Table S1: Abbreviations.

Abbreviations	Description	Abbreviations	Description
WQ	Water quality	MODWT	Maximum overlap discrete wavelet transformation
AI	Artificial intelligence		
DT	Diction tree	ACO_R	Ant colony optimisation
DL	Deep learning	WA	Wavelet analysis
ANN	Artificial neural network	RS	Random Subspace
ANFIS	Adaptive neural-based fuzzy inference system	AIG	Algorithm of the innovative gunner
ARIMA	Auto-regressive integrated moving average	CSO	Chicken swarm optimisation
SVM	Support victor Machin	PSO	Particle swarm optimisation
SVR	Support vector regression	RC	Random Committee
FFNN	Feed forward neural network	EKF	Extended kalman filter
LSTM	Long short-term memory	AR	Additive regression
		FFA	Firefly algorithm
RBFNN	Radial basis function neural network	SAE	Sparse Auto-Encoder
GRNN	General regression neural network model	IABC	Improved artificial bee colony
DBN	Deep belief network model	ABC	Artificial bee colony
CNN	Convolutional neural network	GBO	Gradient-based optimiser
RNN	Recurrent neural network	kPCA	Kernal Principal Component Analysis
BNN	Bayesian neural network	CS	Cuckoo search
QRF	Quantile regression forest	CEEMDAN	Complete ensemble empirical mode decomposition algorithm with adaptive noise
RF	Random forest	GBO	Gradient-based optimiser
GEP	Gene expression programming	kPCA	Kernal Principal Component Analysis
GBM	Stochastic Gradient Boosting	CS	Cuckoo search
GBM_H2O	Gradient Boosting Machines	EEMD	Ensemble empirical mode decomposition
M5P	M5 prime	IGRA	Improved Grey Relational Analysis
XGB	Extreme Gradient Boosting	SWT	Synchrosqueezed wavelet transform
ELM	Extreme Learning Machine	LR	Linear regression
MLR	Multi-linear regression	OBL	Opposition-based learning

Table S1: Abbreviations.

Abbreviations	Description	Abbreviations	Description
MARS	Multivariate adaptive regression splines	BWO	Black widow optimisation
LSSVR	Least squares support vector regression	ISSA	Improved sparrow search algorithm
XGBoost	Extreme gradient boosting	ISA	Improved simulated annealing
LSSVM	Least square support vector machine	DE	Differential Evolution
LWLR	Locally weighted linear regression	DWT	Discrete wavelet transform
cForest	Conditional random forests	U1	Theil U statistic 1
Ranger	Random forest geneRator	U2	Theil U statistic 2
R^2	Determination coefficient	W index	Willmott's index of agreement
RMSE	Root mean square error	MCC	Matthews correlation coefficient
MAE	Mean absolute error	FM index	Fowlkes-Mallows index
MPE	Mean percentage error	TOC	Total organic carbon
CE	Coefficient of efficiency	NSE	Nash-Sutcliffe efficiency
R	Correlation coefficient		
MSRE	mean squared relative error	RAE	Relative absolute error
WL	Water level	WT	Water temperature
TP	Total phosphorus	NH_3-N	Ammonia nitrogen
COD_{Mn}	Potassium permanganate index	ORP	Oxidation-reduction potential
HW	Hammerstein-Weiner	WS	Wind speed
DO	Dissolved oxygen	AP	Atmospheric pressure
EC	Electrical conductivity	Am	Ammonia nitrogen content
T	Temperature	Hu	Air humidity
TDS	Total dissolved solid	TN	Total nitrogen concentration
TH	Total hardness	CH	Carbonate hardness
Chl-a	Chlorophyll-a	SAR	Sodium adsorption ratio
pH	Hydrogen ion concentration	TAlk	Total alkalinity
Q	Discharge	BOD	Biochemical oxygen demand
CC	Correlation coefficient	VMD	Variational mode decomposition
SSA	Sparrow search algorithm	GRU	Gated recurrent unit
HAB	Harmful algal blooms	GDM	Gradient Descent with Momentum
CEEMDAN-LZC	Combination of complete ensemble empirical mode decomposition with an adaptive noise Lempel-Ziv complex	FCM	Fuzzy clustering method
GP	Grid partitioning	IDW	Inverse distance weighting
NF	Neuro-fuzzy	CGA	Continuous genetic algorithm
BN	Bayesian network	WGP	Wavelete gentic programation

Table S1: Abbreviations.

Abbreviations	Description	Abbreviations	Description
KGE	Kling–Gupta efficiency	md	Modified index of agreement
mNSE	Modified Nash–Sutcliffe efficiency coefficient	RSR	Ratio of the RMSE to the standard deviation
GM	Grey model	IGA	Improved genetic algorithm
DE	Differential evolution	DOM	Dissolved Organic Matter
		WVP	Water vapor pressure
FWNN	Fuzzy wavelet neural network	WNN	Wavelet neural network
MLE	Maximal Lyapunov exponent	MOGA	Multi-objective Gentic algorithm
MRE	Maximum relative error		

Table S2: Review of researchers who used data pre-processing.

References	Normalisation	Cleaning	Select the best model input
[4]	✓	✗	✓
[59]	✗	✓	✓
[69]	✓	✓	✓
[60]	✓	✓	✗
[86]	✗	✗	✓
[57]	✓	✗	✗
[58]	✗	✗	✗
[97]	✗	✓	✓
[98]	✗	✓	✓
[105]	✗	✓	✓
[90]	✗	✗	✓
[99]	✗	✓	✓
[16]	✗	✓	✓
[100]	✓	✗	✓
[85]	✓	✓	✓
[38]	✓	✓	✗
[28]	✗	✗	✓
[91]	✗	✓	✗
[83]	✗	✗	✗
[81]	✓	✓	✗
[23]	✗	✓	✓
[94]	✓	✓	✓
[67]	✓	✗	✗
[68]	✗	✗	✓
[95]	✓	✗	✓
[96]	✓	✓	✓
[73]	✗	✗	✓
[78]	✓	✗	✓
[80]	✗	✗	✗
[87]	✗	✗	✓
[92]	✓	✓	✓
[93]	✗	✓	✓
[101]	✓	✓	✓
[102]	✗	✓	✗
[103]	✓	✗	✓
[104]	✗	✓	✗
[106]	✗	✗	✗
[107]	✓	✓	✓
[74]	✗	✗	✗
[108]	✓	✗	✗
[75]	✗	✗	✗
[109]	✓	✗	✓
[110]	✓	✓	✓
[111]	✗	✗	✓
[112]	✓	✓	✓
[113]	✓	✗	✓
[114]	✗	✓	✓
[56]	✗	✗	✓
[115]	✗	✓	✗

Table S2: Review of researchers who used data pre-processing.

References	Normalisation	Cleaning	Select the best model input
[116]	✓	✓	✓
[89]	✓	✓	✓

Reference

4. Melesse, A.M.; Khosravi, K.; Tiefenbacher, J.P.; Heddam, S.; Kim, S.; Mosavi, A.; Pham, B.T. River Water Salinity Prediction Using Hybrid Machine Learning Models. *Water* **2020**, *12*, 2951. https://doi.org/10.3390/w12102951.

16. Banadkooki, F.B.; Ehteram, M.; Panahi, F.; Sh. Sammen, S.; Othman, F.B.; El-Shafie, A. Estimation of total dissolved solids (TDS) using new hybrid machine learning models. *J. Hydrol.* **2020**, *587*, 124989. https://doi.org/10.1016/j.jhydrol.2020.124989.

23. Zhang, Y.-F.; Fitch, P.; Thorburn, P.J. Predicting the Trend of Dissolved Oxygen Based on the kPCA-RNN Model. *Water* **2020**, *12*, 585. https://doi.org/10.3390/w12020585.

28. Tiyasha, T.; Tung, T.M.; Bhagat, S.K.; Tan, M.L.; Jawad, A.H.; Mohtar, W.H.M.W.; Yaseen, Z.M. Functionalization of remote sensing and on-site data for simulating surface water dissolved oxygen: Development of hybrid tree-based artificial intelligence models. *Mar. Pollut. Bull.* **2021**, *170*, 112639. https://doi.org/10.1016/j.marpolbul.2021.112639.

38. Lu, H.; Ma, X. Hybrid decision tree-based machine learning models for short-term water quality prediction. *Chemosphere* **2020**, *249*, 126169. https://doi.org/10.1016/j.chemosphere.2020.126169.

56. Lola, M.S.; Zainuddin, N.H.; Abdullah, M.T.; Ponniah, V.; Ramlee, M.N.A.; Zakariya, R.; Idris, M.S.; Khalili, I. Improving the performance of ann-arima models for predicting water quality in the offshore area of kuala terengganu, terengganu, malaysia. *J. Sustain. Sci. Manag.* **2018**, *13*, 27–37.

57. Barzegar, R.; Aalami, M.T.; Adamowski, J. Short-term water quality variable prediction using a hybrid CNN–LSTM deep learning model. *Stoch. Environ. Res. Risk Assess.* **2020**, *34*, 415–433. https://doi.org/10.1007/s00477-020-01776-2.

58. Baek, S.-S.; Pyo, J.; Chun, J.A. Prediction of Water Level and Water Quality Using a CNN-LSTM Combined Deep Learning Approach. *Water* **2020**, *12*, 3399. https://doi.org/10.3390/w12123399.

59. Yan, J.; Liu, J.; Yu, Y.; Xu, H. Water Quality Prediction in the Luan River Based on 1-DRCNN and BiGRU Hybrid Neural Network Model. *Water* **2021**, *13*, 1273. https://doi.org/10.3390/w13091273.

60. Hien Than, N.; Dinh Ly, C.; Van Tat, P. The performance of classification and forecasting Dong Nai River water quality for sustainable water resources management using neural network techniques. *J. Hydrol.* **2021**, *596*, 126099. https://doi.org/10.1016/j.jhydrol.2021.126099.

67. Aghel, B.; Rezaei, A.; Mohadesi, M. Modeling and prediction of water quality parameters using a hybrid particle swarm optimization–neural fuzzy approach. *Int. J. Environ. Sci. Technol.* **2018**, *16*, 4823–4832. https://doi.org/10.1007/s13762-018-1896-3.

68. Azad, A.; Karami, H.; Farzin, S.; Mousavi, S.-F.; Kisi, O. Modeling river water quality parameters using modified adaptive neuro fuzzy inference system. *Water Sci. Eng.* **2019**, *12*, 45–54. https://doi.org/10.1016/j.wse.2018.11.001.

69. Shah, M.I.; Javed, M.F.; Alqahtani, A.; Aldrees, A. Environmental assessment based surface water quality prediction using hyper-parameter optimized machine learning models based on consistent big data. *Process Saf. Environ. Prot.* **2021**, *151*, 324–340. https://doi.org/10.1016/j.psep.2021.05.026.

73. Azad, A.; Karami, H.; Farzin, S.; Saeedian, A.; Kashi, H.; Sayyahi, F. Prediction of Water Quality Parameters Using ANFIS Optimized by Intelligence Algorithms (Case Study: Gorganrood River). *KSCE J. Civ. Eng.* **2017**, *22*, 2206–2213. https://doi.org/10.1007/s12205-017-1703-6.

74. Stajkowski, S.; Kumar, D.; Samui, P.; Bonakdari, H.; Gharabaghi, B. Genetic-Algorithm-Optimized Sequential Model for Water Temperature Prediction. *Sustainability* **2020**, *12*, 5374. https://doi.org/10.3390/su12135374.

75. Jin, T.; Cai, S.; Jiang, D.; Liu, J. A data-driven model for real-time water quality prediction and early warning by an integration method. *Environ. Sci. Pollut. Res. Int.* **2019**, *26*, 30374–30385. https://doi.org/10.1007/s11356-019-06049-2.

78. Raheli, B.; Aalami, M.T.; El-Shafie, A.; Ghorbani, M.A.; Deo, R.C. Uncertainty assessment of the multilayer perceptron (MLP) neural network model with implementation of the novel hybrid MLP-FFA method for prediction of biochemical oxygen demand and dissolved oxygen: A case study of Langat River. *Environ. Earth Sci.* **2017**, *76*, 503. https://doi.org/10.1007/s12665-017-6842-z.

80. Chatterjee, S.; Sarkar, S.; Dey, N.; Ashour, A.S.; Sen, S.; Hassanien, A.E. Application of cuckoo search in water quality prediction using artificial neural network. *Int. J. Comput. Intell. Stud.* **2017**, *6*, 229–244, doi:doi.org/10.1504/IJCISTUDIES.2017.089054.

81. Li, Z.; Peng, F.; Niu, B.; Li, G.; Wu, J.; Miao, Z. Water Quality Prediction Model Combining Sparse Auto-encoder and LSTM Network. *IFAC PapersOnLine* **2018**, *51–17*, 831–836. https://doi.org/10.1016/j.ifacol.2018.08.09.
83. Chen, S.; Fang, G.; Huang, X.; Zhang, Y. Water Quality Prediction Model of a Water Diversion Project Based on the Improved Artificial Bee Colony–Backpropagation Neural Network. *Water* **2018**, *10*, 806. https://doi.org/10.3390/w10060806.
85. Zhou, J.; Wang, Y.; Xiao, F.; Wang, Y.; Sun, L. Water Quality Prediction Method Based on IGRA and LSTM. *Water* **2018**, *10*, 1148. https://doi.org/10.3390/w10091148.
86. Kadkhodazadeh, M.; Farzin, S. A Novel LSSVM Model Integrated with GBO Algorithm to Assessment of Water Quality Parameters. *Water Resour. Manag.* **2021**, *35*, 3939–3968. https://doi.org/10.1007/s11269-021-02913-4.
87. Dehghani, R.; Torabi Poudeh, H.; Izadi, Z. Dissolved oxygen concentration predictions for running waters with using hybrid machine learning techniques. *Modeling Earth Syst. Environ.* **2021**, *8*, 2599–2613. https://doi.org/10.1007/s40808-021-01253-x.
89. Solgi, A.; Pourhaghi, A.; Bahmani, R.; Zarei, H. Improving SVR and ANFIS performance using wavelet transform and PCA algorithm for modeling and predicting biochemical oxygen demand (BOD). *Ecohydrol. Hydrobiol.* **2017**, *17*, 164–175. https://doi.org/10.1016/j.ecohyd.2017.02.002.
90. Al-Sulttani, A.O.; Al-Mukhtar, M.; Roomi, A.B.; Farooque, A.A.; Khedher, K.M.; Yaseen, Z.M. Proposition of New Ensemble Data-Intelligence Models for Surface Water Quality Prediction. *IEEE Access* **2021**, *9*, 108527–108541. https://doi.org/10.1109/access.2021.3100490.
91. Bi, J.; Lin, Y.; Dong, Q.; Yuan, H.; Zhou, M. Large-scale water quality prediction with integrated deep neural network. *Inf. Sci.* **2021**, *571*, 191–205. https://doi.org/10.1016/j.ins.2021.04.057.
92. Ahmed, A.A.M.; Chowdhury, M.A.I.; Ahmed, O.; Sutradhar, A. Development of Dissolved Oxygen Forecast Model Using Hybrid Machine Learning Algorithm with Hydro-Meteorological Variables. 2021. https://doi.org/10.21203/rs.3.rs-1100147/v1.
93. Ahmadianfar, I.; Jamei, M.; Chu, X. A novel Hybrid Wavelet-Locally Weighted Linear Regression (W-LWLR) Model for Electrical Conductivity (EC) Prediction in Surface Water. *J. Contam. Hydrol.* **2020**, *232*, 103641. https://doi.org/10.1016/j.jconhyd.2020.103641.
94. Eze, E.; Halse, S.; Ajmal, T. Developing a Novel Water Quality Prediction Model for a South African Aquaculture Farm. *Water* **2021**, *13*, 1782. https://doi.org/10.3390/w13131782.
95. Yan, J.; Gao, Y.; Yu, Y.; Xu, H.; Xu, Z. A Prediction Model Based on Deep Belief Network and Least Squares SVR Applied to Cross-Section Water Quality. *Water* **2020**, *12*, 1929. https://doi.org/10.3390/w12071929.
96. Wang, Y.; Yuan, Y.; Pan, Y.; Fan, Z. Modeling Daily and Monthly Water Quality Indicators in a Canal Using a Hybrid Wavelet-Based Support Vector Regression Structure. *Water* **2020**, *12*, 1476. https://doi.org/10.3390/w12051476.
97. Song , C.; Yao, L.; Hua , C.; Ni, Q. A novel hybrid model for water quality prediction based on synchrosqueezed wavelet transform technique and improved long short-term memory. *J. Hydrol.* **2021**, *603*, 126879. https://doi.org/10.1016/j.jhydrol.2021.126879.
98. Jamei, M.; Ahmadianfar, I.; Karbasi, M.; Jawad, A.H.; Farooque, A.A.; Yaseen, Z.M. The assessment of emerging data-intelligence technologies for modeling Mg(+2) and SO4(-2) surface water quality. *J. Environ. Manag.* **2021**, *300*, 113774. https://doi.org/10.1016/j.jenvman.2021.113774.
99. Sha, J.; Li, X.; Zhang, M.; Wang, Z.-L. Comparison of Forecasting Models for Real-Time Monitoring of Water Quality Parameters Based on Hybrid Deep Learning Neural Networks. *Water* **2021**, *13*, 1547. https://doi.org/10.3390/w13111547.
100. Yan, J.; Xu, Z.; Yu, Y.; Xu, H.; Gao, K. Application of a Hybrid Optimized BP Network Model to Estimate Water Quality Parameters of Beihai Lake in Beijing. *Appl. Sci.* **2019**, *9*, 1863. https://doi.org/10.3390/app9091863.
101. Abba, S.I.; Linh, N.T.T.; Abdullahi, J.; Ali, S.I.A.; Pham, Q.B.; Abdulkadir, R.A.; Costache, R.; Nam, V.T.; Anh, D.T. Hybrid Machine Learning Ensemble Techniques for Modeling Dissolved Oxygen Concentration. *IEEE Access* **2020**, *8*, 157218–157237. https://doi.org/10.1109/access.2020.3017743.
102. Song, C.; Yao, L.; Hua, C.; Ni, Q. A water quality prediction model based on variational mode decomposition and the least squares support vector machine optimized by the sparrow search algorithm (VMD-SSA-LSSVM) of the Yangtze River, China. *Environ. Monit. Assess* **2021**, *193*, 363. https://doi.org/10.1007/s10661-021-09127-6.
103. Deng, T.; Chau, K.W.; Duan, H.F. Machine learning based marine water quality prediction for coastal hydro-environment management. *J. Environ. Manag.* **2021**, *284*, 112051. https://doi.org/10.1016/j.jenvman.2021.112051.
104. Huang, J.; Liu, S.; Hassan, S.G.; Xu, L.; Huang, C. A hybrid model for short-term dissolved oxygen content prediction. *Comput. Electron. Agric.* **2021**, *186*, 106216. https://doi.org/10.1016/j.compag.2021.106216.
105. Jiang, J.; Tang, S.; Liu, R.; Sivakumar, B.; Wu, X.; Pang, T. A hybrid wavelet-Lyapunov exponent model for river water quality forecast. *J. Hydroinformatics* **2021**, *23*, 864–878. https://doi.org/10.2166/hydro.2021.023.

106. Maroufpoor, S.; Jalali, M.; Nikmehr, S.; Shiri, N.; Shiri, J.; Maroufpoor, E. Modeling groundwater quality by using hybrid intelligent and geostatistical methods. *Environ. Sci. Pollut. Res. Int.* **2020**, *27*, 28183–28197. https://doi.org/10.1007/s11356-020-09188-z.
107. Jafari, H.; Rajaee, T.; Kisi, O. Improved Water Quality Prediction with Hybrid Wavelet-Genetic Programming Model and Shannon Entropy. *Nat. Resour. Res.* **2020**, *29*, 3819–3840. https://doi.org/10.1007/s11053-020-09702-7.
108. Ye, Q.; Yang, X.; Chen, C.; Wang, J. River Water Quality Parameters Prediction Method Based on LSTM-RNN Model. In Proceedings of the 31th Chinese Control and Decision Conference (2019 CCDC), Nanchang, China, 3–5 June 2019. . https://doi.org/10.1109/CCDC.2019.8832885.
109. Li, L.; Jiang, P.; Xu, H.; Lin, G.; Guo, D.; Wu, H. Water quality prediction based on recurrent neural network and improved evidence theory: A case study of Qiantang River, China. *Environ. Sci. Pollut. Res. Int.* **2019**, *26*, 19879–19896. https://doi.org/10.1007/s11356-019-05116-y.
110. Parmar, K.S.; Makkhan, S.J.S.; Kaushal, S. Neuro-fuzzy-wavelet hybrid approach to estimate the future trends of river water quality. *Neural Comput. Appl.* **2019**, *31*, 8463–8473. https://doi.org/10.1007/s00521-019-04560-8.
111. Kisi, O.; Azad, A.; Kashi, H.; Saeedian, A.; Hashemi, S.A.A.; Ghorbani, S. Modeling Groundwater Quality Parameters Using Hybrid Neuro-Fuzzy Methods. *Water Resour. Manag.* **2018**, *33*, 847–861. https://doi.org/10.1007/s11269-018-2147-6.
112. Fijani, E.; Barzegar, R.; Deo, R.; Tziritis, E.; Skordas, K. Design and implementation of a hybrid model based on two-layer decomposition method coupled with extreme learning machines to support real-time environmental monitoring of water quality parameters. *Sci. Total Environ.* **2019**, *648*, 839–853. https://doi.org/10.1016/j.scitotenv.2018.08.221.
113. Shao, D.; Nong, X.; Tan, X.; Chen, S.; Xu, B.; Hu, N. Daily Water Quality Forecast of the South-To-North Water Diversion Project of China Based on the Cuckoo Search-Back Propagation Neural Network. *Water* **2018**, *10*, 1471. https://doi.org/10.3390/w10101471.
114. Montaseri, M.; Zaman Zad Ghavidel, S.; Sanikhani, H. Water quality variations in different climates of Iran: Toward modeling total dissolved solid using soft computing techniques. *Stoch. Environ. Res. Risk Assess.* **2018**, *32*, 2253–2273. https://doi.org/10.1007/s00477-018-1554-9.
115. Huang, M.; Tian, D.; Liu, H.; Zhang, C.; Yi, X.; Cai, J.; Ruan, J.; Zhang, T.; Kong, S.; Ying, G. A Hybrid Fuzzy Wavelet Neural Network Model with Self-Adapted Fuzzy c-Means Clustering and Genetic Algorithm for Water Quality Prediction in Rivers. *Complexity* **2018**, *2018*, 8241342. https://doi.org/10.1155/2018/8241342.
116. Barzegar, R.; Asghari Moghaddam, A.; Adamowski, J.; Ozga-Zielinski, B. Multi-step water quality forecasting using a boosting ensemble multi-wavelet extreme learning machine model. *Stoch. Environ. Res. Risk Assess.* **2017**, *32*, 799–813. https://doi.org/10.1007/s00477-017-1394-z.