

Review

A Review of Hybrid Soft Computing and Data Pre-Processing Techniques to Forecast Freshwater Quality's Parameters: Current Trends and Future Directions

Zahraa S. Khudhair ¹, Salah L. Zubaidi ¹, Sandra Ortega-Martorell ², Nadhir Al-Ansari ^{3,*}, Saleem Ethaib ⁴ and Khalid Hashim ^{5,6}

¹ Department of Civil Engineering, Wasit University, Wasit 52001, Iraq; zahra.asadik2@gmail.com (Z.S.K.); salahlafta@uowasit.edu.iq (S.L.Z.)

² Department of Applied Mathematics, Liverpool John Moores University, Liverpool L3 3AF, UK; s.ortegamartorell@ljmu.ac.uk

³ Department of Civil Environmental and Natural Resources Engineering, Lulea University of Technology, 971 87 Lulea, Sweden

⁴ Department of Civil Engineering, College of Engineering, University of Thi-Qar, Al-Nassiriya 64001, Iraq; dr.saleem@utq.edu.iq

⁵ School of Civil Engineering and Built Environment, Built Environment and Sustainable Technologies (BEST) Research Institute, Liverpool John Moores University, Liverpool L3 3AF, UK; k.s.hashim@ljmu.ac.uk

⁶ Department of Environment Engineering, Babylon University, Babylon 51001, Iraq

* Correspondence: nadhir.alansari@ltu.se

Citation: Khudhair, Z.S.; Zubaidi, S.L.; Ortega-Martorell, S.; Al-Ansari, N.; Ethaib, S.; Hashim, K. A Review of Hybrid Soft Computing and Data Pre-Processing Techniques to Forecast Freshwater Quality's Parameters: Current Trends and Future Directions. *Environments* **2022**, *9*, 85. <https://doi.org/10.3390/environments9070085>

Academic Editor: Yu-Pin Lin

Received: 11 May 2022

Accepted: 30 June 2022

Published: 2 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Water quality has a significant influence on human health. As a result, water quality parameter modelling is one of the most challenging problems in the water sector. Therefore, the major factor in choosing an appropriate prediction model is accuracy. This research aims to analyse hybrid techniques and pre-processing data methods in freshwater quality modelling and forecasting. Hybrid approaches have generally been seen as a potential way of improving the accuracy of water quality modelling and forecasting compared with individual models. Consequently, recent studies have focused on using hybrid models to enhance forecasting accuracy. The modelling of dissolved oxygen is receiving more attention. From a review of relevant articles, it is clear that hybrid techniques are viable and precise methods for water quality prediction. Additionally, this paper presents future research directions to help researchers predict freshwater quality variables.

Keywords: water quality parameters; hybrid model; metaheuristic algorithms; machine learning

1. Introduction

The growing scarcity of fresh, clean water is one of the most pressing concerns confronting civilization in the twenty-first century [1]. Recent research has proven climate change will have a significant impact on freshwater supplies due to the probable reduction in rainfall [2]. In addition to projected droughts in various river basins throughout the world due to climate change, several studies have shown potential water quality (WQ) degradation due to dilution or concentration of soluble chemicals [3]. Additionally, multiple studies have indicated that pollution has a negative impact on freshwater resources in general [2]. The decline in river WQ has irreversible consequences for the environment and human health as more than one billion people do not have access to clean potable water [4]. Hence, it is necessary to estimate and make predictions regarding water quality in an attempt to anticipate how WQ will change over time. Additionally, forecasting future variations in WQ is very important for future aquaculture control intelligence. As a

result, WQ forecasting is quite useful for anticipating WQ and estimating future supply. Robust, reliable, and flexible models are critically needed [5].

Conventional approaches for time series analysis, such as auto-regressive integrated moving average (ARIMA, abbreviations are collected in Table S1 in the Supplementary Materials) and multiple linear regression (MLR) models have been shown to be limited in terms of carefully determining WQ due to the intricacy and sophistication of the WQ time series. Machine learning (ML) methods such as artificial neural networks (ANN) [6–8], support vector machines (SVM) [9,10], deep neural networks (Deep NN) [11], and k-nearest neighbours (KNN) [12] have also been applied to simulate WQ [13]. Artificial intelligence (AI) techniques are superior to traditional models and achieve better results due to the ability of AI to deal with non-linear and complex properties [14,15]. Additionally, several combined techniques have been widely employed for WQ modelling because combined techniques are better than standalone models, and this is improving forecasting accuracy [16]. The increasing trend in applying hybrid ML methods can be seen in recent years, as revealed in Figure 1.

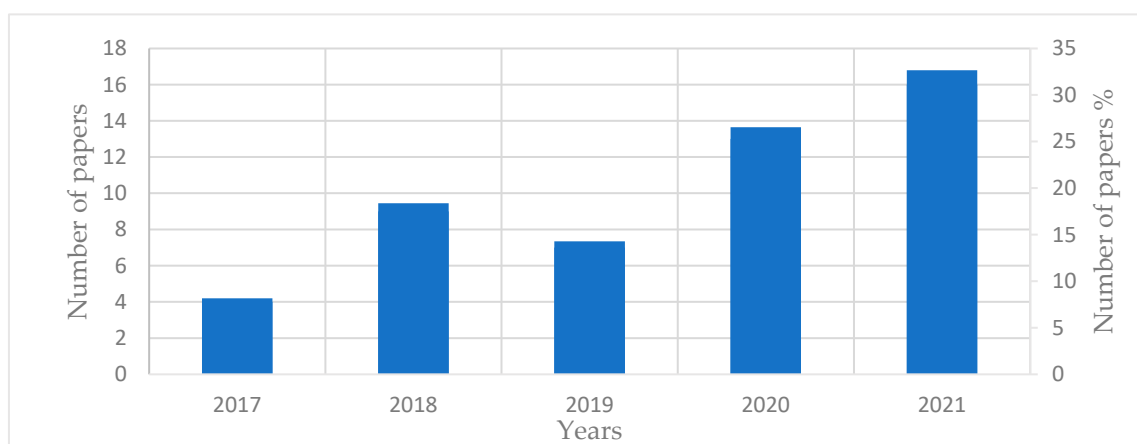


Figure 1. Studies' number of hybrid ML models for WQ parameters prediction over the last four years.

Additionally, several other review papers have introduced the applications of the soft computer to forecast WQ [5,15,17–21], whose keywords and crucial aspects are summarised in Table 1.

Table 1. Summaries of related review papers.

Reference	Keywords	Summary
[19]	River water quality, state of the art, literature assessment and evaluation, AI, hybrid model.	A survey on river water quality modelling using AI models: 2000–2020
[15]	Neural networks, water quality, environment, BPNN, CNN, LSTM.	A review of ANN techniques for environmental issues prediction
[17]	AI, ANFIS, ANN, river, water quality.	AI for surface water quality monitoring and assessment: a systematic literature analysis
[18]	Pollutant, sediment load, ML tool, ANN, discharge prediction	Applications of IoT and AI in Water Quality Monitoring and Prediction: A Review
[5]	ANNs, feed-forward, recurrent, hybrid, water quality prediction.	A Review of the ANN Models for Water Quality Prediction
[20]	Water quality criteria, climate change, Urbanisation, eutrophication, best management practices, critical source areas, water quality index, ML algorithms, remote sensing.	Water quality prospective in Twenty First Century: Status of water quality in major river basins, contemporary strategies and impediments: A review

Reference	Keywords	Summary
[14]	AI; hybrid model; Wavelet transform; river water quality; prediction; review.	AI-based Single and Hybrid Models for Prediction of Water Quality in Rivers: A Review

The literature on WQ forecasting can be seen from a variety of perspectives. Emphasizing the supply side of the problem, Tiyyasha, et al. [19] reviewed papers on AI applications for studying river WQ prediction strategies, including the ANN, kernel-based, fuzzy-based complementary models, and hybrid models. In addition, model architecture, input variability, performance criteria, regional generalisation investigation, and comprehensive evaluations of AI approaches have progressed in river quality research. Han and Wang [15] published a study on how an ANN model can estimate WQ dynamics and compare with other approaches such as radial basis function neural network (RBFNN), long short-term memory (LSTM), and convolutional neural network (CNN) to find precise outcomes and explain their benefits. Additionally, the study focused on how many parameters of prediction and which country used the ANN model. Ighalo et al. [17] reviewed papers on neural networks, WQ parameters, location of study, and model accuracy. Mustafa et al. [18] gave an overview of the internet of things (IoT) in WQ monitoring. Furthermore, their study briefly explained an ANN model with its advantages, limitations, and its recent application. Chen et al. [5] focused on an ANN model and basic model architectures in WQ forecast, such as feed-forward, recurrent, and hybrid structures in addition to data collection, output strategy, input selection, data dividing, and data pre-processing (normalisation, missing data imputation, data correct, data abnormal). Giri [20] presented a holistic assessment of WQ decline in key river basins worldwide as shown in this review. In addition, nine modern methods, including field-scale assessment, optimisation strategies for placement of best management practices, a social component in watershed modelling, ML algorithms to discuss WQ issues in complex natural devices concomitant with spatial heterogeneity, and remote sensing in monitoring WQ were included. The existing constraints on improving WQ are then divided into major and secondary barriers. Rajaei et al. [14] reviewed different kinds of single and combined AI approaches including ANNs, Fuzzy Logic (FL), Genetic Programming (GP), SVM, hybrid ANN-ARIMA, hybrid Genetic Algorithm-Neural Networks (GA-NN), hybrid neuro-fuzzy (NF), and wavelet-based combined techniques such as wavelet-neuro fuzzy (WNF), wavelet-neural networks (WANN), wavelet-support vector regression (WSVR), and wavelet-linear genetic programming (WLGp) models were examined for the prediction of WQ in rivers.

Despite their comprehensive surveys of recent applications of AI methods to the WQ field, few researchers have included studies on hybrid algorithms and how they work step-by-step, and in detail, so we focused on hybrid ML techniques and their classification power, including data pre-processing methods. The reason to study these hybrid models in detail is that they have several advantages, such as (a) enhanced predictive performance due to increased capacity for pattern detection and simulation, (b) reduced risk of employing a sub-optimal technique (if used in isolation), and (c) a simplified procedure for model choice due to the utilisation of various components [21]. Hajirahimi and Khashei [22] classified hybrid models into several categories and explained the unique characteristics of the models. Based on this literature review, the goal of the paper is to categorise the hybrid models suggested for WQ modelling and forecasting into four main classes (the components combination-based hybrid models (CBH), parameter optimisation-based hybrid models (OBH), pre-processing-based hybrid models (PBH), and hybridisation of hybrid models).

2. Water Quality Parameters

The nature and amount of industrial, agricultural, and other anthropogenic activity within a region's catchments considerably influences surface WQ [23]. The WQ parameters are categorised into three primary groups: physical, chemical, and biological. Different WQ factors that have been modelled are reported in this paper. Physical WQ

parameters such as temperature (T), total dissolved solid (TDS), electrical conductivity (EC), salinity, and hydrogen ion concentration (pH) are often of concern as well. Dissolved oxygen (DO), chemical oxygen demand (COD), and biochemical oxygen demand (BOD) are examples of chemical sensors. Figure 2 shows various WQ factors modelled in the previous studies that used a hybrid model for prediction. It can be seen that most studies have been carried out to simulate DO and EC parameters in water.

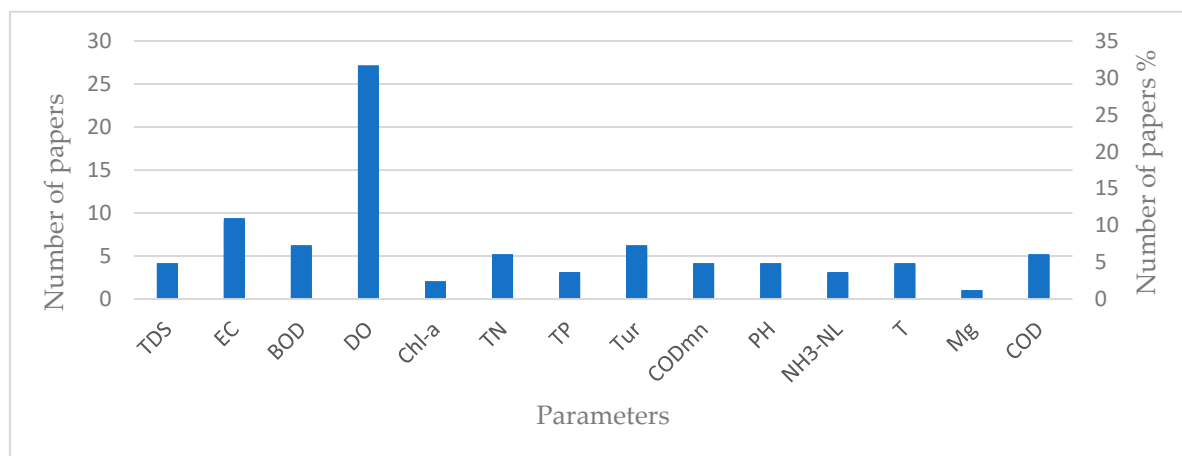


Figure 2. Number of studies employing each parameter of WQ over the years.

3. Machine Learning (ML)

ML has been applied for a long time and has received considerable attention over the last few years. It can handle a huge volume of data and permit non-linear constructions by utilizing complex mathematical calculations [24]. Additionally, ML are categorised as unsupervised and supervised learning. Supervised learning is employed to learn the primary relationship between input and output values. Unsupervised learning, in contrast, gives the learning algorithms no labels or known outcomes [25]. Several ML approaches have been promoted for modelling WQ parameters. The ML models applied include ANN [10,26–28], adaptive neuro-fuzzy inference system (ANFIS) [7,29,30], (SVR) [31–33], random forest (RF) [34,35], k-nearest neighbours (KNN) [36], Naive Bayes [37], decision tree (DT) [38,39], and extreme gradient boosting (XGB) [40]. The advantages and disadvantages of the most used ML techniques are summarised in Table 2.

Table 2. Advantages and disadvantages of the ANN, ANFIS, and SVR models.

Model	Advantage	Disadvantage	References
ANN	It can handle non-linear data series and complicated hydrological processes. Increase the accuracy of WQ forecasting by training and testing data series continuously without understanding the relationship between input and output.	Over parameterisation and overfitting difficulties are common in ANNs, especially when the approaches are based on optimal input selection, and the model is regarded as a black-box model. In addition, because no consistent principles control proper ANN model development and construction, it is not easy to prioritise a suitable model.	[18,41,42]
ANFIS	It can be used when the system input data is confusing and imprecise. It can manage non-linear data series and allow the modelling process to have the least possible uncertainty level.	When the number of fuzzy rules grows, it might become computationally expensive and may risk overfitting.	[42–44]

SVR	Its increased generalisation ability, unique and globally optimum structures, and ability to be quickly trained. And SVR's flexibility is one of its strongest features, dependent on several types of kernel functions such as linear, polynomial, and radial basis function (RBF) kernels.	Hyper-parameters like the penalty factor, accuracy, and kernel function variance significantly impact the performance of the SVR model.	[45,46]
RF	It is able to manage large datasets with several features, and the accuracy of modelling improves when the number of trees increases.	The training process is slowed when using the model with a high number of trees.	[47,48]

4. Data Pre-Processing Techniques

Data pre-processing techniques are considered essential to the data mining process [49]. Data preparation is vital to ensuring that all predictors receive equal attention during the learning phase and helps speed up the procedure [50,51]. These methods play an essential role in models by fostering high accuracy and minimal computational costs at the learning phase, as noisy and unreliable information that could exist in data records will adversely impact the training stage and outcome in a poor model [49]. The pre-processing data method consists of three approaches: normalisation, cleaning, and model input determination, as in Zubaidi et al. [52]. Previous studies used one or two pre-processing steps (Table S2 in the Supplementary Materials). In this study, only 48% of the researchers employed data normalisation, 53% utilised data cleaning, and 67% used best model selection.

1. Data Normalisation

The goal of data normalisation is to have the same range of values for each of the ANN model's inputs and to obtain the time series normally or nearly normally distributed, as this will aid in the stable convergence of the weights and biases and limit the impact of noise [2].

2. Data Cleaning

The cleaning strategy aims to determine and eliminate noise from raw data to reduce the error scale and improve the regression coefficient [2]. Data cleaning is required to discover and treat unwanted values, because the noise and outliers negatively impact data analysis and then the suggested model's performance [51,53].

3. Selecting appropriate descriptors

One of the most critical steps in data pre-processing is selecting the best model input [2]. The selection of explanatory factors influencing WQ metrics as model input data is vital in creating any successful model [54].

5. Hybrid Models

A hybrid model combines two or more methods, one serving as the primary model and the others as pre- or post-processing approaches [2]. In recent years, combined models have arisen as a way to construct flexible and efficient models and improve the forecasting accuracy of individual algorithms [5,55]. The hybrid models can be classified into four types, namely: the components combination-based hybrid models (CBH), parameter optimisation-based hybrid models (OBH), pre-processing-based hybrid models (PBH), and hybridisation of hybrid models as in Hajirahimi and Khashei [22]. There are different studies in the hybrid models shown in Figure 3.

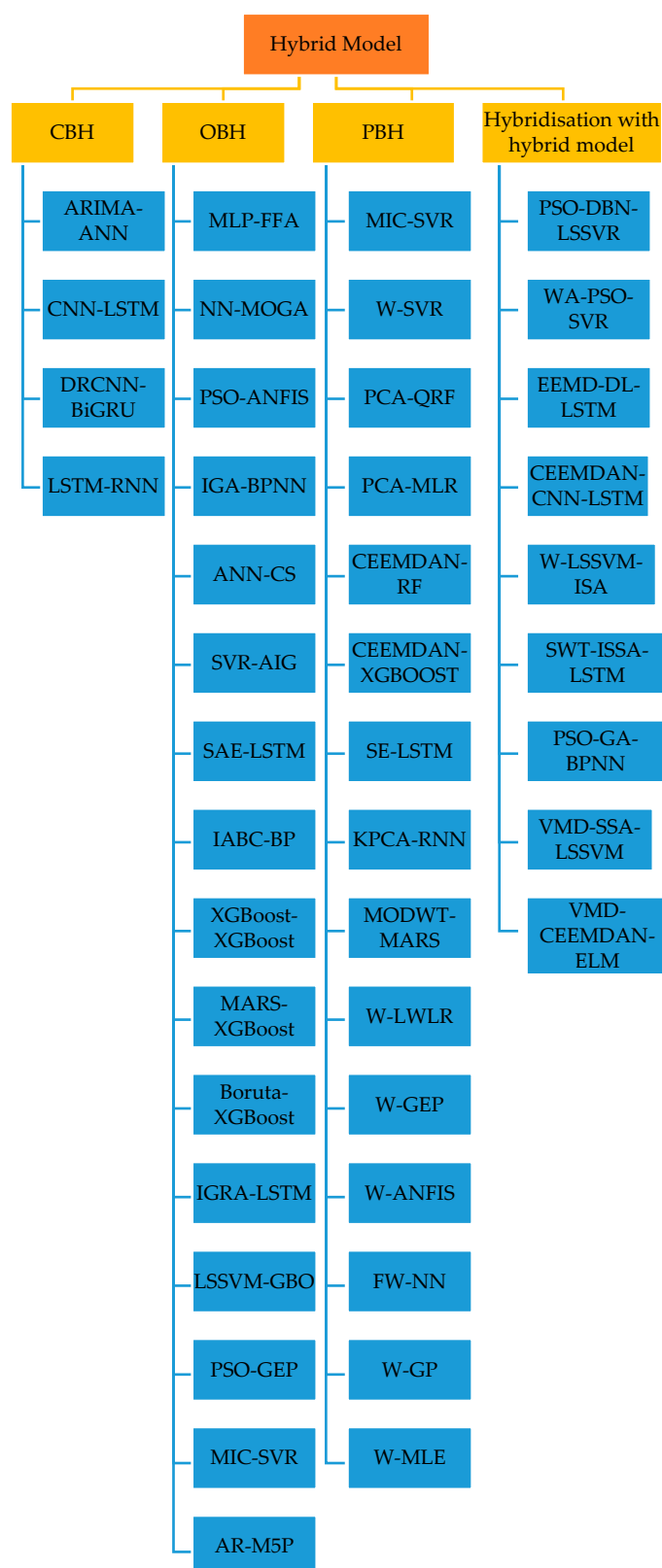


Figure 3. Hierarchy chart to a taxonomy of reviewed hybrid models.

5.1. Components Combination Based Hybrid Models (CBH)

In this section, ML models were combined to correct the relative incompetency of the individual models. The CBH models aim to improve prediction performance by enabling the remarkable capacity of individual prediction models regardless of combination

structures [22]. For example, Lola et al. [56] developed a combined technique to forecast daily WQ data (DO, water T, pH, and salinity) using ARIMA and ANN. When compared to stand-alone ARIMA and ANN, the results of the experiments demonstrate that the suggested model can be a viable and effective strategy to increase prediction precision with high correlation coefficients and decrease the error percentage for all indicators up to the maximum of 87.87% in both mean absolute error (MAE) and root mean square error (RMSE).

Barzegar et al. [57] investigated the predictive capability of two single deep learning (DL) models, the LSTM and CNN models, along with their combined CNN_LSTM technique to forecast short-term WQ. Two conventional ML methods, (SVR) and (DT), were also used, and their results were compared with DL models. Various statistical criteria were considered to assess the models. The results show that both DL models have similar performance for predicting Chlorophyll-a (Chl_a), and LSTM is better than CNN for simulating DO. Generally, the combined technique CNN_LSTM was superior to LSTM, CNN, SVR, and DT models, and it was able to simulate the high and low levels of WQ parameters, especially for the DO concentration. Similarly, Baek et al. [58] also suggested a composite model LSTM with the DL model to forecast the water level (WL) and quality parameters (Total phosphorus TP, total nitrogen TN, total organic carbon TOC). The outcomes showed that the hybrid model's performance was more precise according to the Nash–Sutcliffe efficiency (NSE).

Yan et al. [59] suggested using the one-dimensional residual convolutional neural networks (1-DRCNN) and bi-directional gated recurrent units (BiGRU), GRU, LSTM, and combined 1_DRCNN with BiGRU models, to forecast TN, TP, and potassium permanganate index (COD_{Mn}). The outcomes demonstrate that the combined technique has greater forecasting precision and generalisation to predict WQ than standalone models (LSTM, GRU, and BiGRU) based on statistical metrics, such as MAPE and the determination coefficient (R^2).

Hien Than et al. [60] investigated the LSTM-MA model to forecast DO, PH, COD, BOD, TSS, Tur, ammonia nitrogen oxidation-reduction potential ($\text{NH}_3\text{-NL}$), and Coliform variables and classified WQ. The LSTM-MA combined approach was employed to classify WQ, and this model is dependable and effective. The results revealed that the LSTM-MA was superior to the ARIMA, NAR, NAR-MA, and LSTM models according to the RMSE. According to these reviews, combined approaches can be customised by coupling two ML models together to suit the researchers' needs.

5.2. Parameter Optimisation-Based Hybrid Models (OBH)

Metaheuristics are commonly employed in WQ forecasting models to modify the parameters of other approaches, estimate the coefficients of a function, or train an intelligent agent and are a method for finding a good (near-optimal) answer at a reasonable computational cost [61].

Numerous approaches and algorithms have been developed to allow AI modelers to employ the computing system in hydrology, predicting and optimizing storage systems. The tasks are becoming more complex as the management of water resources improves to a broader scope, with the need to deal with the whims of climate change and more. Aside from AI models, other areas of research include optimisation algorithms and so-called evolutionary computing approaches, which can be utilised as a single algorithm for forecasting or combined with traditional methods to create a hybrid model.

5.2.1. Particle Swarm Optimisation (PSO)

This is a tool for computationally iterative search and optimisation [49]. It is scientifically inspired by social behaviour in animal societies, such as flocking birds or schools of fish. This technique utilises a swarm of particles, each of which represents a potential solution [47]. The PSO is evolved depending on two significant aspects of bird flocks' movement behaviour: their velocity and position [62]. It is applied to obtain the best forecast

technique coefficients that offer the lowest error between measured and forecasted values. So, it has been effectively used recently in various fields to select the optimal solution, such as in intelligent agriculture [63], WL [64], streamflow [62,65], drought [66], and WQ [67,68].

Aghel et al. [67] adopted two AI methods, ANFIS and ANFIS-PSO. The results showed that using two models to forecast inorganic markers of WQ is extremely effective. The flexibility of the PSO-ANFIS approach in modelling, on the other hand, is superior to the standalone ANFIS approach based on performance criteria (i.e., MRE%, MAE, RMSE, R and t statistics).

Azad et al. [68] applied the ANFIS model in conjunction with PSO and ant colony optimisation for continuous domains (ACO_R) in predicting WQ parameters. The ANFIS approach, which uses least squares and gradient descent as training algorithms, was applied and compared with ANFIS_PSO and ANFIS-ACO_R. The research revealed that ANFIS-PSO was the best model to forecast EC, TDS, TH, sodium adsorption ratio SAR, and carbonate hardness CH parameters. However, PSO may be a suitable strategy for optimizing and learning the aforementioned technique.

Shah et al. [69] proposed the hybrid feed forward neural network (PSO-FFNN) and combined gene expression programming (PSO-GEP) to forecast DO and TDS levels. The more essential input factors for TDS and DO forecasting were determined using principal component analysis (PCA). The fallouts show that the PSO-GEP model outperforms the PSO-FFNN model in terms of precision with statistical metrics.

5.2.2. Genetic Algorithm (GA)

This is a robust, powerful, optimised method based on natural selection and evolutionary principles [28]. GA was inspired by natural processes of biological evolution and has been widely employed to generate high-quality solutions to optimisation issues [70]. In the early twentieth century, genetic algorithms found their way into the field of hydrology [47]. The GA algorithm is applied in several areas, such as water flow [71,72] and WQ [73,74].

Stajkowski et al. [74] utilised the GA-LSTM technique to forecast the river water temperature (WT), and an RNN model as a benchmark to check the robustness of the suggested technique. The goal of using GA is to improve the ANN design process. The results showed that the GA-LSTM model outperformed the RNN, and the fundamental issue of identifying the ideal time frame and number of memory cell units was overcome. According to the findings, the GA-LSTM can be applied as an advanced DL approach for time series analysis.

Azad et al. [73] implemented GA, ACO_R, and differential evolution (DE) to improve the performance of an ANFIS. The most appropriate inputs for each model were first determined utilizing sensitivity analysis, and then all of the quality characteristics were forecasted using the aforementioned models. The most acceptable models for simulating EC and TH were ANFIS-DE, but both the ANFIS-DE and ANFIS-GA techniques showed improved performance compared to ANFIS in forecasting river WQ parameters.

Jin et al. [75] investigated a hybrid approach known as an improved genetic algorithm (IGA) back-propagation neural network (BPNN) to forecast variations in surface WQ for real-time early warning for NH₃-N, TURB, and EC parameters. IGA optimises the reasonable initial weight parameters and prevents the evolved method from choosing an optimal local outcome. BPNN is used to adjust suitable connection structures and find the features of WQ variation. The findings revealed that the created AI technique could significantly increase forecasting accuracy and dependability and provide effective real-time early warnings for emergency response. The proposed model outperformed BPNN according to statistical criteria.

5.2.3. Other Optimisation Algorithms

The firefly algorithm (FFA) proposed by Yang [76] in 2010 is a heuristic optimisation algorithm that is biologically inspired, and it depends on a specific behavioural pattern, especially the fireflies' light flashing characteristic [77].

Raheli et al. [78] evaluated the ability of a newly suggested combined prediction technique that depends on the FFA as a heuristic optimiser, coupled with the MLP. The model was applied to forecast monthly WQ (i.e., BOD, DO, COD, K, EC, PH, PO₄, Cl, Na, and NH₄N). Considering the performance criteria outcomes, the MLP-FFA technique outperforms the corresponding MLP model.

The cuckoo search (CS) was proposed by Yang and Deb. It is effective in tackling global optimisation issues [79]. Chatterje, et al. [80] used CS to increase support in the classification technique to predict WQ. To identify the best weight vector for the ANN model, the suggested approach (NN-CS) gradually diminishes an objective function (RMSE). The suggested technique was compared to other well-established approaches, such as NN-GA and NN-PSO, concerning the precision, Matthews correlation coefficient (MCC), recall, Fowlkes–Mallows index (FM index), and f-measure. The simulation outcomes showed that NN-CS outperformed the other models.

Li et al. [81] applied a combined approach that depends on LSTM and sparse auto-encoder (SAE) to enhance the forecasting precision of DO in aquaculture. SAE pre-trained the hidden layer data containing deep latent WQ aspects and then fed it into the LSTM to improve forecast precision. The outcomes showed that SAE-LSTM outperforms LSTM and SAE-BPNN.

The artificial bee colony (ABC) was proposed by Karaboga [82]. It has ushered in a new technique of thinking about optimisation algorithms. It was inspired by the study of the life cycle of bees and included two core concepts: self-organisation and division of labour [82]. The ABC optimisation approach has not been employed broadly in hydrology issues. However, there have been limited attempts to adopt it in optimizing WQ variables, such as Chen et al. [83], which used an improved artificial bee colony (IABC) algorithm with BPNN to predict DO, BOD, and COD_M parameters. The IABC algorithm optimised the connection weight values between network layers and the threshold of each layer using a BP neural network. When compared to the regular BP, ABC-BP, and PSO-BP neural network models, it was revealed that the IABC-BP neural network has better prediction capability and could reach considerably higher accuracy—about 25% higher than the BP neural network. The new technique is beneficial for predicting WQ in a water diversion project and might be quickly used in this area.

Grey Relational Analysis (GRA) is a subdivision of the grey system method that deals with ambiguous or uncertain problems and circumstances involving discrete data and inadequate knowledge [84]. Zhou et al. [85] proposed three models (LSTM, BPNN, and ARIMA) to forecast DO concentrations. Additionally, the improved grey relational analysis (IGRA) method was used for the feature selection of WQ information. The result revealed that LSTM outperformed the other models, and the hybrid IGRA-LSTM technique was the best.

Melesse et al. [4] proposed ten approaches: M5 prime M5P, bagging-M5P, AR-RF, random subspace (RS)-M5P, RF, RC-RF, random committee (RC)-M5P, bagging-RF, RS-RF, and additive regression (AR)-M5P to forecast salinity. The results revealed that the AR-M5P exceeded other models according to performance criteria. The combination of ML algorithms enhanced model performance in terms of capturing extreme salinity values, which is critical in managing water resources.

Tiyasha et al. [28] suggested four tree-based predictive models: RF, random forest geneRator (Ranger), conditional random forests (cForest), and XGBoost compared with algorithms, XGBoost, multivariate adaptive regression splines (MARS), and Boruta, GA. Additionally, four feature selector techniques (GA, Boruta, XGBoost, and MARS) were used to determine the optimum independent variables employed to forecast DO changes. The outcomes show that the performance of all predictive approaches was good as per

the features selected by the algorithms MARS and XGBoost. Additionally, the XGBoost predictive technique recorded the best performance when combined with MARS and XGBoost algorithms in terms of applied various statistical criteria.

Kadkhodazadeh and Farzin [86] explored a novel gradient-based optimiser (GBO) algorithm coupled with a least square support vector machine (LSSVM) technique for the evaluation of WQ parameters. The LSSVM-GBO method's performance is examined using three benchmark datasets to demonstrate its superiority (Housing, LVST, Servo). The novel hybrid algorithm's findings were then compared to ANN, ANFIS, and LSSVM techniques. The modelling results based on evaluation criteria revealed that LSSVM-GBO outperformed all other benchmark datasets and techniques. Then, EC and TDS modelling was done at varying time delays using the best input combination and the best algorithm. The Gotvand station has the highest modelling accuracy for EC and TDS parameters.

Dehghan, et al. [87] used SVR in stand-alone and hybrid versions. SVR was integrated with four metaheuristic algorithms, such as chicken swarm optimisation (CSO), social skidriver (SSD) optimisation, black widow optimisation (BWO), and the algorithm of the innovative gunner (AIG) to predict sufficient monthly DO. All the hybrid models produced good performance based on the different statistical criteria, and SVR–AIG offered better results. Moreover, combined techniques improved the precision of the stand-alone SVR method by 6.52–1.75%.

5.3. Preprocessing-Based Hybrid Models (PBH)

In this method, the input data are pre-processed using various methods such as decomposition-based, filter-based, denoising-based, feature selection, and data cleaning approaches. Following this, the appropriate individual model forecasts the screened time series [88].

Solg, et al. [89] investigated two models: SVR and ANFIS. The wavelet transform approach was used to clean raw data from noise and analyse the data set into sub-series. Additionally, principal component analysis (PCA) is applied to determine the best predictors. The outcomes showed that the SVR was better than the ANFIS model, the wavelet transform approach improved data quality, and the hybrid W-PCA-SVR is the best technique.

Zhang et al. [23] designed Kernel PCA (kPCA) with a recurrent neural network (RNN) model to estimate the trend of DO. The kPCA technique is used to reconstruct WQ variables, which tries to minimise the noise in raw sensory data while preserving actionable information. The model can use previous knowledge to forecast future trends because of the RNN's recurrent connections. When compared to present AI techniques such as FFNN, SVR, and the general regression neural network model (GRNN), the kPCA-RNN model attained the predicted accuracy and outperformed the comparative models.

Al-Sulttani et al. [90] proposed five various hybrid ML techniques, including Gradient Boosting Machines (GBM H2O), RF, Quantile regression forest (QRF), radial SVM, and Stochastic Gradient Boosting (GBM). Furthermore, the techniques were integrated by employing two various algorithms for identifying features, e.g., GA and PCA, to predict monthly BOD values. GA was used to select the best-fitting predictions based on their evolutionary potential. The findings show that the combined PCA-QRF approach was the best performing approach to predict WQ compared to the other models.

Bi et al. [91] suggest ANN, SVR, ARIMA, XBoost, and LSTM models to forecast DO and COD_{mn}. The outcomes reveal that the SE-LSTM technique is superior to the other methods based on statistical metrics. Hence, The Savitzky–Golay filter can remove possible noise from the WQ time series, and the LSTM can examine non-linear properties in a complex water environment.

Ahmed et al. [92] created a hybrid model by combining the MARS model with the maximum overlap discrete wavelet transformation (MODWT) (i.e., MODWT-MARS). The suggested model was also compared against various ML techniques (MARS, CEEMDAN-

MARS, CEEMDAN-SVR, SVR, KRR, KNN, RF) to estimate daily WQ parameters. The results revealed that the combined algorithm (i.e., MODWT-MARS) was superior to the other methods according to statistical criteria. This hybrid approach could be used to anticipate WQ characteristics using fewer predictor factors in the future.

Ahmadianfa, et al. [93] proposed a novel hybrid model discrete wavelet transform coupled with locally weighted linear regression (LWLR) and employing the mother wavelet Bior 6.8 to analyse data into two levels. The outcomes reveal that the W-LWLR technique outperforms other methods such as LWLR, MLR, SVR, ARIMA, W-MLR, W-ARIMA, and W-SVR.

Eze et al. [94] developed a new combined forecast approach that depends on hybrid empirical mode decomposition (EEMD) and an LSTM neural network. Initially, the integrity of the datasets is improved by using moving average filtering and linear interpolation techniques to pre-treat the WQ indicator datasets in this combined EEMD-DL-LSTM technique. Then, the EEMD technique decomposes the dataset of measured real sensor WQ characteristics. Finally, a multi-feature selection procedure is used to carefully choose a collection of IMFs that are substantially linked with the measured real-world WQ parameter datasets and integrate them as inputs to the DL-LSTM neural network. The innovative hybrid prediction model's performance is validated by comparing the results to real datasets. Various measurement criteria, such as (MAE, MAPE, RMSE, and MSE), were utilised to assess the overall precision of the unique hybrid prediction technique.

5.4. Hybridisation of Hybrid Models

The hybridisation of hybrid models is a novel idea proposed to improve forecasting precision over traditional hybrid classes [22].

In 2020, several researchers used a combined hybrid model with a pre-processing algorithm, such as Ya, et al. [95], who suggested a technique for forecasting WQ parameters (TN) that depends on the deep belief network (DBN) method. The deep belief network's network is optimised using the PSO algorithm, which extracts feature vectors from WQ data at several scales. The PSO-DBN WQ prediction model is then integrated with the least squares support vector regression (LSSVR) machine, which is used as the top forecast layer of the approach. When comparing the proposed model (PSO-DBN-LSSVR) to the classic back propagation (BP) neural network, the DBN neural network, LSSVR, and the DBN-LSSVR hybrid technique, the outcomes display that the model can accurately forecast the WQ parameters and has good robustness based on statistical metrics.

Wang et al. [96] established a combined assembly wavlet analysis (WA-PSO-SVR) to simulate three WQ metrics: $\text{KMnO}_4(\text{COD}_{\text{Mn}})$, $(\text{NH}_3\text{-N})$, and (DO). The results showed that the combined WA-PSOSVR technique outperformed two other methods (PSO-SVR and a single SVR) in predicting non-linear stationary and non-stationary time series, particularly for extreme value prediction. Daily forecasts were more precise than monthly forecasts, indicating that the combined technique was better suited to short-term forecasting in this case.

In 2021, Son, et al. [97] suggested a novel hybrid technique (SWT-ISSALSTM). An improved LSTM model was presented to overcome the gradient disappearance or explosion in standard RNNs, as well as the inability to handle the issue of long-time dependence and enhance the model's performance. Additionally, a hybrid model using synchrosqueezed wavelet transform (SWT) to clean the raw data was used to resolve the non-stationarity, unpredictability, and nonlinearity of the WQ parameters data. The improved sparrow search algorithm (ISSA), a novel heuristic optimisation technique integrating Cauchy mutation and opposition-based learning (OBL), was also used to obtain the optimum hyperparameter values for the LSTM method. The suggested combined system was assessed utilising weekly WQ parameters. The results show that the addressed model, which combines the SWT's strong noise-resistant resilience and the LSTM's non-linear mapping, outperforms the peer models (stand-alone LSTM, BPNN, SVR, SWT-LSTM, and

ISSA-LSTM) at two gauging stations. The suggested combined technique (SWT-ISSA-LSTM) can be utilised as a replacement framework for predicting WQ.

Jamei et al. [98] aimed to find two novel wavelet-complementary intelligence methodologies: the wavelet least square support vector machine coupled with improved simulated annealing (W-LSSVM-ISA) and the wavelet extended Kalman filter integrated with an artificial neural network (W-EKF-ANN), to predict monthly Mg and SO₄ metrics. The findings showed that both novel complementary paradigms could provide acceptable accuracy for WQP prediction based on correlation coefficient R and RMSE.

Sha et al. [99] evaluated various DL approaches such as CNN, LSTM, and CNN-LSTM models. Moreover, they employed a complete ensemble empirical mode decomposition algorithm (EEMD) with adaptive noise (CEEMDAN) to decompose and reduce the intricacy of DO and TN concentration. The outcomes reveal that the CNN-LSTM performed better than the stand-alone CNN and LSTM models, the techniques using CEEMDAN-based input data performed significantly better than the techniques using original input data, and the technique precision incrementally reduced with the rise of forecasting stages, while the original input data decayed more rapidly than the CEEMDAN-based input data, indicating that the input data pre-processed by the CEEMDAN method could significantly enhance.

Yan et al. [100] suggested four stand-alone models (GA-BPNN), (PSO-BPNN), (PSO-GA-BPNN), and (BPNN) to forecast DO concentration. The finding indicated that the PSO-GA-BPNN technique had enhanced forecasting precision and robustness compared with other methods. The connection weight and threshold of BPNN were optimised using PSO and GA in this work. This hybrid PSO and GA algorithm are based on the PSO algorithm, with the GA inserted during the PSO method's execution. It combines the benefits of both algorithms, resulting in less processing, faster convergence, and better global convergence performance.

The details of the selected papers, including authors, and the location, time scale, methods, input variables, output prediction, and evaluation criteria, are given in Table 3.

An analysis of several reviewed articles on optimisation algorithms revealed the following:

- The general optimisation approaches demonstrated their ability to tune all AI models to achieve a far higher score on various evaluations as compared to a single model, which does not use any optimisation technique. In addition, when compared to a trial-and-error procedure, the probability of achieving ideal values is substantially higher.
- The most commonly employed algorithm in the WQ area and paired with AI approaches to forming a combined model is the PSO algorithm.
- Several studies used pre-processing algorithms to overcome the data's non-stationarity, randomness, and nonlinearity of the WQ indicators. However, all pre-processing data steps were not used in most papers.
- The trend of using hybrid models has increased in recent years.

Table 3. Summary of application of different type hybrid models in WQ monitoring.

Authors	River	Location	Scale	Predictors	Target	Models Used	Best Model	Measures of Accuracy
[4]	Babol-Rood River	Northern iran	Monthly	PH, HCO ₃ , CL, SO ₄ , EC, Na, Mg, Ca, Q, TDS,		M5P, RF, bagging-M5P, bagging-RF, RS-M5P, RS-RF, RC-M5P, RC-RF, AR-M5P, AR-RF	AR-M5P	RMSE, MAE, NSE, BPIAS
[59]	Luan	Tangshan City	Every 4 h	T, PH, DO, BOD, Tur, COD-Mn, NH ₄ -N, TP, TN, COD-Mn TN		1-DRCNN, BiGRU, GRU, LSTM	Combined (1-DRCNN-BiGRU)	MAE, MAPE, RMSE, R ²
[31]	Pearl	China	Used six different time scale	PH, EC, Tur, DO, NH ₃ -N, TP, COD-Mn, TN, DO WL, WT		SVR	MIC-SVR	NSE, R ² , RMSE
[69]	Indus river	Asia	monthly	Ca, Mg, Na, Cl, SO ₄ , HCO ₃ , PH, EC, WT, DO, TDS DO, TDS		PSO-FFNN, PSO-GEP	PSO-GEP	NSE, RMSE, RRMSE, P, R
[60]	Dong Nai River	Vietnam	Month	DO, PH, COD, BOD, DO, PH, COD, BOD, TSS, Tur, NH ₃ -NL, Col-TSS, Tur, NH ₃ -NL, iform Coliform		ARIMA, NAR, NAR-MA, LSTM, and LSTM-MA	LSTM-MA	MSE, RMSE, MAPE
[86]	Karun	Iran	Monthly	Ca, Cl, Mg, Na, SO ₄ , SAR, Sum.C, Sum.A, TDS, EC PH, Q, HCO ₃		ANN, ANFIS, LSSVM, LSSVM-GBO	LSSVM-GBO	MAE, RRMSE, R, R ²
[57]	Greece's Prespa Lake	Small south-eastern rope	Eu-Every min	15-PH, ORP, T, EC, DO, Chl-a	DO, Chl-a	LSTM, CNN, SVR, and DT, CNN-LSTM	CNN-LSTM	R, RMSE, MAE, PBIAS, NSE, WI, and graphical plots (Taylor diagram, box plot and spider diagram)
[58]	Nakdong	South Korea	Monthly	WL, TOC, TP, TN	TOC, TP, TN	CNN-LSTM	CNN-LSTM	NSE, R ² , MSE
[97]	Yongding River and Gangnan gauging stations in the Haihe River Basin,	Chain	weekly	DO	DO	SWT-LSTM, ISSA-LSTM, SWT-SSA-LSTM, SVR, BPNN, and single LSTM	SWT-ISSA-LSTM	AE _{max} , MAE, MAPE, RMSE, R ² , CC, NSE, IA, 1.96 Se
[98]	Maroon	Southwest Iran	monthly	Q, EC, Mg, SO ₄	Mg, SO ₄	LSSVM-ISA, EKF-ANN W-LSSVM-ISA, W-EKF-W-LSSVM-ISA ANN		R, RMSE, KGE

[90]	the Euphrates River	Iraq	monthly	T, PH, EC, TSS, BOD, ALK, Ca, COD, SO ₄ BOD TDS, TSS, Tur		(QRF), (RF), (SVM), (GBM) (GBM_H2O) PCA-QRF	R ² , RMSE, AE, NSE, W index, PBIAS
[99]	Xin'anjiang River	Huangshan City,	4-h	DO, TN	DO, TN	CNN, LSTM, CNN-LSTM, CEEMDAN-CNN-LSTM	CE, RMSE, MAPE
[100]	Beihai Lake	Beijing	Hourly	PH, CAHL-A, NH ₄ H, BOD, EC	DO	BPNN, PSO-BPNN, GA-BPNN, PSO-GA-BPNN	APE _{max} , MAPE, RMSE, R ²
[85]	Tai Lake, Victoria Bay	China.	Monthly in Tai lake, every two weeks in Victoria Bay	Tai lake (TN, TP, NH ₃ -N, SS, WT, DO, PH, Transparency, CL, Precipitation in Victoria Bay (E.coli, BOD5, NH3-N, Nitrite, phosphate, PH, WT, salinity	DO	LSTM, BP, ARIMA IGRA -LSTM	RMSE,
[38]	Tualatin	Oregon, USA	Hourly	T, DO, PH, Specific conductance, Tur, fluorescent dissolved organic matter	T, DO, PH, Specific conductance, Tur, fluorescent dissolved organic matter	RF, XGboost, CEEMDAN-RF, CEEMDAN-XGBoost, or-PSO-SVM, RBFNN, LSSVM and LSTM	MAPE, MAE, RMSE, RMSPE, U1, U2
[28]	Klang	Malaysia	Monthly daily	15 WQ parameters, 7 hydrological components		XGBoost- XGBoost MARS-XGBoost Boruta- XGBoost GA- XGBoost Boruta-Ranger GA-Ranger MARS-Ranger XGBoost-Ranger	R ² , RMSE, MAE, NSE, MD
[91]	GuBeiKou,	Beijing, China.	Every hour	4-DO, CODmn	DO, CODmn	ANN, SVR, ARIMA, XBoost, LSTM, SE-LSTM	MAE, MAPE, RMSE
[83]	Yangtze river	China	Daily	DO, BOD, CODmn, T, PH, NH3-N	DO, BOD, CODmn	BP, ABC-BP, PSO-BP, IABC-BP	R ² , NSE, RE,

[81]	Shrimp pond	China	Every 10 min	DO, WT, Am, PH, AT, DO Hu, AP, WS	SAE-LSTM, SAE-BPNN, SAE-LSTM LSTM, BPNN	MSE, RMSE, MAPE
[23]	Burnett river	Australia	Hourly	T, EC, DO, PH, Chl-a DO	KPCA-RNN, FFNN, SVR, KPCA-RNN GRNN	MAE, R ² , RMSE
[94]	Abalone farm	South African	Monthly	DO, T, Tur, PH DO, T, Tur, PH	BP, SAE-BP, DL-LSTM, SAE-LSTM, EEMD-DL-EEMD-DL-LSTM LSTM	RMSE, MAE, MSE, MAPE
[96]	Grand Canal	China	Daily and Monthly	CODMn, NH3-N, DO COD	SVR, PSO-SVR, WA-PSO-SVR SVR	RMSE, NSE, MAPE, R ²
[68]	Zayandehrood River	Iran	(2001-2015)	TDS, EC, pH, HCO ₃ , Cl, SO ₄ , Mg, Na, K, CO ₂ , Ca, CH, and TH EC, TDS, SAR, CH, and TH	ANFIS, ANFIS-PSO, ANFIS-PSO FIS-ACOR	MAPE, RMSE, R ² , d
[87]	Cumberland River	Southern States	United Monthly	T, Q DO	SVR, SVR-CSO, SVR-SSD, SVR-BWO, SVR-SVR-AIG AIG	RMSE, R ² , MAE, NSE, BIAS
[80]	Hooghly River	West Bengal, India	Monthly	H, Cl, TH, total alkalinity, Turbidity and Residual Chlorine H, Cl, TH, TALK, Tur and Residual Chlorine	NN-CS, NN-GA, NN-PSO NN-CS	RMSE, accuracy, precision, recall, f-measure, (MCC) (FM index)
[67]	Kermanshah Province	Iran	Monthly	pH, T, SC, SA TALK, TH, TDS, EC	ANFIS, PSO-ANFIS PSO-ANFIS	MRE, RMSE, R
[95]	Juhe River	China	Every 4 h	T, pH, DO, conductivity, NTU, COD _{mn} , TP, TN NH ₄ N	BPNN, LSSVR, DBN, DBN-LSSVR, PSO-DBN-PSO-DBN-LSSVR LSSVR	R ² , RMSE, MAE, MAPE
[78]	Langat Rive	Malaysia	Monthly	COD, PO ₄ , TS, K, Na, Cl, EC, PH, NH ₄ -N BOD, DO	MLP, MLP-FFA MLP-FFA	RMSE, R, WI
[92]	Surma River	Bangladesh	Monthly	Humidity, WT, rainfall, TDS, pH, turb, AT DO	MARS, CEEMDAN-MARS, CEEMDAN-SVR, MODWT-MARS SVR, KRR, KNN, RF	R, WI, RMSE, MAE
[93]	Sefidrud River	Iran	Monthly	EC, Q EC	SVR, W-SVR, ARIMA, W-ARIMA, MLR, and W-W-LWLR MLR, LWLR, W-LWLR	RMSE, NSE, MAE, RAE, MSRE

[101]	Kinta River	Malaysia	Monthly	DO, BOD, COD, Temp, DO NH3, TS, Cl, Ca, PH Na		LSTM, ELM, HW, GRNN, SAE, WAE, LSTM-RF, ELM-RF, GRNN-RF and HW-RF	NSE, WI, RMSE, MAE, MSE, CC
[102]	Yangtze River	China	Weekly	DO	DO	LSSVM, SSA-LSSVM, VMD-LSSVM, SVR, BPNN, VMD-SSA- LSSVM	NSE, RMSE, MAE, MAPE, CC, R ²
[103]	Tolo Harbour	China	bi-weekly/monthly	BOD, TIN, DO, PO ₄ , Temp, Chl-a, SDD, pH	HAB	ANN (LM-PSO), ANN(LM-GA), ANN (GDM-PSO) ANN (GDM-GA), SVM	ANN (LM-PSO) RMSE, CC
[104]	crab culture ponds	China	10 min	DO	DO	CEEMDAN-LZC-GOB- LPSO-GRU, CEEMDAN- GOBLPSO-GRU, GRU, CEEMDAN-LZC-GOB- LPSO-LSTM, CEEMDAN-GOBLPSO- LSTM, CEEMDAN-LZC-GOB- LPSO-RNN, CEEMDAN- GOBLPSO-RNN, RNN, BPNN	CEEMDAN-LZC- GOBLPSO-GRU MAPE, RMSE R ²
[105]	Huaihe River, Poto- mac River	China, US	Weekly, every 15 min	COD, DO, NH ₃ -N	COD, DO, NH ₃ -N	ANN, ARIMA, MLE, W- MLE	W-MLE ARE, MRE
[106]	Bam Plain	Normashir, Iran	Monthly	EC, Cl, Na, Ca, Mg, SAR	Cl, EC, SAR	FCM, GP, ANN, ANN- PSO, IDW, RBF, kriging, NF-GP, NF-MCF	NF-GP RMSE, MAE, CC
[107]	Karaj River	Iran	Monthly	BOD, Q	BOD	WANN, ANN, GP, DT, BN, WGP	WGP MAE, RMSE, R
[74]	Credit River	Canada	hourly	WT	WT	GA-LSTM, LSTM, RNN	GA-LSTM R ² , MAE, RMSE, RSR, mNSE, md, KGE
[108]	River of Shanghai	Shanghai	Daily	P, N, BOD, NH ₄ -NO ₃ , COD	index COD	GM, RNN, LSTM-RNN	LSTM-RNN RMSE, MAPE
[75]	Ashi River	China	Every 4 h	NH ₃ -N, TURB, EC	NH ₃ -N, TURB, EC	BPNN, IGA-BPNN	IGA-BPNN RMSE, MAE, MRE, R ²

[109]	Qiantang River, Zhejiang Province	China	Every 4 h	permanganate index, pH, TP, DO	permanganate index, pH, TP, DO	BPNN, SVR, LSTM, GRU, SRN, RNNs-DS	RNNs-DS	RMSE MAE MAPE
[110]	Yamuna	India	Monthly	BOD	BOD	ANFIS, ANN, W-ANFIS	W-ANFIS	MAE
[111]	Isfahan-Borkhar	Iran	Monthly	SO ₄ , Cl, HCO ₃ , K, Na, Mg, Ca	EC, SAR, TH	ANFIS-CGA, ANFIS-ACO _R , ANFIS-DE, AN-ANFIS-CGA	FIS-PSO, ANFIS	R ² , RMSE, MAPE, SI
[112]	Small Prespa Lake	Greece	Daily	Chl-a, DO	Chl-a, DO	LSSVM, CEEMDAN-LSSVM, VMD-CEEMDAN-ELM, CEEMDAN-ELM, VMD-CEEMDAN-ELM		R, RMSE, MAE, BIAS
[113]	South-to-North-Water Diversion Project	China	Daily	PI, Ph, TN, WT, turb, EC, Chl, DO, DOM	TN, WT, DOM, DO, WVP, AT, PM2.5	BPNN, CS-BP, PSO-BP, GRNN, CS-BP		RMSE, MAPE
[114]	Nazlu Chay, Tajan, Zayandeh Rud and Helleh	Iran	Seasonal	TDS, Cl, EC, Na	TDS	ANN, ANFIS-GP, ANFIS-SC, GEP, WANN, WANFIS, GP, WANFIS-SC, WGEP	WGEP	R, RMSE and MAE
[56]	offshore of Kuala	Terengganu	Daily	WT, pH, salinity, DO	WT, pH, salinity, DO	ARIMA, ANN, ARIMA-ANN	ARIMA-ANN	RMSE, MAE
[115]	Pearl River	China	Daily	COD, NH ₄ N, DO, EC, WT, pH, TU	COD, Tur	WNN, ANN, FWNN	FWNN	R, R ² , MAPE, RMSE, MSE
[116]	Aji-Chay River	Iran	Monthly	EC	EC	ELM, ANFIS, WA-ELM, WA-ANFIS	boosting multi-WA-ELM, multi-WA-ANFIS	RMSE, R ² , NSE
[89]	Karun River	Iran	Monthly	DO, Q, WT, BOD	BOD	SVR, ANFIS, WSVR, WANFIS	WSVR	RMSE, R ²

6. Future Research Directions

Azad et al. [68] suggested employing modified algorithms to enhance other types of ML methods that suffer comparable shortcomings and comparing these changed hybrid models to different physical and soft computing models. Shah et al. [69] proposed that other studies should employ extra AI models, such as ensemble forecasting combined with PSO, to further develop their performance with optimum parameters in modelling WQ factors. Li et al. [81] recommended that it is possible to create a deep network through layer-wise pre-training to collect deeper latent features to investigate the impact of raising the network layers of SAE (sparse auto-encoder) on predictive performance. Tiyyasha et al. [28] mentioned that the MARS algorithm as a feature selector and the XGBoost algorithm as both a feature selector and a predictive method should be investigated to create various types of WQ data. In addition, the Boruta algorithm should be used to create scenarios to determine the best predictors' cutoff value. Furthermore, an examination of uncertainty is required to determine the stochasticity of the data application using the suggested AI techniques (RF, cFores, Ranger) and XGBoost. Song et al. [97] stated that more effective pre-processing procedures for WQ data should be investigated to increase the model's precision. Jamei et al. [98] stated that, in the future, an ensemble multi-wavelet transform (EMWT) paradigm could be employed to utilise the wavelets simultaneously. On the other hand, an ensemble tree-based method could be effective for combining the benefits of each complementary strategy to estimate surface water WQPs. Additionally, combined versions that incorporate more than one training technique for predictability improvement are recommended for such an issue of WQ parameters.

Additionally, all of the studies reviewed here support the suggestions below:

- It is recommended that the three data pre-processing steps be applied to avoid outliers and noise and to select the most reliable and precise data to be employed as predictors later.
- Other techniques for pre-treatment data, such as EEMD and singular spectrum analysis, are proposed.
- Selection predictors are significant in determining the model's performance and precision. Accordingly, it is advised that more efforts be made to select the optimal predictors' combination; consequently, it is proposed that other techniques be used to choose the predictors, such as feature extraction methods, feature selection, and dimensionality reduction methods.
- Applying hybrid metaheuristic algorithms and soft computing techniques in WQ parameter prediction has grown considerably in recent years. Nevertheless, there is still room for improvement concerning WQ parameter prediction.

7. Conclusions

This work attempted to review papers that employed hybrid methodologies to simulate WQ parameters. The selected papers in this review revealed that there has been an increasing tendency toward employing these methods in the area of WQ modelling in recent years. Combining data pre-processing techniques with metaheuristic algorithms and soft computing models has enhanced WQ prediction accuracy among the many modelling approaches. Therefore, hybrid models are the most effective techniques that must be used to enhance the precision of WQ parameter predictions. A comprehensive hybrid model incorporates both pre-processing techniques and metaheuristic algorithms. Accordingly, a key strength of the current study is that it represents a comprehensive examination of all the above factors.

Most of the previous research used the WQ parameters as predictors, and few of them applied other factors such as weather. For this type of data, models that incorporate only factors that have been proven effective are more precise than models that incorporate all factor data without testing variables' efficiency. Additionally, most previous studies used one or two steps of pre-processing, which impacted the accuracy of prediction

models. Therefore, in future studies, the efficiency of the factors should be tested (predictors) before applying all of the data as input to the forecast models and using normalisation and cleaning. Furthermore, although significant advances in hybrid model techniques have been made recently, no new techniques have emerged as the best forecasting model. Consequently, WQ parameter forecasting remains a research problem, which leaves room for scholars to improve hybrid techniques for specific applications.

Supplementary Materials: The following supporting information can be downloaded at: www.mdpi.com/article/10.3390/environments9070085/s1, Table S1: Abbreviations; Table S2: Review of researchers who used data pre-processing [4], [16], [23], [28], [38], [56], [57], [58], [59], [60], [67], [68], [69], [73], [74], [75], [78], [80], [81], [83], [85], [86], [87], [89], [90], [91], [92], [93], [94], [95], [96], [97], [98], [99], [100], [101], [102], [103], [104], [105], [106], [107], [108], [109], [110], [111], [112], [113], [114], [115], [116].

Author Contributions: Conceptualization, Z.S.K. and S.L.Z.; methodology, Z.S.K., S.L.Z. and N.A.-A.; investigation, Z.S.K.; resources, N.A.-A., S.E. and K.H.; writing—original draft preparation, Z.S.K.; writing—review and editing, Z.S.K., S.L.Z. and S.O.-M.; visu-alization, S.O.-M., S.E. and K.H.; supervision, S.L.Z.; project administration, S.L.Z.; funding acquisition, N.A.-A. All authors have read and agreed to the published version of the manuscript.

Funding: The APC was funded by Lulea University of Technology.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Banerjee, A.; Chakrabarty, M.; Bandyopadhyay, G.; Roy, P.K.; Ray, S. Forecasting environmental factors and zooplankton of Bakreswar reservoir in India using time series model. *Ecol. Inform.* **2020**, *60*, 101157, 1–37. <https://doi.org/10.1016/j.ecoinf.2020.101157>.
- Zubaidi, S.L.; Ortega-Martorell, S.; Al-Bugharbee, H.; Olier, I.; Hashim, K.S.; Gharghan, S.K.; Kot, P.; Al-Khaddar, R. Urban Water Demand Prediction for a City That Suffers from Climate Change and Population Growth: Gauteng Province Case Study. *Water* **2020**, *12*, 1885. <https://doi.org/10.3390/w12071885>.
- van Vliet, M.T.H.; Franssen, W.H.-P.; Yearsley, J.R.; Ludwig, F.; Haddeland, I.; Lettenmaier, D.P.; Kabat, P. Global river discharge and water temperature under climate change. *Glob. Environ. Change* **2013**, *23*, 450–464. <https://doi.org/10.1016/j.gloenvcha.2012.11.002>.
- 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>.
- Chen, Y.; Song, L.; Liu, Y.; Yang, L.; Li, D. A Review of the Artificial Neural Network Models for Water Quality Prediction. *Appl. Sci.* **2020**, *10*, 5776. <https://doi.org/10.3390/app10175776>.
- Seo, I.w.; Yun, S.H.; Choi, S.Y. Forecasting Water Quality Parameters by ANN Model Using Pre-processing Technique at the Downstream of Cheongpyeong Dam. *Procedia Eng.* **2016**, *154*, 1110–1115. <https://doi.org/10.1016/j.proeng.2016.07.519>.
- Abba, S.I.; Hadi, S.J.; Abdullahi, J. River water modelling prediction using multi-linear regression, artificial neural network, and adaptive neuro-fuzzy inference system techniques. *Procedia Comput. Sci.* **2017**, *120*, 75–82. <https://doi.org/10.1016/j.procs.2017.11.212>.
- Tahraoui, H.; Belhadj, A.-E.; Hamitouch, A.-e.; Bouhedda, M.; Amrane, A. Predicting the concentration of sulfate (SO₄²⁻) in drinking water using artificial neural networks: A case study: Médéa-Algeria. *Desalination Water Treat.* **2021**, *217*, 181–194. <https://doi.org/10.5004/dwt.2021.26813>.
- Mohammadpour, R.; Shaharuddin, S.; Chang, C.K.; Zakaria, N.A.; Ab Ghani, A.; Chan, N.W. Prediction of water quality index in constructed wetlands using support vector machine. *Environ. Sci. Pollut. Res. Int.* **2015**, *22*, 6208–6219. <https://doi.org/10.1007/s11356-014-3806-7>.
- Stamenkovic, L.J. Application of ANN and SVM for prediction nutrients in rivers. *J. Environ. Sci. Health A Tox. Hazard Subst. Environ. Eng.* **2021**, *56*, 867–873. <https://doi.org/10.1080/10934529.2021.1933325>.
- Aldhyani, T.H.H.; Al-Yaari, M.; Alkahtani, H.; Maashi, M. Water Quality Prediction Using Artificial Intelligence Algorithms. *Appl. Bionics. Biomech.* **2020**, *2020*, 6659314. <https://doi.org/10.1155/2020/6659314>.
- Hmoud Al-Adhaileh, M.; Wasellallah Alsaade, F. Modelling and Prediction of Water Quality by Using Artificial Intelligence. *Sustainability* **2021**, *13*, 4259. <https://doi.org/10.3390/su13084259>.
- Ahmed, U.; Mumtaz, R.; Anwar, H.; Shah, A.A.; Irfan, R.; García-Nieto, J. Efficient Water Quality Prediction Using Supervised Machine Learning. *Water* **2019**, *11*, 2210. <https://doi.org/10.3390/w11112210>.
- Rajae, T.; Khani, S.; Ravansalar, M. Artificial intelligence-based single and hybrid models for prediction of water quality in rivers: A review. *Chemom. Intell. Lab. Syst.* **2020**, *200*, 103978. <https://doi.org/10.1016/j.chemolab.2020.103978>.

15. Han, K.; Wang, Y. A review of artificial neural network techniques for environmental issues prediction. *J. Therm. Anal. Calorim.* **2021**, *145*, 2191–2207. <https://doi.org/10.1007/s10973-021-10748-9>.
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>.
17. Ighalo, J.O.; Adeniyi, A.G.; Marques, G. Artificial intelligence for surface water quality monitoring and assessment: A systematic literature analysis. *Modeling Earth Syst. Environ.* **2020**, *7*, 669–681. <https://doi.org/10.1007/s40808-020-01041-z>.
18. Mustafa, H.M.; Mustapha, A.; Hayder, G.; Salisu, A. Applications of IoT and Artificial Intelligence in Water Quality Monitoring and Prediction: A Review. In Proceedings of the 2021 6th International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 20–22 January 2021; pp. 968–975.
19. Tiyasha; Tung, T.M.; Yaseen, Z.M. A survey on river water quality modelling using artificial intelligence models: 2000–2020. *J. Hydrol.* **2020**, *585*, 124670. <https://doi.org/10.1016/j.jhydrol.2020.124670>.
20. Giri, S. Water quality prospective in Twenty First Century: Status of water quality in major river basins, contemporary strategies and impediments: A review. *Environ. Pollut.* **2021**, *271*, 116332. <https://doi.org/10.1016/j.envpol.2020.116332>.
21. Hajirahimi, Z.; Khashei, M. Hybrid structures in time series modeling and forecasting: A review. *Eng. Appl. Artif. Intell.* **2019**, *86*, 83–106. <https://doi.org/10.1016/j.engappai.2019.08.018>.
22. Hajirahimi, Z.; Khashei, M. Hybridization of hybrid structures for time series forecasting: A review. *Artif. Intell. Rev.* **2022**. <https://doi.org/10.1007/s10462-022-10199-0>.
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>.
24. Malek, N.H.A.; Wan Yaacob, W.F.; Md Nasir, S.A.; Shaadan, N. Prediction of Water Quality Classification of the Kelantan River Basin, Malaysia, Using Machine Learning Techniques. *Water* **2022**, *14*, 1067. <https://doi.org/10.3390/w14071067>.
25. Sundararajan, K.; Garg, L.; Srinivasan, K.; Kashif Bashir, A.; Kaliappan, J.; Pattukandan Ganapathy, G.; Kumaran Selvaraj, S.; Meena, T. A Contemporary Review on Drought Modeling Using Machine Learning Approaches. *Comput. Modeling Eng. Sci.* **2021**, *128*, 447–487. <https://doi.org/10.32604/cmescs.2021.015528>.
26. He, B.; Oki, T.; Sun, F.; Komori, D.; Kanae, S.; Wang, Y.; Kim, H.; Yamazaki, D. Estimating monthly total nitrogen concentration in streams by using artificial neural network. *J. Environ. Manag.* **2011**, *92*, 172–177. <https://doi.org/10.1016/j.jenvman.2010.09.014>.
27. Sami, B.H.Z.; Jee khai, W.; Sami, B.F.Z.; Ming Fai, C.; Essam, Y.; Ahmed, A.N.; El-Shafie, A. Investigating the reliability of machine learning algorithms as a sustainable tool for total suspended solid prediction. *Ain Shams Eng. J.* **2021**, *12*, 1607–1622. <https://doi.org/10.1016/j.asej.2021.01.007>.
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>.
29. Barzegar, R.; Adamowski, J.; Moghaddam, A.A. Application of wavelet-artificial intelligence hybrid models for water quality prediction: A case study in Aji-Chay River, Iran. *Stoch. Environ. Res. Risk Assess.* **2016**, *30*, 1797–1819. <https://doi.org/10.1007/s00477-016-1213-y>.
30. Shah, M.I.; Abunama, T.; Javed, M.F.; Bux, F.; Aldrees, A.; Tariq, M.A.U.R.; Mosavi, A. Modeling Surface Water Quality Using the Adaptive Neuro-Fuzzy Inference System Aided by Input Optimization. *Sustainability* **2021**, *13*, 4576. <https://doi.org/10.3390/su13084576>.
31. Li, W.; Fang, H.; Qin, G.; Tan, X.; Huang, Z.; Zeng, F.; Du, H.; Li, S. Concentration estimation of dissolved oxygen in Pearl River Basin using input variable selection and machine learning techniques. *Sci. Total Environ.* **2020**, *731*, 139099. <https://doi.org/10.1016/j.scitotenv.2020.139099>.
32. Li, X.; Cheng, Z.; Yu, Q.; Bai, Y.; Li, C. Water-Quality Prediction Using Multimodal Support Vector Regression: Case Study of Jialing River, China. *Am. Soc. Civ. Eng.* **2017**, *143*, 04017070. [https://doi.org/10.1061/\(ASCE\)](https://doi.org/10.1061/(ASCE)).
33. Nouraki, A.; Alavi, M.; Golabi, M.; Albaji, M. Prediction of water quality parameters using machine learning models: A case study of the Karun River, Iran. *Environ. Sci Pollut. Res. Int.* **2021**, *28*, 57060–57072. <https://doi.org/10.1007/s11356-021-14560-8>.
34. Ali Khan, M.; Izhar Shah, M.; Faisal Javed, M.; Ijaz Khan, M.; Rasheed, S.; El-Shorbagy, M.A.; Roshdy El-Zahar, E.; Malik, M.Y. Application of random forest for modelling of surface water salinity. *Ain. Shams Eng. J.* **2022**, *13*, 101635. <https://doi.org/10.1016/j.asej.2021.11.004>.
35. Alqahtani, A.; Shah, M.I.; Aldrees, A.; Javed, M.F. Comparative Assessment of Individual and Ensemble Machine Learning Models for Efficient Analysis of River Water Quality. *Sustainability* **2022**, *14*, 1183. <https://doi.org/10.3390/su14031183>.
36. Sattari, M.T.; Joudi, A.R.; Kusiak, A. Estimation of Water Quality Parameters With Data-Driven Model. *J. Am. Water Work. Assoc.* **2016**, *108*, E232–E239. <https://doi.org/10.5942/jawwa.2016.108.0012>.
37. Babbar, R.; Babbar, S. Predicting river water quality index using data mining techniques. *Environ. Earth Sci.* **2017**, *76*, 504. <https://doi.org/10.1007/s12665-017-6845-9>.
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>.
39. Jeihooni, M.; Toomanian, A.; Mansourian, A. Decision Tree-Based Data Mining and Rule Induction for Identifying High Quality Groundwater Zones to Water Supply Management: A Novel Hybrid Use of Data Mining and GIS. *Water Resour. Manag.* **2019**, *34*, 139–154. <https://doi.org/10.1007/s11269-019-02447-w>.

40. Naghibi, S.A.; Hashemi, H.; Berndtsson, R.; Lee, S. Application of extreme gradient boosting and parallel random forest algorithms for assessing groundwater spring potential using DEM-derived factors. *J. Hydrol.* **2020**, *589*, 125197. <https://doi.org/10.1016/j.jhydrol.2020.125197>.
41. Anshuka, A.; van Ogtrop, F.F.; Willem Vervoort, R. Drought forecasting through statistical models using standardised precipitation index: A systematic review and meta-regression analysis. *Nat. Hazards* **2019**, *97*, 955–977. <https://doi.org/10.1007/s11069-019-03665-6>.
42. Fung, K.F.; Huang, Y.F.; Koo, C.H.; Soh, Y.W. Drought forecasting: A review of modelling approaches 2007–2017. *J. Water Clim. Change* **2020**, *11*, 771–799. <https://doi.org/10.2166/wcc.2019.236>.
43. Zhang, Z.; Zhang, Q.; Singh, V.P. Univariate streamflow forecasting using commonly used data-driven models: Literature review and case study. *Hydrol. Sci. J.* **2018**, *63*, 1091–1111. <https://doi.org/10.1080/02626667.2018.1469756>.
44. Mohammadi, B. A review on the applications of machine learning for runoff modeling. *Sustain. Water Resour. Manag.* **2021**, *7*, 98. <https://doi.org/10.1007/s40899-021-00584-y>.
45. Zhang, Y.; Dai, X.; Wan, R.; Yang, G.; Li, B. Comparison of random forests and other statistical methods for the prediction of lake water level: A case study of the Poyang Lake in China. *Hydrol. Res.* **2016**, *47*, 69–83. <https://doi.org/10.2166/nh.2016.264>.
46. Zhang, Y.; Sun, H.; Guo, Y. Wind Power Prediction Based on PSO-SVR and Grey Combination Model. *IEEE Access* **2019**, *7*, 136254–136267. <https://doi.org/10.1109/access.2019.2942012>.
47. Ibrahim, K.S.M.H.; Huang, Y.F.; Ahmed, A.N.; Koo, C.H.; El-Shafie, A. A review of the hybrid artificial intelligence and optimization modelling of hydrological streamflow forecasting. *Alex. Eng. J.* **2022**, *61*, 279–303. <https://doi.org/10.1016/j.aej.2021.04.100>.
48. Dikshit, A.; Pradhan, B.; Alamri, A.M. Temporal Hydrological Drought Index Forecasting for New South Wales, Australia Using Machine Learning Approaches. *Atmosphere* **2020**, *11*, 585. <https://doi.org/10.3390/atmos11060585>.
49. Zubaidi, S.L.; Dooley, J.; Alkhaddar, R.M.; Abdellatif, M.; Al-Bugharbee, H.; Ortega-Martorell, S. A Novel approach for predicting monthly water demand by combining singular spectrum analysis with neural networks. *J. Hydrol.* **2018**, *561*, 136–145. <https://doi.org/10.1016/j.jhydrol.2018.03.047>.
50. Zubaidi, S.L.; Al-Bugharbee, H.; Ortega-Martorell, S.; Gharghan, S.K.; Olier, I.; Hashim, K.S.; Al-Bdairi, N.S.S.; Kot, P. A Novel Methodology for Prediction Urban Water Demand by Wavelet Denoising and Adaptive Neuro-Fuzzy Inference System Approach. *Water* **2020**, *12*, 1628. <https://doi.org/10.3390/w12061628>.
51. Zubaidi, S.L.; Hashim, K.; Ethaib, S.; Al-Bdairi, N.S.S.; Al-Bugharbee, H.; Gharghan, S.K. A novel methodology to predict monthly municipal water demand based on weather variables scenario. *J. King Saud. Univ. Eng. Sci.* **2022**, *34*, 163–169. <https://doi.org/10.1016/j.jksues.2020.09.011>.
52. Zubaidi, S.L.; Ortega-Martorell, S.; Kot, P.; Alkhaddar, R.M.; Abdellatif, M.; Gharghan, S.K.; Ahmed, M.S.; Hashim, K. A Method for Predicting Long-Term Municipal Water Demands Under Climate Change. *Water Resour. Manag.* **2020**, *34*, 1265–1279. <https://doi.org/10.1007/s11269-020-02500-z>.
53. Tabachnick, B.G.; Fidell, L.S. *Using Multivariate Statistics*, 6th ed.; Pearson Education, Inc: Pearson: Boston, MA, USA, 2013.
54. Najah, A.; Elshafie, A.; Karim, O.A.; Jaffar, O. Prediction of Johor River Water Quality Parameters Using Artificial Neural Networks. *Eur. J. Sci. Res.* **2009**, *28*, 422–435.
55. Apaydin, H.; Taghi Sattari, M.; Falsafian, K.; Prasad, R. Artificial intelligence modelling integrated with Singular Spectral analysis and Seasonal-Trend decomposition using Loess approaches for streamflow predictions. *J. Hydrol.* **2021**, *600*, 126506. <https://doi.org/10.1016/j.jhydrol.2021.126506>.
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>.
61. Zubaidi, S.L.; Gharghan, S.K.; Dooley, J.; Alkhaddar, R.M.; Abdellatif, M. Short-Term Urban Water Demand Prediction Considering Weather Factors. *Water Resour. Manag.* **2018**, *32*, 4527–4542. <https://doi.org/10.1007/s11269-018-2061-y>.
62. Adnan, R.M.; Mostafa, R.R.; Kisi, O.; Yaseen, Z.M.; Shahid, S.; Zounemat-Kermani, M. Improving streamflow prediction using a new hybrid ELM model combined with hybrid particle swarm optimization and grey wolf optimization. *Knowl. Based Syst.* **2021**, *230*, 107379. <https://doi.org/10.1016/j.knosys.2021.107379>.
63. Jawad, H.M.; Jawad, A.M.; Nordin, R.; Gharghan, S.K.; Abdullah, N.F.; Ismail, M.; Abu-AlShaeer, M.J. Accurate Empirical Path-Loss Model Based on Particle Swarm Optimization for Wireless Sensor Networks in Smart Agriculture. *IEEE Sens. J.* **2020**, *20*, 552–561. <https://doi.org/10.1109/jsen.2019.2940186>.

64. Panyadee, P.; Champrasert, P.; Aryupong, C. Water Level Prediction using Artificial Neural Network with Particle Swarm Optimization Model. In Proceedings of the 2017 Fifth International Conference on Information and Communication Technology (ICoICT), Melaka, Malaysia, 17–19 May 2017. <https://doi.org/10.1109/ICoICT.2017.8074670>.
65. Peng, T.; Zhou, J.; Zhang, C.; Fu, W. Streamflow Forecasting Using Empirical Wavelet Transform and Artificial Neural Networks. *Water* **2017**, *9*, 406. <https://doi.org/10.3390/w9060406>.
66. Nabipour, N.; Dehghani, M.; Mosavi, A.; Shamshirband, S. Short-Term Hydrological Drought Forecasting Based on Different Nature-Inspired Optimization Algorithms Hybridized With Artificial Neural Networks. *IEEE Access* **2020**, *8*, 15210–15222. <https://doi.org/10.1109/access.2020.2964584>.
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>.
70. Yudina, E.; Petrovskaya, A.; Shadrin, D.; Tregubova, P.; Chernova, E.; Pukalchik, M.; Oseledets, I. Optimization of Water Quality Monitoring Networks Using Metaheuristic Approaches: Moscow Region Use Case. *Water* **2021**, *13*, 888. <https://doi.org/10.3390/w13070888>.
71. Tripura, J.; Roy, P.; Barbhuiya, A.K. Simultaneous streamflow forecasting based on hybridized neuro-fuzzy method for a river system. *Neural Comput. Appl.* **2020**, *33*, 3221–3233. <https://doi.org/10.1007/s00521-020-05194-x>.
72. Hassan, M.; Hassan, I. Improving Artificial Neural Network Based Streamflow Forecasting Models through Data Preprocessing. *KSCE J. Civ. Eng.* **2021**, *25*, 3583–3595. <https://doi.org/10.1007/s12205-021-1859-y>.
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>.
76. Yang, X.-S. Firefly algorithm, stochastic test functions and design optimisation. *Int. J. Bio-Inspired Comput.* **2010**, *2*, 78–84. <https://doi.org/10.1504/IJBIC.2010.032124>.
77. Li, J.; Abdulmohsin, H.A.; Hasan, S.S.; Kaiming, L.; Al-Khateeb, B.; Ghareb, M.I.; Mohammed, M.N. Hybrid soft computing approach for determining water quality indicator: Euphrates River. *Neural Comput. Appl.* **2017**, *31*, 827–837. <https://doi.org/10.1007/s00521-017-3112-7>.
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>.
79. Yang, X.-S.; Deb, S. Cuckoo search: Recent advances and applications. *Neural Comput. Appl.* **2013**, *24*, 169–174. <https://doi.org/10.1007/s00521-013-1367-1>.
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>.
82. Karaboga, D. *An Idea Based on Honey Bee Swarm for Numerical Optimization; Technical Report tr06; Erciyes University, Engineering Faculty, Computer Engineering Department: Kayseri, Turkey*; 2005.
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>.
84. Hashemi, S.H.; Karimi, A.; Tavana, M. An integrated green supplier selection approach with analytic network process and improved Grey relational analysis. *Int. J. Prod. Econ.* **2015**, *159*, 178–191. <https://doi.org/10.1016/j.ijpe.2014.09.027>.
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>.
88. He, Q.; Wang, J.; Lu, H. A hybrid system for short-term wind speed forecasting. *Appl. Energy* **2018**, *226*, 756–771. <https://doi.org/10.1016/j.apenergy.2018.06.053>.

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>. (accessed on 1st May 2022)
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.; Rajaei, 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>.