


## Research progress of the artificial intelligence application in wastewater treatment during 2012–2022: a bibliometric analysis

Xiaoman Yu<sup>a</sup>, Shuai Chen<sup>a,b</sup>, Xiaojiao Zhang<sup>a</sup>, Hongcheng Wu<sup>c</sup>, Yaoguang Guo<sup>a</sup> and Jie Guan <sup>a,\*</sup>

<sup>a</sup> School of Resources and Environmental Engineering, Shanghai Polytechnic University, Shanghai 201209, China

<sup>b</sup> Anhui International Joint Research Center for Nano Carbon-based Materials and Environmental Health, Huainan 232001, China

<sup>c</sup> Shanghai Wobai Environmental Development Co. Ltd, Shanghai 201209, China

\*Corresponding author. E-mail: guanjie@sspu.edu.cn

 JG, 0000-0003-1197-3142

### ABSTRACT

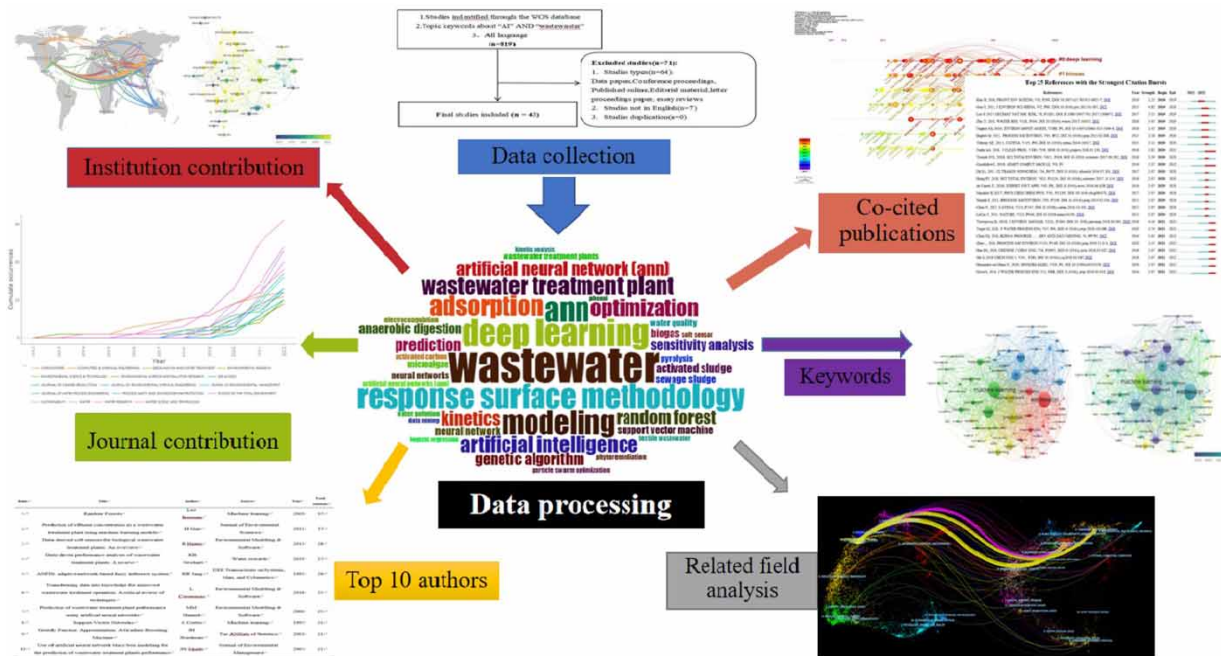
This study identified literatures from the Web of Science Core Collection on the application of artificial intelligence in wastewater treatment from 2011 to 2022, through bibliometrics, to summarize achievements and capture the scientific and technological progress. The number of papers published is on the rise, and especially, the number of papers issued after 2018 has increased sharply, with China contributing the most in this regard, followed by the US, Iran and India. The University of Tehran has the largest number of papers, WATER is the most published journal, and Nasr M has the largest number of articles. Collaborative network has been developed mainly through cooperation between European countries, China and the US. Remote sensing in developing countries needs to be further integrated with water quality monitoring programs. It is worth noting that artificial neural network is a research hotspot in recent years. Through keyword clustering analysis, 'machine learning' and 'deep learning' are hot keywords that have emerged since 2019. The use of neural networks for predicting the effectiveness of treatment of difficult to degrade wastewater is a future research trend. The rapid advancement of deep learning provides the opportunity to build automated pipeline defect detection systems through image recognition.

**Key words:** artificial intelligence, artificial neural network, deep learning, machine learning, wastewater treatment

### HIGHLIGHTS

- Using bibliometrics to explore the application of artificial intelligence in wastewater treatment.
- The combination of deep learning has been developed.
- Remote sensing technology has become a research hotspot.
- Machine learning or deep learning modeling is widely used to build data-driven predictive models.

## GRAPHICAL ABSTRACT



## 1. INTRODUCTION

Water is the most crucial resources for organisms' survival. In the past decades, the demand for and abuse of water resources have caused enormous pressure on the water supply. Moreover, many countries are facing serious water pollution issues. In contemporary industry, water is necessary. However, in the midst of development, many companies use water in inefficient ways and even sometimes pollute it, for example, by discharging wastewater unwisely (Elsey-Quirk *et al.* 2022). Wastewater treatment (WWT) is an efficacious way to address water scarcity and water quality deterioration (Abdeldayem *et al.* 2022). Further treatment or polishing of treated wastewater can provide high-quality water, such as drinking water (Altowayti *et al.* 2022). Therefore, the requirements for sustainable sources of clean water are therefore essential for wastewater treatment plants (WWTPs), as well as for many other natural and industrial organizations that depend on water availability (Faherty 2021). In addition to meeting consumer demand and providing the necessary quality-of-life upgrades to infrastructure, WWTPs must cope with complex regulatory measures to meet rising quality criteria (Lowe *et al.* 2022).

Artificial intelligence (AI) focuses on the study of computers to mimic some human thought processes and intelligent behaviors, thus enabling computers to achieve higher-level applications. After training, AI systems perform intelligent control and optimization of various processes to enhance the product conformity rate. It is used extensively in such sectors as industrial control (Seo *et al.* 2021), optimal design (Matheri *et al.* 2022), fault diagnosis (Fuente & Represa 1997) and intelligent detection (Chen & Chen 2020). Machine learning (ML) is at the heart of AI and includes almost all of the most influential methods in AI. It includes deep learning (DL), random forests (RF), support-vector machine (SVM) algorithms, etc. With state-of-the-art advances in computational algorithms and a modern computer computing power, ML has made significant contributions to the development and application of various types of AI. At the meantime, ML models have showcased their capabilities in optimizing, modeling and automating water supply and WWT applications (Yaqub *et al.* 2020; Zhu & Piotrowski 2020). In a way which provides computer assistance for complex problems in WWT using chemical, physical and biological processes, AI can optimize water-based applications and reduce capital expenditure (Torregrossa *et al.* 2018; Zhang *et al.* 2021). In the background of the progress for AI, it is a necessity to probe research hotspots in order to help the investigation and progress in the domain of WWT. Bibliometrics measure the size, authorship and lexicon by using mathematics, statistics and other analytical quantitative research methods, to study the distribution of information in the literature. The bibliometric survey

can conclude the historical research results of AI in WWT, undertake hotspot analysis and predict the futuristic research trends (Xu *et al.* 2022).

A bibliometric survey allows for a summary of historical research results in WWT. Carmona & Abejón (2023) used the Scopus database as a source to examine the scientific literature on the removal of heavy metals from wastewater using electro dialysis, membrane distillation and forward osmosis. Ola *et al.* (2023) studied the progress of research on chitosan-based composite photocatalyst systems for WWT from a bibliometric perspective and developed new and efficient methods to synthesize chitosan from fish by-products (waste) with high adsorption efficiency for the removal of various types of pollutants. The econometric literature can also be used to conduct a hotspot analysis and predict future research trends. Zhang *et al.* (2023) used CiteSpace to systematically evaluate the structure, trends, research hotspots and frontiers of greenhouse gas emissions from WWTPs from a bibliometric perspective. Kamilya *et al.* (2022) explored the technical progress and future scope of research trends on nutrient recovery from waste streams. In contrast, under the global trend of rapid development of the information technology revolution, there is no article about systematically analyzing the application of AI in the field of WWT from a bibliometric perspective.

In this paper, a bibliometric analysis method is applied to search the papers in the Web of Science Core Collection (WOS) from October 2012 to October 2022, matching the keywords 'artificial intelligence' and 'wastewater treatment'. Countries and institutions are analyzed according to filtering and downloading needs. Collaborations, publications and citations are visualized by means of maps and charts. The subject knowledge base is examined by extracting author information (number of institutions, h-index, etc.), and cited literature data. Finally, the cited references are extracted, and key words are used to identify topical and current research trends.

## 2. MATERIALS AND METHODS

### 2.1. Data collection

Data were gathered from WOS. To eliminate the bias caused by database updates and to ensure the integrity of the search, the main search terms used were 'wastewater', 'effluent', 'treatment', 'handover', 'machine learning' and 'artificial intelligence'.

Zhu *et al.* (2021) pointed out that synonyms would be found in the recalled keywords during data collection, such as ('wastewaters' and 'sewage') (Table S1). In this endeavor, English was used as the only language, and publications like editorials, letters, case reports, etc., were excluded. Details such as titles, authors, (author-defined) keywords, abstracts and references were extracted from the eligible documents and exported them in TXT format. Due to the nature of CiteSpace, the data collector does not drop articles that appear to be ineligible, as they may interfere with potential data linkages of the text and reduce the accuracy/sensitivity of the analysis results (Li & Chen 2016).

### 2.2. Data analysis

CiteSpace 6.1.R6 software (Drexel University, Philadelphia, PA, USA) is used as a citation visualization and parsing software. During analysis using CiteSpace, data are saved as 'down\_xxx' and stored and loaded according to the program's specific data format requirements. In this study, the slice length was always set to 1. The VOSviewer (Leiden University, Leiden, Netherlands) tool is adept at generating text maps of any type, visualizing the relationships between documents in stunning graphics to quickly target key documents in the subject area. Data were visualized using CiteSpace and Vosviewer for country and institution. The Bibliometrix package was loaded by using RStudio (version 1.4.1717) for co-occurrence of keywords and RStudio was used to visualize temporal trends in keywords.

## 3. RESULTS

### 3.1. Article distributions in various years

A total of 748 records were analyzed in this study, including 678 research papers (90.64%) and 70 reviews (9.36%). Based on the number of publications published from 2012 to 2022, the annual average percentage growth rate from 2012 to 2017 was only 36.34%. From 2017 onwards, the number of publications per year has surged, averaging 73.91% from 2012 to 2022 (Figure 1). It is evident that AI in WWT has attracted the attention of scholars since 2018. Based on the trend in the number of published papers, it can be predicted that this field will remain hot in the future.



**Figure 1** | Trends in the number of AI publications in WWT from 2012 to 2022.

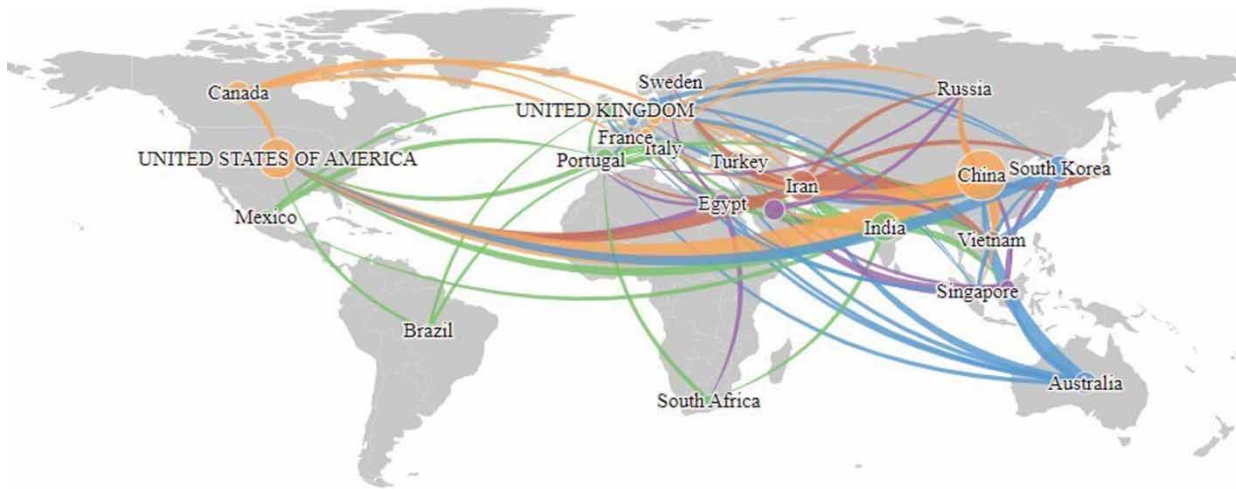
### 3.2. Countries, institutions and regions

The publications come from 82 countries, among which 10 countries with the highest number of publications were statistically represented. China has the greatest number ( $n = 210$ ), accounting for 28.07% of the total number of published articles, followed by the US ( $n = 170$ , 17.38%), Iran ( $n = 77$ , 10.29%) and India ( $n = 70$ , 9.37%). Of these, the US (0.33) and China (0.31) had the highest centrality, much higher than the other countries. This indicates that both China and the US have made a contribution and have a certain impact in this research area (Table 1). The top 10 countries with the most published articles were cited a total of 12,081 times (excluding self-citations). China is the most cited country (3,174, 26.27%). The US and Iran are the two most commonly cited countries on average. This indicated that there is no strong link between average article citations and centrality.

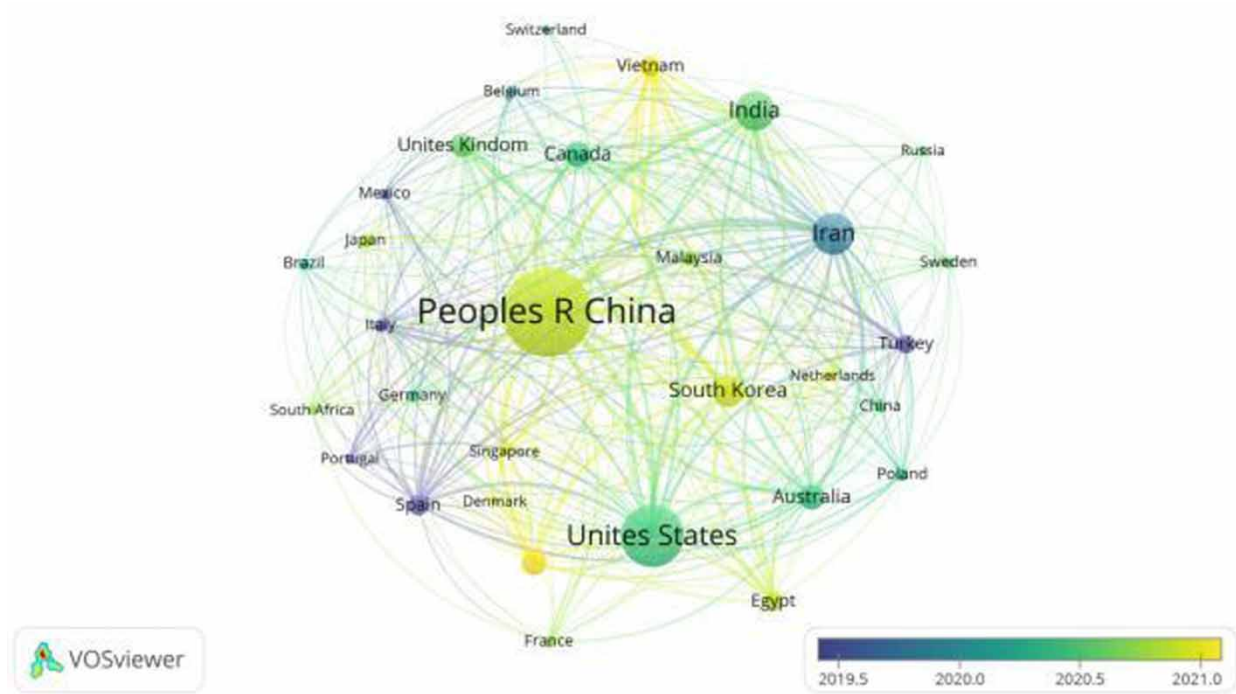
Map visualization was made by a combination of VOSviewer and Scimago Graphic software (Figure 2). It can be seen that China and the US, as the countries with the highest number of publications, tend to have a lot of interchanges and cooperation with countries such as Canada and Russia. Developing countries, like Mexico, South Africa and India, tend to cooperate with developed European countries such as France and Italy. Moreover, Australia has widespread cooperation with many countries. In terms of the temporal distribution, the size of the circle is in proportion to the number of publications. Spain, Turkey and Italy were the first countries to start conducting new research in the field of WWT and AI from 2019, followed by Iran, China, South Korea and Vietnam (Figure 3).

**Table 1** | Top 10 countries/regions of the publications

No.	Countries/regions	Number of publications	Number of citations	Average article citations	Centrality
1	China	210	3,174	15.11428571	0.31
2	Unites States	130	1,521	11.7	0.33
3	Iran	77	2,011	26.11688312	0.15
4	India	70	1,180	16.85714286	0.2
5	South Korea	52	534	10.26923077	0.02
6	Australia	40	630	15.75	0.04
7	Canada	40	594	14.85	0.01
8	Saudi Arabia	37	293	7.918918919	0.12
9	United Kingdom	36	771	21.41666667	0.07
10	Spain	29	431	14.86206897	0.2



**Figure 2** | Visual map of country cooperation publications (publications  $\geq 10$ ). The thickness of lines and the size of circles represent the number of publications, and the type of color represents the clustering of countries that regularly cooperate.



**Figure 3** | Network of inter-country cooperative relations over time (publications  $\geq 10$ ). The clearer the color, the earlier the country starts research in the linked field.

In total, all records encompassed 1,242 organizations. The top 10 organizations and co-cited authors are as shown in Table 2. University of Tehran (Iran) has the largest number of publications, with 19 publications, accounting for 2.5% of the total, and the highest centrality of 0.08, followed by Duy Tan University (18 publications, Vietnam) and Chinese Academy of Sciences (16 publications, China). Duy Tan University (Vietnam) was the most frequently cited, with a total of 743 citations (6.18%), noticeably more than any other institution. In the top 10 organizations publishing articles, 4 are from China, 1 is from the US and the remaining 5 are from developing Asia.



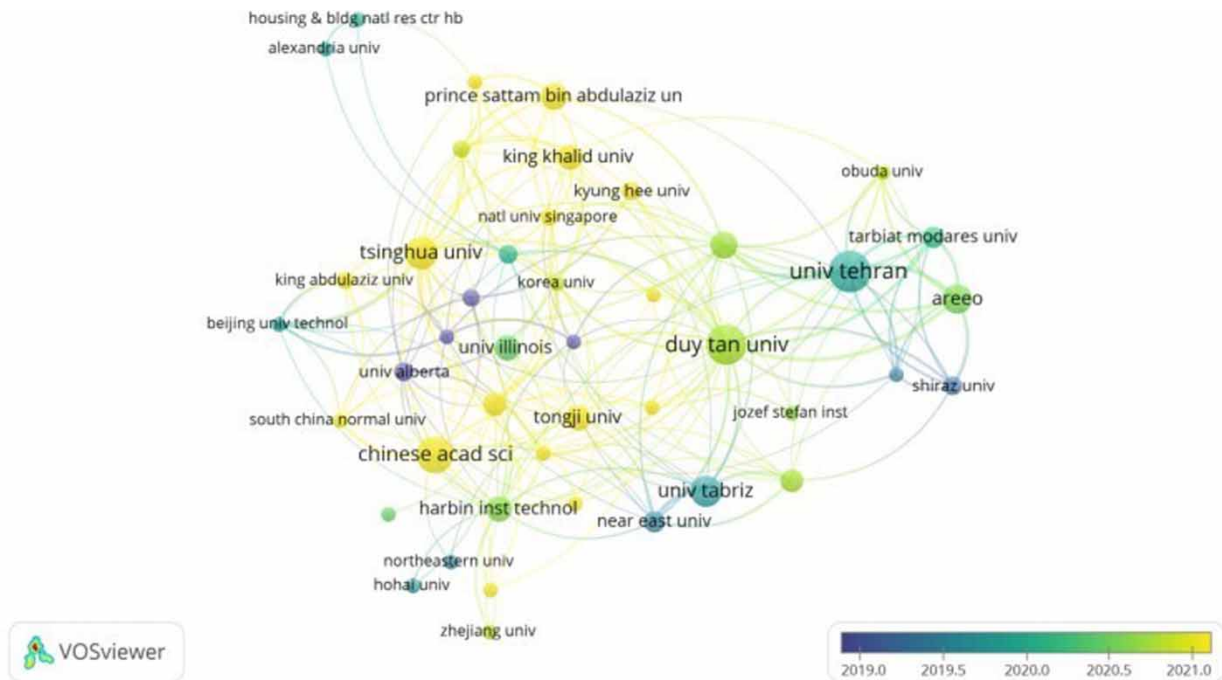
**Table 2** | Top 10 relevant institutions of publications

No.	Institutions	Number of publications	Number of citations	Centrality
1	Univ Tehran (Iran)	19	560	0.08
2	Duy Tan Univ (Vietnam)	18	743	0.06
3	Chinese Acad Sci (China)	16	218	0.05
7	Tsinghua Univ (China)	14	184	0.07
4	Univ Tabriz (Iran)	13	249	0.01
6	AREEO (Iran)	12	514	0.04
5	Harbin Inst Technol (China)	10	139	0
8	Tongji Univ (China)	10	58	0
9	King Khalid Univ (Saudi Arabia)	10	28	0
10	Univ Illinois (US)	10	60	0

More institutions could be included to analyze the tendency of the research in this area (Figure 4). The color changes from dark blue to bright yellow, the former color indicates the institutions with early publication time and the later color indicates the ones with late publication time. Through a brief analysis, it can be seen that the Chinese universities, like Tongji University, Harbin Institute of Technology and Tsinghua University are in yellow color. It can be inferred that this field has been a growing field of research in Chinese universities.

### 3.3. Journals

Considering the top 10 journals from 2012 to 2022, the top three were *WATER* ( $n = 31$ ), *WATER RESEARCH* ( $n = 24$ ) and *SCIENCE OF THE TOTAL ENVIRONMENT* ( $n = 23$ ). In terms of citations, *SCIENCE OF THE TOTAL ENVIRONMENT*



**Figure 4** | Number of inter-institutional collaborative publications. The size of the circle indicates the volume of publications of the institution, and the color indicates the time of the publication.

had the most citations ( $n = 812$ ), followed by *WATER* ( $n = 495$ ) and *WATER RESEARCH* ( $n = 278$ ). Of the top 10 publications, only *WATER SCIENCE AND TECHNOLOGY* had a centrality greater than 0.1, which suggests that this journal has a certain level of authority in this field (Table 3).

The number of journals has risen significantly since 2018, it can be forecasted that the number of publications will continually grow (Figure S1). Of these journals, *SUSTAINABILITY* is one that has shone brightly in recent years. From 2013 to 2019, the annual number of articles published in it has increased from less than 300 initially to over 7,000. In 2013, the value was 0.861, equivalent to 80% of the impact factor (1.077). In 2019, the value (1.711) was equivalent to just 66.4% of the impact factor (2.576). Since 2019, *WATER* has been the most published journal in the field and the journal has been able to quickly capture and research hotspots.

### 3.4. Authors of the top 10 publications

A total of 3,295 authors published articles in this area. The top 10 authors, including the author's country, affiliation and other information as shown in Table 4. The two authors with the largest number of articles are Nasr M and Pourghasemi HR, both published seven articles (0.94%). The most cited author is Choubin B ( $n = 267$ , 2.21%). However, only three authors have an h-index above 5, Pourghasemi HR is the author with the top h-index of 7.

**Table 3** | Top 10 journals in terms of number of publications

No.	Journals	Number of publications	Number of citations	Centrality
1	<i>WATER</i>	31	495	0.02
2	<i>WATER RESEARCH</i>	24	278	0.05
3	<i>SCIENCE OF THE TOTAL ENVIRONMENT</i>	23	812	0.06
4	<i>SUSTAINABILITY</i>	20	117	0.01
5	<i>WATER SCIENCE AND TECHNOLOGY</i>	18	205	0.11
6	<i>JOURNAL OF CLEANER PRODUCTION</i>	17	241	0.01
7	<i>JOURNAL OF ENVIRONMENTAL MANAGEMENT</i>	16	194	0.02
8	<i>ENVIRONMENTAL SCIENCE AND POLLUTION RESEARCH</i>	13	82	0.02
9	<i>JOURNAL OF WATER PROCESS ENGINEERING</i>	13	67	0.01
10	<i>PROCESS SAFETY AND ENVIRONMENTAL PROTECTION</i>	13	284	0.02

**Table 4** | Top 10 journals in terms of number of publications

No.	Authors	Countries	Affiliations	No. of publications	Number of citations	H-index
1	Nasr M	Egypt	Alexander Academy	7	90	6
2	Pourghasemi HR	Iran	Shiraz University	7	66	7
3	Choubin B	Iran	University of Tehran	5	267	4
4	Harrou F	Saudi Arabia	King Abdullah University of Science and Technology (KAUST)	5	119	5
5	Jang A	South Korea	Sungkyunkwan University (SKKU)	5	28	3
6	Mahmoud AS	US	Virginia Tech	5	38	3
7	Qiao JF	China	Beijing University of Technology	5	152	5
8	LY QV	Korea	Sejong University	5	25	3
9	Sun Y	US	Utah State University	5	119	6
10	Wang GM	China	Chinese Academy of Medical Sciences	5	152	5

A total of 37 publications have been published since 2018. Among them, the most cited articles were about the assessment of environmental efficiency and technologies for phytoremediation and activated carbon treatment of wastewater (Ansari *et al.* 2018). Under the premise of sufficient data, they established a predictive model based on ANN to estimate the cost of removing chlorothalonil pesticides by activated carbon and fermenting biohydrogen by *Pistia stratiotes*. Pourghasemi HR's publications are related to agricultural engineering geology, environmental science and ecology (Alalm & Nasr 2018; Mthethwa *et al.* 2018). He has published a total of 26 publications. He was awarded the titles of 'Highly Cited Interdisciplinary Researcher' and 'Highly Cited Interdisciplinary Scientist' in 2019 and 2020, consecutively. Choubin B's publications focus on the use of ML modeling for flood risk and sensitivity assessment (Taromideh *et al.* 2022; Choubin *et al.* 2023), using integrated learning modeling to explore the sensitivity of the model to groundwater salinity/hardness. He was the first person who analyzed the performance of 10 advanced ML techniques, including ANN, boosted regression trees (BRT), classification and regression trees (CART) and generalized linear models, to illustrate the central importance of ML modeling for landslide susceptibility and assessment variables in GIS and R open-source software (Pourghasemi & Rahmati 2018). Subsequently, landslide sensitivity modeling, adaptive neuro-fuzzy inference systems (Youssef *et al.* 2022) and a new set of meta-heuristics for flood sensitivity mapping applied to ML methods were developed (Chen *et al.* 2018). In 37 articles published by Choubin *et al.* (2020), models were constructed using ML for a multitude of areas of prediction, including climate prediction, flood sensitivity prediction (Choubin *et al.* 2023), groundwater hardness sensitivity and other topics (Mosavi *et al.* 2021).

### 3.5. Co-cited top 10 articles

Co-citation profiling can be used not only to show researchers' concern and to reveal developments in the structure of science but also to reveal cutting edge and domain analysis. Table 5 shows the top 10 most cited papers from 2012 to 2022. The most frequently cited paper is RF ( $n = 37$ ), published in the journal *MACHINE LEARNING* in 2001. RF is an ensemble learning method which can handle classification and regression problems very successfully. It is highly sought after for its simplicity. A total of three review articles in the 10 co-cited articles examine and discuss the usage of data-driven soft sensing technologies in wastewater biological treatment plants. The contributions focus on the available state of the art of these soft sensing technologies and the specific tasks and potential they may face in practical applications. Further, there are four articles on ML, two of which introduce RF and SVM algorithms, and the other one optimizes the algorithms. Also, there is one

**Table 5** | Top 10 most cited articles of the publications

No.	Titles	Authors	Sources	Year	Total citations
1	Random Forests	Breiman L	Machine learning	2001	37
2	Prediction of effluent concentration in a wastewater treatment plant using machine learning models	Guo H	Journal of Environmental Sciences	2015	37
3	Data-derived soft-sensors for biological wastewater treatment plants: An overview	Haimi H	Environmental Modelling & Software	2013	28
4	Data-driven performance analyses of wastewater treatment plants: A review	Newhart KB	Water research	2019	27
5	ANFIS: adaptive-network-based fuzzy inference system	Jang JSR	IEEE Transactions on Systems, Man, and Cybernetics	1993	26
6	Transforming data into knowledge for improved wastewater treatment operation: A critical review of techniques	Corominas L	Environmental Modelling & Software	2018	25
7	Prediction of wastewater treatment plant performance using artificial neural networks	Hamed MM	Environmental Modelling & Software	2004	25
8	Support-Vector Networks	Cortes C	Machine learning	1995	21
9	Greedy function approximation: A gradient boosting machine	Friedman JH	The Annals of Statistics	2001	21
10	Use of artificial neural network black-box modeling for the prediction of wastewater treatment plants performance	Mjalli FS	Journal of Environmental Management	2007	21



paper that constructs an adaptive network-based fuzzy inference system (ANFIS) with promising applications in automatic control and signal processing.

Generating co-citation clusters can help to scientifically measure the relationships between references and thus infer research hotspots and trends (Table 6). In the top left corner, Modularity  $Q = 0.9028$  ( $Q > 0.5$  indicates significant clustering structure) and Weighted Mean Silhouette  $S = 0.9435$  ( $S > 0.7$  suggests convincing clustering) show good performance, indicating very good clustering (Figure 5). All co-cited literatures were grouped into seven clusters, namely #0 'machine learning', #1 'biogas', #2 'hazards', #3 'ditch erosion', #4 'sorption', #5 'wastewater treatment' and #6 'artificial nerves'.

Considering the 25 most cited articles (Figure S2), 11 of them centered on the use of AI technologies to improve plant performance in terms of reducing operational costs, as well as enhancing effluent quality. Eight of the 11 articles are dedicated to the construction of predictive wastewater concentration models based on different ML methods such as ANN (Guo *et al.* 2015; Kim *et al.* 2016), KNN (Chen & Schmidt 2016; Nadiri *et al.* 2018), FNN (Han *et al.* 2018; Hernández-del-Olmo *et al.* 2019) and DL (Bagheri *et al.* 2015; Naghibi *et al.* 2016)).

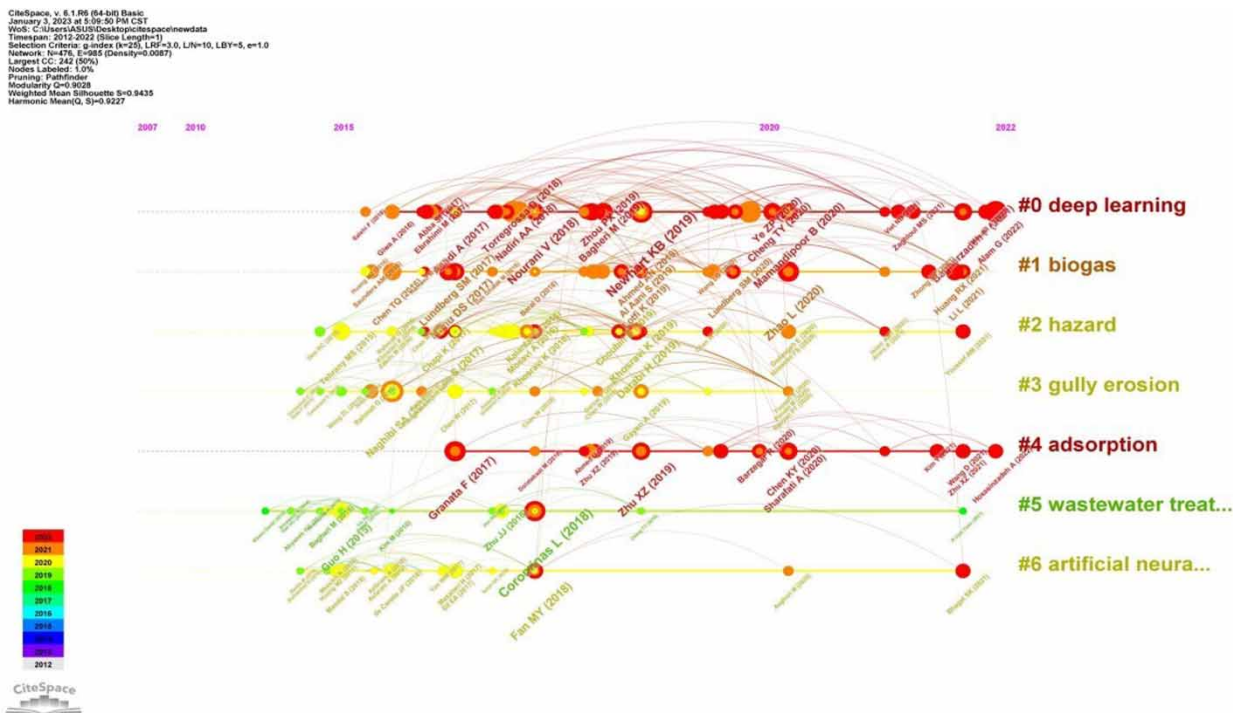
Torregrossa *et al.* (2018) built a new approach using energy prices as model parameters, in which an ML algorithm-based approach was proposed, and its relevance was evaluated for the first time. A high-performance energy cost model was generated for WWT works, which has been used in a database of 317 WWTPs in northwestern Europe. Bagheri *et al.* (2015) mixed liquid volatile suspended solids, pH, dissolved oxygen, temperature, total suspended solid, chemical oxygen demand and total nitrogen (TN) as inputs to a neural network, and developed a hybrid artificial neural network-genetic algorithm model to accurately determine the sludge volume index for better prediction of sludge expansion in WWTPs. Another paper developed an ML soft sensor that can predict unobservable measurements from existing data based on ML and models of past influent states at WWTPs. This soft sensor can predict weather conditions while the operator is monitoring WWTPs (Hernández-del-Olmo *et al.* 2019).

### 3.6. Key words

There are four clusters around the key words 'machine learning', 'effluent', 'models' and 'prediction' (Figure 6). As an emerging key word, 'machine learning', a branch of AI, has also been seen a large number of articles published, which warrants further study.

**Table 6** | Top 10 co-cited publications

No.	Titles	Authors	Sources	Year	Total citations
1	Random forests	Breiman L	Machine learning	2001	37
2	Prediction of effluent concentration in a wastewater treatment plant using machine learning models	Guo H	Journal of Environmental Sciences	2015	37
3	Data-derived soft-sensors for biological wastewater treatment plants: An overview	Haimi H	Environmental Modelling & Software	2013	28
4	Data-driven performance analyses of wastewater treatment plants: A review	Newhart KB	Water research	2019	27
5	ANFIS: adaptive-network-based fuzzy inference system	Jang JSR	IEEE Transactions on Systems, Man, and Cybernetics	1993	26
6	Transforming data into knowledge for improved wastewater treatment operation: A critical review of techniques	Corominas L	Environmental Modelling & Software	2018	25
7	Prediction of wastewater treatment plant performance using artificial neural networks	Hamed MM	Environmental Modelling & Software	2004	25
8	Support-Vector Networks	Cortes C	Machine learning	1995	21
9	Greedy Function Approximation: A Gradient Boosting Machine	Friedman JH	The Annals of Statistics	2001	21
10	Use of artificial neural network black-box modeling for the prediction of wastewater treatment plants performance	Mjalli FS	Journal of Environmental Management	2007	21



**Figure 5** | Co-cited clustered view of AI in WWT 2012–2021.

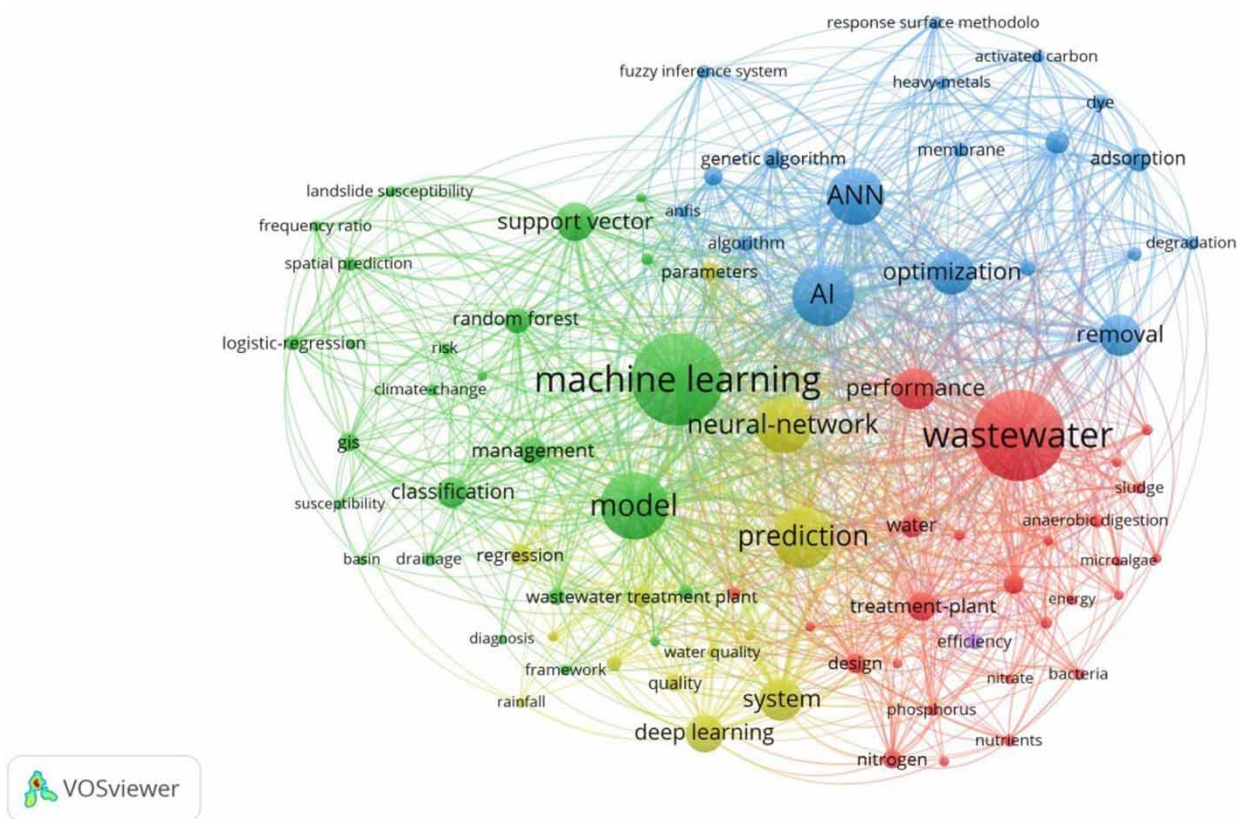
The key words include the core ideas, research topics and research methods of the literature, which can explore the hot-spots and emerging trends in this research field. To extract the 17 key words with the strongest citation burst, the key words include title, abstract, author keywords, key words plus (Table S1). An analysis of the outbreak years of key words in the figure shows that the burst key words from 2013 to 2019 are mainly artificial neural network (2013–2019), optimization (2014–2019) and system (2014–2019).

Tümer & Edebalı (2015) used 4-month daily records from the Konya WWTP to simulate the Konya WWTP using multiple linear regression and artificial neural networks with different architectures in SPSS and MATLAB software. By comprehensively considering the input values of pH, temperature, chemical oxygen demand (COD), total suspended solids (TSS) and biochemical oxygen demand, the treatment efficiency of the device is determined, and it is proved that ANN is more satisfactory than the multiple linear regression model. Ozkan *et al.* (2009) used the 3-year daily input data of the same WWTP to define the output biochemical oxygen demand concentration of the Kaiseri WWTP to estimate the data through the training of a multi-layer artificial neural network model. The Levenberg–Marquardt algorithm is used to train the artificial neural network structure with five inputs and two hidden layers, and the best structure is obtained. Under this structure, the mean square error is 0.45 (MSE = 0.45), the average absolute error is 0.445,  $R^2 = 0.915$  and the effect is good.

Skoczko *et al.* (2017) studied the Bystre WWTP near Gitycko for 2 years and used the chemical parameters of sewage and the amount of sewage flowing into the facility as input variables, to establish an artificial neural network model to approximate the concentration and value of the basic quality parameters of sewage treatment. It was found that the concentration of TN and total phosphorus (TP) had the greatest influence on the variables and could best reflect the change of total suspended solids (TSM) in the treated sewage.

The key words that have appeared since 2019 are predictive control (2019–2021), adaptive regression spline (2019–2021), classification (2020–2021). It is easy to see that these words belong to the professional vocabulary in the field of ML. Due to the increasing concern about eutrophication, Ly *et al.* (2022) examined and compared six different ML algorithms from shallow to DL architectures and introduced the potential applications of ML to predict sewage quality. These models were developed to detect TP in the outlet, providing a reliable method for predicting sewage effluent quality.

As a statistical learning theory, SVM is suitable for small sample space classification. (Fang *et al.* 2019) proposed a geometric feature and environmental pollution factor advanced feature fusion algorithm based on multi-kernel function SVM,



**Figure 6** | Co-occurrence analysis of all key words in publications on the application of AI in wastewater. The gradient from blue to yellow means the time when the keyword appeared, with yellow representing the most recent popular keywords. The size of the circles is proportional to the quantity of articles published.

which improved the DL algorithm in the analysis of sewage membrane penetration. It requires a lot of calculation and training samples, but it shows unsatisfactory and unfavorable conditions in the case of a small sample field.

ML approach can also diagnose faults in WWT process. *Xu et al. (2018)* built an integrated classifier with a weighted limit learning machine as the base classified machine for the Bagging integrated framework. An improved Bagging composite WWT fault diagnosis modeling method based on a weighted limit learning machine was proposed to augment the accuracy of fault class recognitions.

### 3.7. Dual-map overlay

The double-figure overlay of the journal can illuminate the overall direction of AI research in the field of WWT from a macro perspective of the discipline. *Figure 7* can be divided into two parts. According to the analysis, the cited literature in this study is primarily concentrated in the fields of medicine, mathematics and chemistry. Studies in these fields are usually published in journals in the fields of computer chemistry and the environment, reflecting the strong relevance of journals from these fields in the areas of AI and WWT.

## 4. DISCUSSION

### 4.1. Reasons for the contribution differences in various countries

In terms of overall publication quantity, the top three countries are China, the US and Iran. Collecting training data is labor-intensive and time-consuming. As the most populous country, China has an advantaged role in the field of big data, which is crucial for the development of ML applications (*Sommerville et al. 2021*). China started late in AI, and lags behind some other developed countries at the basic levels. However, thanks to the vigorous support of national policies, the technology





Some of the top-ranked institutions do not have high centrality. Greater collaboration between agencies can improve the quality of their publications, thus greatly benefits this research area. The model is considered to be a formidable tool in the field of WWT. Activated sludge models and anaerobic digestion models have been used quite a bit since the beginning of the 21st century (Chen *et al.* 2022). Since 2012, AI has been used mostly in ANN algorithms for WWT, mainly for predicting wastewater parameters, simulating WWT processes and sensors (Sommerville *et al.* 2021). Although ANN is strong, it cannot replace the application of discharge mechanism models. In the last 2 years, ML has developed a standardized and applicable system of model parameter measurement methods for relevant mechanism models. Control methods which are applicable to different process control parameters in different WWT processes have been established, and eventually a combination of ANN control and model control has been found. This has greatly compensated for the shortcomings of ANN (El-Din *et al.* 2004). As of 2019, ML can be used to solve complex problems involving a large number of nonlinear processes or combinatorial spaces and can address typical problems in natural and engineered water systems (Hwang & Tu 2021). It could be predicted that using neural network ML algorithms to assess the pollutant removal in synthetic or real industrial wastewater using isotherms, kinetic and thermodynamic mechanisms models will become a research hotspot (Da Silva Medeiros *et al.* 2021).

#### 4.2. Research frontiers and hotspots

ML can make aggregate predictions of water resources and has advantages over traditional statistical tools. Nelson *et al.* (2021) developed a framework using ML to calculate monthly-resolution isotope time series using available climate and location data. The model's predictions can now be used for the past 70 years (1950–2019) at any location in Europe, to determine the source history of water or to understand the application of hydrological or meteorological processes that determine these values. This addresses that existing precipitation isotope models are often less accurate for examining features such as long-term trends or interannual variability.

Remote sensing as a driver of water management decisions in developing countries needs to be further integrated with water quality monitoring programs. Rahman *et al.* (2021) used a hybrid combination of locally weighted linear regression (LWLR), stochastic subspace (RSS), reduced error pruned tree (REP Tree), random forest (RF) and M5P model tree algorithms, to assess the probability of multiple types of floods across the country. Historical flood data (1988–2020), remotely sensed imagery (MODIS, Landsat 5–8 and Sentinel-1) as well as topographic, hydrogeological and environmental datasets were used to train and validate the proposed algorithms. Arias-Rodriguez *et al.* (2021) combined field measurements from the Mexican National Water Quality Monitoring System (RNMCA) (2013–2019) with Landsat-8 OLI, Sentinel-3 OLCI and Sentinel-2 MSI data to train an extreme learning machine (ELM), support-vector regression machine (SVR) and linear regression machine (LR) to estimate chlorophyll a (Chl-a), turbidity, TSM and transparency (SDD) efficacy assessments of available sensors to complement the often limited *in situ* measurements in such programs to build models that support monitoring tasks.

Water-related data collection and model interpretation rich and complex field water data have been collected for many years, including flow rates, temperatures, DO concentrations, turbidity and chlorine levels. Guo *et al.* (2022) used spatial autocorrelation, environmental kuznets curve (EKC) and logarithmic mean Divisia index model to study the spatial characteristics and driving factors of industrial wastewater discharge. As these data are often incomplete or weakly correlated, it is hard to capture the time-varying or nonlinear behavior of dynamic water systems using traditional statistics-based models. Chys *et al.* (2017) proposed a novel robust correlation real-time monitoring and control model for trace organic pollutants (Troc) removal by ozonation based on UVA and fluorescent substitutes, and the kinetic information was taken into account. ML models are better able to cope with rapidly changing conditions. Moghadam *et al.* (2021) chose Fanno Creek (Oregon, USA) to make a case study. Daily values of water temperature, specific conductivity, stream discharge, pH and DO concentration were used as input variables for a study describing the implementation of a DL approach applied to a recurrent neural network (RNN) algorithm. Ji *et al.* (2017) developed a promising AI model, known as SVM, to predict the DO concentration in a hypoxic river in southeastern China. Four different calibration models, being multiple linear regression, back propagation neural network, general regression neural network and SVM, were established. Meanwhile, their prediction accuracy was systemically investigated and compared. The result proposed that SVM model can effectively predict water quality, especially for highly impaired and anoxic river systems.

Most WWTPs use activated sludge and other processes to remove pollutants (organic carbon, nitrogen, phosphorus). Tens of thousands of different microbial species may be present in each wastewater system. Deterministic models based on the



biokinetics of the activated sludge process are not particularly practical due to the complexity of the biological reactions. The variability of the treatment plant and ML techniques can predict sludge expansion in WWTPs with much higher accuracy without the burden of calibration. *Su et al. (2022)* studied neural network and convolutional neural network (CNN) algorithms in DL and then built a target detection system. Finally, the treated wastewater was detected and the comparison results with the traditional target detection system showed that the target detection system of the CNN algorithm was more stable in identifying the treated wastewater than the target detection system of the neural network algorithm. *Zhao et al. (2022)* developed an ML model to predict pH, total nitrogen (TN), total organic carbon (TOC) and total phosphorus (TP) for WWTPs. The results show that gradient-boosted decision trees (mean test  $R$  of 0.85–0.96) can accurately predict the above wastewater characteristics for both single and multi-objective models. *Li et al. (2022)* proposed a new hybrid modeling approach to predict NO emissions by integrating the first master model and DL techniques. The results show that the hybrid model has higher accuracy in modeling NO emissions from WWTPs compared to mechanical or pure DL models. Furthermore, the hybrid model is more applicable than a pure DL model with a pure mechanical model due to lower data requirements.

The rapid advancement of DL, especially in the last 5 years, provides a wonderful opportunity to build automated pipeline defect detection systems through image recognition. Accumulation of fat, oil and grease (FOG) in the sumps of wastewater pumping stations is a common failure cause for these facilities. *Moreno-Rodenas et al. (2021)* present a low-cost camera-based automated system for the observation of FOG layer dynamics in wastewater pumping stations at high-frequency over extended time windows. Optical imagery is processed through a deep-learning computer vision routine that allows describing FOG layer dynamics and various hydraulic processes in the pump sump. When deployed at remote locations, compressed processed datasets can be transmitted (edge AI computing), which could be very useful to the hydro-ecological monitoring community. *Pham et al. (2022)* proposed and implemented a novel DL approach that outperformed random forest and SVM methods for monitoring various ecosystems in estuaries. By integrating spatial and background paths into a novel bilateral segmentation network (BiSeNet), the processing speed and accuracy of ordinary neural networks are enhanced by more than 10 times.

## 5. CONCLUSIONS

An objective, systematic and comprehensive bibliometric analysis of WWT scientific research related to ML, DL and ANN, which belong to AI, was conducted. Between 2012 and 2022, the number of papers showed an upward trend, with a significant increase in the number of papers published after 2018. China ( $n = 210$ , 28.07%) contributed the most publications. Spain, Turkey and Italy are the first countries which started the research on using AI in WWT in 2019. Duy Tan University (Vietnam) is the most cited university ( $n = 743$ ) and Tehran University (Iran) has the highest number, being 19, of published papers. Nasr M and Pourghasemi HR published the most articles, being 7. Among these, the most cited articles focus on the use of AI and metal insulator metal (MIM) in improving factory performance, as well as on the construction of predictive models of wastewater concentration by using ML-based ANN-based methods. *WATER* is the most published journal.

Major research trends at this stage include the combination of ML or DL modeling with remote sensing techniques to predict flooding or groundwater sensitivity in specific areas, implementing controls and building data-driven predictive models. Neural networks are used to predict the effectiveness of treatment for difficult-to-degrade wastewater, to build models for different pollutants, to select dominant algal species with high nutrient value and effectiveness.

## FUNDING

This work is funded by Foundation of Shanghai Science and Technology Commission (20ZR1449700), National Natural Science Foundation of China (51678353, 52070127), Chunhui Program of Ministry of Education of China (HZKY20220059), Cooperation project between China and Central and Eastern European Countries (2022192), Cooperation project from Shanghai Chimbusco Marine Bunker Co., Ltd (C80ZH236014), and Open Fund of Anhui International Joint Research Center for Nano Carbon-based Materials and Environmental Health (NCMEH2022Y02).

## 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

- Abdeldayem, O. M., Dabbish, A. M., Habashy, M. M., Mostafa, M. K., Elhefnawy, M., Amin, L., Al-Sakkari, E. G., Ragab, A. & Rene, E. R. 2022 Viral outbreaks detection and surveillance using wastewater-based epidemiology, viral air sampling, and machine learning techniques: a comprehensive review and outlook. *Science of the Total Environment* **803**, 149834.
- Alalm, M. G. & Nasr, M. 2018 Artificial intelligence, regression model, and cost estimation for removal of chlorothalonil pesticide by activated carbon prepared from casuarina charcoal. *Sustainable Environment Research* **28**, 101–110.
- Altowayti, W. A. H., Shahir, S., Othman, N., Eisa, T. A. E., Yafooz, W. M., Al-Dhaqm, A., Soon, C. Y., Yahya, I. B., Che Rahim, N. A. N. B. & Abaker, M. 2022 The role of conventional methods and artificial intelligence in the wastewater treatment: a comprehensive review. *Processes* **10**, 1832.
- Ansari, F. A., Gupta, S. K., Nasr, M., Rawat, I. & Bux, F. 2018 Evaluation of various cell drying and disruption techniques for sustainable metabolite extractions from microalgae grown in wastewater: a multivariate approach. *Journal of Cleaner Production* **182**, 634–643.
- Arias-Rodriguez, L., Duan, Z., Díaz-Torres, J., Basilio Hazas, M., Huang, J., Kumar, B., Tuo, Y. & Disse, M. 2021 Integration of remote sensing and Mexican water quality monitoring system using an extreme learning machine. *Sensors* **21**, 4118.
- Bagheri, M., Mirbagheri, S. A., Bagheri, Z. & Kamarkhani, A. M. 2015 Modeling and optimization of activated sludge bulking for a real wastewater treatment plant using hybrid artificial neural networks-genetic algorithm approach. *Process Safety and Environmental Protection* **95**, 12–25.
- Carmona, B. & Abejón, R. 2023 Innovative membrane technologies for the treatment of wastewater polluted with heavy metals: perspective of the potential of electrodialysis, membrane distillation, and forward osmosis from a bibliometric analysis. *Membranes* **13**, 385.
- Chen, T. & Chen, C. 2020 Hierarchical intelligence platform designed for wastewater management systems: information technology integration. *Desalination and Water Treatment* **183**, 167–172.
- Chen, T. Q. & Schmidt, M. 2016 Fast patch-based style transfer of arbitrary style. arXiv preprint arXiv:1612.04337.
- Chen, W., Pourghasemi, H. R. & Naghibi, S. A. 2018 A comparative study of landslide susceptibility maps produced using support vector machine with different kernel functions and entropy data mining models in China. *Bulletin of Engineering Geology and the Environment* **77**, 647–664.
- Chen, X., Chen, H., Yang, L., Wei, W. & Ni, B.-J. 2022 A comprehensive analysis of evolution and underlying connections of water research themes in the 21st century. *Science of the Total Environment* **835**, 155411.
- Choubin, B., Abdolshahnejad, M., Moradi, E., Querol, X., Mosavi, A., Shamshirband, S. & Ghamisi, P. 2020 Spatial hazard assessment of the PM10 using machine learning models in Barcelona, Spain. *Science of the Total Environment* **701**, 134474.
- Choubin, B., Hosseini, F. S., Rahmati, O. & Youshanloei, M. M. 2023 A step toward considering the return period in flood spatial modeling. *Natural Hazards* **115**, 431–440.
- Chys, M., Audenaert, W., Deniere, E., Mortier, S., Van Langenhove, H., Nopens, I., Demeestere, K. & Van Hulle, S. 2017 Surrogate-based correlation models in view of real-time control of ozonation of secondary treated municipal wastewater-model development and dynamic validation. *Environmental Science & Technology* **51** (24), 14233–14243.
- Da Silva Medeiros, D. C. C., Nzediegwu, C., Benally, C., Messele, S. A., Kwak, J.-H., Naeth, M. A., Ok, Y. S., Chang, S. X. & El-Din, M. G. 2021 Pristine and engineered biochar for the removal of contaminants co-existing in several types of industrial wastewaters: a critical review. *Science of the Total Environment* **809**, 151120.
- Guo, H., Zhang, Y. & Yang, Z. 2022 Quantification of industrial wastewater discharge from the major cities in Sichuan province, China. *Environmental Science and Pollution Research international* **29** (34), 51567–51577.
- El-Din, A. G., Smith, D. W. & El-Din, M. G. 2004 Application of artificial neural networks in wastewater treatment. *Journal of Environmental Engineering and Science* **3**, S81–S95.
- Elsley-Quirk, T., Watson, E., Raper, K., Kreeger, D., Paudel, B., Haaf, L., Maxwell-Doyle, M., Padeletti, A., Reilly, E. & Velinsky, D. 2022 Relationships between ecosystem properties and sea-level rise vulnerability of tidal wetlands of the U.S. Mid-Atlantic. *Environmental Monitoring and Assessment* **194**, 292.
- Faherty, A. 2021 Tapped out: how Newark, New Jersey's lead drinking water crisis illuminates the inadequacy of the federal drinking water regulatory scheme and fuels environmental injustice throughout the nation. *Environmental Claims Journal* **33**, 304–327.
- Fang, A., Cui, L. & Zhang, Z. 2019 Multi-kernel SVM based multi-feature fusion algorithm for analyzing wastewater membrane permeation in environmental pollution. *Desalination and Water Treatment* **163**, 415–420.
- Fuente, M. & Represa, C. 1997 A comparative study of neural networks based approach for fault detection. *IFAC Proceedings Volumes* **30**, 471–476.
- Guo, H., Jeong, K., Lim, J., Jo, J., Kim, Y. M., Park, J.-P., Kim, J. H. & Cho, K. H. 2015 Prediction of effluent concentration in a wastewater treatment plant using machine learning models. *Journal of Environmental Sciences* **32**, 90–101.
- 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**, 2093–2101.

- Hernández-del-Olmo, F., Gaudioso, E., Duro, N. & Dormido, R. 2019 Machine learning weather soft-sensor for advanced control of wastewater treatment plants. *Sensors* **19**, 3139.
- Hwang, G.-J. & Tu, Y.-F. 2021 Roles and research trends of artificial intelligence in mathematics education: a bibliometric mapping analysis and systematic review. *Mathematics* **9**, 584.
- Ji, X., Shang, X., Dahlgren, R. & Zhang, M. 2017 Prediction of dissolved oxygen concentration in hypoxic river systems using support vector machine: a case study of Wen-Rui Tang river, China. *Environmental Science and Pollution Research* **24**, 16062–16076.
- Kamiliya, T., Gautam, R., Muthukumaran, S., Navaratna, D. & Mondal, S. 2022 Technical advances on current research trends and explore the future scope on nutrient recovery from waste-streams: a review and bibliometric analysis from 2000 to 2020. *Environmental Science and Pollution Research* **29** (33), 49632–49650.
- Kim, M., Kim, Y., Kim, H., Piao, W. & Kim, C. 2016 Evaluation of the k-nearest neighbor method for forecasting the influent characteristics of wastewater treatment plant. *Frontiers of Environmental Science & Engineering* **10**, 299–310.
- Li, J. & Chen, C. 2016 *Citespace: Text Mining and Visualization in Scientific Literature*. Capital University of Economics & Business Press, Beijing, China, pp. 149–152.
- Li, K., Duan, H., Liu, L., Qiu, R., van den Akker, B., Ni, B., Chen, T., Yin, H., Yuan, Z. & Ye, L. 2022 An integrated first principal and deep learning approach for modeling nitrous oxide emissions from wastewater treatment plants. *Environmental Science & Technology* **56**, 2816–2826.
- Lowe, M., Qin, R. & Mao, X. 2022 A review on machine learning, artificial intelligence, and smart technology in water treatment and monitoring. *Water* **14**, 1384.
- Ly, Q. V., Truong, V. H., Ji, B., Nguyen, X. C., Cho, K. H., Ngo, H. H. & Zhang, Z. 2022 Exploring potential machine learning application based on big data for prediction of wastewater quality from different full-scale wastewater treatment plants. *Science of the Total Environment* **832**, 154930.
- Matheri, A. N., Mohamed, B., Ntuli, F., Ngila, J. C. & Zvinowanda, C. 2022 Computational modelling of geochemical speciation of the trace metals in the wastewater treatment process optimization. *Applied Water Science* **12**, 1–12.
- Moghadam, S., Sharafati, A., Feizi, H., Marjaie, S., Asadollah, S. & Motta, D. 2021 An efficient strategy for predicting river dissolved oxygen concentration: application of deep recurrent neural network model. *Environmental Monitoring and Assessment* **193**, 798.
- Moreno-Rodenas, A., Duinmeijer, A. & Clemens, F. 2021 Deep-learning based monitoring of fog layer dynamics in wastewater pumping stations. *Water Research* **202**, 117482.
- Mosavi, A., Sajedi Hosseini, F., Choubin, B., Taromideh, F., Ghodsi, M., Nazari, B. & Dineva, A. A. 2021 Susceptibility mapping of groundwater salinity using machine learning models. *Environmental Science and Pollution Research* **28**, 10804–10817.
- Mthethwa, N. P., Nasr, M., Bux, F. & Kumari, S. 2018 Utilization of *Pistia stratiotes* (aquatic weed) for fermentative biohydrogen: electron-equivalent balance, stoichiometry, and cost estimation. *International Journal of Hydrogen Energy* **43**, 8243–8255.
- Nadiri, A. A., Shokri, S., Tsai, F. T.-C. & Moghaddam, A. A. 2018 Prediction of effluent quality parameters of a wastewater treatment plant using a supervised committee fuzzy logic model. *Journal of Cleaner Production* **180**, 539–549.
- Naghbi, S. A., Pourghasemi, H. R. & Dixon, B. 2016 Gis-based groundwater potential mapping using boosted regression tree, classification and regression tree, and random forest machine learning models in Iran. *Environmental Monitoring and Assessment* **188**, 1–27.
- Nelson, D., Basler, D. & Kahmen, A. 2021 Precipitation isotope time series predictions from machine learning applied in Europe. *Proceedings of the National Academy of Sciences of the United States of America* **118**, e2024107118.
- Ola, A., Heryanto, H., Armynah, B. & Tahir, D. 2023 Bibliometric analysis of chitosan research for wastewater treatment: a review. *Environmental Monitoring and Assessment* **195** (4), 474.
- Ozkan, O., Ozdemir, O. & Azgin, S. 2009 Prediction of biochemical oxygen demand in a wastewater treatment plant by artificial neural networks. *Asian Journal of Chemistry* **21**, 4821–4830.
- Pham, H., Dang, K., Nguyen, T., Tran, N., Ngo, X., Nguyen, D., Phan, T., Nguyen, T., Guo, W. & Ngo, H. 2022 A new deep learning approach based on bilateral semantic segmentation models for sustainable estuarine wetland ecosystem management. *Science of the Total Environment* **838**, 155826.
- Pourghasemi, H. R. & Rahmati, O. 2018 Prediction of the landslide susceptibility: which algorithm, which precision? *Catena* **162**, 177–192.
- Rahman, M., Chen, N., Elbeltagi, A., Islam, M., Alam, M., Pourghasemi, H., Tao, W., Zhang, J., Shufeng, T., Faiz, H., Baig, M. & Dewan, A. 2021 Application of stacking hybrid machine learning algorithms in delineating multi-type flooding in Bangladesh. *Journal of Environmental Management* **295**, 113086.
- Rios-Miguel, A., Jhm van Bergen, T., Zillien, C., Mj Ragas, A., van Zelm, R., Sm Jetten, M., Jan Hendriks, A. & Welte, C. 2023 Predicting and improving the microbial removal of organic micropollutants during wastewater treatment: a review. *Chemosphere* **333**, 138908.
- Seo, G., Yoon, S., Kim, M., Mun, C. & Hwang, E. 2021 Deep reinforcement learning-based smart joint control scheme for on/off pumping systems in wastewater treatment plants. *IEEE Access* **9**, 95360–95371.
- Skoczko, I., Struk-Sokolowska, J. & Ofman, P. 2017 Modelling changes in the parameters of treated sewage using artificial neural networks. *Rocznik Ochrona Srodowiska* **19**, 633–650.
- Sommerville, R., Zhu, P., Rajaeifar, M. A., Heidrich, O., Goodship, V. & Kendrick, E. 2021 A qualitative assessment of lithium ion battery recycling processes. *Resources, Conservation and Recycling* **165**, 105219.
- Su, B., Lin, Y., Wang, J., Quan, X., Chang, Z. & Rui, C. 2022 Deep learning target detection system for sewage treatment. *Computational Intelligence and Neuroscience* **2022**, 2743781.

- Taromideh, F., Fazloula, R., Choubin, B., Emadi, A. & Berndtsson, R. 2022 Urban flood-risk assessment: integration of decision-making and machine learning. *Sustainability* **14** (8), 4483.
- Torregrossa, D., Leopold, U., Hernández-Sancho, F. & Hansen, J. 2018 Machine learning for energy cost modelling in wastewater treatment plants. *Journal of Environmental Management* **223**, 1061–1067.
- Tümer, A. E. & Edebal, S. 2015 Prediction of wastewater treatment plant performance using multilinear regression and artificial neural networks. In: *2015 International Symposium on Innovations in Intelligent Systems and Applications (INISTA)*. IEEE, p. 15278013.
- Xu, Y., Lai, C. & Luo, F. 2018 Bagging ensemble fault diagnosis modeling with imbalanced classification in wastewater treatment plant. *Journal of South China University of Technology* **46** (8), 107–115.
- Xu, D., Liu, B., Wang, J. & Zhang, Z. 2022 Bibliometric analysis of artificial intelligence for biotechnology and applied microbiology: exploring research hotspots and frontiers. *Frontiers in Bioengineering and Biotechnology* **10**, 998298.
- 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**, 101388.
- Youssef, A. M., Mahdi, A. M. & Pourghasemi, H. R. 2022 Optimal flood susceptibility model based on performance comparisons of LR, EGB, and RF algorithms. *Natural Hazards* **115**, 1071–1096.
- Zhang, S., Wang, H. & Keller, A. A. 2021 Novel machine learning-based energy consumption model of wastewater treatment plants. *ACS ES&T Water* **1**, 2531–2540.
- Zhang, Y., Ni, X. & Wang, H. 2023 Visual analysis of greenhouse gas emissions from sewage treatment plants based on citeSpace: from the perspective of bibliometrics. *Environmental Science and Pollution Research* **30** (16), 45555–45569.
- Zhao, G., Suzuki, S., Deng, J. & Fujita, M. 2022 Machine learning estimation of biodegradable organic matter concentrations in municipal wastewater. *Journal of Environmental Management* **323**, 116191.
- Zhu, S. & Piotrowski, A. P. 2020 River/stream water temperature forecasting using artificial intelligence models: a systematic review. *Acta Geophysica* **68**, 1433–1442.
- Zhu, J.-J., Dressel, W., Pacion, K. & Ren, Z. J. 2021 ES&t in the 21st century: a data-driven analysis of research topics, interconnections, and trends in the past 20 years. *Environmental Science & Technology* **55**, 3453–3464.

First received 21 May 2023; accepted in revised form 11 September 2023. Available online 22 September 2023