

Article

Artificial Intelligence and Wastewater Treatment: A Global Scientific Perspective through Text Mining

Abdelhafid El Alaoui El Fels ¹, Laila Mandi ^{1,2,*}, Aya Kammoun ^{1,2}, Naaila Ouazzani ^{1,2}, Olivier Monga ^{3,4} and Moulay Lhassan Hbid ⁴

¹ National Center for Studies and Research on Water and Energy (CNEREE), Cadi Ayyad University, P.O. Box 511, Marrakech 40000, Morocco; elalaoui.abdelhafid@gmail.com (A.E.A.E.F.); kamouneaya2@gmail.com (A.K.); ouazzani@uca.ac.ma (N.O.)

² Laboratory of Water, Biodiversity and Climate Change (EauBiodiCc), Faculty of Sciences Semlalia, Cadi Ayyad University, P.O. Box 2390, Marrakech 40000, Morocco

³ Institut de Recherche pour le Développement (IRD), Unité de Modélisation Mathématique et Informatique des Systèmes Complexes (UMMISCO), F-93143 Bondy, France; olivier.monga@ird.fr

⁴ Department of Mathematics, UMMISCO, Faculty of Sciences Semlalia, Cadi Ayyad University, P.O. Box 2390, Marrakech 40000, Morocco; hbid@uca.ac.ma

* Correspondence: mandi@uca.ma

Abstract: The concept of using wastewater as a substitute for limited water resources and environmental protection has enabled this sector to make major technological advancements and, as a result, has given us an abundance of physical data, including chemical, biological, and microbiological information. It is easier to comprehend wastewater treatment systems after studying this data. In order to achieve this, a number of studies use machine learning (ML) algorithms as a proactive approach to solving issues and modeling the functionalities of these processing systems while utilizing the experimental data gathered. The goal of this article is to use textual analysis techniques to extract the most popular machine learning models from scientific documents in the “Web of Science” database and analyze their relevance and historical development. This will help provide a general overview and global scientific follow-up of publications dealing with the application of artificial intelligence (AI) to overcome the challenges faced in wastewater treatment technologies. The findings suggest that developed countries are the major publishers of articles on this research topic, and an analysis of the publication trend reveals an exponential rise in numbers, reflecting the scientific community’s interest in the subject. As well, the results indicate that supervised learning is popular among researchers, with the Artificial Neural Network (ANN), Random Forest (RF), Support Vector Machine (SVM), Linear Regression (LR), Adaptive Neuro-Fuzzy Inference System (ANFIS), Decision Tree (DT), and Gradient Boosting (GB) being the machine learning models most frequently employed in the wastewater treatment domain. Research on optimization methods reveals that the most well-known method for calibrating models is genetic algorithms (GA). Finally, machine learning benefits wastewater treatment by enhancing data analysis accuracy and efficiency. Yet challenges arise as model training demands ample, high-quality data. Moreover, the limited interpretability of machine learning models complicates comprehension of the underlying mechanisms and decisions in wastewater treatment.

Keywords: wastewater treatment; machine learning; textual analysis; artificial intelligence; optimization methods



Citation: El Alaoui El Fels, A.; Mandi, L.; Kammoun, A.; Ouazzani, N.; Monga, O.; Hbid, M.L. Artificial Intelligence and Wastewater Treatment: A Global Scientific Perspective through Text Mining. *Water* **2023**, *15*, 3487. <https://doi.org/10.3390/w15193487>

Academic Editors: Hong Chen and Antonio Panico

Received: 12 July 2023

Revised: 23 August 2023

Accepted: 7 September 2023

Published: 5 October 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Machine learning (ML) is a buzzword these days. Clearly, machine learning is a branch of artificial intelligence (AI). It is a technology aimed at producing many powerful and practical tools for modeling complex problems in various fields and simplifying difficult systems into manageable applications. In the literature, several researchers in diverse fields

have used this technology because of its flexibility of use, speed of operation, and allowable accuracy without the need to understand the behavior of natural phenomena or physical systems [1,2].

In geoscience, machine learning is used in geology for lithological discrimination of facies [3–5] and in hydrology for the prediction of extreme events and calibration of satellite-based precipitation products [6,7]. In the financial sector, AI is a hot topic that has attracted several authors to use it in risk management and financial services [8–11]. In healthcare, intelligence has taken its place in several areas: in medicine, and more specifically in the medical field, machine learning algorithms are heavily involved [12–16], in epidemiological behavior analysis [17,18], and also in the pharmaceutical field [19–21].

ML has been used in various environmental disciplines due to its ability to simulate the behavior of aquatic systems, such as wastewater treatment plants [22,23]. Wastewater treatment plays a key role in the economy and conservation of water resources, as well as in the environmental protection of aquatic environments. The complication of natural systems and the diversity of wastewater treatment conditions lead to a variation in the treatment behavior of the water and a fluctuation in operating costs.

The variety of wastewater, the characteristics of the tributaries, the concentrations of pollutants from multiple sources, and the geospatial and climatic conditions vary considerably from one treatment plant to another. These factors make wastewater treatment a complex process governed by a strong oscillation of chemical, physical, and microbiological parameters. The stochastic notion of natural phenomena coupled with anthropogenic activities leads to a multitude of uncertainties in the wastewater treatment process [24]. This complexity leads many scientists to search for simple approaches to understand and control the behavior of these systems through mathematical models in the future. There are several models in the literature that allow the separate modeling of chemical reactions, physical processes, and biological behavior, as well as other processes supported in the treatment system. This variety in the typology of the models requires the coupling of the latter in order to have complex global modeling. The complexity of this modeling has led researchers to call another type of model into the category of ML capable of encompassing all these processes in a structure, namely machine learning models [25].

Overall, there are a multitude of applications of machine learning in wastewater treatment that have proven the potential and effectiveness of these approaches in confronting problems linked to wastewater treatment [26,27].

One of the important applications of machine learning in wastewater treatment is improving the operating efficiency of treatment processes by optimizing the energy consumption of a wastewater treatment plant [28]. Machine learning can also be used to optimize treatment processes by adjusting them based on the prediction of wastewater quality [29]. Another area where machine learning can be applied to water treatment is wastewater reuse. Based on the analysis of wastewater characteristics by machine learning models, optimal conditions for wastewater reuse can be identified. Konstantina et al. [30] showed the feasibility of using machine learning for optimal irrigation purposes. Machine learning can also be used to discover the performance of wastewater treatment processes, such as biological nutrient removal, with a high degree of accuracy [31,32]. These models can be used in the modification of processing processes by identifying optimal conditions for maximum performance while minimizing the use of resources.

On the other hand, it is essential to highlight that the preprocessing of big data holds crucial significance in the field of artificial intelligence applied to wastewater treatment. This indispensable and frequently required step plays a pivotal role in ensuring the accuracy and reliability of analyses and results obtained through machine learning models [33,34]. Therefore, it is imperative to devote special attention to this essential subject in order to ensure the quality and relevance of the conclusions drawn from these complex datasets.

Ultimately, there are many uses for machine learning in wastewater treatment, and each one has a different goal depending on the issues it encounters and the goals it sets out to achieve. A variety of ML models are emerging as a result of the intense research being

done on the use of artificial intelligence methods. The goal of this work is to apply textual analysis techniques to extract the information required for using ML models in the field of wastewater treatment and to provide a general overview of the most popular models and algorithms currently in use. The bibliometric and textual analysis of the Web of Science database served as the foundation for this study.

2. Material and Methods

The sheer number of scientific documents makes it difficult to search for specific topics. The Web of Science platform provides filters that focus on the target documents based on keywords or author names. However, this platform remains insufficient to extract the maximum amount of information. The application of statistical and data analysis approaches specific to the processing of large volumes of textual information allows the development of decision-support tools.

The methodology used in this paper (Figure 1) is based on the import of data from the Web of Science platform while using keywords (Wastewater, Machine Learning, Artificial Intelligence) to filter our scientific niche. The obtained data are of a semi-structured type. They can be used directly to perform bibliometric analyses or processed through text mining methods to transform them into structured data and then draw the desired relevant information, such as the geographical distribution of the target publications, the most commonly used models, the chronological timeline of the topic selected, the word clouds, etc.

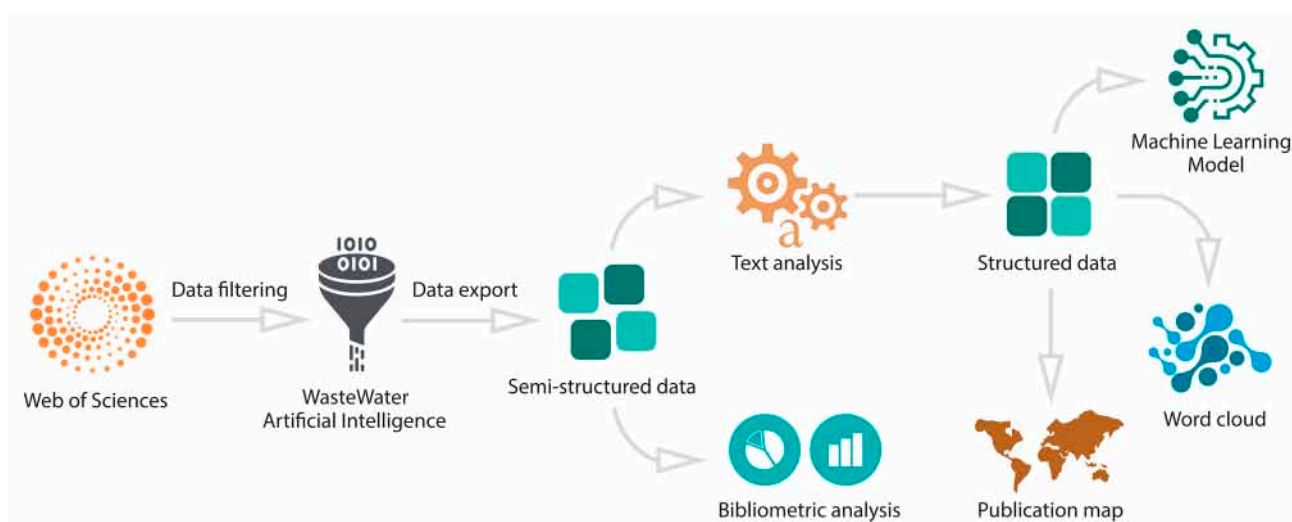


Figure 1. Schema of the Web of Science database processing and analysis methodology.

2.1. Database: Web of Science

The data used in this document is extracted from the world's most credible global citation database, Web of Science. The Web of Science platform is considered the most robust search engine, providing a library of the best publication and citation data, classified according to several categorical criteria to provide browsers with flexible access, discovery, and evaluation with confidence. The Web of Science platform is multi-disciplinary and tracks temporal developments in science across all disciplines from over 1.7 billion cited references in over 159 million records. The credibility of this platform is built on the trust provided by over 9000 academics, corporate and institute institutions, and millions of researchers to the Web of Science to produce well-regarded research in the scientific community, obtain the necessary information, and make clear decisions that guide the future of their institution and their research strategy.

2.2. Text Mining

Text mining, or text analytics, is a set of techniques and tools for mining unstructured or semi-structured documents. It is the process of extracting ideas and information from a textual dataset. Text mining is used to help researchers discover hidden relationships and patterns between words and find relevant information from the textual content.

This content can be in the form of a Word document, publications on social networks such as Twitter, libraries in the form of semi-structured Excel databases, as in our case in this article, or others. The text mining technique allows us to schematize the links between words and also to show, in a visual representation, the clouds of the most frequently used words in a textual database.

3. Results and Discussion

3.1. Publishing Trend

The trend in scientific production is shown in Figure 2. In total, more than 600 papers have been published on the topic of wastewater treatment by numerical methods, proving the importance of this research field. From 1991 to 2016, the annual number of publications on this topic was extremely low, with few publications per year. Meanwhile, a significant increase in this field's publications has occurred from 2016 to 2020. However, from 2020 up until now, the findings showed an exponential increase in the annual number of publications, with a maximum number of 198 reached in 2022. The analysis of the publication trend prior to 2016 revealed that the focus on the application of numerical methods and Machine Learning models to solve the problems faced in wastewater treatment was not widely given so much importance by the world's researchers. The increase in scientific output after 2017 is strongly related to the evolution of science, the invention of new numerical and statistical approaches, and the progress in the performance of computing machines, data processing, and analysis software and programs.

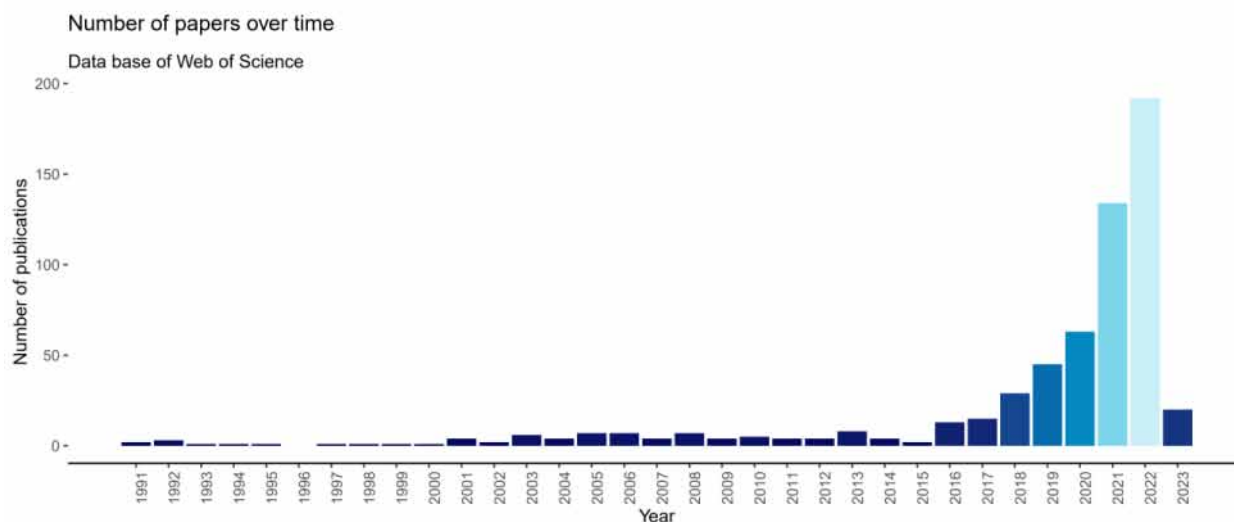


Figure 2. Temporal trend in scientific literature that involves the use of artificial intelligence as an approach to solving problems faced in wastewater treatment.

3.2. Publications in the World

The distribution map of the global paper network was generated in order to thoroughly examine how scientific articles are dispersed throughout the world based on an examination of author and co-author addresses and international scientific cooperation (Figure 3). The map revealed that 81 countries have published at least one paper about the application of numerical modeling to wastewater treatment, with a total exceeding 605 papers in the last 30 years. The number of papers is unusual among countries; it was much higher in developed countries than in developing countries. For the past 30 years, all continents,

with the exception of Antarctica, have published papers on the topic of applying artificial intelligence to wastewater treatment, demonstrating its popularity globally. Moreover, Figure 3 indicates that the USA, Canada, Spain, China, India, Saudi Arabia, Iran, Egypt, Turkey, and Australia have published more papers than others.

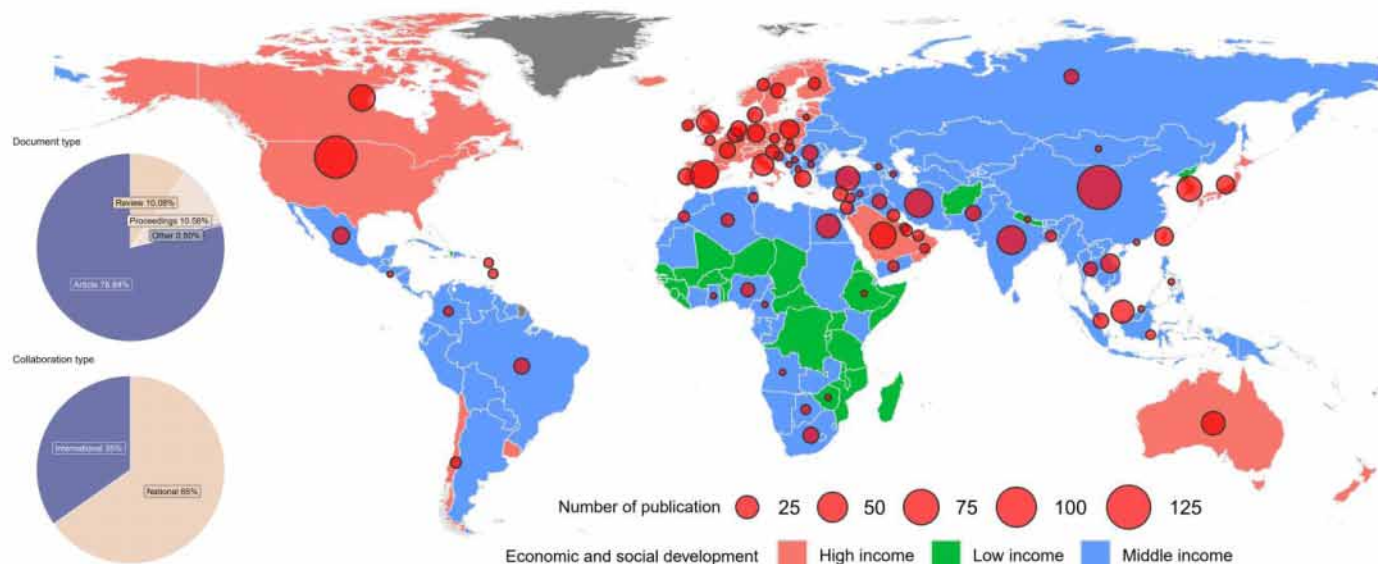


Figure 3. Worldwide distribution of scientific papers involving the application of artificial intelligence to wastewater treatment issues.

Further analysis revealed that the number of international and national collaborative papers was 35% and 65%, respectively. These results depict that the international collaboration was widely known for its application of machine learning algorithms in the wastewater treatment domain.

Regarding the types of papers, a total of 605 publications published between 1991 and 2023 are distributed over 8 types of papers. Articles (477) accounted for 78.84% of the total contributions, followed by proceedings (10.58%) and review documents (10.08%) while the remaining documents, including database reviews and editorial content, were less expressive (0.5%).

Concerning the linguistic side of the documents, the analysis shows that the overwhelming majority of the publications are published in English, representing 99.7% of the total publications, while the other languages were less used, by a minimal percentage of 0.03%.

3.3. Collaborations

In order to accelerate scientific progress, provide a wide range of perspectives that can lead to innovative solutions, and accomplish a common goal in scientific research, international collaborations within a particular discipline or across many scientific fields between at least two scientists from different nations are crucial.

Figure 4 depicts the cooperative relationships among the countries in each region. Each country is represented by a segment, whose size reflects the number of publications it has produced. A cooperative relationship between two linked countries is symbolized by a curve connecting the country segments, with a thicker curve indicating a stronger partnership. Overall, the findings showed that the USA, Canada, Spain, China, India, Saudi Arabia, Iran, Egypt, and Turkey had the strongest cooperation, the highest number of publications worldwide, and clearly promoted scientific production in the application of machine learning models in the wastewater field. Regarding the scientific collaboration in each region, the East Asia and Pacific region had the largest number of cooperating countries, compared to the Sub-Saharan Africa, Latin America, and Caribbean regions,

which had the lowest number of countries. This difference could therefore be explained by the collaborative network, the level of development, the scientific interest of each country, and the number of countries in each region.

A cooperative relationship between two linked countries is symbolized by a curve connecting the country segments, with a thicker curve indicating a stronger partnership.

Furthermore, the spatial pattern of international collaboration is dominated by the following countries: in the Islamic world, by Saudi Arabia, Iran, Pakistan, and Egypt; in Asia, by China and India; on the American continent, by the United States and Canada; and in Europe, by Spain and England. Meanwhile, sub-Saharan and Latin American countries are characterized by a low rate of collaboration with other regions of the world.

In general, there are various areas where scientific cooperation is crucial when applying machine learning models to wastewater:

- Knowledge exchange: it allows researchers to share their knowledge and expertise, allowing them to learn from each other, access tools and information not otherwise available, and thus promote the development of the field in question;
- Higher research quality: collaboration leads to a more rigorous and objective research process. Peer review and collaboration help to ensure the reliability, reproducibility, and accuracy of results. Moreover, it allows researchers to validate their results by replicating them independently, thereby enhancing the evidence base;
- Increased funding opportunities: funding agencies are often more willing to support collaborative research because it is perceived as more impactful and innovative;
- Resources accessibility: it affords researchers access to a wider range of resources, such as specialized equipment, databases, and research facilities. This access can be particularly beneficial for early-career researchers who may not have the resources to conduct certain types of research;
- Diverse perspectives and innovation: the emergence of new study fields, new methodologies, and innovative approaches requires the collaboration of researchers from different disciplines and backgrounds, bringing different perspectives and ideas to the table;
- Addressing environmental and climate challenges: A call for interdisciplinary and international cooperation is crucial in order to address challenges facing humanity, including climate change, water scarcity, and pollution.

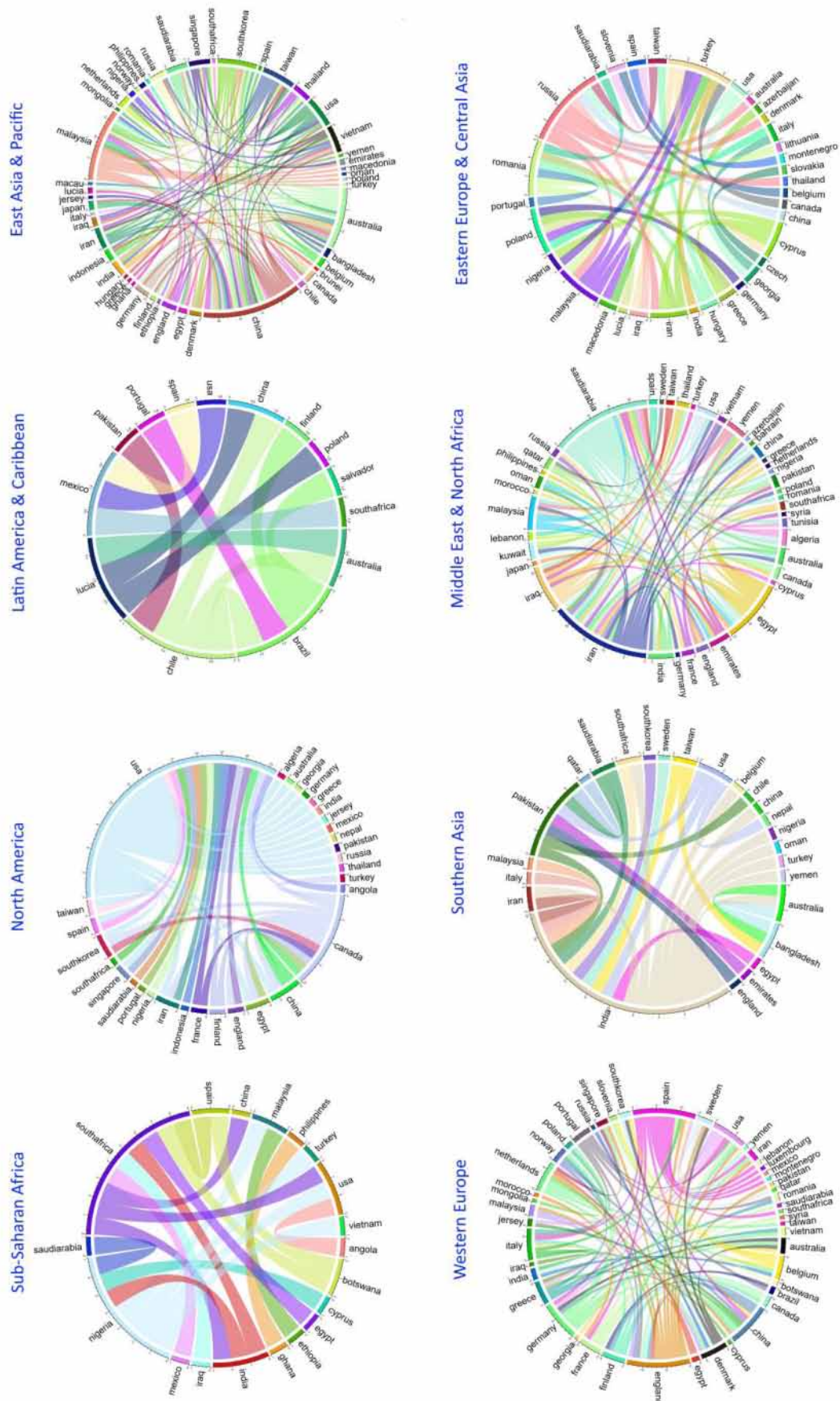


Figure 4. Network visualization for country collaboration by world region.

3.4. Publishing Journals

The published papers on the topic of the application of artificial intelligence in wastewater treatment came from 261 different journals. The analysis of the cumulative number of articles published in each journal showed that 5.34 percent of the journals have had fewer than 10 research papers published in the last three decades. The top 30 publishing journals are presented in Figure 5a. These journals published almost a total of 245 papers, which represents 40.49% of the total publications during the period 1991–2023, with an average of 12. This mean value appeared to be quite low, indicating that the application of artificial intelligence in the wastewater treatment domain is new and trending.

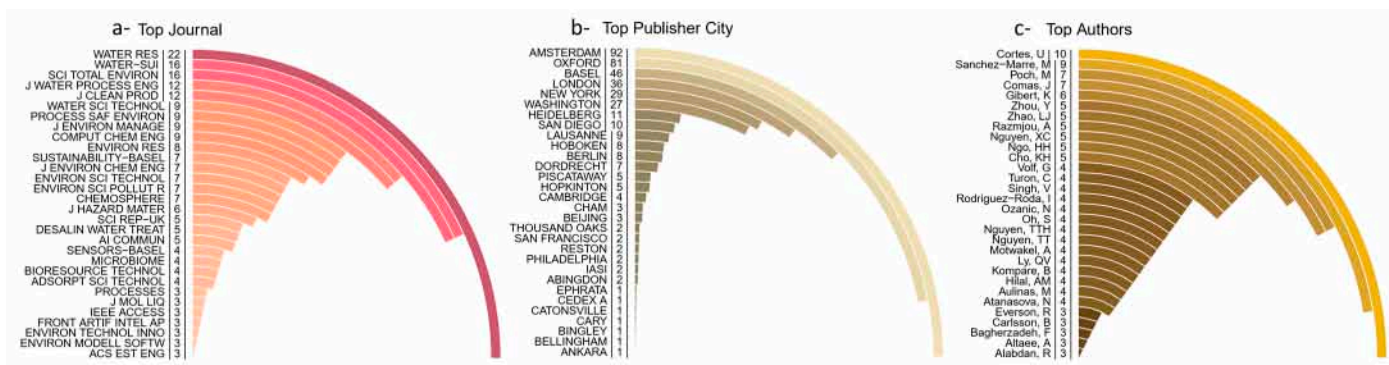


Figure 5. Most productive journals (a), publisher city (b), and authors in the application of artificial intelligence in wastewater treatment (c).

These 30 journals are relatively distributed in five countries: the Netherlands, Switzerland, the United Kingdom, the United States, and Germany, where the United States occupies the first place with a percentage of 43% (Figure 5b). Furthermore, these five countries are developed countries, which perhaps indicates that developing countries do not value the creation of well-reputed journals in this field.

This section has been concluded with Figure 5c, which suggests the 30 most active authors applying artificial intelligence in the wastewater domain. As displayed, the well-ranked authors are: Cortes, U.; Sanchez-Marre, M.; Poch, M.; Comas, J.; and Gibert, K., with a number of 6 to 10 publications or collaborations. The findings also showed that the majority of the authors belong to Spain, which highlights their interest in this research topic.

3.5. Optimization Methods

Optimization methods are one of the key elements in machine learning, as they play a crucial role in calibrating models capable of learning the mechanics of a system or a phenomenon from the actual data obtained and subsequently making accurate predictions. In machine learning, optimization consists of finding the optimal values of the model parameters according to the minimization or maximization of a cost function.

However, the optimization phase of the model parameters can be arduous and challenging to deal with due to the model parameters, the high-dimensional search space, or complex cost functions. Stochastic optimization methods provide a powerful framework to address these challenges by combining stochastic modeling and optimization techniques.

The choice of the optimization method depends on various factors, including the type of model, the magnitude of the data, and the available computational resources. Hence, researchers in the field of machine learning continue to innovate and develop optimization techniques to refine and improve model performance and reduce calibration time.

Optimization methods have attracted progressive attention in recent years due to their widespread application in various research themes, such as wastewater treatment. As shown in Figure 6, there are several optimization algorithms used in the literature for wastewater treatment. Among these methods, Genetic Algorithm (GA) is the most famous algorithm; it is cited in 20 articles with 27 repetitions, followed by Optimization of Particle

Swarms (PSO) in 5 articles with 5 repetitions, and Gene Expression (GE) in 2 articles with 2 repetitions. The rest of the algorithms are marked by a single appearance.

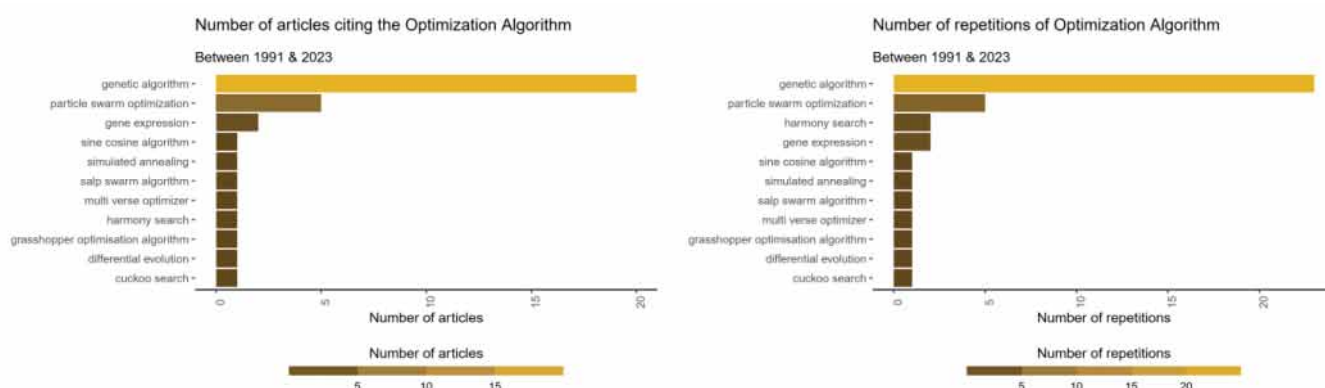


Figure 6. The most used Optimization Algorithm method in the field of wastewater.

- The GA is a probabilistic optimization method inspired by the natural selection process and genetics [35,36]. It involves the use of an initial population of potential solutions that undergo a series of selection, reproduction, and mutation operations.
- The PSO is a metaheuristic optimization approach that simulates the behavior of social animals such as birds and fish [37,38]. This involves the use of a swarm of particles that move through the research space and update their positions according to their own experience and that of their neighboring particles.
- The GE is an automatic programming method that evolves computer programs using a combination of genetic algorithms and symbolic regression [39,40].

3.6. Machine Learning Models

Machine learning is becoming a priority in many scientific research themes. They want to model the behavior of a complex system using large volumes of data. The choice of the right model depends on the objectives set forth and the maturity of the data manipulators. Among the research topics is the application of these algorithms to find their place in the modeling of wastewater treatment systems. The major goal of this study is to exploit textual analysis to discover the most commonly used machine learning models in the wastewater domain. After the preprocessing and cleaning of the abstract text, a list of the artificial intelligence models is prepared in order to perform an extraction, followed by a unique count of publications citing each model and an overall count of the repetition of the citations of each model.

The obtained results are depicted in Figure 7; reading the results reveals that seven models are the most confronted in the field of wastewater: Artificial Neural Network (ANN), Random Forest (RF), Support Vector Machine (SVM), Linear Regression (LR), Adaptive Neuro-Fuzzy Inference System (ANFIS), Decision Tree (DT), and Gradient Boosting (GB). The analysis also demonstrates that after 2018, the temporal evolution of the use of these models is enhanced. This evolution is marked by an increase in the level of application by 75% for some models, while others do not exceed this value. In this part, we propose a review of the main machine learning models that are used in practice in the field of wastewater treatment. All of them are described with very intuitive information about the most common models used in Web of Science articles. Each model description ends with a set of citations as an example of research served or exploited by the model concerned.

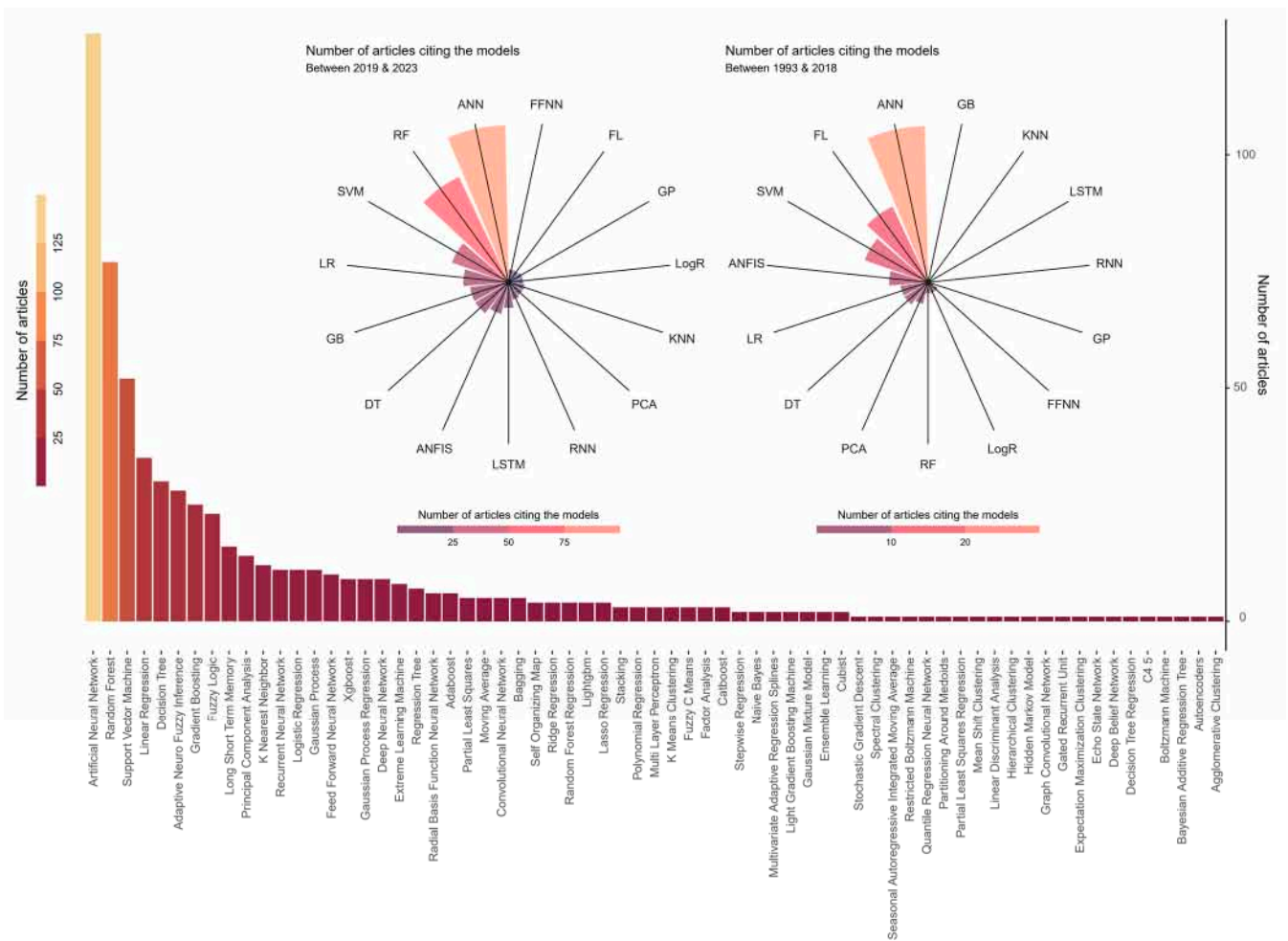


Figure 7. The most used machine learning models in the field of wastewater.

Therefore, machine learning models are grouped into two categories: supervised and unsupervised. If the model is a supervised model, it can belong to one of two classes or sub-categories: a regression or classification model.

3.6.1. Artificial Neural Networks (ANN)

ANN is a specific branch of research in computer science and neurocomputing. Its concept is based on the structure of biological neurons in the human brain. There are different types of this model, offering different paths of information processing.

Globally, ANN can be schematized as a set of neural systems composed of at least two layers: an input layer and an output layer, generally including intermediate layers (“hidden layers”). The more complex the problem to be solved, the more layers the neural network must have, where each layer contains enough artificial neurons. The model is used in several research topics in the field of wastewater, citing us as an example: monitoring and design of biological wastewater treatment systems [41,42]; energy consumption optimization in the sludge incineration unit of wastewater treatment systems [43]; predicting the performances of wastewater treatment plants [26,44]; etc.

3.6.2. Random Forest (RF)

RF is an approach to ensemble learning that is based on decision trees. This model is based on the creation of multiple decision trees involving split data sets from the original data. A random selection of a subset of variables at each stage of the decision tree is required. The model then selects the mode for all predictions in each decision tree. RF can be used in

wastewater-based epidemiology applications [45], investigating the energy consumption in wastewater treatment plants [46], modeling and evaluating the performance of wastewater treatment systems [47], predicting inflow at wastewater treatment systems [48], etc.

3.6.3. Support Vector Machine (SVM)

A SVM is a supervised classification algorithm that allows for the treatment of non-linear problems by reformulating them into quadratic optimization problems. Its purpose is to facilitate these resolutions. This classification technique is based on two main axes: (i) the notion of maximum margin, and (ii) the notion of kernel function. Briefly, the maximum margin is the data separation boundary that maximizes the distance between the separation boundary and the nearest data, where the kernel function is a kind of alternative to a scalar product in a very high-dimensional space. This model was applied to evaluate the efficacy of a large-scale aerobic biological wastewater treatment system in removing nitrogen [49] as well as to present an indirect approach that relies on some drainage basin criteria for the calculation of the major wastewater quality indicators [50]. Other searches employ this model with the usual tuning parameters to find industrial discharge [51] and, additionally, to simulate the unpredictability of chemical phosphorus removal procedures in wastewater treatment facilities [52].

3.6.4. Linear Regression (LR)

LR is undoubtedly still the simplest statistical modeling method to use. The idea of this model is simply to find a line that best fits the output data. The LR penetrates the multiple linear regression and the polynomial regression. There are generally two types of regression: simple regression (one explanatory variable) and multiple regression (several explanatory variables), although the conceptual framework and calculation methods are identical. The model is applied to several research topics in the field of wastewater, citing as examples: Altowayti et al. [53], who used linear regression to explain the adsorption of Zn^{2+} from synthesized water; Nourani et al. [54], who also used it to predict the performance of the Nicosia wastewater treatment plant; and Nguyen et al. [55], who worked on the adsorption mechanisms of pharmaceuticals onto biochar by applying the same model.

3.6.5. Decision Tree (DT)

DT is a popular model used in operations research, strategic planning, and machine learning. Its principle is based on a diagram that conceptualizes the possible outcomes of a series of interconnected choices. The model usually begins with a node from which several possible outcomes flow. Each of these outcomes leads to other nodes, from which other possibilities arise. The resulting diagram conceptualizes the shape of a tree. The decision tree model is used in different application areas, such as the prediction of effluent total nitrogen from a full-scale wastewater treatment plant using a combined trickling filter-activated sludge system [56]; wastewater network maintenance and optimization [57]; predicting water quality [58]; and controlling sewage discharge in wastewater treatment systems [59].

3.6.6. Logistic Regression

Logistic regression is a supervised classification algorithm popular in machine learning that is similar to linear regression. Otherwise, it is a generalized linear model using a logistic function as a link function. It is utilized to model the probability of a finite number of outcomes. It predicts the probability that an event will or will not occur based on the optimization of the regression coefficients. This outcome always varies between 0 and 1. When the predicted value is above a threshold, the event is likely to occur, while when this value is below the same threshold, it is not. The model is applied to several wastewater topics. Suchetana et al. [60] use logistic regression to predict total inorganic nitrogen in treated wastewaters, while Robles-Velasco et al. [61] work on the estimation of a failure in a sewer pipe. Szelag et al. [62] applied a logistic regression model to identify the best

wastewater parameter as a suitable indicator for operational monitoring, process control, and simulation purposes. Bencke et al. [63] develop models that would automatically filter citizen messages into categories based on the variety of city service dimensions (e.g., water, wastewater), and Liu et al. [64] utilize machine learning to predict groundwater flooding (or infiltration) into sewer networks.

3.6.7. k-Nearest Neighbors (KNN)

The k-Nearest Neighbors algorithm, also known as KNN or k-NN, is a non-parametric supervised learning discriminant that uses proximity to classify or predict groups of individual data points. Although it can be used in regression or classification problems, it is often used as a classification algorithm, assuming that similar adjacent points can be found. The approach has been used on a variety of topics. Ravi et al. [65] applied the model to understand and examine the underlying factors that could cause septic failures. Zidan et al. [31] used the KNN model in the prediction of total coliform removal from urban wastewater by an on-site, two-stage multi-soil-layering plant. The KNN algorithm was used by Hmoud and Waselallah [66] to predict the water quality as well as the effluent total nitrogen from a large-scale wastewater treatment facility [56].

3.7. Timeline of Machine Learning Model in Wastewater

An in-depth analysis in chronological terms of the machine learning models used in wastewater treatment is summarized in Figure 8. This study gives an idea of the timeline of evolution and appearance of the first citations or uses of the machine learning models, as well as the total number of publications citing each model by year.

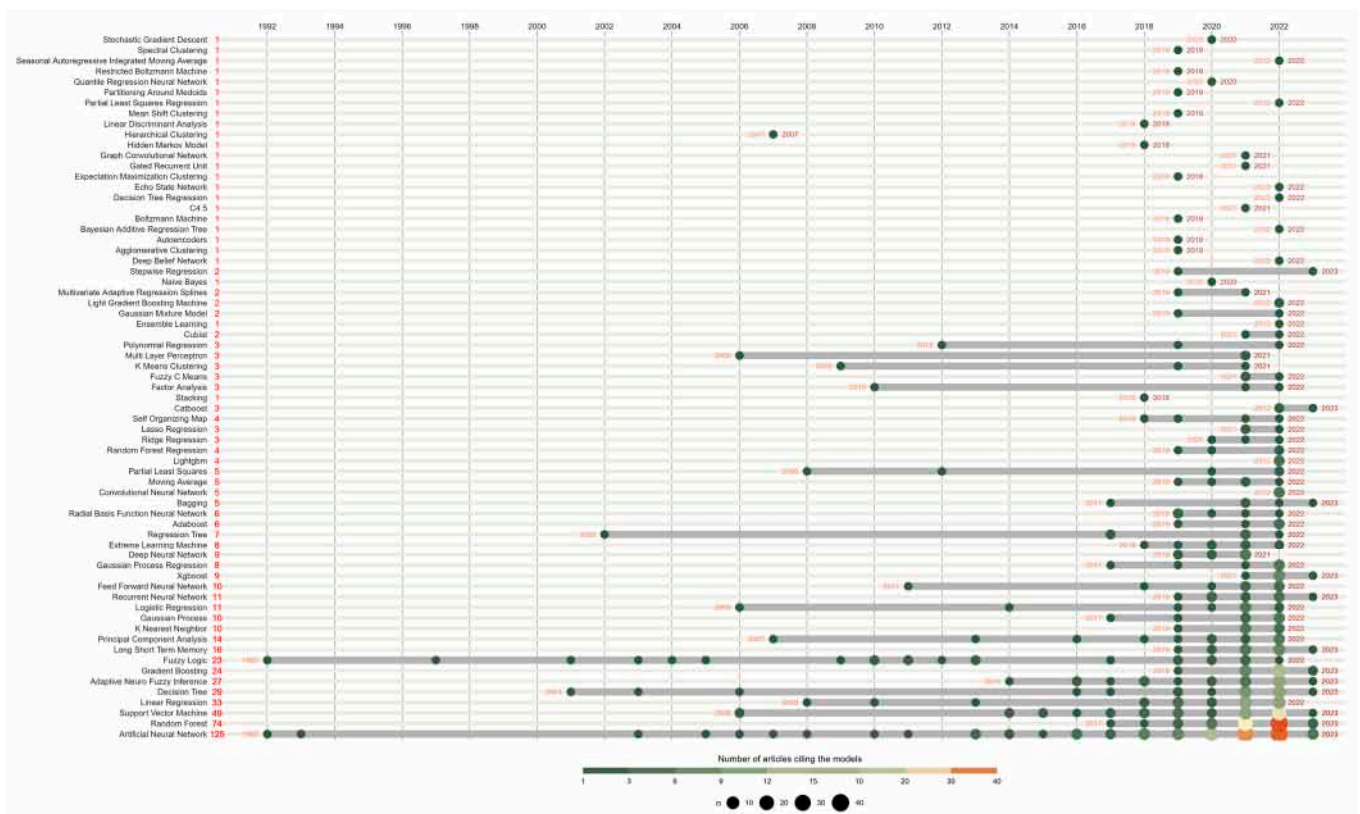


Figure 8. Timeline of machine learning models’ appearance applied in the field of wastewater.

Despite the austerity given to the use of ML models in our days, the appearance of these approaches in wastewater was marked in the 1990s. Figure 8 shows that the first models used are dated 1992, which are ANN and fuzzy logic (FL). These models are the oldest in the history of the involvement of artificial intelligence in the field of wastewater.

Both models are characterized by a more remarkable lacunar chronological appearance, while the FL model reveals a stationary chronological evolution. However, the ANN model is absent from the documents for 10 years from its first use, followed by an appearance marked by an exponential chronological evolution. The timeline 2001–2012 is marked by the appearance of a set of models, which are: Hierarchical Clustering, Factor Analysis, K-Means Clustering, Multi-Layer Perceptron, Polynomial Regression, Feed-Forward Neural Network, Partial Least Squares, Logistic Regression, PCA, Nearest Neighbor, Linear Regression, Regression Tree, and Support Vector Machine.

The analysis also shows the new machine learning models performed in the wastewater domain, such as Cubist, Ensemble Learning, Gaussian Mixture Model, Light Gradient, Boosting Machine, Multivariate Adaptive Regression Splines, Naive Bayes, Stepwise Regression, Deep Belief Network, Agglomerative Clustering, Fuzzy C Means, Autoencoders, Bayesian Additive Regression Tree, Boltzmann Machine, C4.5, Decision Tree Regression, Echo State Network, Expectation, Maximization Clustering, Gated Recurrent Unit, Graph Convolutional Network, Hidden Markov Model, Linear Discriminant Analysis, Mean Shift Clustering, Partial Least Squares Regression, Partitioning Around Medoids, Quantile Regression Neural Network, Restricted Boltzmann Machine, Seasonal Autoregressive Integrated Moving Average, Spectral Clustering, Stochastic Gradient Descent, Recurrent Neural Network, Xgboost, Gaussian Process Regression, Deep Neural Network, Extreme Learning Machine, Adaboost, Radial Basis Function Neural Network, Bagging, Convolutional Neural Network, Lightgbm, Random Forest Regression, Ridge Regression, Lasso Regression, Self-Organizing Map, and Catboost.

Generally, the involvement of these models has increased in the last few years, which shows the adaptation of the wastewater treatment domain to the trend of scientific research.

3.8. Wastewater Treatment Technologies Involving Machine Learning Methods

Figure 9 summarizes the number of articles and repetitions for each technology using machine learning as a modeling method. The analysis of the results shows that more than 30 articles have been published on the adsorption technique. Activated sludge has been the subject of more than 25 studies, followed by more than 10 articles on screening and constructed wetland systems. On the other hand, there are treatment methods that rarely involve machine learning. As an illustration, we can cite emerging technologies and chemical precipitation, which together account for little more than five articles.

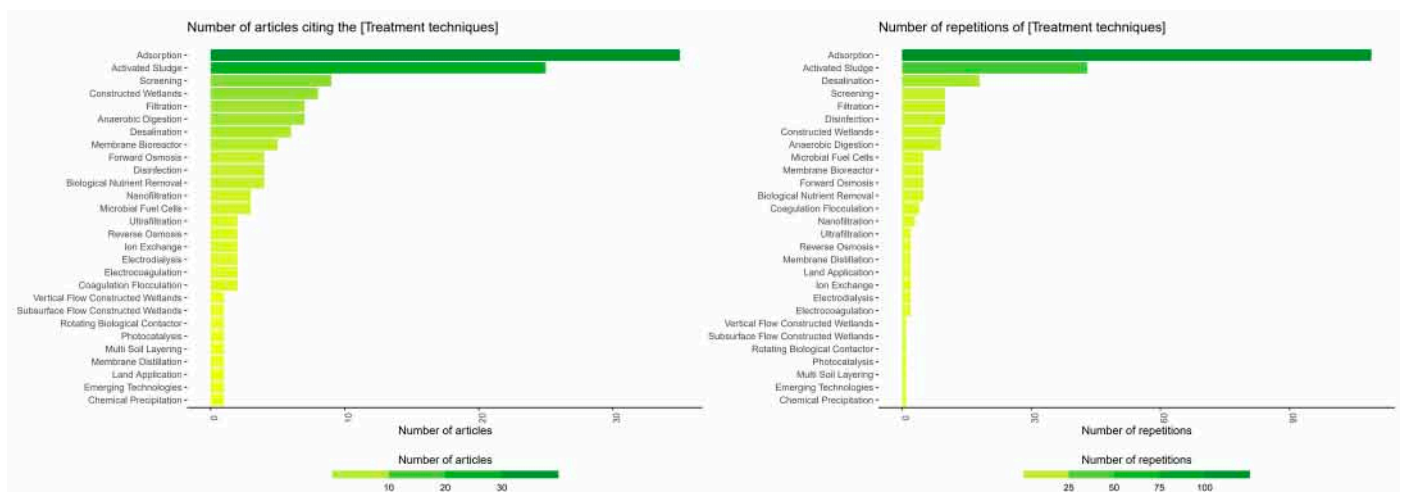


Figure 9. Wastewater treatment technologies treated by machine learning models.

In order to study the link between the techniques and technologies used in wastewater treatment and the machine learning models in our database, a cross-reference of the variables is necessary. This statistical technique is commonly used to analyze the relationship between two categorical variables, in our case, machine learning models and process-

ing techniques. The technique consists of creating a contingency table that summarizes the frequency of each category combination among the two variables studied. Figure 10 summarizes the information in the table (cf. Supplementary Data) as a graph to visually analyze the crossing. Reading the graph shows that the ANN, SVM, RF, and LR are the models most closely related to with the majority of techniques and technologies used in wastewater treatment. The most important frequencies are identified between the crossing of the activated sludge and adsorption techniques and the ANN, LR, RF, and SVM models. Concerning the techniques most closely related to the maximum of the models, we note activated sludge, adsorption, biological treatment, screening, and membrane bioreactor.

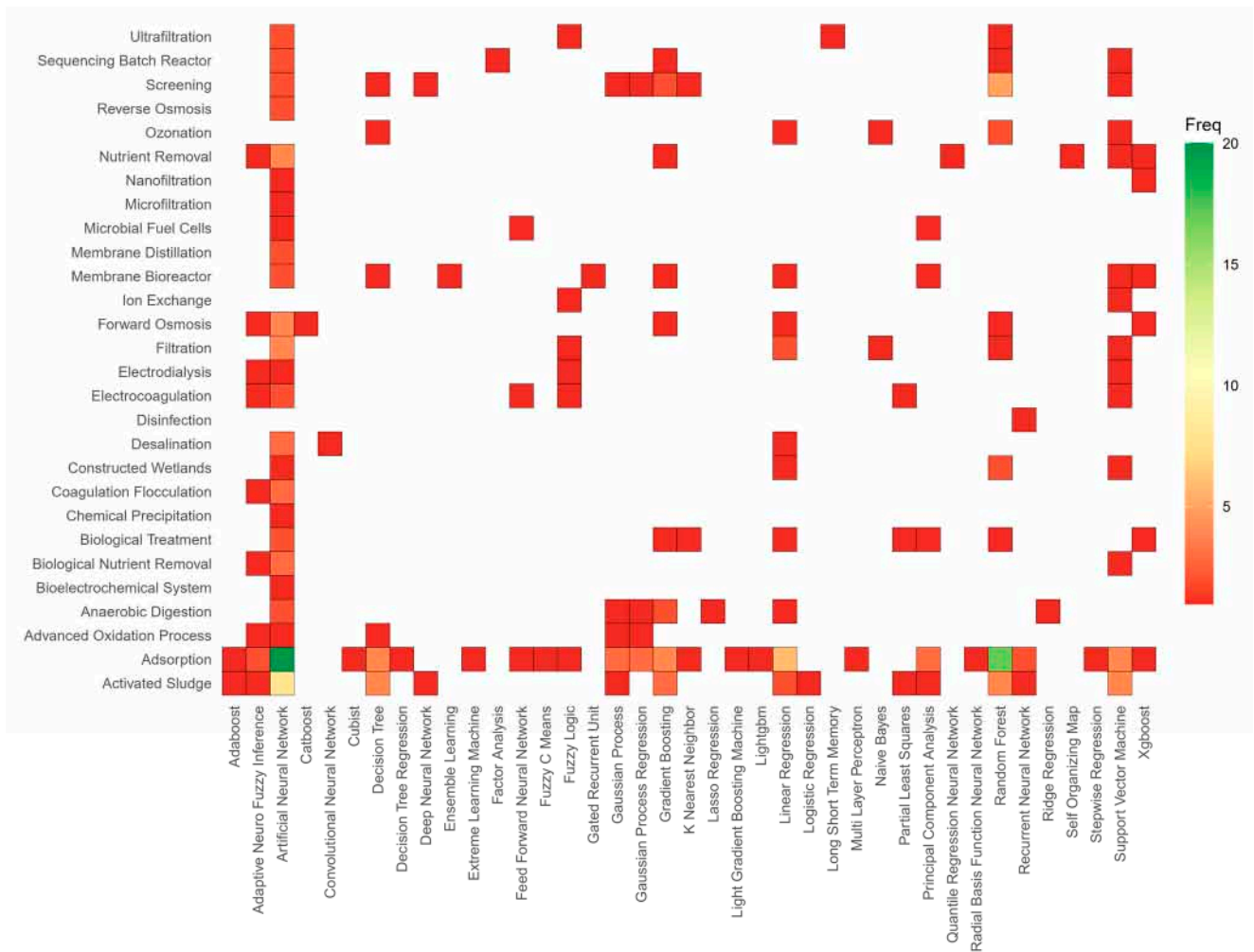


Figure 10. Crossing of wastewater treatment techniques and machine learning models.

Choosing the right machine learning model for a wastewater problem is a difficult task. Several factors need to be considered, such as the data typology, the size and quantity of the data set, the number of features, the complexity of the problem, and the evaluation criteria and performance measures.

3.9. High-Frequency Keywords and Groups

To highlight the central purposes of documents and facilitate mapping for readers, keywords are essential [67]. In order to extract the information residing in the abstracts from our database, it is necessary to perform pretreatment to eliminate irrelevant information and noise. This process typically involves cleaning up data by eliminating punctuation, numbers, blank words, and any other non-informative text. In addition, the text requires segmentation into separate words or phrases and standardization to ensure consistency of

use of these models varies from one study to another based on the challenges faced. This variability in model usage is strongly controlled by the typology of the data. Generally, these models can be classified into four categories: regression, classification, dimensionality reduction, and clustering models.

The first category, classification models, is concerned with assigning input data to predefined classes or categories. Numerous studies in the literature focus on the application of classification models in the field of wastewater treatment. Choi et al. [68], for example, used DT and RF models to identify odor sources in urban areas. Ribalta et al. [69] employed the GB model to reduce the frequency of sewer maintenance and cleaning. In a different case, Wodecka et al. [70] applied SVM to predict the variability of wastewater quality at the inlet of wastewater treatment plants.

The second category is regression models, which are used to predict continuous numerical values based on input factors. Among the studies that used the Zhao et al. [71] regression model, a comparison was conducted using six machine learning models: an ANN, SVR, an ANN-based AdaBoost, a SVR-based AdaBoost, DT, and RF to select the most appropriate model for predicting the easily biodegradable chemical oxygen demand (RB-COD) and the slowly biodegradable chemical oxygen demand (SB-COD) in municipal wastewater based on anoxic water oxidation-reduction potential (ORP) data. In another case study, Granata et al. [72] applied the Regression Tree M5P model, the Bagging algorithm, and the Random Forest algorithm to address complex wastewater engineering problems, notably predicting energy loss and forecasting lateral flow in a spillway.

Time series models belong to the class of regression models. This type of model captures and predicts patterns and dependencies in sequential data over time and is commonly applied in the wastewater field. Time series analysis usually uses algorithms such as Autoregressive Integrated Moving Average (ARIMA), Recurrent Neural Networks (RNNs), and Long Short-Term Memory (LSTM) networks. Among the examples of studies using this type of model, Nourani et al. [54] utilized the ARIMA linear model to predict the parameters of biological oxygen demand (BOD_{eff}) and chemical oxygen demand (COD_{eff}) to compare the capabilities of linear and nonlinear models in predicting complex processes. Kang et al. [73] utilize bidirectional long-short-term memory (LSTM) to predict wastewater flow rates for a municipal wastewater treatment plant; another study by Dairi et al. [74] proposes an approach based on coupling deep learning methods (recurrent neural networks) and clustering algorithms to detect anomalies in wastewater treatment plants.

The third category is dimensionality reduction. These types of models are utilized to reduce the number of features in a dataset while retaining essential information, assisting in overcoming the curse of dimensionality and improving computational efficiency. Principal Component Analysis (PCA) is a prominent dimensionality reduction method used in wastewater treatment. Murshid et al. [75] employed this method, PCA, to evaluate the adsorption efficiency of different metal oxide nanoparticles (MONPs) composite hydrogels for wastewater dye treatment. The PCA was also used to select inputs for other machine-learning models [76].

The fourth category is clustering models, which are unsupervised learning techniques used to group similar data based on their common characteristics. The goal of clustering is to identify intrinsic structures or hidden relationships in the data without any prior labels or categorization. In the literature, there are numerous applications of clustering models in the field of wastewater treatment. For instance, Alsulaili et al. [77] utilized a self-organizing map (SOM) to group and prepare input data for the ANN model, with the ultimate aim of predicting the quality of effluents from a wastewater treatment plant. In another study, Navato et al. [78] evaluated four models: K-means, GMM, hierarchical, and DBSCAN clustering, to identify anomalies in the complex chemical context of wastewater.

To understand the scientific orientation of applying machine learning models in the field of wastewater treatment, a typological classification of articles is conducted. Figure 12 summarizes the results obtained.

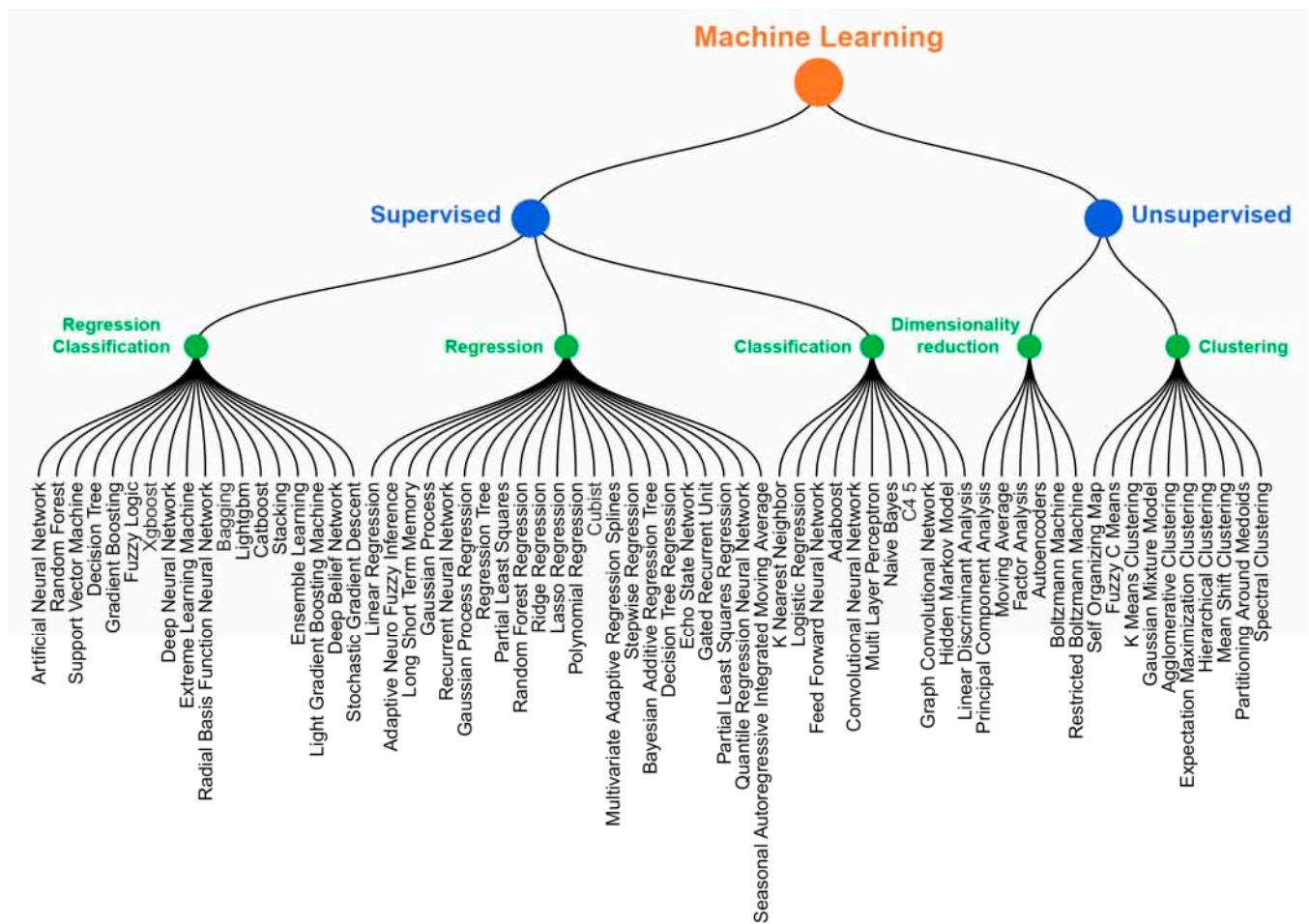


Figure 12. Classification of machine learning models.

Broadly, machine learning models can be classified into two large groups based on the availability of labeled data. The models that require labeled data for training are considered supervised learning models, and the models that do not require labeled data are classified as unsupervised learning models. Among the 67 models in Figure 9, 52 are labeled as supervised learning models, while the remaining 15 are categorized as unsupervised learning models.

In each group, we can enhance the specification of machine learning models by organizing them into distinct classes based on their characteristics, such as regression, classification, clustering, and dimensionality reduction models. Regression models are the ones that predict continuous values, while classification models predict categorical values. Clustering algorithms group the data into subsets or clusters, whereas dimensionality reduction techniques reduce the number of features while preserving the structure of the data. Out of the 67 models, 22 are regression models, 18 are models that could merge regression and classification, 11 are classification models, 10 are clustering models, and 6 are dimensionality reduction techniques.

As depicted in Figure 12, supervised model learning groups are the most popular models used among researchers. Meanwhile, deep learning models such as artificial neural networks, convolutional neural networks, and recurrent neural networks are the most prolific models with a high number of articles. The analysis of the graph also shows that SVM and RF have achieved significant success in various classification problems and are popular among researchers. Regarding unsupervised learning models, particularly in the clustering class, it is observed that K-means clustering and self-organizing maps are the

most commonly encountered models in this category. For dimensionality reduction, the PCA model has taken first place among models, with 14 articles citing it.

Understanding these typologies is critical for selecting the best machine-learning strategy based on the nature of the data and the problem at hand. Researchers and practitioners can effectively analyze data, generate accurate predictions, and get useful insights by employing regression, classification, dimensionality reduction, or time series analysis techniques.

3.11. Critical Analysis of Machine Learning Model Usage

The integration of new technologies, particularly machine learning models, has brought about a revolutionary transformation in the field of wastewater treatment, providing substantial contributions in several axes of this field. However, it is important to recognize that, such as any technology, the integration of these modeling approaches into wastewater treatment has both advantages and challenges that need to be carefully considered.

One way to assess the advantages and disadvantages of using machine learning in wastewater treatment is to compare it with the conventional deterministic or classical models typically employed in this field. By conducting a comparative analysis between these two types of models, we can better understand the strengths and limitations of machine learning in addressing wastewater-related problems.

In a comprehensive manner, machine learning models rely on data to conceptualize patterns and make predictions. They are trained using large datasets, allowing them to extract information and identify complex relationships that would not be evident with traditional modeling techniques. Conversely, classical modeling models, such as physical and conceptual models, are based on theoretical principles or empirical observations.

Furthermore, machine learning models have the ability to learn automatically from new data and adapt their predictions accordingly. They can continuously improve their performance by incorporating new information. Classical or deterministic models, on the other hand, generally require manual adjustments or updates when confronted with new data.

Thus, machine learning models aim to generalize the relationships in the training data to unseen data. They strive to capture underlying concepts and make accurate predictions about new instances. Classical models represent specific cases and may not be generalizable to different conditions or contexts.

On the other hand, machine learning models offer flexibility and high adaptability, making them suitable for handling large and diverse datasets. Deterministic models, while providing tangible representations, may be limited in terms of adaptability and scalability.

One weakness of machine learning models is the complexity of model conceptualization. For example, neural network families are often criticized for their lack of interpretability. It can be challenging to understand how these models arrive at their predictions, rendering their decision-making process inexplicable. In contrast, classical models are typically designed to be interpretable, allowing users to understand the mechanisms, assumptions, and underlying relationships explicitly.

Another limitation of machine learning models is that they often require substantial computational resources, especially for training complex systems or processing vast datasets. These models may require robust and powerful hardware, storage, and processing capabilities. In contrast, deterministic models, being abstract representations, have minimal resource requirements compared to machine learning models.

Despite numerous important studies on the application of machine learning models and the integration of artificial intelligence in wastewater treatment, the crucial problem remains: AI-assisted systems are not yet widely deployed in treatment plants, despite sustained efforts to improve their design and operation. To bridge this gap, close collaboration between researchers and industry is essential. Together, they must develop AI-assisted systems specifically tailored to the daily operations of these facilities. This approach will address the technical, economic, and regulatory challenges that currently hinder the inte-

gration of AI in this critical field of water management. By working together, researchers and industry professionals can make significant progress towards better management and sustainable preservation of our water resources.

4. Conclusions

This work has provided a general overview of the application of models and machine learning algorithms in wastewater treatment. According to this background, plenty of literature analysis techniques were used to assess the trend of publications, their global ranking, the most influential models, and their evolution over time. This study was based on the exploitation of the Web of Science Database by data mining techniques to determine the most applied machine learning algorithms in wastewater treatment and their most prolific countries.

The analysis of the spatial distribution of publications has shown that scientific production is strongly linked to the level of development in a country. Thus, the countries of the USA, China, Canada, England, Spain, Egypt, India, Iran, and Saudi Arabia were determined to be the most scientifically productive countries.

Enhancing scientific progress concerning the application of machine learning in the wastewater sector requires fostering a cooperative relationship between countries worldwide. Scientific collaboration enables researchers to establish a forum for knowledge exchange, improve the quality of research, gain access to new materials, amplify and generate innovative solutions and new perspectives, and address challenges in the wastewater sector.

The findings revealed that the application of machine learning models in wastewater treatment has been increasing recently. This variation could be explained by the adaptation of this topic to the trend of science and also by the practicality and flexibility of these models in deconstructing and modeling the complex systems faced in this field.

Among the most evolved models in this research field, we find ANN, RF, SVM, LR Model, ANFIS, DT, and GB. Despite the enormous machine learning algorithms developed in the field of AI, the choice of the most suitable model depends on the problem encountered, the orientation of the data processors and analysts, as well as on the quantity and quality of the data used in this study and their typology.

The cluster analysis of the models reveals that several reasons explain the difference in the number of articles and models for each class of machine learning models. Wastewater modeling is primarily concerned with predicting the behavior of a complex, non-linear system. Classification models are well-suited to this task, as they can be used to predict discrete outcomes such as the presence or absence of a pollutant. Regression models are also useful, as they can be used to predict continuous outcomes such as pollutant concentrations or flow rates. Clustering and dimensionality reduction models, on the other hand, are less commonly employed in wastewater modeling. Clustering algorithms are useful for identifying patterns in large datasets; however, the smaller size of many wastewater datasets may not necessitate their use.

However, this study is still nascent and has gaps. Hence, the dataset was derived from the WoS in order to have access to high-quality scientific publications and well-known models. Yet, the treated publications are limited, and the new models are not taken into consideration in this analysis.

The use of machine learning models involves and requires a detailed examination of how each model works, as there are several potential shortcomings of these models in their applications. Machine learning models can become overly complex as they force data into simulation, expressing their impressive performance on training data and their unsatisfactory performance on new and unseen data. These computer-aided models have the potential to discriminate against bias in the simulation phase when trained with biased data. Otherwise, some machine learning models, such as those in the neural network category, can be difficult to interpret, making it difficult to understand how they simulated the data.

Overall, choosing the most suitable machine learning model requires careful consideration of the problem, data, interpretability, performance metrics, and scalability. By deconstructing the strengths and weaknesses of different machine learning models, one can choose the one that best suits their needs and can provide the most accurate predictions.

Over the years, the use of AI in wastewater treatment has become increasingly widespread. Machine learning and modeling techniques have played a crucial role in this field. Overall, the application of AI in wastewater management has played a crucial role in promoting sustainability, reducing costs, and improving the quality of life for communities.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/w15193487/s1>, Supplementary S1: List of Machine Learning models covered in this study.

Author Contributions: Conceptualization, methodology, software, validation, data curation, writing—original draft, A.E.A.E.F. and L.M.; investigation, writing—original draft, A.E.A.E.F. and A.K.; writing—review and editing, N.O., O.M. and M.L.H.; visualization, supervision, funding acquisition, L.M. and M.L.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data will be made available on request.

Acknowledgments: This work was supported by the project I-MAROC ‘Artificial Intelligence/ Applied Mathematics, Health/Environment: Simulation for Decision Support’, the APRD Program (Morocco), and the National Center for Studies and Research on Water and Energy, Cadi Ayyad University, Morocco.

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

References

1. Barrera-Animas, A.Y.; Oyedele, L.O.; Bilal, M.; Akinosho, T.D.; Delgado, J.M.D.; Akanbi, L.A. Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting. *Mach. Learn. Appl.* **2021**, *7*, 100204. [[CrossRef](#)]
2. Li, K.-Q.; Liu, Y.; Kang, Q. Estimating the thermal conductivity of soils using six machine learning algorithms. *Int. Commun. Heat Mass Transf.* **2022**, *136*, 106139. [[CrossRef](#)]
3. David, N.; Mathew, R. Review of Machine Learning in Geosciences and Remote Sensing. In *Proceedings of the International Conference on Computer Networks, Big Data and IoT (ICCB-2019)*; Pandian, A.P., Palanisamy, R., Ntalianis, K., Eds.; Springer International Publishing: Berlin/Heidelberg, Germany, 2019; Volume 49, pp. 195–204. ISBN 978-3-030-43191-4/978-3-030-43192-1.
4. Gil, Y.; Pierce, S.A.; Babaie, H.; Banerjee, A.; Borne, K.; Bust, G.; Cheatham, M.; Ebert-Uphoff, I.; Gomes, C.; Hill, M.; et al. Intelligent systems for geosciences: An Essential Research Agenda. *Commun. ACM* **2018**, *62*, 76–84. [[CrossRef](#)]
5. Alaoui El Fels, A.; El Ghorfi, M. Using remote sensing data for geological mapping in semi-arid environment: A machine learning approach. *Earth Sci. Inform.* **2022**, *15*, 485–496. [[CrossRef](#)]
6. Alaoui El Fels, A.; Saidi, M.E.; Bin Alam, J. Rainfall Frequency Analysis Using Assessed and Corrected Satellite Precipitation Products in Moroccan Arid Areas. The Case of Tensift Watershed. *Earth Syst. Environ.* **2022**, *6*, 391–404. [[CrossRef](#)]
7. El Alaoui El Fels, A.; Saidi, M.E.M.; Bouijji, A.; Benrhanem, M. Rainfall regionalization and variability of extreme precipitation using artificial neural networks: A case study from western central Morocco. *J. Water Clim. Chang.* **2020**, *12*, 1107–1122. [[CrossRef](#)]
8. Mashrur, A.; Luo, W.; Zaidi, N.A.; Robles-Kelly, A. Machine Learning for Financial Risk Management: A Survey. *IEEE Access* **2020**, *8*, 203203–203223. [[CrossRef](#)]
9. Aziz, S.; Dowling, M. Machine Learning and AI for Risk Management. In *Disrupting Finance*; Lynn, T., Mooney, J.G., Rosati, P., Cummins, M., Eds.; Springer International Publishing: Berlin/Heidelberg, Germany, 2019; pp. 33–50. ISBN 978-3-030-02329-4/978-3-030-02330-0.
10. Damrongsakmethee, T.; Neago, V.-E. Data Mining and Machine Learning for Financial Analysis. *Indian J. Sci. Technol.* **2017**, *10*, 1–7. [[CrossRef](#)]
11. Schmitt, M. *Artificial Intelligence in Business Analytics, Capturing Value with Machine Learning Applications in Financial Services*; University of Strathclyde: Glasgow, UK, 2020.
12. Latif, J.; Xiao, C.; Imran, A.; Tu, S. Medical Imaging Using Machine Learning and Deep Learning Algorithms: A Review. In *Proceedings of the 2019 2nd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, Sukkur, Pakistan, 30–31 January 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–5.
13. Mujumdar, A.; Vaidehi, V. Diabetes Prediction using Machine Learning Algorithms. *Procedia Comput. Sci.* **2019**, *165*, 292–299. [[CrossRef](#)]

14. Senturk, Z.K. Early diagnosis of Parkinson's disease using machine learning algorithms. *Med. Hypotheses* **2020**, *138*, 109603. [[CrossRef](#)] [[PubMed](#)]
15. Sharma, S.; Aggarwal, A.; Choudhury, T. Breast Cancer Detection Using Machine Learning Algorithms. In Proceedings of the 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS), Belgaum, India, 21–22 December 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 114–118.
16. Khanam, J.J.; Foo, S.Y. A comparison of machine learning algorithms for diabetes prediction. *ICT Express* **2021**, *7*, 432–439. [[CrossRef](#)]
17. Punn, N.S.; Sonbhadra, S.K.; Agarwal, S. COVID-19 Epidemic Analysis Using Machine Learning and Deep Learning Algorithms. *MedRxiv* **2020**. [[CrossRef](#)]
18. Akour, I.; Alshurideh, M.; Al Kurdi, B.; Al Ali, A.; Salloum, S. Using Machine Learning Algorithms to Predict People's Intention to Use Mobile Learning Platforms During the COVID-19 Pandemic: Machine Learning Approach. *JMIR Med. Educ.* **2021**, *7*, e24032. [[CrossRef](#)]
19. Xin, D.; Gonnella, N.C.; He, X.; Horspool, K. Solvate Prediction for Pharmaceutical Organic Molecules with Machine Learning. *Cryst. Growth Des.* **2019**, *19*, 1903–1911. [[CrossRef](#)]
20. Ghosh, A.; Louis, L.; Arora, K.K.; Hancock, B.C.; Krzyzaniak, J.F.; Meenan, P.; Nakhmanson, S.; Wood, G.P.F. Assessment of machine learning approaches for predicting the crystallization propensity of active pharmaceutical ingredients. *CrystEngComm* **2018**, *21*, 1215–1223. [[CrossRef](#)]
21. Unnikrishnan, S.; Donovan, J.; Macpherson, R.; Tormey, D. Machine Learning for Automated Quality Evaluation in Pharmaceutical Manufacturing of Emulsions. *J. Pharm. Innov.* **2019**, *15*, 392–403. [[CrossRef](#)]
22. Sbahi, S.; Ouazzani, N.; Hejjaj, A.; Mandi, L. Nitrogen modeling and performance of Multi-Soil-Layering (MSL) bioreactor treating domestic wastewater in rural community. *J. Water Process. Eng.* **2021**, *44*, 102389. [[CrossRef](#)]
23. Zidan, K.; Mandi, L.; Hejjaj, A.; Sbahi, S.; El Alaoui El Fels, A.; Ouazzani, N.; Assabbane, A. Efficiency of a new hybrid multi-soil-layering eco-friendly technology for removing pollutants from domestic wastewater under an arid climate. *J. Water Process. Eng.* **2023**, *51*, 103482. [[CrossRef](#)]
24. Maleki, R.; Jahromi, A.M.; Ghasemy, E.; Khedri, M. Smart Sensing Technologies for Wastewater Treatment Plants. In *Artificial Intelligence and Data Science in Environmental Sensing*; Elsevier: Amsterdam, The Netherlands, 2022; pp. 1–17. ISBN 978-0-323-90508-4.
25. Southern, J.; Pitt-Francis, J.; Whiteley, J.; Stokeley, D.; Kobashi, H.; Nobes, R.; Kadooka, Y.; Gavaghan, D. Multi-scale computational modelling in biology and physiology. *Prog. Biophys. Mol. Biol.* **2008**, *96*, 60–89. [[CrossRef](#)] [[PubMed](#)]
26. Sbahi, S.; Ouazzani, N.; Hejjaj, A.; Mandi, L. Neural network and cubist algorithms to predict fecal coliform content in treated wastewater by multi-soil-layering system for potential reuse. *J. Environ. Qual.* **2020**, *50*, 144–157. [[CrossRef](#)] [[PubMed](#)]
27. Sundui, B.; Calderon, O.A.R.; Abdeldayem, O.M.; Lázaro-Gil, J.; Rene, E.R.; Sambuu, U. Applications of machine learning algorithms for biological wastewater treatment: Updates and perspectives. *Clean Technol. Environ. Policy* **2021**, *23*, 127–143. [[CrossRef](#)]
28. Oulebsir, R.; Lefkir, A.; Safri, A.; Bermad, A. Optimization of the energy consumption in activated sludge process using deep learning selective modeling. *Biomass Bioenergy* **2019**, *132*, 105420. [[CrossRef](#)]
29. Wang, D.; Thunell, S.; Lindberg, U.; Jiang, L.; Trygg, J.; Tysklind, M.; Souihi, N. A machine learning framework to improve effluent quality control in wastewater treatment plants. *Sci. Total. Environ.* **2021**, *784*, 147138. [[CrossRef](#)]
30. Banti, K.; Karampelia, I.; Tsipis, E.; Dimakis, T.; Papadimitriou, N.; Boulogeorgos, A.-A.A.; Kyriakidis, T.; Karamitsou, T.; Karapiperis, C.; Lappos, S.; et al. Intelligent Mechanisms for Irrigation Optimization via Treated Wastewater Management in Precision Agriculture—The AUGELAS Example. *Environ. Sci. Proc.* **2022**, *21*, 50. [[CrossRef](#)]
31. Zidan, K.; Sbahi, S.; Hejjaj, A.; Ouazzani, N.; Assabbane, A.; Mandi, L. Removal of bacterial indicators in on-site two-stage multi-soil-layering plant under arid climate (Morocco): Prediction of total coliform content using K-nearest neighbor algorithm. *Environ. Sci. Pollut. Res.* **2022**, *29*, 75716–75729. [[CrossRef](#)] [[PubMed](#)]
32. Zaghloul, M.S.; Achari, G. Application of machine learning techniques to model a full-scale wastewater treatment plant with biological nutrient removal. *J. Environ. Chem. Eng.* **2022**, *10*, 107430. [[CrossRef](#)]
33. Wongburi, P.; Park, J.K. Prediction of Sludge Volume Index in a Wastewater Treatment Plant Using Recurrent Neural Network. *Sustainability* **2022**, *14*, 6276. [[CrossRef](#)]
34. Wongburi, P.; Park, J.K. Big Data Analytics from a Wastewater Treatment Plant. *Sustainability* **2021**, *13*, 12383. [[CrossRef](#)]
35. Mirjalili, S. Genetic Algorithm. In *Evolutionary Algorithms and Neural Networks*; Springer International Publishing: Berlin/Heidelberg, Germany, 2019; Volume 780, pp. 43–55. ISBN 978-3-319-93024-4/978-3-319-93025-1.
36. Katoch, S.; Chauhan, S.S.; Kumar, V. A review on genetic algorithm: Past, present, and future. *Multimed. Tools Appl.* **2021**, *80*, 8091–8126. [[CrossRef](#)] [[PubMed](#)]
37. Banks, A.; Vincent, J.; Anyakoha, C. A review of particle swarm optimization. Part I: Background and development. *Nat. Comput.* **2007**, *6*, 467–484. [[CrossRef](#)]
38. Jain, N.K.; Nangia, U.; Jain, J. A Review of Particle Swarm Optimization. *J. Inst. Eng. Ser. B* **2018**, *99*, 407–411. [[CrossRef](#)]
39. Jędrzejowicz, J.; Jędrzejowicz, P. Gene Expression Programming as a data classification tool. A review. *J. Intell. Fuzzy Syst.* **2019**, *36*, 91–100. [[CrossRef](#)]

40. Shishegaran, A.; Boushehri, A.N.; Ismail, A.F. Gene expression programming for process parameter optimization during ultrafiltration of surfactant wastewater using hydrophilic polyethersulfone membrane. *J. Environ. Manag.* **2020**, *264*, 110444. [[CrossRef](#)] [[PubMed](#)]
41. Oruganti, R.K.; Biji, A.P.; Lanuyanger, T.; Show, P.L.; Sriariyanun, M.; Upadhyayula, V.K.K.; Gadhamshetty, V.; Bhattacharyya, D. Artificial intelligence and machine learning tools for high-performance microalgal wastewater treatment and algal biorefinery: A critical review. *Sci. Total. Environ.* **2023**, *876*, 162797. [[CrossRef](#)]
42. Singh, N.K.; Yadav, M.; Singh, V.; Padhiyar, H.; Kumar, V.; Bhatia, S.K.; Show, P.-L. Artificial intelligence and machine learning-based monitoring and design of biological wastewater treatment systems. *Bioresour. Technol.* **2023**, *369*, 128486. [[CrossRef](#)]
43. Adibimanesh, B.; Polesek-Karczewska, S.; Bagherzadeh, F.; Szczuko, P.; Shafighard, T. Energy consumption optimization in wastewater treatment plants: Machine learning for monitoring incineration of sewage sludge. *Sustain. Energy Technol. Assess.* **2023**, *56*, 103040. [[CrossRef](#)]
44. Hamed, M.M.; Khalafallah, M.G.; Hassanien, E.A. Prediction of wastewater treatment plant performance using artificial neural networks. *Environ. Model. Softw.* **2004**, *19*, 919–928. [[CrossRef](#)]
45. Kanneganti, D.; Reinersman, L.E.; Holm, R.H.; Smith, T. Estimating sewage flow rate in Jefferson County, Kentucky, using machine learning for wastewater-based epidemiology applications. *Water Supply* **2022**, *22*, 8434–8439. [[CrossRef](#)]
46. Zhang, S.; Wang, H.; Keller, A.A. Novel Machine Learning-Based Energy Consumption Model of Wastewater Treatment Plants. *ACS ES&T Water* **2021**, *1*, 2531–2540. [[CrossRef](#)]
47. Salem, M.; EL-Sayed Gabr, M.; Mossad, M.; Mahanna, H. Random Forest modelling and evaluation of the performance of a full-scale subsurface constructed wetland plant in Egypt. *Ain Shams Eng. J.* **2022**, *13*, 101778. [[CrossRef](#)]
48. Zhou, P.; Li, Z.; Snowling, S.; Baetz, B.W.; Na, D.; Boyd, G. A random forest model for inflow prediction at wastewater treatment plants. *Stoch. Environ. Res. Risk Assess.* **2019**, *33*, 1781–1792. [[CrossRef](#)]
49. Manu, D.S.; Thalla, A.K. Artificial intelligence models for predicting the performance of biological wastewater treatment plant in the removal of Kjeldahl Nitrogen from wastewater. *Appl. Water Sci.* **2017**, *7*, 3783–3791. [[CrossRef](#)]
50. Granata, F.; Papirio, S.; Esposito, G.; Gargano, R.; De Marinis, G. Machine Learning Algorithms for the Forecasting of Wastewater Quality Indicators. *Water* **2017**, *9*, 105. [[CrossRef](#)]
51. Thompson, K.A.; Branch, A.; Nading, T.; Dziura, T.; Salazar-Benites, G.; Wilson, C.; Bott, C.; Salveson, A.; Dickenson, E.R.V. Detecting industrial discharges at an advanced water reuse facility using online instrumentation and supervised machine learning binary classification. *Front. Water* **2022**, *4*, 1014556. [[CrossRef](#)]
52. Tabatabaei, T.S.; Farooq, T.; Guergachi, A.; Krishnan, S. Support Vector Machines Based Approach for Chemical Phosphorus Removal Process in Wastewater Treatment Plant. In Proceedings of the 2006 Canadian Conference on Electrical and Computer Engineering, Ottawa, ON, Canada, 7–10 May 2006; IEEE: Piscataway, NJ, USA, 2006; pp. 318–322. [[CrossRef](#)]
53. Altowayti, W.A.H.; Othman, N.; Al-Gheethi, A.; Dzahir, N.H.B.M.; Asharuddin, S.M.; Alsharif, A.F.; Nasser, I.M.; Tajarudin, H.A.; Al-Towayti, F.A.H. Adsorption of Zn²⁺ from Synthetic Wastewater Using Dried Watermelon Rind (D-WMR): An Overview of Nonlinear and Linear Regression and Error Analysis. *Molecules* **2021**, *26*, 6176. [[CrossRef](#)] [[PubMed](#)]
54. Nourani, V.; Elkiran, G.; Abba, S.I. Wastewater treatment plant performance analysis using artificial intelligence—An ensemble approach. *Water Sci. Technol.* **2018**, *78*, 2064–2076. [[CrossRef](#)]
55. Nguyen, X.C.; Ly, Q.V.; Nguyen, T.T.H.; Ngo, H.T.T.; Hu, Y.; Zhang, Z. Potential application of machine learning for exploring adsorption mechanisms of pharmaceuticals onto biochars. *Chemosphere* **2021**, *287*, 132203. [[CrossRef](#)]
56. Moghaddam, S.S.; Mesghali, H. A new hybrid ensemble approach for the prediction of effluent total nitrogen from a full-scale wastewater treatment plant using a combined trickling filter-activated sludge system. *Environ. Sci. Pollut. Res.* **2022**, *30*, 1622–1639. [[CrossRef](#)]
57. Bailey, J.; Harris, E.; Keedwell, E.; Djordjevic, S.; Kapelan, Z. Developing Decision Tree Models to Create a Predictive Blockage Likelihood Model for Real-World Wastewater Networks. *Procedia Eng.* **2016**, *154*, 1209–1216. [[CrossRef](#)]
58. Anmala, J.; Turuganti, V. Comparison of the performance of decision tree (DT) algorithms and extreme learning machine (ELM) model in the prediction of water quality of the Upper Green River watershed. *Water Environ. Res.* **2021**, *93*, 2360–2373. [[CrossRef](#)]
59. Stańczyk, J.; Burszta-Adamiak, E.; Kajewska-Szkudlarek, J.; Kurkiewicz, R.; Przerwa, A. Intelligent sewage discharge control in a wastewater treatment plant during rainfall periods. *Urban Water J.* **2023**, *20*, 380–393. [[CrossRef](#)]
60. Suchetana, B.; Rajagopalan, B.; Silverstein, J. Investigating regime shifts and the factors controlling Total Inorganic Nitrogen concentrations in treated wastewater using non-homogeneous Hidden Markov and multinomial logistic regression models. *Sci. Total Environ.* **2019**, *646*, 625–633. [[CrossRef](#)]
61. Robles-Velasco, A.; Cortés, P.; Muñuzuri, J.; Onieva, L. Estimation of a logistic regression model by a genetic algorithm to predict pipe failures in sewer networks. *OR Spectr.* **2021**, *43*, 759–776. [[CrossRef](#)]
62. Szelaq, B.; Mehrani, M.J.; Drewnowski, J.; Majewska, M.; Łagód, G.; Kumari, S.; Bux, F. Assessment of wastewater quality indicators for wastewater treatment influent using an advanced logistic regression model. *Desalin. Water Treat.* **2021**, *232*, 421–432. [[CrossRef](#)]
63. Bencke, L.; Cechinel, C.; Munoz, R. Automated classification of social network messages into Smart Cities dimensions. *Future Gener. Comput. Syst.* **2020**, *109*, 218–237. [[CrossRef](#)]
64. Liu, T.; Ramirez-Marquez, J.E.; Jagupilla, S.C.; Prigiobbe, V. Combining a statistical model with machine learning to predict groundwater flooding (or infiltration) into sewer networks. *J. Hydrol.* **2021**, *603*, 126916. [[CrossRef](#)]

65. Ravi, N.; Johnson, D.P. Artificial intelligence-based monitoring system for onsite septic systems failure. *Process. Saf. Environ. Prot.* **2021**, *148*, 1090–1097. [[CrossRef](#)]
66. Al-Adhaileh, M.H.; Alsaade, F.W. Modelling and Prediction of Water Quality by Using Artificial Intelligence. *Sustainability* **2021**, *13*, 4259. [[CrossRef](#)]
67. Jin, Q.; Liu, J.; Lu, X. Deep Contextualized Biomedical Abbreviation Expansion. In *Proceedings of the 18th BioNLP Workshop and Shared Task*; Association for Computational Linguistics: Florence, Italy, 2019; pp. 88–96.
68. Choi, Y.; Kim, K.; Kim, S.; Kim, D. Identification of odor emission sources in urban areas using machine learning-based classification models. *Atmos. Environ. X* **2022**, *13*, 100156. [[CrossRef](#)]
69. Ribalta, M.; Mateu, C.; Bejar, R.; Rubión, E.; Echeverria, L.; Alegre, F.J.V.; Corominas, L. Sediment Level Prediction of a Combined Sewer System Using Spatial Features. *Sustainability* **2021**, *13*, 4013. [[CrossRef](#)]
70. Wodecka, B.; Drewnowski, J.; Białek, A.; Łazuka, E.; Szulżyk-Cieplak, J. Prediction of Wastewater Quality at a Wastewater Treatment Plant Inlet Using a System Based on Machine Learning Methods. *Processes* **2022**, *10*, 85. [[CrossRef](#)]
71. Zhao, G.; Suzuki, S.; Deng, J.; Fujita, M. Machine learning estimation of biodegradable organic matter concentrations in municipal wastewater. *J. Environ. Manag.* **2022**, *323*, 116191. [[CrossRef](#)] [[PubMed](#)]
72. Granata, F.; de Marinis, G. Machine learning methods for wastewater hydraulics. *Flow Meas. Instrum.* **2017**, *57*, 1–9. [[CrossRef](#)]
73. Kang, H.; Yang, S.; Huang, J.; Oh, J. Time Series Prediction of Wastewater Flow Rate by Bidirectional LSTM Deep Learning. *Int. J. Control Autom. Syst.* **2020**, *18*, 3023–3030. [[CrossRef](#)]
74. Dairi, A.; Cheng, T.; Harrou, F.; Sun, Y.; Leiknes, T. Deep learning approach for sustainable WWTP operation: A case study on data-driven influent conditions monitoring. *Sustain. Cities Soc.* **2019**, *50*, 101670. [[CrossRef](#)]
75. Murshid, N.; Mouhtady, O.; Abu-Samha, M.; Obeid, E.; Kharboutly, Y.; Chaouk, H.; Halwani, J.; Younes, K. Metal Oxide Hydrogel Composites for Remediation of Dye-Contaminated Wastewater: Principal Component Analysis. *Gels* **2022**, *8*, 702. [[CrossRef](#)]
76. Bhagat, S.K.; Pyrgaki, K.; Salih, S.Q.; Tiyyasha, T.; Beyaztas, U.; Shahid, S.; Yaseen, Z.M. Prediction of copper ions adsorption by attapulgite adsorbent using tuned-artificial intelligence model. *Chemosphere* **2021**, *276*, 130162. [[CrossRef](#)]
77. Alsulaili, A.; Refaie, A. Artificial neural network modeling approach for the prediction of five-day biological oxygen demand and wastewater treatment plant performance. *Water Supply* **2020**, *21*, 1861–1877. [[CrossRef](#)]
78. Navato, A.P.; Mueller, A.V. Enabling Automatic Detection of Anomalies in Wastewater: A Highly Simplified Approach to Defining “Normal” in Complex Chemical Mixtures. *Front. Water* **2021**, *3*, 734361. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.