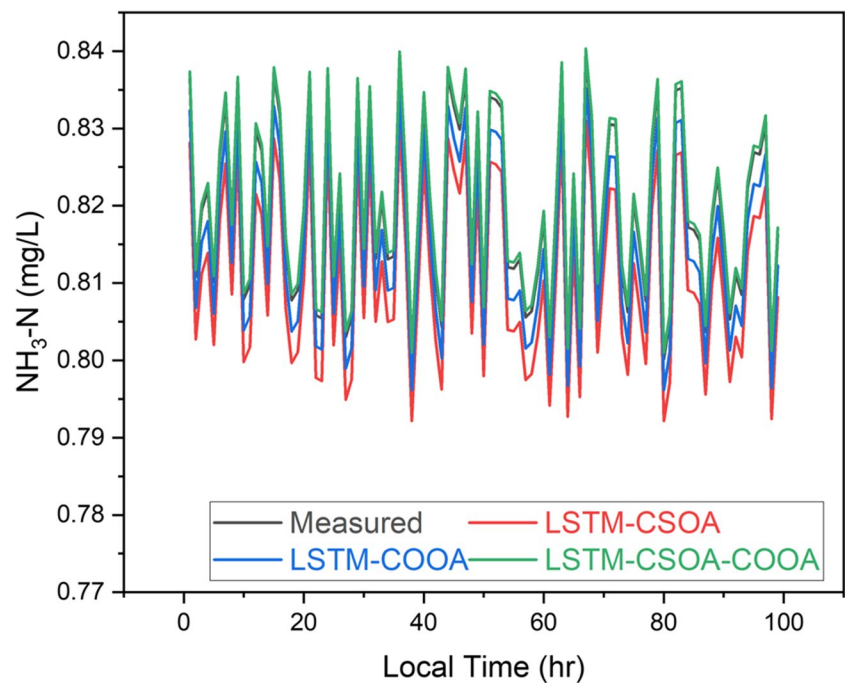


Fig. 6 Prediction of $\text{NH}_3\text{-N}$ in the Vaigai River with LSTM-CSOA, LSTM-COOA and LSTM-CSOA-COOA



model with the original data one by one. The desired goal is for the RMSE and MAPE values to be smaller as the results approach 1, which is evident. This proves that the residual feedback network design mitigates non-strictly stationary multivariate data systematic mistakes. A hybrid prediction model's network topology is straightforward, the model's execution time is minimal, and the model's statistical parameters may be roughly estimated from the data's feature. This simplifies the process of theoretical derivation and allows for residual feedback to fix prediction errors, which improves the algorithm's operational efficiency and generalizability. The model also fully accounts for sequence correlations for improved prediction accuracy and concurrent multivariate data processing.

The best performing model is determined using three assessment metrics: RMSE, MAPE and MSE. These metrics are applied to the anticipated and observed data for 20 weeks' worth of water quality parameters by the SVR model, LSTM-COOA, LSTM-CSOA and LSTM-COOA. The study's three techniques for predicting water quality measurements are displayed in Table 1, along with their respective anticipated inaccuracies. The results demonstrate that the LSTM neural network model is unable to produce more accurate predictions of multivariate water quality indicators and has poor prediction accuracy when working with limited data sets. For the majority of water quality data, the LSTM-COOA, LSTM-CSOA and LSTM-COOA models provide good forecast accuracy. Nevertheless, a notable inaccuracy in parameter prediction exists for $\text{NH}_3\text{-N}$. When it comes to multiparameter prediction jobs, the SVR model falls short since it is not sensitive to certain

characteristic factors and does not capture all the aspects of the water quality data. All four data points show that the LSTM-CSOA, LSTM-COOA and LSTM-CSOA-COOA models produce identical prediction results to the hybrid model. Nonetheless, the hybrid model's generalizability is diminished due to the increased difficulty in applying and deploying it in practice caused by its more substantial input data needs. As shown in Table 1, the hybrid prediction model achieves MAPEs RMSE and MAE for every water quality indicator and practically flawless data matching rates. It should be noted that the model's performance improves with decreasing RMSE, and vice versa. The prediction error will be less than or equal to 0.01, according to the hybrid prediction model. Table 2 shows the geographically reported study.

Table 2 shows the geographically reported study. Water quality parameters refer to various physical, chemical and biological characteristics used to assess the condition of water. These parameters are crucial for evaluating the suitability of water for multiple purposes, such as drinking, agriculture, industrial use and aquatic habitat support. Some standard water quality parameters include pH, which measures the acidity or alkalinity of water. It is an important parameter as it affects the solubility of minerals and metals and the survival of aquatic organisms.

Water temperature influences various biological and chemical processes in aquatic ecosystems. It can affect dissolved oxygen levels, metabolic rates of aquatic organisms and the solubility of gases and minerals. Dissolved oxygen (DO): DO is the amount of oxygen dissolved in water, which is essential for the survival of aquatic organisms. It

Table 2 Globally reported water quality index study

Method	Parameters	Performance metrics		References
		RMSE	MAPE	
Ant bee colony (ABC) BPNN	DO	0.683	9.43	L. Chen et al. (2023b)
	COD	9.43	27.07	
LSTM	DO	0.08	9.7	Zhao et al. (2021)
	COD	0.03	1.7	
LSTM-Random forest	DO	0.09	10.5	Rezaie-Balf et al. (2020)
	COD	0.045	2	
LSTM-SOD-VGG	DO	0.225	0.014	Wan et al. (2022b)
	COD	0.051	0.046	
LSTM-CSOA-COOA	DO	0.0495	0.0396	Present study
	COD	0.0099	0.072	
	NH ₃ -N	0.00405	0.1152	

is influenced by temperature, salinity and the presence of organic matter (L. Chen et al. 2023b).

Even though LSTMs are designed to mitigate the vanishing gradient problem compared to vanilla RNNs, they are not immune. The gradients can still diminish during backpropagation, especially with very long sequence. Conversely, gradients can also become excessively large, leading to unstable training and divergence of the model. To prevent exploding gradients, clip the gradients during backpropagation. Use dropout regularization to prevent overfitting. Adjust the learning rate dynamically during training to improve convergence.

Turbidity measures the cloudiness or haziness of water caused by suspended solids, such as silt, clay and plankton. High turbidity can affect light penetration, aquatic plant growth and the health of aquatic ecosystems. Conductivity: Conductivity measures water's ability to conduct electrical current, which is influenced by the concentration of dissolved ions such as salts and minerals. It assesses the salinity and total dissolved solids (TDSs) in water (Zhao et al. 2021). Environmental agencies, water utilities, researchers and other stakeholders often monitor these parameters to assess water quality, identify pollution sources and implement appropriate management measures (Wan et al. 2022b).

Conclusion

The preservation of river water ecosystems is greatly influenced by water quality, which affects all living things and human endeavours. The data was culled from 300 weeks of monitoring in the Vaigai River, which flows through Madurai, Tamil Nadu, India. This work utilised various AI techniques, including baseline, modelling and hybrid methods. As baseline models, it experimentally evaluated and examined the performance of LSTM, LSTM-CSOA, LSTM-COOA and LSTM-CSOA-COOA hybrid models in predicting river water quality. The prediction results demonstrate

that it outperforms the competition and produces findings more in line with real-world indicators. The study found that the algorithm achieved accuracy greater than 94%. The wavelet neural network demonstrated higher prediction precision and stronger learning and generalization abilities compared to the traditional BP neural network and the Elman neural network. The proposed model meets management requirements for intensive freshwater systems. Since water quality is significantly influenced by hydrological and meteorological factors, future research should focus on developing different predictive models for various weather conditions and combining these models to enhance prediction accuracy.

Author contribution Sankarasubbu Rajagopal and Sundaram Sankar Ganesh: conceptualization; Alagar Karthick Supervision, Alagar Karthick, Sankarasubbu Rajagopal and Sundaram Sankar Ganesh: methodology; Alagar Karthick and Thangavel Sampradeepraj: investigations; Thangavel Sampradeepraj: writing (original draft); Thangavel Sampradeepraj and Alagar Karthick: writing (original draft); Thangavel Sampradeepraj and Alagar Karthick: validation.

Data availability Not applicable.

Declarations

Ethical approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

References

- Adaryani FR, Jamshid Mousavi S, Jafari F (2022) Short-term rainfall forecasting using machine learning-based approaches of PSO-SVR, LSTM and CNN. *J Hydrol* 614. <https://doi.org/10.1016/j.jhydrol.2022.128463>

- Asadollah SBHS, Sharafati A, Motta D, Yaseen ZM (2021) River water quality index prediction and uncertainty analysis: a comparative study of machine learning models. *J Environ Chem Eng* 9(1). <https://doi.org/10.1016/j.jece.2020.104599>
- Chen C, Xue X (2023) A novel coupling preprocessing approach for handling missing data in water quality prediction. *J Hydrol*, 617. <https://doi.org/10.1016/j.jhydrol.2022.128901>
- Chen L, Wu T, Wang Z, Lin X, Cai Y (2023a) A novel hybrid BPNN model based on adaptive evolutionary Artificial Bee Colony Algorithm for water quality index prediction. *Ecol Indic*, 146. <https://doi.org/10.1016/j.ecolind.2023.109882>
- Chen L, Wu T, Wang Z, Lin X, Cai Y (2023b) A novel hybrid BPNN model based on adaptive evolutionary Artificial Bee Colony Algorithm for water quality index prediction. *Ecol Indic* 146. <https://doi.org/10.1016/j.ecolind.2023.109882>
- Ding X, Zhu Q, Zhai A, Liu L (2019) Water quality safety prediction model for drinking water source areas in Three Gorges Reservoir and its application. *Ecol Ind* 101:734–741. <https://doi.org/10.1016/j.ecolind.2019.01.068>
- Fahad S, Su F, Khan SU, Naeem MR, Wei K (2023) Implementing a novel deep learning technique for rainfall forecasting via climatic variables: an approach via hierarchical clustering analysis. *Sci Total Environ*, 854. <https://doi.org/10.1016/j.scitotenv.2022.158760>
- Feng J, Chen Y, Wang T, Jian Tan P, Li C, Shuang (2022) Online prediction of loader payload based on a multi-stage progressive model. *Automation in Construction*, 144. <https://doi.org/10.1016/j.autcon.2022.104621>
- Fernandes de Souza JA, Silva MM, Rodrigues SG, Machado Santos S (2022) A forecasting model based on ARIMA and artificial neural networks for end-of-life vehicles. *J Environ Manag* 318. <https://doi.org/10.1016/j.jenvman.2022.115616>
- Fowdar HS, Neo TH, Ong SL, Hu J, McCarthy DT (2022) Performance analysis of a stormwater green infrastructure model for flow and water quality predictions. *J Environ Manag*, 316. <https://doi.org/10.1016/j.jenvman.2022.115259>
- He S, Guo S, Zhang J, Liu Z, Cui Z, Zhang Y, Zheng Y (2022) Multi-objective operation of cascade reservoirs based on short-term ensemble streamflow prediction. *J Hydrol*, 610. <https://doi.org/10.1016/j.jhydrol.2022.127936>
- Ho JY, Afan HA, El-Shafie AH, Koting SB, Mohd NS, Jaafar WZB, Lai Sai H, Malek MA, Ahmed AN, Mohtar WHMW, Elshorbagy A, El-Shafie A (2019) Towards a time and cost effective approach to water quality index class prediction. *J Hydrol* 575:148–165. <https://doi.org/10.1016/j.jhydrol.2019.05.016>
- Islam N, Irshad K (2022) Artificial ecosystem optimisation with Deep Learning Enabled Water Quality Prediction and Classification model. *Chemosphere*, 309. <https://doi.org/10.1016/j.chemosphere.2022.136615>
- Kruk M (2023) Prediction of environmental factors responsible for chlorophyll a-induced hypereutrophy using explainable machine learning. *Ecological Informatics*, 75. <https://doi.org/10.1016/j.ecoinf.2023.102005>
- Lao T, Sun Y (2022) Predicting the production and consumption of natural gas in China by using a new grey forecasting method. *Math Comput Simul* 202:295–315. <https://doi.org/10.1016/j.matcom.2022.05.023>
- Lap BQ, Phan TTH, Nguyen H, Du Quang LX, Hang PT, Phi NQ, Hoang VT, Linh PG, Hang BTT (2023) Predicting water quality index (WQI) by feature selection and machine learning: a case study of An Kim Hai irrigation system. *Ecol Inform* 74. <https://doi.org/10.1016/j.ecoinf.2023.101991>
- M, G. J. (2024) Secure water quality prediction system using machine learning and blockchain technologies. *J Environ Manag* 350:119357. <https://doi.org/10.1016/j.jenvman.2023.119357>
- Naruei I, Keynia, F (2021) A new optimisation method based on COOT bird natural life model. *Expert Syst Appl* 183. <https://doi.org/10.1016/j.eswa.2021.115352>
- Paul V, Ramesh R, Sreeja P, Jarin T, Sujith Kumar PS, Ansar S, Ashraf GA, Pandey S, Said Z (2022) Hybridisation of long short-term memory with Sparrow Search Optimization model for water quality index prediction. *Chemosphere*, 307. <https://doi.org/10.1016/j.chemosphere.2022.135762>
- Rezaie-Balf M, Attar NF, Mohammadzadeh A, Murti MA, Ahmed AN, Fai CM, Nabipour N, Alaghmand S, El-Shafie A (2020) Physico-chemical parameters data assimilation for efficient improvement of water quality index prediction: comparative assessment of a noise suppression hybridisation approach. *J Clean Product* 271. <https://doi.org/10.1016/j.jclepro.2020.122576>
- Singh D, Chaudhary R, Karthick A, Patil PP, Kaliappan S (2024) Economic and life cycle cost analysis of building-integrated photovoltaic system for composite climatic conditions. *Environ Sci Pollut Res*. <https://doi.org/10.1007/s11356-023-31781-1>
- Song C, Yao L, Hua C, Ni Q (2021) A novel hybrid model for water quality prediction based on synchrosqueezed wavelet transform technique and improved long short-term memory. *J Hydrol*, 603. <https://doi.org/10.1016/j.jhydrol.2021.126879>
- Souaissi Z, Ouarda TBMJ, St-Hilaire A (2023) Non-parametric, semi-parametric, and machine learning models for river temperature frequency analysis at ungauged basins. *Ecol Inform* 75. <https://doi.org/10.1016/j.ecoinf.2023.102107>
- Uddin MG, Nash S, Rahman A, Olbert AI (2022) A comprehensive method for improvement of water quality index (WQI) models for coastal water quality assessment. *Water Res*, 219. <https://doi.org/10.1016/j.watres.2022.118532>
- Wan H, Xu R, Zhang M, Cai Y, Li J, Shen X (2022a) A novel model for water quality prediction caused by non-point sources pollution based on deep learning and feature extraction methods. *J Hydrol* 612. <https://doi.org/10.1016/j.jhydrol.2022.128081>
- Wan H, Xu R, Zhang M, Cai Y, Li J, Shen X (2022b) A novel model for water quality prediction caused by non-point sources pollution based on deep learning and feature extraction methods. *J Hydrol* 612. <https://doi.org/10.1016/j.jhydrol.2022.128081>
- Yao S, Chen C, He M, Cui Z, Mo K, Pang R, Chen Q (2023) Land use as an important indicator for water quality prediction in a region under rapid urbanisation. *Ecol Indic* 146. <https://doi.org/10.1016/j.ecolind.2022.109768>
- Yogambal Jayalakshmi N, Shankar R, Subramaniam U, Baranilingesan I, Karthick A, Stalin B, Rahim R, Ghosh A (2021) Novel multi-time scale deep learning algorithm for solar irradiance forecasting. *Energies*, 14(9). <https://doi.org/10.3390/en14092404>
- Zhao S, Zhang S, Liu J, Wang H, Zhu J, Li D, Zhao R (2021) Application of machine learning in intelligent fish aquaculture: a review. *In Aquaculture* (Vol. 540). Elsevier B.V. <https://doi.org/10.1016/j.aquaculture.2021.736724>
- Zheng Z, Ding H, Weng Z, Wang L (2023). Research on a multiparameter water quality prediction method based on a hybrid model. *Ecol Inform* 76. <https://doi.org/10.1016/j.ecoinf.2023.102125>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.