Water Science & Technology

Water Science & Technology Vol 88 No 6, 1447 doi: 10.2166/wst.2023.276

Artificial neural networks for performance prediction of full-scale wastewater treatment plants: a systematic review

Marina Salim Dantas [a,*], Cristiano Christofaro [b] and Sílvia Corrêa Oliveira [a]

[a] Department of Sanitary and Environmental Engineering, Federal University of Minas Gerais, Av. Presidente Antônio Carlos, 6627, Belo Horizonte, MG CEP 31270-901, Brazil

[b] Department of Forestry Engineering, Federal University of Jequitinhonha and Mucuri Valleys, Road MG 367, 5000, Diamantina, MG CEP 39100-000, Brazil

*Corresponding author. E-mail: marina-dantas@hotmail.com

MSD, 0000-0002-7084-7798; CC, 0000-0002-9957-202X; SCO, 0000-0003-4286-3667

ABSTRACT

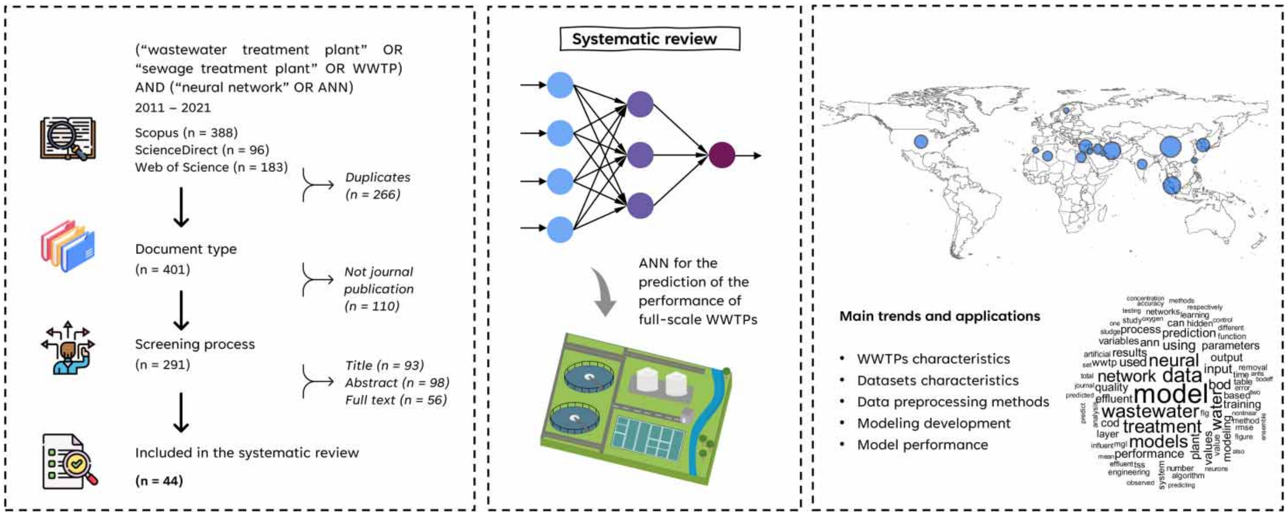
Wastewater treatment plants (WWTPs) are complex systems that must maintain high levels of performance to achieve adequate effluent quality to protect the environment and public health. Artificial intelligence and machine learning methods have gained attention in recent years for modeling complex problems, such as wastewater treatment. Although artificial neural networks (ANNs) have been identified as the most common of these methods, no study has investigated the development and configuration of these models. We conducted a systematic literature review on the use of ANNs to predict the effluent quality and removal efficiencies of full-scale WWTPs. Three databases were searched, and 44 records of the 667 identified were selected based on the eligibility criteria. The data extracted from the papers showed that the majority of studies used the feedforward neural network model with a backpropagation training algorithm to predict the effluent quality of plants, particularly in terms of organic matter indicators. The findings of this research may help in the search for an optimum design modeling process for future studies of similar prediction problems.

Key words: ANN, artificial intelligence, data science, literature review, machine learning, WWTP

HIGHLIGHTS

- Machine learning approaches are effective for modeling wastewater treatment plants (WWTPs).
- Artificial neural networks (ANNs) are the most employed in the wastewater treatment sector.
- The various ANN structures used in the sector have not been adequately studied.
- The systematic review focused on the use of ANN for performance prediction of WWTPs.
- The findings are beneficial for future studies with similar prediction problems.

GRAPHICAL ABSTRACT



ABBREVIATIONS

ANFIS	Adaptive neuro-fuzzy inference system
ANFIS-GA	Adaptive neuro-fuzzy inference system coupled with genetic algorithm
NH_4-N	Ammonia nitrogen
BOD	Biochemical oxygen demand
CBOD	Carbonaceous biochemical oxygen demand
COD	Chemical oxygen demand
R^2	Coefficient of determination
R	Correlation coefficient
DCB	Deep cascade-forward backpropagation networks
DFFNN	Deep feedforward neural network
DSAE-NN-GA	Deep learning which combines stacked autoencoders with neural network and genetic algorithm
EC	Electrical conductivity
ELM	Extreme learning machine
FFNN	Feedforward neural network
Q	Flow rate
GA	Genetic algorithm
GRNN	Generalized regression neural networks
HELM	Hierarchical extreme learning machine
LSTM	Long short-term memory
LSTM-AM	Long short-term memory based on attention mechanism
MSE	Mean square error
MLP	Multilayer perceptron network
MLP-GA	Multilayer perceptron network coupled with genetic algorithm
NO_3-N	Nitrate nitrogen
NO_2-N	Nitrite nitrogen
NARX	Nonlinear autoregressive with exogenous neural network
PO_4	Phosphate/orthophosphate
RBF	Radial basis function neural network
RBF-GA	Radial basis function neural network coupled with genetic algorithm
RVFL	Random vector functional link networks
RHONN	Recurrent high-order neural network
RMSE	Root mean square error
SO-RBF	Self-organizing radial basis function neural network
SWNN	Small-world neural network
T	Temperature
TKN	Total Kjeldahl nitrogen

TN	Total nitrogen
TP	Total phosphorus
TSS	Total suspended solids
VSS	Volatile suspended solids

1. INTRODUCTION

Recent concerns regarding environmental issues have induced specialists to focus their attention on the efficient operation and control of wastewater treatment plants (WWTPs) (Mjalli *et al.* 2007; Pham *et al.* 2020). WWTPs are highly complex and dynamic systems that require consistent high performance despite hourly, daily, and seasonal fluctuations (Corominas *et al.* 2018).

The treatment of wastewater is affected by several chemical, physical, and microbiological factors. The complexity of wastewater treatment technology results in uncertainty and variation in the treatment system, leading to fluctuations in effluent quality and environmental risks to the receiving water (Zhao *et al.* 2020; Zhang *et al.* 2023). Hence, proper operation and control are essential for safeguarding public health and protecting the environment (Nourani *et al.* 2018).

Safe operation and control of WWTPs can be achieved through the development of a robust and appropriate mathematical model for predicting plant performance based on past observations of key quality parameters (Hamed *et al.* 2004; Singh *et al.* 2010; Nasr *et al.* 2012). Modeling is widely used to assess the performance of WWTPs (Hamed *et al.* 2004; Mjalli *et al.* 2007; Singh *et al.* 2010); however, the complexity and dynamics of treatment systems make it difficult to perform predictions and simulations using traditional linear methods (Nourani *et al.* 2018).

Artificial intelligence (AI) has become a powerful tool for minimizing the complexities in wastewater treatment (Zhao *et al.* 2020; Malviya & Jaspal 2021; Zhang *et al.* 2023). Zhao *et al.* (2020) conducted a bibliometric analysis of the trends in AI technology as applied to wastewater treatment. Those authors found that the number of published articles utilizing AI in wastewater treatment research was 19 times greater in 2019 than that in 1995. Most AI techniques have been modeled using experimental data to simulate, predict, confirm, and optimize contaminant removal in wastewater treatment processes (Zhao *et al.* 2020).

Machine learning is a central subfield of AI. Machine learning algorithms are increasingly used and play a fundamental role in the operation of WWTPs (de Canete *et al.* 2021). Machine learning approaches have become powerful tools for dealing with the complexities of uncertain and dynamic problems. Therefore, these techniques are becoming common for modeling complex environmental problems, such as that of wastewater treatment and optimization of wastewater (Guo *et al.* 2015; Ye *et al.* 2020; Zhao *et al.* 2020). These approaches maximize the knowledge obtained from data and operational experience and help strengthen the management and control of WWTPs, thereby improving the performance of these facilities (Zhao *et al.* 2020).

Machine learning methods can be supervised or unsupervised. Supervised methods are used to build predictive models that characterize the link between explanatory and response variables. These models predict the response variable of interest (output) using the explanatory variables (inputs) of the dataset (Lantz 2013; Corominas *et al.* 2018; Newhart *et al.* 2019; Newhart *et al.* 2022). Supervised machine learning includes models such as naïve Bayes, regression trees, artificial neural networks (ANNs), and support vector machines (Lantz 2013). Unsupervised methods are used to build descriptive models. They are applied when the goal is to identify patterns in the data without any advanced knowledge of the possible relationships involved (Newhart *et al.* 2019).

Previous literature reviews have identified ANNs as the most employed in the wastewater treatment sector. Hadjimichael *et al.* (2016) conducted a literature review on the application of AI methods (mainly machine learning) to the urban water sector. Those authors found 1,394 papers on wastewater published between 1935 and 2016, and ANNs were found to be the most common method used in various sectors of water-related research, including that of wastewater treatment (Hadjimichael *et al.* 2016). ANNs have emerged as an attractive option for predicting and classifying water systems as well as for modeling and optimizing performance (Hadjimichael *et al.* 2016).

Corominas *et al.* (2018) performed a literature review of computer-based techniques for data analysis to improve the operation of WWTPs. Those authors described various methods that enable the transformation of data into pertinent information. According to Corominas *et al.* (2018), the European Union is the leading region in this field with the largest number of studies (61%), followed by Asia-Oceania (34%) and North America (12%). A minority of studies (less than 4%) have been conducted by South American or African research groups. Among the 340 selected papers (published up to 2015), ANN was the most commonly used technique, particularly for predicting process performance, soft sensing, and control (Corominas *et al.* 2018).

Zhao *et al.* (2020) conducted a bibliometric analysis covering 1995–2019 of trends in applying AI technology to wastewater treatment. According to those authors, research has mainly focused on AI technology in relation to pollutant removal. The majority of studies utilized ANN models to simulate and predict the performance of biological WWTP, and there has been an increase in the number of publications using this technique in recent years (Zhao *et al.* 2020).

Soft measurement estimates variables that are difficult to measure by correlating them with available variables that are more readily measured (Osman & Li 2020). Ching *et al.* (2021) conducted a literature review covering 102 studies on the development of soft sensors for wastewater treatment. Those authors showed that neural networks were the most common modeling approach. These methods have remained the dominant methodologies for soft sensor development since the early 2000s, and it appears that ANNs will continue to predominate in the coming years (Ching *et al.* 2021).

Bahramian *et al.* (2023) conducted a comprehensive literature review on the state-of-the-art in the application of data-driven models in WWTPs. They searched publications from 2000 to 2021 and selected 281 studies for qualitative assessment. The ANNs were identified as the most popular model among the studies and were commonly used as a prediction model focusing on the removal of pollutants (Bahramian *et al.* 2023).

Zhang *et al.* (2023) provided a summary of the status and trends in AI research as applied to wastewater treatment, based on published papers and patents from 2000 to 2022. According to the authors, ANN is the most common and widely used model for AI in wastewater treatment (Zhang *et al.* 2023).

The parameters of wastewater treatment monitoring data tend to share nonlinear and complex chemical relationships (Ching *et al.* 2021). The nonlinear nature of an ANN can accurately predict pollutant removal in WWTPs (Ye *et al.* 2020). The wide usage of ANNs in water-related research relates to their ability to learn (through the training process) complex non-linear and multi-input/output relationships between process parameters using historical data (Madić & Radovanović 2011). ANNs can also be applied when there is insufficient knowledge of the process to construct a mechanistic model of the wastewater treatment system, which relies on fundamental material and energy balances and empirical correlations that are often inaccurate (Mjalli *et al.* 2007). Many simplifications and assumptions are required to ensure that mechanistic models are tractable and computable, and accordingly, they have many limitations (Wang *et al.* 2021).

ANN models consist of predefined mathematical functions that effectively capture the nonlinear relationships between variables in complex systems (Civelekoglu *et al.* 2009). ANNs require historical data during training, after which they should have the ability to extrapolate correlations to new data (Palani *et al.* 2008). The ANN learns from the training data and captures the relationships between data points, which can be used for simulation, prediction, and optimization (Zhao *et al.* 2020).

The concept of an ANN was based on the biological human brain and its learning processes. ANNs are numerical structures comprising nodes (neurons) and connections (weights) (Mjalli *et al.* 2007; Nezhad *et al.* 2016). The ANN architecture is the overall structure and manner in which information flows from one layer to another (Chen *et al.* 2020). The architecture consists mainly of the number of neurons and the manner in which they are interconnected (Mjalli *et al.* 2007). An ANN includes a variety of hyperparameters that must be tuned during model development, including the number of hidden layers, number of neurons in each hidden layer, and activation functions that are applied (Ching *et al.* 2021).

The main task in designing a robust neural network is to determine the appropriate model architecture to minimize the overall model error (Madić & Radovanović 2011; Nezhad *et al.* 2016). Selecting a network structure (e.g., a feedforward neural network (FFNN) with one hidden layer and five neurons in the hidden layer that are connected by a sigmoid activation function, or a deep neural network with multiple hidden layers and multiple parameters) is a crucial step in the design of ANNs. The structure must be optimized for reducing computer processing, achieving adequate performance, and avoiding overfitting (Mjalli *et al.* 2007).

There is a limited theoretical and practical background to assist in the systematic selection of ANN hyperparameters through model development and training processes (Madić & Radovanović 2011). Therefore, most studies choose the appropriate ANN model structure using a trial-and-error approach (Mjalli *et al.* 2007; Palani *et al.* 2008; Madić & Radovanović 2011; Chen *et al.* 2020), whereby several networks are trained and compared (Mjalli *et al.* 2007; Madić & Radovanović 2011), which is challenging and time-consuming (Lee *et al.* 2011). Choosing the ANN architecture and selecting the training algorithm (which is used to minimize the error between the observed and predicted output) and related parameters is primarily related to the experience of the designer (Madić & Radovanović 2011).

Although previous literature reviews have identified the ANN as the data-driven technique and machine learning model most applied in the wastewater sector (Hadjimichael *et al.* 2016; Corominas *et al.* 2018; Zhao *et al.* 2020; Malviya & Jaspal 2021; Bahramian *et al.* 2023), no studies have identified the model structures adopted in this research. No specific literature review has been found on the use of ANN in the wastewater treatment sector. Therefore, the current investigation may improve the configuration of models based on studies in this field. Understanding the hyperparameter tuning process from datasets of WWTPs might improve the efficiency of determining the optimum setting and the performance of future models.

2. METHODS

2.1. Review objective and research question

With the increased use of neural network methods for predictions, it is important to study their role in predicting WWTP performance. The various ANN structures and hyperparameters used in the wastewater treatment sector have not been adequately studied. Therefore, a systematic review was conducted to develop an understanding of WWTP performance predictions using an ANN.

A systematic review is a literature review based on clearly formulated questions. It identifies relevant studies and summarizes evidence using an explicit methodology (Khan *et al.* 2003). A systematic review differs from a traditional general review, as it adopts a replicable, scientific, and transparent process (Qazi *et al.* 2015). The current study followed the guidelines and protocols for systematic reviews (Khan *et al.* 2003; Pullin & Stewart 2006; Page *et al.* 2021).

The first step in a systematic review is to formulate a specific question. The following research question was the basis of this review: 'What are the main architectures and hyperparameters of ANN models used to predict the performance of different types of full-scale WWTPs?'

2.2. Search strategy

The next step is to identify relevant studies by formulating a formal search strategy. The systematic review design reported here was initiated in August 2021. After several refinements and improvements, the publication search began in February 2022. The ScienceDirect, Scopus, and Web of Science databases were searched, and the results restricted to peer-reviewed articles published in journals from 2011 through 2021 in English. Pilot searches were performed to refine the keywords, and the following final search strategy was used, based on document titles, abstracts, and author-specified keywords: ('wastewater treatment plant' OR 'sewage treatment plant' OR WWTP) AND ('neural network' OR ANN).

2.3. Selection criteria

The study selection criteria flow directly from the review questions and should be previously specified. The reasons for inclusion and exclusion were recorded (Khan *et al.* 2003). The eligibility criteria were designed to focus exclusively on the use of ANNs for predicting the performance of WWTPs in terms of effluent quality or removal efficiencies. The goal was to gather a comprehensive set of studies specifically focused on the application of ANNs in this context. The selection process was structured as follows:

Inclusion Criteria:

a. Studies using ANNs: Only studies that employed ANNs as the modeling tool for predicting the effluent quality or removal efficiencies of WWTPs were considered for inclusion. Other machine learning algorithms and modeling techniques were excluded to maintain a specific focus on ANNs.
b. Full-scale WWTPs: Only studies involving full-scale WWTPs were included in the review. Pilot- and bench-scale plants were excluded to ensure relevance to real operational conditions.
c. Domestic effluent treatment: The review was limited to studies that focused on WWTPs specifically designed to treat domestic effluent. Industrial plants were excluded.

Exclusion Criteria:

a. Studies using ANNs for other purposes: Studies utilizing ANNs for purposes other than predicting WWTP performance in terms of effluent quality or removal efficiencies (e.g., energy consumption control, process optimization) were excluded, as they deviated from the primary research focus.
b. Non-journal publications: Publications such as book chapters, conference papers, and lecture notes were excluded from the review as journal publications were the focus.

Language and data criteria:

a. Language: Articles published in languages other than English were excluded.
b. Data availability: During the full-text screening, an additional criterion was applied to assess whether the selected papers contained the necessary data and information to effectively answer the main research question (Espinosa *et al.* 2020). Studies lacking relevant data were excluded.

After selecting documents based on the search strategy, duplicates were removed using Mendeley software. Then, non-journal publications were excluded. Subsequently, articles were screened for exclusion criteria based on their titles. The abstracts were then evaluated for the inclusion and exclusion criteria, and the remaining articles were subsequently screened based on their full text for the eligibility criteria.

2.4. Data extraction and analysis

The next step was to extract data from the final selected papers by identifying relevant information related to the research question (Qazi *et al.* 2015). A detailed investigation was conducted, and data from papers were extracted and presented in a table with the following fields: (i) reference (author(s), year, journal, and paper title); (ii) country of the study; (iii) wastewater treatment technology and inflow rate/design flow of the facility; (iv) monitoring frequency and period; (v) number of samples; (vi) data division into training/validation/testing datasets (%); (vii) input and output variables; (viii) data preprocessing methods; (ix) neural network architectures and hyperparameters (ANN methods, training algorithms, number of hidden layers, number of neurons in each hidden layer, and activation functions); and (x) metrics of model performance.

3. RESULTS AND DISCUSSION

3.1. Search results

A total of 667 articles were identified by searching the three databases. Duplicates (266 records) and publications from books, book chapters, conference articles, and lecture notes (110 documents) were removed. In the next step, 291 records were screened based on their titles, and 93 were excluded. The main reasons for exclusion at this stage were that the studies focused on the gas and solid phases of WWTPs; other operating conditions of the systems, such as energy consumption, treatment cost, odor, or membrane fouling; or industrial effluents. Following this, 198 papers were screened based on their abstracts, and 98 were excluded. The main exclusions occurred in relation to studies not conducted on full-scale WWTPs (pilot plants, bench-scale, or benchmark simulation models); models of influent conditions (quality and quantity) or other operating conditions (such as aeration control); studies on industrial effluents; or articles that did not use ANNs. Subsequently, 100 papers were screened based on the full text, and 56 were excluded, mainly for not including the appropriate data to answer the research question or not assessing full-scale WWTPs (pilot plants, laboratory scale, or benchmark simulation models). The remaining 44 studies were included in this review (Figure 1).

There was no observed increase in the number of publications selected among the years. However, 13 papers (30%) were published in 2021 (Figure 2), which may be related to the COVID-19 pandemic. Due to the lockdown, researchers in many fields were able to commit more time to writing and submitting papers to peer-reviewed journals. In addition, researchers were hindered from conducting laboratory research, and studies were more focused on statistics and mathematical modeling using secondary data.

Figure 3 shows the word cloud generated from the 44 selected papers using the package 'wordcloud' (Fellows 2018) of the R programming language. The size of a word is proportional to its frequency in the texts. Some terms (artificial, neural, network, ANN, wastewater, treatment, plant, and WWTP) were expected to be in the word cloud, as they were used in the search strategy of the systematic review. The word 'model' was highlighted, which was not used in the search strategy but was the most frequently used term in the papers. Other words related to the modeling process also appeared such as algorithm, modeling, data, predict, predicted, predicting, prediction, training, testing, learning, method, function, hidden, layer, nonlinear, ensemble, accuracy, RMSE, error, neurons, output, input, and ANFIS. Another category included terms related to the wastewater treatment process such as engineering, effluent, influent, quality, system, water, removal, concentration, control, process, oxygen, sludge, BOD, COD, TSS, and BODeff.

The complete spreadsheet with detailed information extracted from the 44 papers is shown in Supplementary Table S1. The following sections discuss the main results presented in Supplementary Table S1.

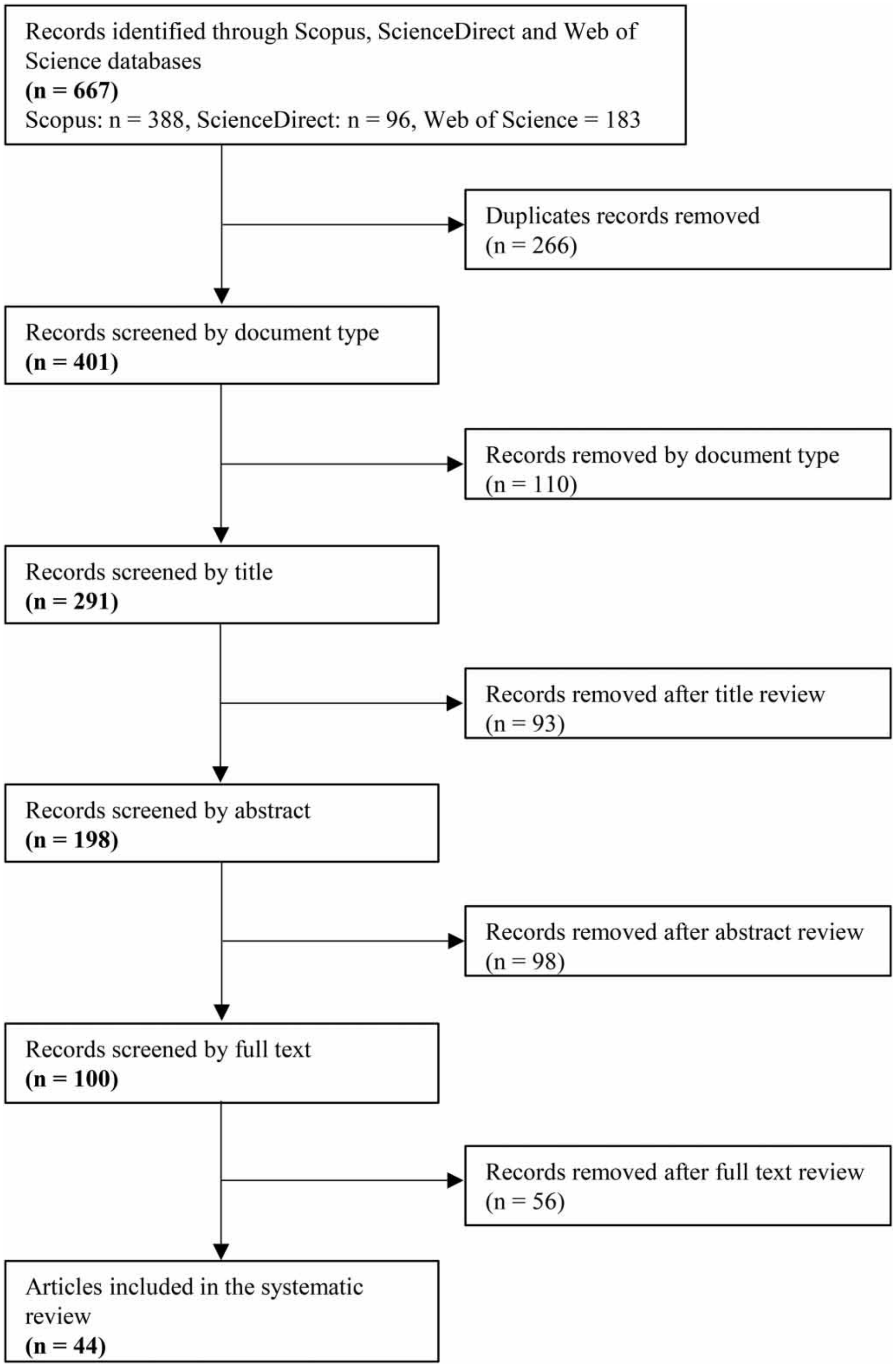


Figure 1 | Flow diagram of the systematic review on the use of ANNs to predict the performance of WWTPs.

3.2. WWTPs characteristics

One study (Ge *et al.* 2020) assessed two WWTPs, while the remaining 43 studies evaluated only one treatment facility. Three papers (7%) did not provide information on the treatment technology adopted in the WWTP under investigation. The conventional activated sludge process was the most common and was found in 18 papers (41%), followed by anaerobic/anoxic/oxic processes in five articles (11%). The remaining 41% of the studies included WWTPs that employed different activated sludge configurations (anoxic/oxic processes, activated sludge with coagulation/flocculation, extended aeration activated sludge, aerated lagoon followed by activated sludge, step-feed activated sludge processes, sequential batch reactors,

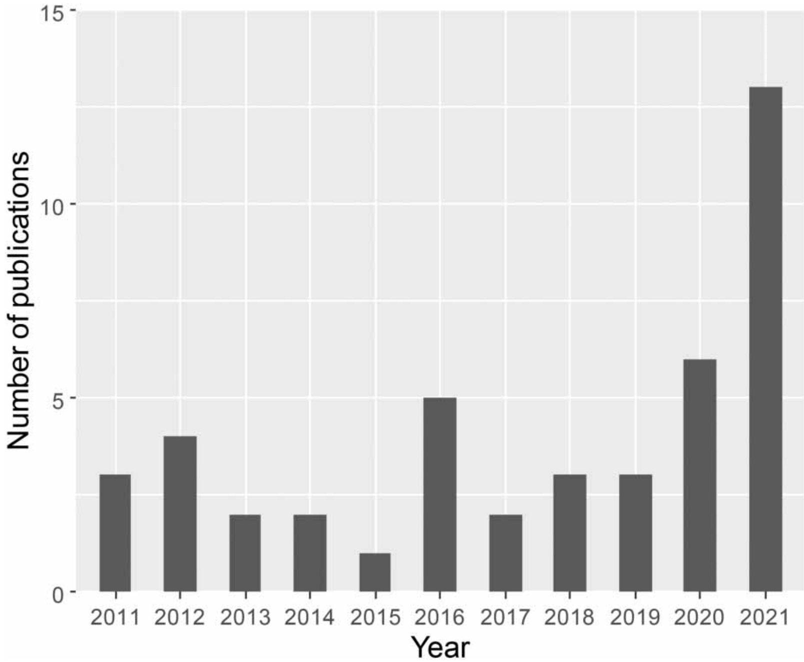


Figure 2 | Number of publications over the years ($n = 44$).

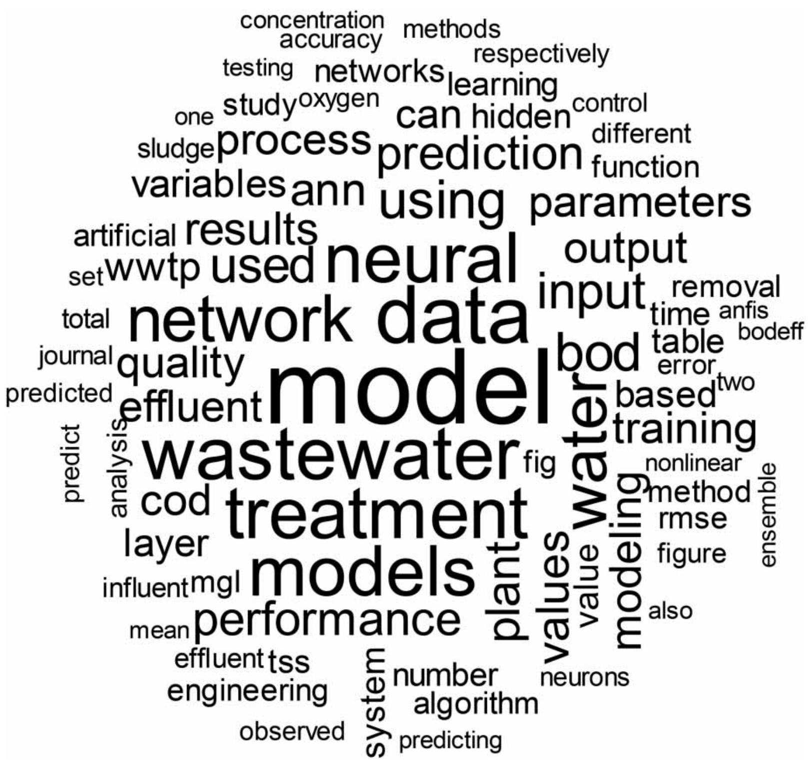


Figure 3 | Word cloud generated from the 44 selected papers of the systematic review.

intermittent cycle extended aeration-sequential batch reactors), two-stage trickling filters, constructed wetlands, and membrane bioreactors. The activated sludge processes had a higher prevalence in the selected studies because this is the most employed wastewater treatment technology globally (Sin & Al 2021).

Sixteen (36%) of the selected papers did not mention the size of the WWTP under study. The remaining 28 papers (64%) reported the inflow rate, design flow of the WWTPs, or both. The sizes of the WWTPs were variable, ranging from 52.1 to 11,574 L/s. However, most studies assessed large WWTPs. Fourteen WWTPs had inflow rates or design flows above 1,000 L/s. The inclusion of large facilities in the studies may be because large systems have better monitoring schemes with more data to train the ANN models. Larger WWTPs also have improved operational control, which encourages the development of models for predicting system performance.

Figure 4 shows the locations of the WWTPs studied. The country of the authors was considered in four papers that did not mention the WWTP site. This was an acceptable criterion, as the WWTPs were located in the same countries as the authors in papers that presented that information. The 44 selected publications for the systematic review originated in 15 countries, with the largest contribution from China (20% of the papers). The publications were concentrated in northern countries. Further research should be conducted in countries from other regions with other socioeconomic and climatic characteristics that lead to different wastewater treatment operational conditions. These distinct conditions may aid in providing important information on the use of ANNs for predicting WWTP performance.

3.3. Datasets characteristics

The WWTP data were collected at various time intervals, from continuous online sensor measurements to quarterly laboratory results (Newhart *et al.* 2019). In the WWTPs under study, 11 (25%) of the publications did not include the monitoring frequency, while three presented more than one frequency, from daily to monthly.

Four papers (9%) had samples collected at a frequent temporal resolution, such as every 10 min, every hour, or three times a day. The most common data collection period was daily (20 papers, 45%). Other studies collected samples every 2 or 3 days, or 3 days a week (three papers, 7%); weekly (four papers, 9%); monthly (five papers, 11%); and biweekly, once or twice a month, or every 2 weeks (two papers, 5%).

Six (14%) studies did not provide the period in which the data were collected. The remaining 38 studies had distinct time frames, from 3 months (Ge *et al.* 2020) to more than 15 years (Hejabi *et al.* 2021). Most studies (50%) assessed 1–2 years of a dataset.

There are no strict standards for the amount of experimental data required to train a prediction model for reliable results (Ye *et al.* 2020). Data size information was not detailed in 15 papers (34%). The remaining 29 reported the number of samples (also called data points, instances, and records), which varied from 21 (Hazali *et al.* 2017) to 105,763 (Wang *et al.* 2021) illustrating that ANN models are capable of dealing with different-sized datasets (Chen *et al.* 2020). However, most studies presented relatively small samples, and the median considering all papers that reported this information was 361.5 (Figure 5).

3.4. Data preprocessing methods

An important preprocessing method is to normalize the data, and most reviewed papers used this step. Twelve studies did not mention whether normalization was performed on the data, while six conducted this step but did not identify the specific

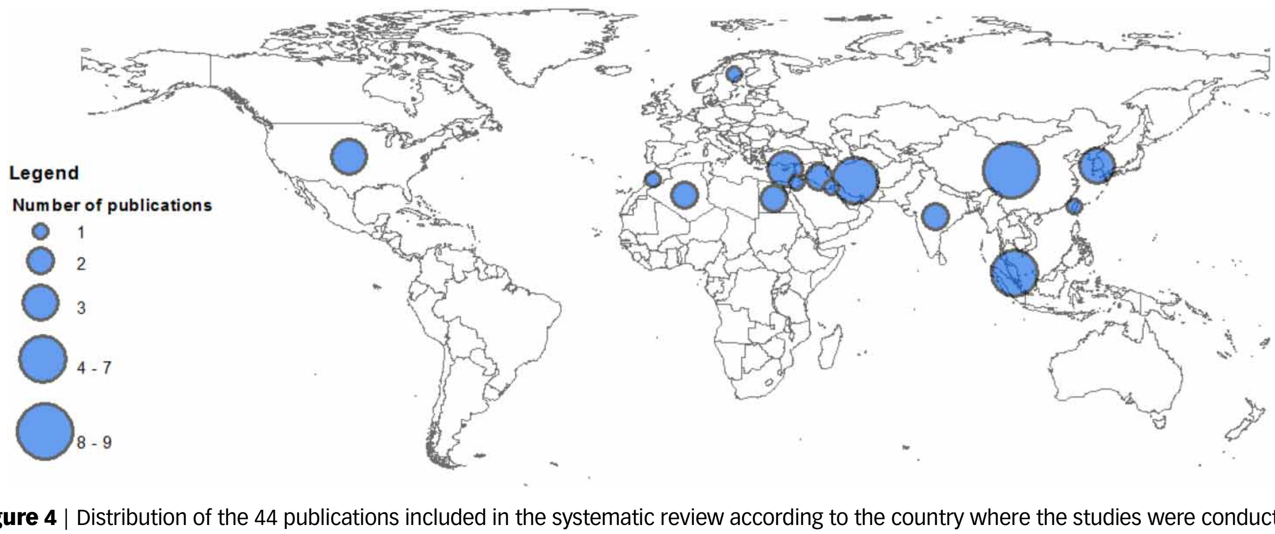


Figure 4 | Distribution of the 44 publications included in the systematic review according to the country where the studies were conducted.

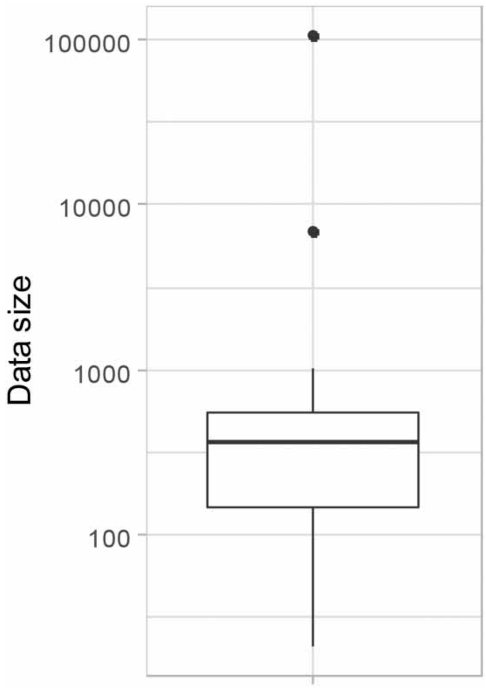


Figure 5 | Data size considering all 29 papers that included the number of samples.

method used. This information should be clearly defined because different methods affect the final result of the model differently (Chen *et al.* 2020). Among the papers that provided details about data normalization, the most used method (16 papers) was min–max normalization in the range [0, 1]. Distinct range values were also used, but less frequently, namely, [−1, 1] (two papers), [−0.9, 0.9] (one paper), [0.1, 0.9] (one paper), and [0.05, 1] (two papers). The min–max technique normalizes the data using Equation (1):

$$y = \left(\frac{x - x_{\min}}{x_{\max} - x_{\min}}\right)(\text{new\_max}_x - \text{new\_min}_x) + \text{new\_min}_x \tag{1}$$

where y is the normalized data, x is the measured data, $x_{\min}$ and $x_{\max}$ are the minimum and maximum values of the measured data, respectively, and [new_min$_x$, new_max$_x$] is the range to which the data are normalized (Ge *et al.* 2020).

Another normalization method is the Z-score normalization, in which the variables are standardized to have a zero mean and unit variance (Jami *et al.* 2011). This approach was used in four papers, and Equation (2) shows the Z-score transformation. The results were in accordance with those of Chen *et al.* (2020), who mentioned that range scaling and standardization are two common categories in data normalization:

$$y = \frac{x - \bar{x}}{\sigma} \tag{2}$$

where y is the normalized data; x is the measured data; and $\bar{x}$ and σ are the mean and the standard deviation of the variable, respectively (Jami *et al.* 2011).

Other methods of preprocessing used were the removal of outliers, abnormal data, noise, or errors in the data (Jami *et al.* 2011, 2012; Zhao *et al.* 2012; Kusiak & Wei 2013; Han *et al.* 2014; Qiao *et al.* 2016; Yaqub *et al.* 2020; Alsulaili & Refaie 2021); the estimation, interpolation, or imputation of missing points (Zhao *et al.* 2012; Han *et al.* 2014; Aldaghi & Javanmard 2021; Liu *et al.* 2021); and the use of multivariate statistical analyses, such as clustering methods and principal component analyses (Qiao *et al.* 2016; Zhao *et al.* 2016; Yasmin *et al.* 2017; Han *et al.* 2018; Sharghi *et al.* 2019; Abba *et al.* 2021b), mainly for the selection of input variables of the models.

3.5. Modeling development

3.5.1. Data dividing

Data division is an important step in modeling (Chen *et al.* 2020). Most studies (30 papers, 68%) divided the dataset into training and testing subsets. The training dataset is used to develop the model, that is, to accomplish network learning and fit the network weights. The testing dataset is used to evaluate how well the model generalizes to unseen data, that is, how accurately the network predicts targets for inputs that are not in the training set (Mjalli *et al.* 2007; Lantz 2013; Zhao *et al.* 2020). Of these 30 papers, 26 mentioned the proportion of data division. The most common allocation (used in eight studies) was 75% for training and 25% for testing.

A different approach was adopted in 12 (27%) articles that divided the dataset into training, validation, and testing subsets, and the validation dataset was used to optimize the model (Zhao *et al.* 2020) by adjusting the hyperparameters (Chen *et al.* 2020). Most of these (seven studies) divided the dataset into 70% for training, 15% for validation, and 15% for testing.

Different approaches have accomplished data division. Nine papers divided data in chronological order, in which the first data points were used for training, and the remainder for validation and testing. Another nine papers randomly divided the dataset.

For larger samples, it was expected that a greater percentage would be destined to train the model. However, there was no significant correlation between the number of samples and the percentage used for training ($p = 0.27$ and Pearson correlation coefficient $= 0.22$). This confirms that there are no uniform rules for dividing the dataset, and most researchers divided the data either by domain knowledge or arbitrarily (Chen *et al.* 2020).

3.5.2. Input and output parameters

Forty papers (91%) used effluent quality indicators as the target parameters, and four (9%) had removal efficiencies as the targets. The majority (28 papers, 64%) of the studies had more than one output parameter in single-output models (20 papers), multi-output models (seven papers), or both (one paper).

Table 1 shows that biochemical oxygen demand (BOD) and chemical oxygen demand (COD) effluent concentrations were the outputs in most papers. Other target parameters commonly used in the models were effluent concentrations of solids (total suspended solids, TSS) and effluent concentrations of nutrients (ammonia nitrogen, NH_4-N, total nitrogen, TN, and total phosphorus, TP). The three most used output variables appeared in the word cloud generated from the 44 selected papers of the systematic review (Figure 3). Among these three, the largest term in the word cloud was BOD, followed by COD and TSS, which is according to Table 1. According to Alsulaili & Refaie (2021), most studies have utilized BOD, COD, and TSS to predict the performance of WWTPs using ANN-based models.

Key variables in wastewater treatment must be evaluated to control pollution (Osman & Li 2020), and their use as targets in the models confirms that they are important for assessing the performance of a WWTP. BOD and COD reflect organic water pollution and are considered the most important parameters for effluent quality control (Nourani *et al.* 2021). BOD is difficult to measure online, and laboratory measurements are time-consuming, as they are calculated by a 5-day off-line delay (Osman & Li 2020; Rahmati *et al.* 2021), which reinforces the importance of the development of predictive models for this parameter. TSS is another important variable, as excess TSS depletes dissolved oxygen in effluent water (Verma *et al.* 2013). There has been a continuous increase in the number of studies concerning nutrient removal (Ching *et al.* 2021) due to the control of effluents to prevent eutrophication of water bodies. According to Ching *et al.* (2021), the various parameters involved in the nitrogen removal process are consistent areas of interest in soft sensor development. In comparison, there are fewer sensor studies on phosphorus removal processes. The significance of phosphorus as a wastewater parameter depends on the local abundance or shortage of this nutrient (Ching *et al.* 2021).

The quality of the treated effluent depends on the influent quality and process parameters of the WWTP (Khatri *et al.* 2020). The explanatory variables (input) of the models were highly changeable in the studies, as many affect WWTP performance. Most papers (52%) had influent wastewater quality and quantity indicators as input variables. This means that the majority of studies used influent characteristics to predict effluent wastewater quality, demonstrating the value of using ANNs to represent the complex and nonlinear relationship between raw influent and treated effluent water quality measurements (Saleh 2021). For example, Bekkari & Zeddouri (2019) used the influent variables pH, temperature (T), TSS, total Kjeldahl nitrogen (TKN), BOD, and COD as inputs. The purpose of that study was to predict the performance of an activated sludge WWTP in Algeria in terms of effluent COD. In evaluating WWTP soft sensors, Ching *et al.* (2021) also found that influent quality parameters were used in most cases as input variables for modeling effluent quality.

Table 1 | Number of publications for each output variable of the ANN models

Target variable	Number of publications
Effluent BOD	25
Effluent COD	21
Effluent TSS	19
Effluent NH_4-N	10
Effluent TN	7
Effluent TP	7
Effluent pH	4
Effluent quality index	2
Removal efficiency of NH_4-N	2
Effluent CBOD	1
Effluent biodegradable dissolved organic nitrogen	1
Effluent total coliform	1
Effluent fecal streptococci	1
Effluent TKN	1
Effluent PO_4	1
Effluent NO_2	1
Effluent NO_3	1
Effluent T	1
Effluent EC	1
Removal efficiency of fecal coliform	1
Removal efficiency of total coliform	1
Removal efficiency of arsenic	1
Removal efficiency of TN	1
Removal efficiency of TP	1
Removal efficiency of TSS	1
Removal efficiency of COD	1
Removal efficiency of BOD	1
Removal efficiency of sulfide	1

Other approaches included using treated effluent quality indicators as input variables to predict a different effluent indicator as the output, wastewater quality indicators sampled at different locations in the treatment train, and combinations of influent quality indicators and operational variables (such as returned sludge flow rate, sludge volume index, food/microorganism ratio, sludge retention time, and energy and chemical products consumption). For example, to predict the effluent concentrations of TP, BOD, COD, TSS, and NH_4-N in a WWTP (Harbin, China), Zhao *et al.* (2016) developed an ANN model using raw wastewater quality data (influent concentrations of TP, BOD, COD, TSS, NH_4-N, and influent pH) and energy consumption parameters (electricity consumption, coagulant, and flocculants) as the input variables.

Table 2 shows the most common input variables, all of which were included in more than 20% of the papers, highlighting their importance as predictors of WWTP performance in the ANN models. The majority of studies included indicators of organic matter, BOD and COD, as both input (influent concentrations, Table 2) and output (effluent concentrations, Table 1) variables. According to Ching *et al.* (2021), COD is one of the strongest estimators for BOD; hence, most studies use COD concentrations as inputs for BOD models.

Other important input parameters in the models were influent TSS concentration, pH, nutrients concentration (NH_4-N, TN, and TP), and flow (Q). The choice of these variables may be related to their ease of measurement (such as pH and

Table 2 | Number of publications with the most used input variables in the ANN models of the selected papers of the systematic review

Input variable	Number of publications
Influent COD	31
Influent TSS	28
Influent BOD	25
Influent pH	20
Influent NH_4-N	17
Influent TN	11
Influent Q	10
Influent TP	9

Q) or the ability to develop models to predict some indicators in the treated effluent using the same indicator measured in the influent as one of the explanatory variables.

3.5.3. ANN methods

There are several different classifications of ANNs (Ye *et al.* 2020), and the most used model structure is the traditional FFNN, which was adopted in 21 papers (48%). This structure consists of one input layer, one or more hidden layers, and one output layer (Figure 6). The term feedforward describes the method in which the output of the neural network is calculated layer by layer from its input throughout the network (Mjalli *et al.* 2007; Palani *et al.* 2008; Corominas *et al.* 2018). Information is transmitted from one layer to another through serial operations (Palani *et al.* 2008; Civelekoglu *et al.* 2009). According to Chen *et al.* (2020), most researchers use the FFNN for water quality prediction in WWTP systems, which may be because this method provides a good analysis of these systems.

Bahramian *et al.* (2023) and Corominas *et al.* (2018) also found that FFNNs were the most popular architecture. These networks serve as universal approximators and can effectively learn complex patterns, making them suitable for solving a wide

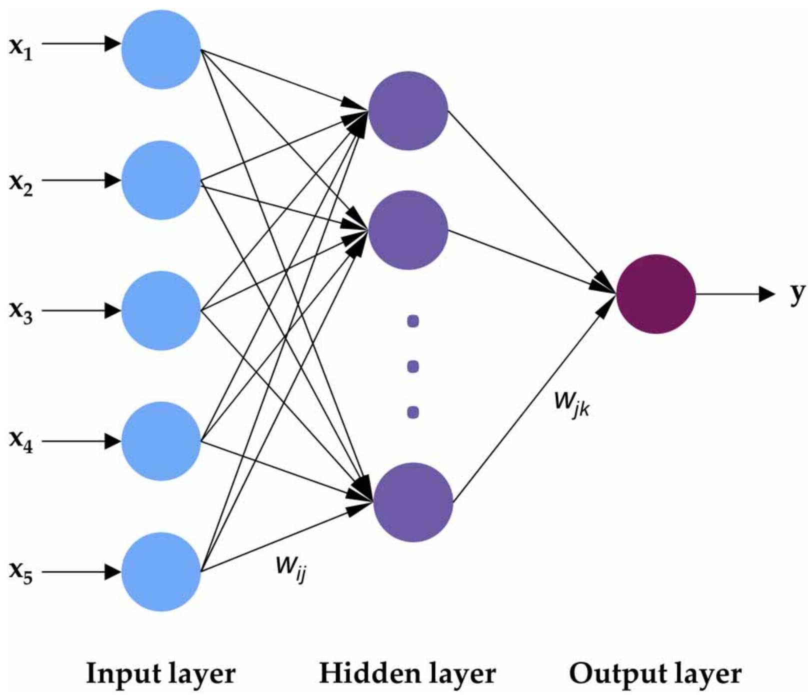


Figure 6 | Typical neural network structure with one hidden layer. Index: x_i is the input variable; w_{ij} is the weight between input i and hidden neuron j; w_{jk} is the weight of the connection of neuron j in the hidden layer to neuron k in the output layer, and y is the output variable.

range of problems. However, it is essential to be cautious of potential overfitting issues, and careful hyperparameter tuning is often required to achieve optimal performance.

The other commonly used neural network types are described next. A multilayer perceptron network (MLP) is a type of FFNN (Bagheri *et al.* 2015) and was used in seven (16%) studies. According to Newhart *et al.* (2022), a neural network that uses sigmoid functions in the hidden layer and a linear function in the output layer is more commonly referred to as an MLP.

A radial basis function neural network (RBF) is another type of FFNN (Bagheri *et al.* 2015) that uses radial basis activation functions in the hidden layer (Chen *et al.* 2020). Although Newhart *et al.* (2022) mentioned that RBF is increasingly used, it was adopted in only three (7%) papers in this systematic review.

An extreme learning machine (ELM) was used in four studies (9%). An ELM consists of a single hidden layer FFNN (Abba *et al.* 2021b) where the values of the weights between the input and hidden layers are randomly selected and the weights between the hidden and output layers are analytically characterized (Pham *et al.* 2020). As an ELM only needs to learn the output weight, it can reduce computation problems because the weights of the input and hidden layers do not require adjustment (Chen *et al.* 2020).

Deep learning refers to the use of multiple hidden layers in a network (Corominas *et al.* 2018) and is suitable for modern applications with highly complex processes (Osman & Li 2020). Deep learning methods were used in three studies (7%). One of these (Osman & Li 2020) was published in 2020, and the other two (El-Rawy *et al.* 2021; Wang *et al.* 2021) in 2021. This result indicates that deep learning is a recent technique. Corominas *et al.* (2018) did not find any advances in the identification of deep learning methods for wastewater treatment applications in papers published up to 2015.

Recurrent neural networks were used in three papers (7%), two of them utilizing long short-term memory (LSTM) methods. Recurrent neural networks are distinguished by their internal memory features, which allow observations to be considered in an ordered sequence (Newhart *et al.* 2022). Recurrent neural networks allow signals to travel in both directions using loops to learn highly complex patterns (Lantz 2013). LSTM is capable of learning sequences of events over a period of time and can capture long-term dependencies in the data. Therefore, LSTM is frequently used to deal with time-series tasks, including those of wastewater data (Liu *et al.* 2021).

An adaptive neuro-fuzzy inference system (ANFIS) is a hybrid learning method that combines neural and fuzzy methods. It integrates the learning capacities of the ANN with fuzzy logic reasoning abilities to map the input–output relationships (Ye *et al.* 2020; Onu *et al.* 2021). ANFIS uses a hybrid of backpropagation and least-squares algorithms to train the parameters and automatically generate 'If/Then' rules (Zhao *et al.* 2020). ANFIS was used in seven papers (16%).

3.5.4. Network structure

As shown in Figure 6, each layer of a neural network structure contains a certain number of neurons, also known as nodes. The numbers of input and output nodes are the number of features in the input data and the number of output variables to be modeled, respectively. The number of hidden layers and neurons in these layer(s) are configured by the user before training the model, and depend on the difficulty of the problem (Saleh 2021). An insufficient number of hidden layer neurons may reduce prediction accuracy, causing underfitting problems. However, an excessive amount of neurons may lead to overfitting, whereby the error on the training set is driven to a small value and the test data are presented to the network with a large error. This implies that the generalization ability of the neural network was affected (Gaya *et al.* 2014; Chen *et al.* 2020; Ye *et al.* 2020).

In most studies (27 papers, 61%), the authors tuned the network structure using a trial-and-error approach, whereby ranges of values for the number of hidden layers and hidden neurons were tested to search for the optimum architecture. In some cases, other configurations were also tested by trial-and-error, such as the proportions of samples allocated to the training, validation, and testing subsets, and the training algorithms and activation functions to be used. In this trial-and-error approach, several ANNs are developed and compared to select the best result. For example, Sharghi *et al.* (2019) developed FFNN models to predict effluent BOD concentrations in an activated sludge WWTP. Those authors adopted one hidden layer, and the optimal hidden layer was determined by varying the number of nodes from 1 to 10. The authors observed the best results in a model with five neurons in the hidden layer.

Five of the 27 papers that adopted the trial-and-error approach established the range of hidden neurons and/or hidden layers to be tested using equations from the literature. To some extent, the use of equations may contribute to determining the model structure as they guide researchers based on previous studies (Chen *et al.* 2020).

Another approach to determine the best network structure was adopted in five (11%) papers that used hybrid learning and combined various neural network methods (MLP, ANFIS, ELM, RBF, or deep learning) with a genetic algorithm (GA). A GA is an efficient search algorithm that can be applied to identify the combination of hyperparameters that will result in the best model performance (Ching *et al.* 2021). These hybrid models use a GA to iteratively optimize the parameters in the neural network to increase the problem-solving ability (Zhao *et al.* 2020).

Table 3 shows the final and complete network structures of the papers that presented this information. The structure column indicates the number of neurons in the input layer, each hidden layer, and the output layer. For example, Jami *et al.* (2011) developed a model using the influent BOD concentration, NH_4-N concentration, pH, and Q as explanatory variables (four input neurons), with 15 neurons in the single hidden layer of the FFNN, to predict the effluent concentrations of NH_4-N (one output neuron) in a sequential batch reactor WWTP in Malaysia.

Although some recent studies have used deep learning, most developed shallow neural networks with a single hidden layer (Table 3). Other review papers have identified that most ANN models use a single hidden layer (Corominas *et al.* 2018; Ye *et al.* 2020) as this is usually sufficient to investigate many problems (Saleh 2021). There was a wide range in the number of neurons in the hidden layer(s) of the studies, from 2 to 256.

Considering the studies that developed single-output models for both BOD and COD effluent concentrations (the two most common target variables in the studies, Table 1), the same network structure for the two variables was adopted in three papers (Jami *et al.* 2012; Nourani *et al.* 2018, 2021). In the other four papers, more complex structures were used to model BOD effluent concentrations, with greater numbers of hidden layers (Lee *et al.* 2011; Saleh 2021) or hidden neurons (Khatri *et al.* 2019; Alsulaili & Refaie 2021). Only one study (Abba *et al.* 2021b) had a larger number of hidden neurons for the COD model. This result shows that modeling BOD concentrations may be more complex than modeling COD concentrations, with more intricate network structures required to map the relationship between the input and output phases.

3.5.5. Activation functions

In a neural network, each artificial neuron in the hidden and output layers calculates the weighted sum of its inputs and produces an output value using predefined activation functions, also known as transfer functions (Mjalli *et al.* 2007; Elfanssi *et al.* 2018). Therefore, the activation function is applied to a certain layer to obtain the output of that layer, which is then used as the input for the next layer (Sharma *et al.* 2020).

Activation functions introduce nonlinearity into the neural network. The choice of the activation function is important because it affects the prediction performance of the neural network (Sharma *et al.* 2020).

From the papers in the systematic review that included this information, the most common activation functions in the hidden layer were the logistic sigmoid (nine studies) and hyperbolic tangent (12 studies) functions. This result is in accordance with Corominas *et al.* (2018), who mentioned hyperbolic tangent and sigmoid functions among the typically applied activation functions in ANN models for nonlinear classification and regression problems in wastewater treatment research. Newhart *et al.* (2022) stated that the most widely used ANN activation function in environmental engineering is the logistic sigmoid function.

The logistic sigmoid function is given by Equation (3), where x is the input of the activation function. The curve resembles an S-shape, and the returned values range from 0 to 1 (Feng & Lu 2019):

$$f(x) = \frac{1}{1 + e^{-x}} \tag{3}$$

The hyperbolic tangent function is given by Equation (4). The output values range from −1 to 1. The function is symmetric around the origin, which makes its outputs more likely to be closer to zero than those of the sigmoid function, leading to faster convergence. For this reason, it is often used in hidden layers of ANNs (Feng & Lu 2019), which may be the reason that it was the most used in the studies of this systematic review:

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} = \frac{2}{1 + e^{-2x}} - 1 \tag{4}$$

The output layer is the layer in the neural network model that directly returns a prediction. The most used activation function in the output layer of the neural network models of this systematic review (15 studies) was the linear activation function,

Table 3 | Neural network structure from 31 papers that presented this information

Reference	Output parameter(s)	Structure
Jami *et al.* (2011)	Effluent NH_4-N	4-15-1[a]
Lee *et al.* (2011)	Effluent BOD	8-19-14-1[b]
	Effluent COD	8-27-1[b]
	Effluent SS	8-3-6-1[b]
	Effluent TN	8-17-23-1[b]
Qiao *et al.* (2011)	Effluent COD, BOD, SS, and NH_4-N (multi-output model)	8-4-8-4[c]
Zhang & Hu (2012)	Effluent BOD	5-2-3-8-1[d]
Chen & Lo (2012)	Effluent Q, BOD, COD, and SS (multi-output model)	4-16-4[e]
Jami *et al.* (2012)	Effluent BOD, SS, COD (single-output models)	1-20-1[a] or 3-30-1[a]
Kusiak & Wei (2013)	Effluent CBOD	5-3-1[e]
	Effluent TSS	5-10-1[e]
Liu *et al.* (2013)	Effluent COD	9-54-6-6-1[f]
Han *et al.* (2014)	Effluent BOD	5-150-1[g] and 5-180-1[g]
Gaya *et al.* (2014)	Effluent COD, SS, NH_4-N (single-output models)	5-10-1[a]
Bagheri *et al.* (2015)	Effluent COD, TN, TSS (single-output models)	5-10-1[b]; 5-5-1[h]
Simsek (2016)	Effluent biodegradable dissolved organic nitrogen	4-10-1[e]
Zhao *et al.* (2016)	Effluent TP, BOD, COD, SS, and NH_4-N (multi-output model)	9-19-5[a], 9-19-5[a], 9-16-5[a], 9-14-5[a], and 9-15-5[a]
Nezhad *et al.* (2016)	Effluent quality index	8-7-1[a]
Hazali *et al.* (2017)	Effluent TN, TP, NH_4-N (single-output models)	6-6-1[i]
Nourani *et al.* (2018)	Effluent BOD, COD, TN (single-output models)	5-3-1[a]
Elfanssi *et al.* (2018)	Effluent TSS, BOD, COD, total coliform, and fecal streptococci (multi-output model)	5-7-8-7-5[a]
Sharghi *et al.* (2019)	Effluent BOD	3-5-1[a]
Khatri *et al.* (2019)	Effluent TSS	7-4-1[a]
	Effluent pH, COD, TKN (single-output models)	7-5-1[a]
	Effluent BOD, NH_4-N, TP (single-output models)	7-6-1[a]
Bekkari & Zeddouri (2019)	Effluent COD	6-50-1[a]
Khatri *et al.* (2020)	Removal efficiency of fecal coliform	10-6-1[a]
	Removal efficiency of total coliform	10-8-1[a]
Ge *et al.* (2020)	Removal efficiency of arsenic	4-3-1[a]
Al-Obaidi (2020)	Effluent quality index	5-3-1[a]
Osman & Li (2020)	Effluent BOD	19-13-13-13-1[j]
El-Rawy *et al.* (2021)	Removal efficiency of TSS, COD, BOD, NH_4-N, sulfide (single-output models)	5-8-1[a]; 5-10-10-10-10-1[k], 5-10-10-10-10-1[l]
Wang *et al.* (2021)	Effluent TSS	32-128-256-128-1[k]
	Effluent PO_4	32-256-128-128-1[k]
Nourani *et al.* (2021)	Effluent BOD, COD (single-output models)	5-3-1[a]
Alsulaili & Refaie (2021)	Effluent BOD	3-17-17-17-1[a]
	Effluent COD	3-13-13-13-1[a]
	Effluent TSS	3-11-11-11-11-1[a]
Aldaghi & Javanmard (2021)	Effluent Q, BOD, COD, TSS, pH, T, TP, NO_3, TN, NO_2, NH_4-N, and EC (multiple-output model)	12-25-12[e]
Saleh (2021)	Effluent COD	9-6-6-1[a]
	Effluent BOD	9-6-6-6-1[a]

(*Continued.*)

Table 3 | Continued

Reference	Output parameter(s)	Structure
	Effluent TSS	9-6-6-6-1[a]
	Effluent COD, BOD, and TSS (multiple-output model)	7-6-6-3[a]
Abba *et al.* (2021b)	Effluent BOD	6-6-1[e]
	Effluent COD, TN, TP (single-output models)	9-10-1[e]

Obs.: Neural network methods. [a]FFNN; [b]MLP-GA; [c]RHONN; [d]SWNN; [e]MLP; [f]ANFIS-GA; [g]HELM; [h]RBF-GA; [i]SO-RBF; [j]DSAE-NN-GA; [k]DFFNN; [l]DCB.

also called 'identity function' or 'no activation'. The linear activation function is given by Equation (5) and is represented by a straight line. It does not change the weighted sum of the previous layer, but only returns the value directly. The outputs can range from $-\infty$ to $+\infty$ (Feng & Lu 2019):

$$f(x) = x \tag{5}$$

3.5.6. Training algorithms

The training of a neural network is performed by adjusting the neurons weights to minimize the error between the observed data and network output (Mjalli *et al.* 2007; Nasr *et al.* 2012). The most common learning algorithm used for this purpose is backpropagation, which involves working backward layer by layer from the output to adjust the weights accordingly and reduce the average error across all layers (Mjalli *et al.* 2007; Nezhad *et al.* 2016; Newhart *et al.* 2019). The backpropagation algorithm was used in 27 papers in this systematic review (61%).

Backpropagation is the most widely used ANN training algorithm (Zhao *et al.* 2016; Chen *et al.* 2020; Ye *et al.* 2020; Newhart *et al.* 2022), and is commonly applied in the field of environmental pollution control (Ye *et al.* 2020). The majority of applications of neural networks in engineering or wastewater treatment problems use the FFNN architecture with a back-propagation training algorithm because of its accuracy and capability (Al-Ghazawi & Alawneh 2021).

The standard backpropagation algorithm uses the gradient descent optimization method to perform calculations (Chen & Lo 2012; Zhao *et al.* 2016). This method involves the network weight value moving along a negative gradient of the performance function. Hence, the weight and bias values are continually renewed to minimize the performance function (Chen & Lo 2012).

3.5.7. Software tools

Sixteen studies (36%) did not specify which software tools were used for model development. Among the papers that provided this information, the most frequently used was MATLAB, which was used in 21 publications (48%). Other tools included R (two studies, 5%), SPSS (two studies, 5%), Python (one study, 2%), MATLAB integrated with C++ (one study, 2%), and MATLAB integrated with C# (one study, 2%).

MATLAB was also found by Corominas *et al.* (2018), Ye *et al.* (2020), and Bahramian *et al.* (2023) to be the most popular software platform in the literature for modeling WWTPs with AI techniques. According to the authors, the wide usage of this software platform is due to its packages and toolboxes, which are user-friendly and convenient for users with minimal knowledge of data science (Ye *et al.* 2020; Bahramian *et al.* 2023).

3.6. Model performance

The model performance indicates the results of a comparison of the experimental data with the predicted data (Zhao *et al.* 2020). The performance of the models in the studies was calculated using various statistical metrics, including error (mainly mean square error (MSE) and root mean square error (RMSE)) and goodness-of-fit (mainly correlation coefficient (R) and coefficient of determination (R^2)). The MSE and RMSE indicators identify the errors between the experimental values and model output, with smaller results signifying higher accuracy. The metrics R and R^2 indicate the degree of correlation between the observed and predicted values, with higher R or R^2 values indicating better prediction performance (Ye *et al.* 2020).

Table 4 | Model performance in terms of goodness-of-fit indicators

Reference	Output parameter(s)	ANN methods	Model performance
Jami *et al.* (2011)	Effluent NH_4-N	FFNN	$R = 0.7980$
Zhao *et al.* (2012)	Effluent BOD	Selective ensemble ELM-GA	$R^2 = 0.7576$
	Effluent COD		$R^2 = 0.7729$
	Effluent SS		$R^2 = 0.5957$
	Effluent NH_4-N		$R^2 = 0.8273$
Chen & Lo (2012)	Effluent Q	MLP	$R = 0.9781$
	Effluent BOD		$R = 0.6963$
	Effluent COD		$R = -0.0178$
	Effluent SS		$R = 0.1031$
Jami *et al.* (2012)	Effluent BOD	FFNN	$R = 0.346948$
	Effluent COD		$R = 0.052622$
	Effluent SS		$R = 0.158717$
Liu *et al.* (2013)	Effluent COD	ANFIS-GA	$R^2 = 0.800$
	Effluent TN		$R^2 = 0.577$
	Effluent TP		$R^2 = 0.284$
Gaya *et al.* (2014)	Effluent COD	FFNN	$R = 0.647$
	Effluent SS		$R = 0.512$
	Effluent NH_4-N		$R = 0.425$
	Effluent COD	ANFIS	$R = 0.847$
	Effluent SS		$R = 0.995$
	Effluent NH_4-N		$R = 0.948$
Bagheri *et al.* (2015)	Effluent COD	MLP-GA	$R^2 = 0.98044$
	Effluent TN		$R^2 = 0.98479$
	Effluent TSS		$R^2 = 0.95484$
	Effluent COD	RBF-GA	$R^2 = 0.97232$
	Effluent TN		$R^2 = 0.98325$
	Effluent TSS		$R^2 = 0.95217$
Simsek (2016)	Effluent biodegradable dissolved organic nitrogen	ANFIS	$R^2 = 0.94$
		MLP	$R^2 = 0.78$
		RBF	$R^2 = 0.66$
		GRNN	$R^2 = 0.97$
Heddam *et al.* (2016)	Effluent BOD	GRNN	$R = 0.922$
Nezhad *et al.* (2016)	Effluent quality index	FFNN	$R = 0.96$
Hazali *et al.* (2017)	Effluent TP	SO-RBF	$R^2 = 0.8442$
	Effluent TN		$R^2 = 0.7282$
	Effluent NH_4-N		$R^2 = 0.2833$
Yasmin *et al.* (2017)	Effluent pH	FFNN	$R = 0.39698$
		ANFIS	$R = 0.70868$
Nourani *et al.* (2018)	Effluent BOD	FFNN	$R^2 = 0.6600$
	Effluent COD		$R^2 = 0.9363$
	Effluent TN		$R^2 = 0.9022$
	Effluent BOD	ANFIS	$R^2 = 0.7640$
	Effluent COD		$R^2 = 0.9260$
	Effluent TN		$R^2 = 0.9410$
Sharghi *et al.* (2019)	Effluent BOD	FFNN	$R^2 = 0.67$
Khatri *et al.* (2019)	Effluent pH	FFNN	$R = 0.816$
	Effluent BOD		$R = 0.649$
	Effluent COD		$R = 0.656$
	Effluent TSS		$R = 0.457$
	Effluent TKN		$R = 0.670$

(*Continued.*)

Table 4 | Continued

Reference	Output parameter(s)	ANN methods	Model performance
	Effluent NH_4-N		$R = 0.493$
	Effluent TP		$R = 0.748$
Bekkari & Zeddouri (2019)	Effluent COD	FFNN	$R = 0.8781$
Khatri *et al.* (2020)	Removal efficiency of fecal coliform	FFNN	$R = 0.986$
	Removal efficiency of total coliform		$R = 0.977$
Ge *et al.* (2020)	Removal efficiency of arsenic	FFNN	$R^2 = 0.851$
Al-Obaidi (2020)	Effluent quality index	FFNN	$R^2 = 0.998$
Osman & Li (2020)	Effluent BOD	DSAE-NN-GA	$R^2 = 0.987$
El-Rawy *et al.* (2021)	Removal efficiency of BOD	FFNN	$R = 0.55564$
	Removal efficiency of COD		$R = 0.90859$
	Removal efficiency of TSS		$R = 0.52105$
	Removal efficiency of NH_4-N		$R = 0.95459$
	Removal efficiency of sulfide		$R = 0.9866$
	Removal efficiency of BOD		$R = 0.76327$
	Removal efficiency of COD	DFFNN	$R = 0.66487$
	Removal efficiency of TSS		$R = 0.70718$
	Removal efficiency of NH_4-N		$R = 0.99427$
	Removal efficiency of sulfide		$R = 0.92402$
	Removal efficiency of BOD		$R = 0.77167$
	Removal efficiency of COD	DCB	$R = 0.94572$
	Removal efficiency of TSS		$R = 0.80847$
	Removal efficiency of NH_4-N		$R = 0.97696$
	Removal efficiency of sulfide		$R = 0.98585$
Al-Ghazawi & Alawneh (2021)	Effluent BOD	FFNN	$R^2 = 0.48$
	Effluent COD		$R^2 = 0.45$
	Effluent SS		$R^2 = 0.44$
	Effluent NH_4-N		$R^2 = 0.26$
Wang *et al.* (2021)	Effluent TSS	DFFNN	$R^2 = 0.920$
	Effluent PO_4		$R^2 = 0.872$
Nourani *et al.* (2021)	Effluent BOD	FFNN	$R^2 = 0.7182$
	Effluent COD		$R^2 = 0.7178$
	Effluent BOD	ANFIS	$R^2 = 0.7203$
	Effluent COD		$R^2 = 0.7148$
Elmaadawy *et al.* (2021)	Effluent BOD	RVFL	$R^2 = 0.924$
	Effluent TSS		$R^2 = 0.917$
Alsulaili & Refaie (2021)	Effluent BOD	FFNN	$R^2 = 0.752$
	Effluent COD		$R^2 = 0.6115$
	Effluent TSS		$R^2 = 0.6308$
Abba *et al.* (2021a)	Effluent TSS	NARX	$R^2 = 0.9846$
	Effluent pH		$R^2 = 0.6293$
Hejabi *et al.* (2021)	Effluent BOD	FFNN	$R^2 = 0.760$
	Effluent COD		$R^2 = 0.715$
	Effluent TSS		$R^2 = 0.632$
Liu *et al.* (2021)	Effluent COD	LSTM-AM	$R^2 = 0.869$
Aldaghi & Javanmard (2021)	Effluent Q, BOD, COD, TSS, pH, T, TP, NO_3-N, TN, NO_2-N, NH_4-N, and EC	MLP	$R = 0.5804$
Saleh (2021)	Effluent BOD	FFNN	$R = 0.99782$
	Effluent COD		$R = 0.77301$
	Effluent TSS		$R = 0.8317$

(*Continued.*)

Table 4 | Continued

Reference	Output parameter(s)	ANN methods	Model performance
Abba *et al.* (2021b)	Effluent BOD	ELM	$R^2 = 0.6341$
	Effluent COD		$R^2 = 0.9742$
	Effluent TN		$R^2 = 0.9656$
	Effluent TP		$R^2 = 0.8807$
	Effluent BOD	MLP	$R^2 = 0.5776$
	Effluent COD		$R^2 = 0.9555$
	Effluent TN		$R^2 = 0.86662$
	Effluent TP		$R^2 = 0.72544$
Rahmati *et al.* (2021)	Effluent BOD	FFNN	$R = 0.897$
		ANFIS	$R = 0.930$

Some papers presented the metrics separately for the training and testing subsets, each target variable being modeled, and each type of model used. For this reason, there are many results for model performance, which can be found in Supplementary Table S1.

The following discusses the results of the performance of the models. As the metrics of errors, RMSE and MSE, depend on the unit of the variable or if they are presented as normalized data, the R and R^2 results are presented in Table 4. These data highlight the large variability in the results, with R ranging from -0.018 to 0.998 and R^2 from 0.260 to 0.998.

It is unfeasible to determine the reasons for the differences between the studies because the context of each application is different, with distinct methods, target parameters, and datasets (Ching *et al.* 2021). Even when a single study developed different types of neural network methods for the same target variable, various situations were observed. For example, Yasmin *et al.* (2017) observed a better prediction accuracy of the ANFIS model compared with the FFNN method when modeling the pH effluent. In contrast, Nourani *et al.* (2021) achieved similar results with ANFIS and FFNN when modeling the same output parameters (effluent BOD and COD concentrations). This highlights that the advantage of one method over another may be due to the context of the application, the differences in the dataset used, and the configuration settings in the model of each study.

3.7. Limitations of the review and future perspectives

The ever-evolving nature of machine learning techniques leads to numerous possibilities for applications in the wastewater treatment sector. This systematic review focused specifically on the use of ANNs for predicting the performance of WWTPs in terms of effluent quality and removal efficiencies. This more focused approach was necessary due to the rigorous methods employed in a systematic review, allowing for thorough selection, screening, and analysis of publications, facilitating a deeper understanding of the main architectures, hyperparameter configurations of the models, and assessment of the studies. It is important to note that the implementation of the models in real-world WWTPs was not the primary focus of this work. However, it is worth mentioning that one of the main challenges in implementing these models remains the availability of high-quality data (Corominas *et al.* 2018; Faisal *et al.* 2023; Ray *et al.* 2023).

Other systematic reviews should be conducted for other specific applications of neural networks and even other machine learning algorithms in the wastewater treatment sector. For instance, neural networks and different machine learning approaches have been utilized for the optimization of WWTPs, including operational cost and energy consumption optimization, automation, control of operational conditions, real-time monitoring, forecasting of membrane fouling or operational failure (Ray *et al.* 2023), fault detection, and multi-objective control strategies that aim to maintain effluent quality while reducing energy consumption (Faisal *et al.* 2023). Each of these applications could serve as the focus of new systematic reviews.

Still considering the constantly evolving nature of machine learning and its applications, according to Zhang *et al.* (2023), future AI research applied to wastewater treatment will continue to focus on the removal of phosphorus, organic pollutants, and emerging contaminants. Promising directions for research include exploring microbial community dynamics, achieving multi-objective optimization, improving the performance of WWTPs to remove various pollutants, and predicting water quality under specific conditions (Zhang *et al.* 2023).

4. CONCLUSIONS

The results of the systematic review of the use of ANN models for the prediction of the performance of full-scale WWTPs, considering 44 relevant papers that were extracted and assessed accordingly, indicated the main trends and applications in the field. Most studies modeled a large activated sludge facility because they have better monitoring and control schemes. The datasets usually included a monitoring period of 1–2 years, with daily samplings, resulting in relatively small datasets (median = 361.5). Prior to training the models, the most common preprocessing method was the min–max normalization in the range [0, 1], and data division was achieved mainly with either 75% for training and 25% for testing the model, or 70% for training, 15% for validation, and 15% for testing.

The publications used influent indicator qualities as the input variables for neural network models to predict WWTP effluent quality, mainly those of organic matter concentrations. Although other methods were utilized, such as MLP, RBF, hybrid learning, and in recent years, deep learning, the FFNN architecture with a backpropagation training algorithm was the most common. In general, shallow networks with single hidden layers were used, and good performance was achieved.

Not all models must be tuned in the same manner, as they vary according to the dataset characteristics and study objectives. However, the findings of this research may act as a starting point and provide highly beneficial information to industry and research practitioners in the search for an optimum design modeling process in future studies with similar prediction problems.

ACKNOWLEDGMENTS

The authors would like to thank Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for their financial support during the course of the research.

FUNDING

This work was supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). The funding sources had no involvement in study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the article for publication.

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Abba, S. I., Abdulkadir, R. A., Gaya, M. S., Sammen, S. S., Ghali, U., Nawaila, M. B., Oğuz, G., Malik, A. & Al-Ansari, N. 2021a Effluents quality prediction by using nonlinear dynamic block-oriented models: a system identification approach. *Desalination and Water Treatment* **218**, 52–62. https://doi.org/10.5004/dwt.2021.26983.

Abba, S. I., Elkiran, G. & Nourani, V. 2021b Improving novel extreme learning machine using PCA algorithms for multi-parametric modeling of the municipal wastewater treatment plant. *Desalination and Water Treatment* **215**, 414–426. https://doi.org/10.5004/dwt.2021.26903.

Aldaghi, T. & Javanmard, S. 2021 The evaluation of wastewater treatment plant performance: a data mining approach. *Journal of Engineering, Design and Technology*. https://doi.org/10.1108/JEDT-07-2021-0394.

Al-Ghazawi, Z. & Alawneh, R. 2021 Use of artificial neural network for predicting effluent quality parameters and enabling wastewater reuse for climate change resilience – a case from Jordan. *Journal of Water Process Engineering* **44**, 1–10. https://doi.org/10.1016/j.jwpe.2021.102423.

Al-Obaidi, B. H. K. 2020 Predicting municipal sewage effluent quality index using mathematical models in the Al-Rustamiya sewage treatment plant. *Journal of Engineering Science and Technology* **15** (6), 3571–3587. Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85097312115&partnerID=40&md5=52f11896f100e75befc67517873499d9.

Alsulaili, A. & Refaie, A. 2021 Artificial neural network modeling approach for the prediction of five-day biological oxygen demand and wastewater treatment plant performance. *Water Supply* **21** (5), 1861–1877. https://doi.org/10.2166/ws.2020.199.

Bagheri, M., Mirbagheri, S. A., Ehteshami, M., Bagheri, Z. & Kamarkhani, A. M. 2015 Analysis of variables affecting mixed liquor volatile suspended solids and prediction of effluent quality parameters in a real wastewater treatment plant. *Desalination and Water Treatment* **57** (45), 1–14. https://doi.org/10.1080/19443994.2015.1125796.

Bahramian, M., Dereli, R. K., Zhao, W., Giberti, M. & Casey, E. 2023 Data to intelligence: the role of data-driven models in wastewater treatment. *Expert Systems with Applications* **217**, 1–20. https://doi.org/10.1016/j.eswa.2022.119453.

Bekkari, N. & Zeddouri, A. 2019 Using artificial neural network for predicting and controlling the effluent chemical oxygen demand in wastewater treatment plant. *Management of Environmental Quality: An International Journal* **30** (3), 593–608. https://doi.org/10.1108/MEQ-04-2018-0084.

Chen, H.-M. & Lo, S. L. 2012 Neural network-based multi-back-propagation prediction model of a domestic wastewater treatment plant for an under-construction sewer system. *Journal of the Chinese Institute of Engineers* **35** (7), 815–826. https://doi.org/10.1080/02533839.2012.708516.

Chen, Y., Song, L., Liu, Y., Yang, L. & Li, D. 2020 A review of the artificial neural network models for water quality prediction. *Applied Sciences* **10** (17), 1–49. https://doi.org/10.3390/app10175776.

Ching, P. M. L., So, R. H. Y. & Morck, T. 2021 Advances in soft sensors for wastewater treatment plants: a systematic review. *Journal of Water Process Engineering* **44**, 1–11. https://doi.org/10.1016/j.jwpe.2021.102367.

Civelekoglu, G., Yigit, N. O., Diamadopoulos, E. & Kitis, M. 2009 Modelling of COD removal in a biological wastewater treatment plant using adaptive neuro-fuzzy inference system and artificial neural network. *Water Science and Technology* **60** (6), 1475–1488. https://doi.org/10.2166/wst.2009.482.

Corominas, L., Garrido-Baserba, M., Villez, K., Olsson, G., Cortés, U. & Poch, M. 2018 Transforming data into knowledge for improved wastewater treatment operation: a critical review of techniques. *Environmental Modelling and Software* **106**, 89–103. https://doi.org/10.1016/j.envsoft.2017.11.023.

de Canete, J. F., del Saz-Orozco, P., Gómez-de-Gabriel, J., Baratti, R., Ruano, A. & Rivas-Blanco, I. 2021 Control and soft sensing strategies for a wastewater treatment plant using a neuro-genetic approach. *Computers and Chemical Engineering* **144**, 1–14. https://doi.org/10.1016/j.compchemeng.2020.107146.

Elfanssi, S., Ouazzani, N., Latrach, L., Hejjaj, A. & Mandi, L. 2018 Phytoremediation of domestic wastewater using a hybrid constructed wetland in mountainous rural area. *International Journal of Phytoremediation* **20** (1), 75–87. https://doi.org/10.1080/15226514.2017.1337067.

Elmaadawy, K., Elaziz, M. A., Elsheikh, A. H., Moawad, A., Liu, B. & Lu, S. 2021 Utilization of random vector functional link integrated with manta ray foraging optimization for effluent prediction of wastewater treatment plant. *Journal of Environmental Management* **298**, 1–9. https://doi.org/10.1016/j.jenvman.2021.113520.

El-Rawy, M., Abd-Ellah, M. K., Fathi, H. & Ahmed, A. K. A. 2021 Forecasting effluent and performance of wastewater treatment plant using different machine learning techniques. *Journal of Water Process Engineering* **44**, 102380. https://doi.org/https://doi.org/10.1016/j.jwpe.2021.102380.

Espinosa, M. F., Sancho, A. N., Mendoza, L. M., Mota, C. R. & Verbyla, M. E. 2020 Systematic review and meta-analysis of time-temperature pathogen inactivation. *International Journal of Hygiene and Environmental Health* **230**, 1–9. https://doi.org/10.1016/j.ijheh.2020.113595.

Faisal, M., Muttaqi, K. M., Sutanto, D., Al-shetwi, A. Q., Ker, P. J. & Hannan, M. A. 2023 Control technologies of wastewater treatment plants: the state-of-the-art, current challenges, and future directions. *Renewable and Sustainable Energy Reviews* **181**, 1–28. https://doi.org/10.1016/j.rser.2023.113324.

Fellows, I. 2018 wordcloud: Word Clouds. R package version 2.6. Available from: https://CRAN.R-project.org/package=wordcloud

Feng, J. & Lu, S. 2019 Performance analysis of various activation functions in artificial neural networks. *Journal of Physics: Conference Series* **1237** (2), 1–6. https://doi.org/10.1088/1742-6596/1237/2/022030.

Gaya, M. S., Abdul Wahab, N., Sam, Y. M. & Samsudin, S. I. 2014 ANFIS modelling of carbon and nitrogen removal in domestic wastewater treatment plant. *Jurnal Teknologi (Sciences and Engineering)* **67** (5), 29–34. https://doi.org/10.11113/jt.v67.2839.

Ge, J., Guha, B., Lippincott, L., Cach, S., Wei, J., Su, T.-L. & Meng, X. 2020 Challenges of arsenic removal from municipal wastewater by coagulation with ferric chloride and alum. *Science of The Total Environment* **725**, 1–9. https://doi.org/10.1016/j.scitotenv.2020.138351.

Guo, H., Jeong, K., Lim, J., Jo, J., Kim, Y. M., Park, J., Kim, J. H. & Cho, K. W. 2015 Prediction of effluent concentration in wastewater treatment plant using machine learning models. *Journal of Environmental Sciences* **32**, 90–101. https://doi.org/10.1016/j.jes.2015.01.007.

Hadjimichael, A., Comas, J. & Corominas, L. 2016 Do machine learning methods used in data mining enhance the potential of decision support systems? A review for the urban water sector. *AI Communications* **29** (6), 747–756. https://doi.org/10.3233/AIC-160714.

Hamed, M. M., Khalafallah, M. G. & Hassanien, E. A. 2004 Prediction of wastewater treatment plant performance using artificial neural networks. *Environmental Modelling and Software* **19** (10), 919–928. https://doi.org/10.1016/j.envsoft.2003.10.005.

Han, H.-G., Wang, L.-D. & Qiao, J.-F. 2014 Hierarchical extreme learning machine for feedforward neural network. *Neurocomputing* **128**, 128–135. https://doi.org/10.1016/j.neucom.2013.01.057.

Han, H., Zhu, S., Qiao, J. & Guo, M. 2018 Data-driven intelligent monitoring system for key variables in wastewater treatment process. *Chinese Journal of Chemical Engineering* **26** (10), 2093–2101. https://doi.org/https://doi.org/10.1016/j.cjche.2018.03.027.

Hazali, N., Wahab, N. A. & Ibrahim, S. 2017 Modelling and evaluation of sequential batch reactor using artificial neural network. *International Journal of Electrical and Computer Engineering* **7** (3), 1620–1627. https://doi.org/10.11591/ijece.v7i3.pp1620-1627.

Heddam, S., Lamda, H. & Filali, S. 2016 Predicting effluent biochemical oxygen demand in a wastewater treatment plant using generalized regression neural network based approach: a comparative study. *Environmental Processes* **3** (1), 153–165. https://doi.org/10.1007/s40710-016-0129-3.

Hejabi, N., Saghebian, S. M., Aalami, M. T. & Nourani, V. 2021 Evaluation of the effluent quality parameters of wastewater treatment plant based on uncertainty analysis and post-processing approaches (case study). *Water Science and Technology* **83** (7), 1633–1648. https://doi.org/10.2166/wst.2021.067.

Jami, M. S., Mujeli, M. & Kabbashi, N. A. 2011 Simulation of ammoniacal nitrogen effluent using feedforward multilayer neural networks. *African Journal of Biotechnology* **10** (81), 18755–18762. https://doi.org/10.5897/AJB11.2748.

Jami, M. S., Husain, I. A. F., Kabashi, N. A. & Abdullah, N. 2012 Multiple inputs artificial neural network model for the prediction of wastewater treatment plant performance. *Australian Journal of Basic and Applied Sciences* **6** (1), 62–69. Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84856678604&partnerID=40&md5=f6fb609d52fee2b555e2c80319f18ded

Khan, K. S., Kunz, R., Kleijnen, J. & Antes, G. 2003 Five steps to conducting a systematic review. *Journal of the Royal Society of Medicine* **96** (3), 118–121. https://doi.org/10.1177/014107680309600304.

Khatri, N., Khatri, K. K. & Sharma, A. 2019 Prediction of effluent quality in ICEAS-sequential batch reactor using feedforward artificial neural network. *Water Science and Technology* **80** (2), 213–222. https://doi.org/10.2166/wst.2019.257.

Khatri, N., Khatri, K. K. & Sharma, A. 2020 Artificial neural network modelling of faecal coliform removal in an intermittent cycle extended aeration system-sequential batch reactor based wastewater treatment plant. *Journal of Water Process Engineering* **37**, 1–8. https://doi.org/10.1016/j.jwpe.2020.101477.

Kusiak, A. & Wei, X. 2013 Optimization of the activated sludge process. *Journal of Energy Engineering* **139** (1), 12–17. https://doi.org/10.1061/(ASCE)EY.1943-7897.0000092.

Lantz, B. 2013 *Machine Learning with R (P. Publishing, Ed.)*. Packt Publishing, Birmingham.

Lee, J.-W., Suh, C., Hong, Y.-S. T. & Shin, H.-S. 2011 Sequential modelling of a full-scale wastewater treatment plant using an artificial neural network. *Bioprocess and Biosystems Engineering* **34** (8), 963–973. https://doi.org/10.1007/s00449-011-0547-6.

Liu, H., Huang, M. & Yoo, C. 2013 A fuzzy neural network-based soft sensor for modeling nutrient removal mechanism in a full-scale wastewater treatment system. *Desalination and Water Treatment* **51** (31–33), 6184–6193. https://doi.org/10.1080/19443994.2013.780757.

Liu, X., Shi, Q., Liu, Z. & Yuan, J. 2021 Using LSTM neural network based on improved PSO and attention mechanism for predicting the effluent COD in a wastewater treatment plant. *IEEE Access* **9**, 146082–146096. https://doi.org/10.1109/ACCESS.2021.3123225.

Madić, M. J. & Radovanović, M. R. 2011 Optimal selection of ANN training and architectural parameters using Taguchi method: a case study. *FME Transactions* **39** (2), 79–86.

Malviya, A. & Jaspal, D. 2021 Artificial intelligence as an upcoming technology in wastewater treatment: a comprehensive review. *Environmental Technology Reviews* **10** (1), 177–187. https://doi.org/10.1080/21622515.2021.1913242.

Mjalli, F. S., Al-Asheh, S. & Alfadala, H. E. 2007 Use of artificial neural network black-box modeling for the prediction of wastewater treatment plants performance. *Journal of Environmental Management* **83** (3), 329–338. https://doi.org/10.1016/j.jenvman.2006.03.004.

Nasr, M. S., Moustafa, M. A. E., Seif, H. A. E. & El Kobrosy, G. 2012 Application of Artificial Neural Network (ANN) for the prediction of EL-AGAMY wastewater treatment plant performance-Egypt. *Alexandria Engineering Journal* **51** (1), 37–43. https://doi.org/10.1016/j.aej.2012.07.005.

Newhart, K. B., Holloway, R. W., Hering, A. S. & Cath, T. Y. 2019 Data-driven performance analyses of wastewater treatment plants: a review. *Water Research* **157**, 498–513. https://doi.org/10.1016/j.watres.2019.03.030.

Newhart, K. B., Hering, A. S. & Cath, T. Y. 2022 Data science tools to enable decarbonized water and wastewater treatment systems. In: *Pathways to Water Sector Decarbonization, Carbon Capture and Utilization*. IWA Publishing, pp. 275–301. https://doi.org/10.2166/9781789061796.

Nezhad, M. F., Mehrdadi, N., Torabian, A. & Behboudian, S. 2016 Artificial neural network modeling of the effluent quality index for municipal wastewater treatment plants using quality variables: south of Tehran wastewater treatment plant. *Journal of Water Supply: Research and Technology* **65** (1), 18–27. https://doi.org/10.2166/aqua.2015.030.

Nourani, V., Elkiran, G. & Abba, S. I. 2018 Wastewater treatment plant performance analysis using artificial intelligence – an ensemble approach. *Water Science and Technology* **78** (10), 2064–2076. https://doi.org/10.2166/wst.2018.477.

Nourani, V., Asghari, P. & Sharghi, E. 2021 Artificial intelligence based ensemble modeling of wastewater treatment plant using jittered data. *Journal of Cleaner Production* **291**, 1–15. https://doi.org/10.1016/j.jclepro.2020.125772.

Onu, C. E., Nwabanne, J. T., Ohale, P. E. & Asadu, C. O. 2021 Comparative analysis of RSM, ANN and ANFIS and the mechanistic modeling in eriochrome black-T dye adsorption using modified clay. *South African Journal of Chemical Engineering* **36**, 24–42. https://doi.org/10.1016/j.sajce.2020.12.003.

Osman, Y. B. M. & Li, W. 2020 Soft sensor modeling of key effluent parameters in wastewater treatment process based on SAE-NN. *Journal of Control Science and Engineering* **2020**, 1–9. https://doi.org/10.1155/2020/6347625.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P. & Moher, D. 2021 The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* **372** (71), 1–9. https://doi.org/10.1136/bmj.n71.

Palani, S., Liong, S.-Y. & Tkalich, P. 2008 An ANN application for water quality forecasting. *Marine Pollution Bulletin* **56** (9), 1586–1597. https://doi.org/10.1016/j.marpolbul.2008.05.021.

Pham, Q. B., Gaya, M. S., Abba, S. I., Abdulkadir, R. A., Esmaili, P., Linh, N. T. T., Sharma, C., Malik, A., Khoi, D. N., Dung, T. N. & Linh, D. Q. 2020 Modeling of Bunus regional sewage treatment plant using machine learning approaches. *Desalination and Water Treatment* **203**, 80–90. https://doi.org/10.5004/dwt.2020.26160.

Pullin, A. S. & Stewart, G. B. 2006 Guidelines for systematic review in conservation and environmental management. *Conservation Biology* **20** (6), 1647–1656. https://doi.org/10.1111/j.1523-1739.2006.00485.x.

Qazi, A., Fayaz, H., Wadi, A., Raj, R. G., Rahim, N. A. & Khan, W. A. 2015 The artificial neural network for solar radiation prediction and designing solar systems: a systematic literature review. *Journal of Cleaner Production* **104**, 1–12. https://doi.org/10.1016/j.jclepro.2015.04.041.

Qiao, J.-f., Yang, W. & Yuan, M. 2011 Recurrent high order neural network modeling for wastewater treatment process. *Journal of Computers* **6** (8), 1570–1577. https://doi.org/10.4304/jcp.6.8.1570-1577.

Qiao, J., Hu, Z. & Li, W. 2016 Soft measurement modeling based on chaos theory for biochemical oxygen demand (BOD). *Water* **8** (12), 1–21. https://doi.org/10.3390/w8120581.

Rahmati, M. G., Tishehzan, P. & Moazed, H. 2021 Determining the best and simple intelligent models for evaluating BOD5 of Ahvaz wastewater treatment plant. *Desalination and Water Treatment* **209**, 242–253. https://doi.org/10.5004/dwt.2021.26481.

Ray, S. S., Verma, R. K., Singh, A., Ganesapillai, M. & Kwon, Y.-N. 2023 A holistic review on how artificial intelligence has redefined water treatment and seawater desalination processes. *Desalination* **546**, 1–14. https://doi.org/10.1016/j.desal.2022.116221.

Saleh, H. A. A. 2021 Wastewater pollutants modeling using artificial neural networks. *Journal of Ecological Engineering* **22** (7), 35–45. https://doi.org/10.12911/22998993/138872.

Sharghi, E., Nourani, V., Aliashrafi, A. & Gökçekuş, H. 2019 Monitoring effluent quality of wastewater treatment plant by clustering based artificial neural network method. *Desalination and Water Treatment* **164**, 86–97. https://doi.org/10.5004/dwt.2019.24385.

Sharma, S., Sharma, S. & Anidhya, A. 2020 Activation functions in neural networks. *International Journal of Engineering Applied Sciences and Technology* **4** (12), 310–316. https://doi.org/10.33564/IJEAST.2020.v04i12.054.

Simsek, H. 2016 Mathematical modeling of wastewater-derived biodegradable dissolved organic nitrogen. *Environmental Technology* **37** (22), 2879–2889. https://doi.org/10.1080/09593330.2016.1167964.

Sin, G. & Al, R. 2021 Activated sludge models at the crossroad of artificial intelligence – a perspective on advancing process modeling. *Clean Water* **4** (16), 1–7. https://doi.org/10.1038/s41545-021-00106-5.

Singh, K. P., Basant, N., Malik, A. & Jain, G. 2010 Modeling the performance of 'up-flow anaerobic sludge blanket' reactor based wastewater treatment plant using linear and nonlinear approaches – a case study. *Analytica Chimica Acta* **658** (1), 1–11. https://doi.org/10.1016/j.aca.2009.11.001.

Verma, A., Wei, X. & Kusiak, A. 2013 Predicting the total suspended solids in wastewater: a data-mining approach. *Engineering Applications of Artificial Intelligence* **26** (4), 1366–1372. https://doi.org/10.1016/j.engappai.2012.08.015.

Wang, D., Thunéll, S., Lindberg, U., Jiang, L., Trygg, J., Tysklind, M. & Souihi, N. 2021 A machine learning framework to improve effluent quality control in wastewater treatment plants. *Science of The Total Environment* **784**, 1–11. https://doi.org/10.1016/j.scitotenv.2021.147138.

Yaqub, M., Asif, H., Kim, S. & Lee, W. 2020 Modeling of a full-scale sewage treatment plant to predict the nutrient removal efficiency using a long short-term memory (LSTM) neural network. *Journal of Water Process Engineering* **37**, 1–11. https://doi.org/10.1016/j.jwpe.2020.101388.

Yasmin, N. S. A., Gaya, M. S., Wahab, N. A. & Sam, Y. M. 2017 Estimation of pH and MLSS using neural network. *Telkomnika (Telecommunication Computing Electronics and Control)* **15** (2), 912–918. https://doi.org/10.12928/TELKOMNIKA.v15i2.6144.

Ye, Z., Yang, J., Zhong, N., Tu, X., Jia, J. & Wang, J. 2020 Tackling environmental challenges in pollution controls using artificial intelligence: a review. *Science of the Total Environment* **699**, 1–28. https://doi.org/10.1016/j.scitotenv.2019.134279.

Zhang, R. & Hu, X. 2012 Effluent quality prediction of wastewater treatment system based on small-world ANN. *Journal of Computers* **7** (9), 2136–2143. https://doi.org/10.4304/jcp.7.9.2136-2143.

Zhang, S., Jin, Y., Chen, W., Wang, J., Wang, Y. & Ren, H. 2023 Artificial intelligence in wastewater treatment: a data-driven analysis of status and trends. *Chemosphere* **336**, 1–8. https://doi.org/10.1016/j.chemosphere.2023.139163.

Zhao, L.-J., Chai, T.-Y. & Yuan, D.-C. 2012 Selective ensemble extreme learning machine modeling of effluent quality in wastewater treatment plants. *International Journal of Automation and Computing* **9** (6), 627–633. https://doi.org/10.1007/s11633-012-0688-3.

Zhao, Y., Guo, L., Liang, J. & Zhang, M. 2016 Seasonal artificial neural network model for water quality prediction via a clustering analysis method in a wastewater treatment plant of China. *Desalination and Water Treatment* **57** (8), 3452–3465. https://doi.org/10.1080/19443994.2014.986202.

Zhao, L., Dai, T., Qiao, Z., Sun, P., Hao, J. & Yang, Y. 2020 Application of artificial intelligence to wastewater treatment: a bibliometric analysis and systematic review of technology, economy, management, and wastewater reuse. *Process Safety and Environmental Protection* **133** (92), 169–182. https://doi.org/10.1016/j.psep.2019.11.014.

First received 25 April 2023; accepted in revised form 20 August 2023. Available online 29 August 2023