OPEN ACCESS

Volume: 5

Issue: 3

Month: July

Year: 2026

ISSN: 2583-7117

Published: 24.07.2026

Citation:

Aayushi Sharma,Dr. Mohammad Faheem "Smart Bioremediation: Role Of Artificial Intelligence-Based Monitoring and Control of Aquatic Plant Systems for Wastewater Treatment" International Journal of Innovations in Science Engineering and Management, vol. 5, no. 3, 2026, pp. 136-152

DOI:

10.69968/ijisem.2026v5i3136-152

Smart Bioremediation: Role Of Artificial Intelligence-Based Monitoring and Control of Aquatic Plant Systems for Wastewater Treatment

Aayushi Sharma[1], Dr. Mohammad Faheem[2]

[1]PhD Research Scholar Botany (Life Science), Glocal University Mirzapur pole, Distt- Sharanpur
Email: aayushijaiwal123@gmail.com

[2]Assistant Professor, School of Life & Allied Health Sciences Glocal University, Saharanpur Uttar Pradesh - 247121
Email: faheembly@gmail.com

Abstract

Significant technological developments in this field have been made possible by the idea of using wastewater as a replacement for scarce water supplies and to protect the environment, which has produced a wealth of physical data, including chemical, biological, and microbiological information. After looking at this data, wastewater treatment systems become more understandable. Several studies use "machine learning (ML) algorithms" as a proactive approach to tackle issues and forecast how these processing systems will function while using gathered experimental data. This article's objective is to extract the most widely used "machine learning models" from scientific publications in the "Web of Science" database using textual analysis techniques, then examine their applicability and evolution over time. This will provide a comprehensive overview of articles that address the application of artificial intelligence (AI) to address issues in "wastewater treatment technologies", as well as a global scientific follow-up. According to an analysis of research publications, the machine learning models most commonly used in the wastewater treatment domain are "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". The findings show that the primary publishers of publications on this study issue are developed countries. Finally, because machine learning requires a large amount of high-quality data, it helps with wastewater treatment. Furthermore, machine learning models' limited interpretability makes it more challenging to comprehend the underlying mechanisms and decisions in wastewater treatment. The AI algorithm for wastewater treatment is also examined in terms of the explainability of data-driven models for transfer learning and reinforcement learning, as well as other significant voids and future prospect. Enhancing energy efficiency, adhering to increasingly strict water quality laws, and optimising resource recovery prospects are just a few of the many issues associated with wastewater treatment. The operational and economic performance of "wastewater treatment plants (WWTPs)" has improved as a result of computational models' growing recognition as effective solutions for solving these difficulties. The paper also covers the use of several "AI algorithms" in "wastewater treatment plants (WWTPs)", such as identifying anomalies, optimising energy utilisation, and forecasting WWTP effluent characteristics and wastewater intakes.

Keywords; Artificial Intelligence, Machine Learning, Wastewater Treatment, Bioremediation, Aquatic Plants, Optimisation Methods, Algorithms, AI-powered.

INTRODUCTION

These days, machine learning is a popular term. it is evident that artificial intelligence (AI) encompasses machine learning (ML). This technology is designed to generate a multitude of practical and potent tools for the modelling of intricate problems in a variety of disciplines, as well as the simplification of complex systems into more manageable applications. Without having to comprehend the behaviour of natural events or physical systems, several researchers from a variety of disciplines have employed this method in the literature because of its adaptability, quickness, and acceptable accuracy.

In geoscience, machine learning is applied in hydrology to forecast severe occurrences and calibrate satellite-based precipitation products, and in geology to distinguish between different lithological facies. The use of AI in financial services and risk management is a hot topic in the financial industry, which has prompted several authors to look into it. In the field of medicine, and particularly in "the medical profession, machine learning algorithms" are crucial in the pharmaceutical industry, behaviour analysis, and epidemiology.

The types of wastewater, the attributes of the tributaries, the quantity of pollutants from various places, and "the geospatial and climate conditions" vary considerably from one treatment plan to the next. This factory transforms wastewater treatment into a complicated process that is controlled by noticeable fluctuations in physical, chemical, and microbiological variables. Many scientists are searching for straightforward methods to comprehend and regulate the behaviour of these systems using mathematical models due to "the stochastic character of natural phenomena and anthropogenic activities". Chemical reaction modelling may be separated using a number of models found in the literature. The treatment systems are supported by biological, physical, and other processes. To accomplish complicated global modelling, these different model typologies must be coupled.

In general, there are many "applications of machine learning in wastewater treatment" that maximise energy usage in wastewater treatment facilities, hence increasing the operational efficiency of wastewater treatment processes. The enhancement of treatment processes' efficacy is a critical "application of machine learning in wastewater treatment". By modifying treatment procedures in response to wastewater quality predictions, "machine learning" may also be utilised to improve treatment processes.

Another application of machine learning in water treatment is the reuse of wastewater. The best conditions for wastewater reuse may be found by using machine-learning analysis of wastewater properties. However, it is crucial to emphasise that when artificial intelligence is applied to wastewater treatment, extensive data preparation is required. This crucial and often necessary phase is essential to guaranteeing the precision and dependability of analyses and outcomes produced by machine learning models. To ensure the quality and applicability of the results derived from these intricate databases, particular attention must be paid to this crucial topic. Ensuring the accessibility of clean water supplies and preserving environmental health depend heavily on wastewater treatment. To solve these issues and improve many elements of wastewater treatment, artificial intelligence (AI) techniques have emerged as a viable option.

Artificial intellect (AI) is the umbrella term for a variety of computational technologies that allow robots, tools, or computers to carry out activities that normally require human intellect, including "learning from data, pattern recognition, decision-making, prediction, and condition adaptation".

In the end, machine learning has several applications in wastewater treatment, each with a distinct purpose based on the problems it tackles and the goals it seeks to accomplish.

DEFINITION OF ARTIFICIAL INTELLIGENCE

The phrase "artificial intelligence" refers to computer systems that imitate human intelligence. Included in these processes are information acquisition and utilisation, reasoning (which involves the application of principles to reach precise or approximate conclusions), and autocorrelation. "Machine learning (ML), natural language processing (NLP), computer vision, robotics, expert systems", and other fields are all included in the broad category of artificial intelligence (AI). The fundamental objective of artificial intelligence is to develop intelligent systems that are capable of performing tasks that typically necessitate human capability, such as the comprehension of natural language, the identification of patterns, the formulation of judgements, and the resolution of complex problems.

The environment, transportation, health care, and education are just a few of the many domains where AI has applications.

MATERIALS AND METHODS

It is challenging to search for certain topics due to the enormous volume of scientific publications. Filters to target publications by authors or keywords are available on "the Web of Science platform". Nevertheless, this platform is still not adequate to extract the most information.

The paper's technique (Figure No. 01) is based on importing data from the Web of Science platform and filtering our scientific niche using terms like wastewater, artificial intelligence, and machine learning. The collected data is semi-structured. They can be processed through text mining techniques to create structured data, from which the desired pertinent information can be obtained, such as "the geographic distribution of the target publications, the most popular models, the chronological timeline of the chosen topic, the word cloud, and so on".

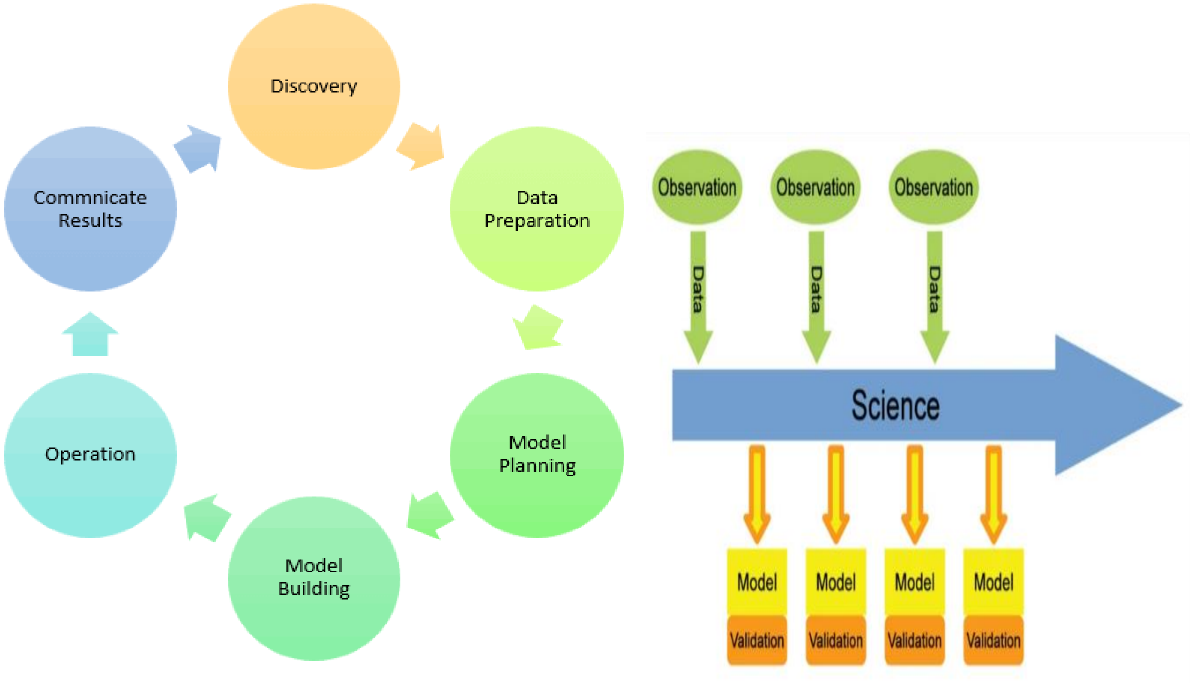

Figure No. 01: "Web of Science database processing and analysis methodology techniques"

Database: Web of Science

The information utilised in these papers was taken from Web of Science, the most reliable worldwide citation database. The most dependable search engine is thought to be "the Web of Science platform". In order to give users flexible access, confident evaluation, and discovery, it provides "a collection of the best publication and citation data", organised into multiple categories and criteria.

Using more than 1.7 billion referenced references in more than 160 million entries, the transdisciplinary Web of knowledge platform traces the evolution of knowledge across time in all fields. The Web of Science's capacity to produce highly regarded research within the scientific community, gather the necessary information, and make transparent decisions that influence the future of their institutions and research strategies is critical to the credibility of institutions and millions of researchers.

Text Mining

Text mining, often known as text analytics, is a group of methods and tools for analysing semi-structured or unstructured information. Taking ideas and information from a textual dataset is what it is called. Researchers can locate pertinent information in textual datasets and uncover hidden correlations and patterns among words by using text mining.

A paper, a post on a social network like Twitter, a library in the form of "a semi-structured Excel spreadsheet" (like in this article), or other forms can all be used to provide this material. The next mining technique allows us to schematise the relationships between terms and view clouds of the most often used terms in a textual database.

RESULT

Publishing Trends: - Figure 02 illustrates the scientific production trend. The significance of this study area is highlighted by the fact that over 700 articles on wastewater treatment using numerical approaches have been published. There were very few publications on this subject per year between 1991 and 2017.

In the meantime, between 2017 and 2022, the number of publications in this sector has increased. However, the results indicate that the yearly number of publications increased exponentially between 2021 and the present, reaching a maximum of 1991 in 2023. Prior to 2017, academics throughout the world did not give significant weight to using machine learning models and numerical techniques to solve wastewater treatment issues, according to an examination of publication patterns. The growth in scientific output after 2018 can be attributed in large part to the development of computing devices, data processing and analysis software, and new numerical and statistical methods.

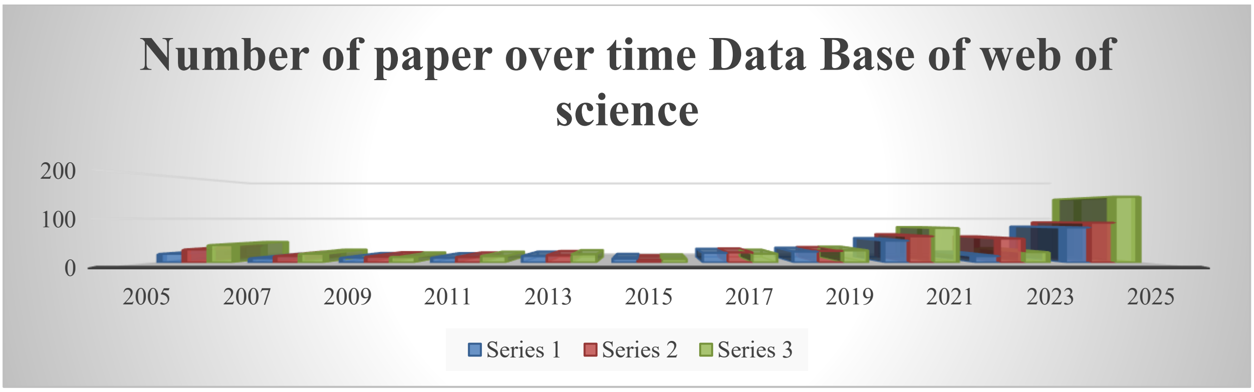

Figure No. 2: "Temporal trend in scientific literature involving the use of AI".

Publications in the world

The global paper network distribution was developed to conduct a comprehensive analysis of the global distribution of scientific articles, using the addresses of the authors and co-authors and international scientific collaboration (Figure 3). According to the map, over 605 publications have been published in the previous 30 years on the use of numerical modelling for wastewater treatment in 81 different nations. The quantity of papers was significantly larger in industrialised nations than in emerging nations, making it unique among nations. Every continent—aside from Antarctica—has published stories about artificial intelligence's application in wastewater treatment over the past 30 years, demonstrating how widely accepted it is. Furthermore, "the USA, Canada, China, India, Saudi Arabia, Iran, Egypt, Turkey, and Australia" have published more articles than any other country, as shown in Figure No. 3.

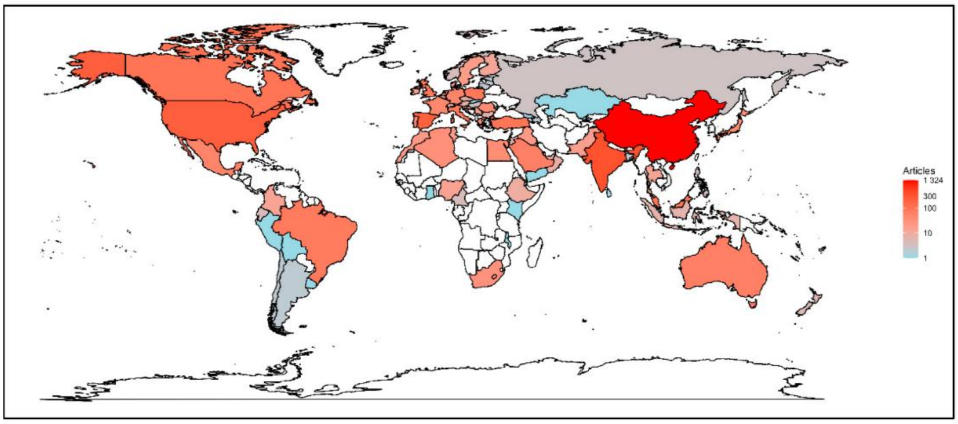

Figure 3: "Worldwide distribution of scientific papers on the application of artificial intelligence to wastewater treatment".

Subsequent investigation showed that 35% and 65%, respectively, of the publications were international and national collaborative. These results demonstrate a well-known international collaboration in the use of "machine learning algorithms in wastewater treatment". In terms of paper kinds, 605 publications made up 78.84% of submissions, with proceedings coming in second at 10.58% and review documents at 10.08%. The remaining papers, which included editorial material and database evaluations, on the other hand, were less expressive (0.5%).

In terms of the language component of the documents, the study reveals that 99.7% of the publications are in English. At the same time, just 0.03% of people speak foreign languages.

Collaborations: - It is imperative to engage in international collaboration with a minimum of two scientists from different countries in order to expedite scientific advancement, offer a diverse range of perspectives that can result in innovative solutions, and achieve a common objective in scientific research. This cooperation might take place across several scientific domains or within a particular subject.

The cooperative connections between the nations in each region are shown in Figure No. 4. Each nation is represented by a section, the size of which indicates how many publications it has generated. A cooperative relationship between two connected countries is represented by a curve connecting the nation segments; a wider curve indicates a more robust cooperation. Overall, the results demonstrated that "Canada, the United States, Spain, India, China, Saudi Arabia, Iran, Egypt, and Turkey" had the biggest number of publications globally, the strongest collaboration, and a clear promotion of scientific production in the use of "machine learning models in the wastewater area". In terms of regional scientific cooperation, Sub-Saharan Africa and the Caribbean had the fewest collaborating nations, while East Asia and the Pacific had the most. Therefore, the number of countries in each region, the level of development, the collaborative network, and each country's scientific interest could all account for this disparity.

A cooperative relationship between two linked nations is represented by a curve connecting the nation segments; a wider curve indicates a more robust cooperation. Additionally, the Islamic curve shows that the following nations dominate the geographical pattern of international collaboration, indicating greater relationships. Additionally, the following nations dominate the geographical structure of international cooperation: "Saudi Arabia, Iran, Pakistan, and Egypt in the Islamic world; China and India in Asia"; the United States and Canada on the American continent; and Spain and England in Europe. Conversely, countries in the Sahara and Latin America show little interaction with other regions of the world.

Scientific cooperation is frequently crucial when using machine learning models for wastewater in several ways:

- **Knowledge exchange:** By sharing their knowledge and skills, researchers may learn from one another, get access to resources and information that would not otherwise be available, and advance the topic in question.
- **Higher research quality:** A more thorough and impartial research process results from collaboration. Collaboration and peer review contribute to the accuracy, repeatability, and dependability of findings. Additionally, it enables researchers to independently replicate their findings, strengthening the foundation of evidence.
- **Increased funding opportunities:** Because collaborative research is seen as more innovative and effective, funding organisations are frequently more inclined to support it:
- **Resource accessibility:** It provides scholars with access to a wider range of resources, such as databases, specialised tools, and research facilities. This access might be highly beneficial for early-career academics who do not have the resources to conduct certain types of research.
- **Diverse perspectives and innovation:** Researchers from various disciplines and backgrounds must work together to bring diverse viewpoints and ideas to the table in order for new study fields, novel methods, and new methodology to emerge:
- **Addressing environmental and climate challenges:** To solve the problems facing humanity, multidisciplinary and global cooperation is essential.

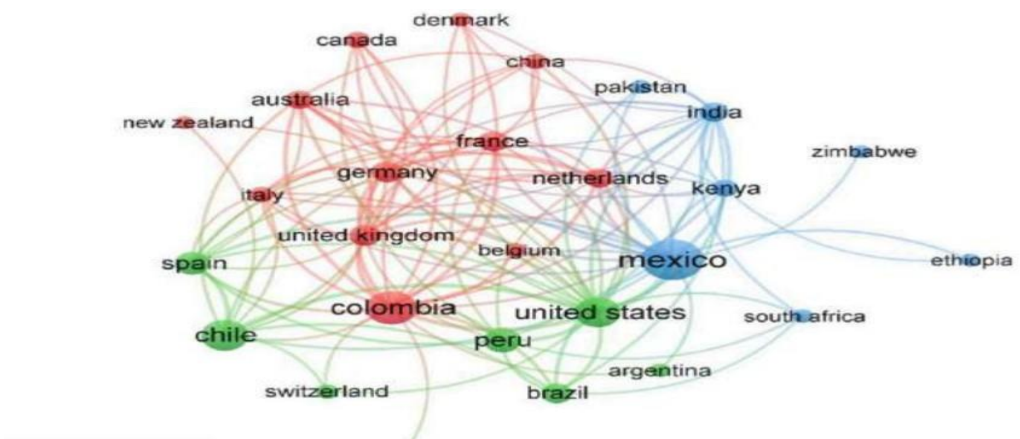

Figure 4: "Network visualisation for country collaboration by world region".

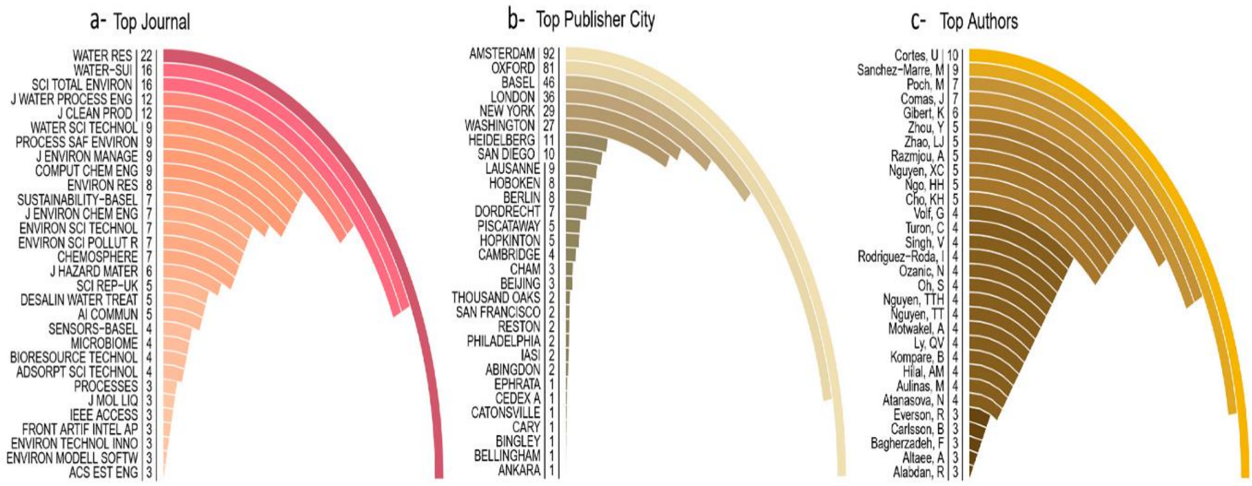

Figure .5: "(a) Most productive journals, (b) publisher city, (c) authors in the application of artificial intelligence in wastewater treatment".

Publishing Journals

There were 261 distinct publications that published articles on the applications of "artificial intelligence in wastewater" after treatment. 5.43% of journals have produced less than 12 research papers in the previous three decades, according to an examination of the total number of articles published in each magazine. Figure No. 5a displays the top 30 publishing journals. With an average of 13 publications, these journals published about 245 papers, or 40.50% of all publications between 1991 and 2004. This mean value appears to be rather low, indicating that the use of AI to wastewater treatment is still relatively new and popular.

The United States accounts for 43% of these 40 journals, which are split rather evenly across five nations: "Switzerland, the Netherlands, the United Kingdom, the United States, and Germany" (Figure 5b). Additionally, the fact that these five nations are developed may indicate that the establishment of reputable journals in this sector is not valued in emerging nations.

The 30 authors who are most active in the wastewater domain application of artificial intelligence are depicted in Figure 5c, which concludes this section. With six to fifteen publications, “Cortes, U., Sanchez-Maree, M., Poch, M., Comas, J., and Gibert, K”. are the highly regarded writers. According to the data, the majority of the authors are from Spain, which highlights their interest in these research issues.

Optimisation Methods

In order to calibrate models that learn the mechanics of a system or phenomena from real data, and so make correct predictions, optimisation techniques are essential components of machine learning. Optimisation in machine learning is maximising or minimising a cost function to determine the ideal values for the model's parameters.

However, the model parameter optimisation phase may be challenging due to "the high-dimensional search space or complex cost functions". By fusing stochastic models and optimisation techniques, stochastic optimisation methods offer a potent framework to tackle these issues. The “type of model, the volume of data, and the computational resources” at hand all influence the optimisation technique selection. In order to improve performance, improve models, and shorten calibration times, machine learning researchers continue to create and develop optimisation strategies.

Due to their extensive use in a variety of research fields, including wastewater treatment, optimisation techniques have garnered more attention in recent years. It is utilised for wastewater treatment in the literature. The most well-known of these techniques is the Genetic Algorithm (GA), which has been referenced in 20 articles with 28 citations. Particle Swarm Optimisation (PSO) has been cited in 6 articles with 7 citations, while Gene Expression (GE) has been cited in 3 articles with 3 citations. One appearance distinguishes the remaining algorithms.

- The GA is a probabilistic optimisation technique that draws inspiration from genetics and natural selection. It uses a starting population of possible solutions that go through a number of processes related to selection, reproduction, and mutation.
- A metaheuristic optimisation method called PSO mimics the behaviour of social animals like fish and birds. This means using a swarm of particles that move through the search space and change their positions based on their experiences as well as those of their neighbours.
- The GE is an autonomous programming technique that combines symbolic regression with genetic algorithms to evolve computer systems.

Machine Learning Models

A growing number of scientific research topics are prioritising machine learning. They wish to use vast amounts of data to model the behaviour of complicated systems. The objectives stated and the level of experience of the data manipulators determine which is best. One area of study is the use of these algorithms to ascertain their function in wastewater treatment system modelling. This study's main objective is to utilise “textual analysis” to determine which machine learning models are most popular in “the wastewater industry”.

A list of AI models is produced for extraction once the abstract text has been preprocessed and cleared, and then a unique count of publications that cite each model is produced. The findings indicate that "Artificial Neural Network (ANN), Random Forest (RF), Support Vector Machine (SVM), Linear Regression (LR), and Decision Tree (DT)" are the most popular models in the wastewater area. Furthermore, the data indicates that after 2019, the temporal development of these models' adoption accelerates.

Some models show a 78% rise in application level throughout this phase, while others do not surpass this threshold. The primary “machine learning models” that are currently in use for wastewater treatment are reviewed in this section.The most popular models utilised in “Web of Science articles” are explained in detail for each of them. The explanation of each model concludes with a list of citations that provide instances of the research that the model has supported or utilised. AI uses artificial algorithms to carry out challenging jobs.

Promising advanced “machine learning algorithms include Random Forests, Support Vector Machines (SVMs), and Artificial Neural Networks (ANNs)". In order to facilitate effective degradation processes, methods like clustering and classification aid in the identification and characterisation of various microbial populations. ANNs and ensemble approaches, which may predict "microbial community dynamics" and explore the relationship between environmental conditions and the microbial community, assist effective bioremediation.

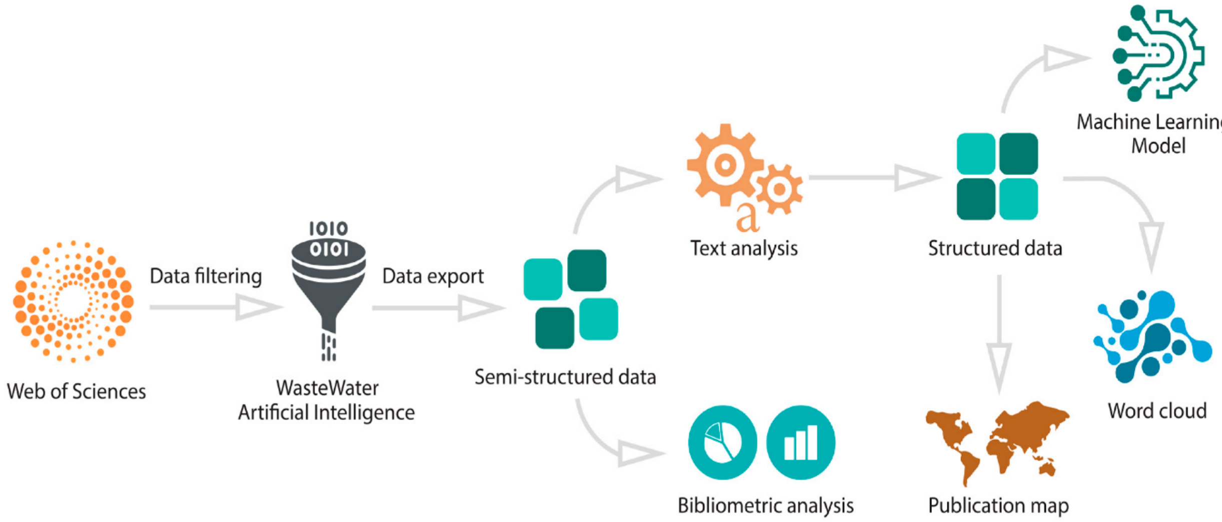

Figure no. 6: "The most used machine learning models in the field of wastewater".

Consequently, there are two types of "machine learning models: supervised and unsupervised". Regression or classification models are both possible if the model is supervised.

Artificial Neural Networks (ANN)

ANN is a subfield of neurocomputing and computer science research. Its concept is based on "the structure of biological neurones in the human brain". This paradigm comes in a variety of forms, each providing a unique information processing approach. An ANN may be conceptualised globally as a collection of neural networks with sufficient artificial neurones in each layer. We are used as an example in several wastewater research initiatives, such as the monitoring and design of biological wastewater treatment systems and "the optimisation of energy consumption in the sludge incineration unit of wastewater treatment systems". Forecasting the effectiveness of associated systems and wastewater treatment facilities.

Random Forest (RF)

RF is a decision tree-based ensemble learning method. The foundation of this approach is the use of split subsets of the original data to create many decision trees. At every decision tree level, a random collection of variables is chosen. Next, for every prediction in every decision tree, the model chooses the mode. In wastewater-based epidemiological applications, RF may be used to estimate and assess the operation of wastewater treatment systems, forecast their input, and examine energy consumption in wastewater treatment facilities.

Support Vector Machine (SVM)

By rephrasing nonlinear issues as quadratic optimisation problems, an SVM, a supervised classification technique, permits their treatment. Its goal is to make these resolutions easier. Two primary axes serve as the foundation for this categorisation method:

- "The notion of maximum margin".
- "The notion of a kernel function".

To put it briefly, "the optimum margin is the data-separation border that maximises the distance to the next data point", and the kernel function serves as a stand-in for a scalar product in a very high-dimensional space. This model was used to assess how well "a large-scale aerobic biological wastewater treatment system removed nitrogen" and to provide an indirect method that computes important wastewater quality indicators using drainage basin criteria. Other searches utilise this model with conventional tuning parameters to identify industrial pollutants, and it also simulates the unpredictable nature of "chemical phosphorus removal processes in wastewater treatment facilities".

Linear Regression (LR)Research

The simplest statistical modelling technique is still LR, without a doubt. Finding the line that best matches the output data is the basic concept behind this approach. Polynomial regression and multiple linear regression are included in the LR. Although the conceptual framework and computation techniques are the same, there are often two forms of "regression: simple regression (one explanatory variable) and multiple regression (many explanatory variables)". Attowatt et al. employed linear regression to elucidate Zn2+ adsorption at "the Nicosia wastewater treatment plant", while Nguyen et al. employed the same model to investigate the adsorption mechanism of "pharmaceuticals onto biochar". The model is applicable to a variety of wastewater research topics.

Decision Tree (DT)

DT is a widely used model in machine learning, "operations research, and strategic planning". Its guiding principles are based on a diagram that conceptualises the potential results of a number of interrelated decisions. Typically, the model starts with a node from which further options emerge. The resultant graphic conceptualises a tree's form. Predicting total nitrogen in effluent from activated sludge systems, maintaining and improving wastewater networks, "forecasting water quality, and regulating sewage discharge in wastewater treatment systems" are just a few of the applications for the decision tree model.

Logistic Regression

Logistic regression is a popular supervised classification method in machine learning, much like linear regression. If not, it is a logistic-linked generalised linear model. It is used to simulate the likelihood of a limited number of possible outcomes. By improving the regression coefficients, it forecasts the likelihood that an event will take place. This result is always in the range of 0 to 1. When the expected value exceeds a threshold, the event is bound to occur; when it falls below that threshold, it is unlikely to occur. Numerous wastewater-related concerns are covered by the model. Robles-Velasco et al. concentrate on the estimation of sewage pipe failure, whereas Suchet Ana et al. use logistic regression to predict "the total inorganic nitrogen in treated wastewater". To determine the best wastewater parameters for process control, modelling, and operational monitoring, "Szelag et al. used a logistic regression model". While Bencke et al. developed models that automatically classify citizen reactions according to the variety of municipal service features (e.g., water, wastewater), Liu et al. employ machine learning to predict "groundwater flooding (or infiltration) into sewage networks".

K- Nearest Neighbours (KNN)

The K-nearest neighbours' algorithm, also referred to as KNN or K-KNN, is a non-parametric supervised learning technique that uses proximity to predict or categorise groups of individual data points. Although it may be used for either regression or classification, it is commonly employed as a classification approach, provided that similar neighbouring points can be found. The method had been used to many other subjects. The model was used by Ravi et al. to investigate and comprehend the underlying causes of septic failures. Zidan et al. used the KNN model to predict the overall removal of coliform from urban wastewater by an on-site, two-stage, multi-soil-layering plant. At a large-scale wastewater treatment plant, Hamoud and Wasel Allah employed the KNN algorithm to forecast the effluent's total nitrogen content and water quality.

Timeline of Machine Learning Model in Wastewater

In Figure 8, a comprehensive, chronological examination of the machine learning model employed in effluent treatment is presented. The evolutionary history, the first machine learning model citations or applications, and the total number of publications referencing each model are all summarised in this paper. These methods first appeared in wastewater in the 1990s, despite the recent austerity in the usage of ML models. Figure 8 illustrates that fuzzy logic (FL) and artificial neural networks (ANN) were the first models utilised in 1992. In the history of "artificial intelligence's" application to wastewater, these models are the earliest. The FL model has a consistent chronological trend, but both models have a more noticeable lacunar chronological pattern. But after being used for the first time, "the ANN model" disappears from the papers for ten years before reappearing with an exponential chronological evolution.

A number of models, including "Hierarchical Clustering, Factor Analysis, K-Means Clustering, Multi-Layer Perceptron Regression, Feed-Forward Neural Network, Partial Least Squares, Logistic Regression, PCA, Nearest Neighbour, Linear Regression, Regression Tree, and Support Vector Machine", emerged between 2002 and 2013.

"Cubist, Ensemble Learning, Gaussian Mixture Model, Light Gradient, Boosting Machine, Multivariate Adaptive Regression Splines, Naïve Bayes, and Stepwise Regression, seasonal auto regression integrated moving average, deep neural network extreme learning machine, radial basis function neural network begging self-organising map, random forest regression, conventional neural network lesson regression, and cat boost" are among "the new machine learning models used in the wastewater domain".

Over the past several years, these models have generally become more involved, which reflects the wastewater treatment industry's adaptability to scientific research trends.

Analysis of Machine Learning Model Usage

Wastewater treatment has undergone a radical change as a result of the incorporation of new technologies, especially "machine learning models", which have made significant advancements in a number of areas. But it's crucial to understand that, like any technology, incorporating these modelling techniques into "wastewater treatment" has benefits and drawbacks that need to be carefully considered.

Machine learning algorithms typically use data to find patterns and make predictions. They are taught using a sizable database, which enables them to extract data and recognise intricate linkages that would not be apparent using conventional modelling methods. On the other hand, traditional modelling models, including "conceptual and physical models, are based on theoretical ideas or empirical data". Therefore, the goal of "machine learning models" is to generalise the relationship found in training data to new data. They aim to accurately forecast new occurrences and capture underlying notions. Classical models may not be applicable to other situations or contexts since they only depict certain occurrences.

However, machine learning algorithms are well-suited to managing big, varied datasets since they are flexible and adaptable. Although deterministic models offer concrete representations, their flexibility and scalability may be constrained. The intricacy of model conceptualisation is one drawback of machine learning models. Neural network families, for instance, are frequently accused for being difficult to understand. These models' decision-making process is unexplained since it might be difficult to comprehend how they arrive at their forecasts. Classical models, on the other hand, are usually made to be interpretable so that users may clearly comprehend the underlying relationships, assumptions, and operations.

The fact that machine learning models frequently need a lot of computing power, particularly when training intricate models or handling massive volumes of data, is another drawback.

Strong and potent hardware, storage, and computing power may be needed for these models. However, because deterministic models are abstract representations, they require fewer assets than machine learning models.

Despite several important studies on the application of artificial intelligence and machine learning models in wastewater treatment, the crucial problem persists. Despite persistent efforts to enhance their design and functionality, "AI-assisted systems" are still not commonly used in treatment facilities. To address this gap, researchers and industry must work closely together. They must work together to create AI-assisted systems that are especially suited to these facilities' everyday operations. The technological, financial, and legal obstacles that presently prevent the application of AI in this crucial area of "water management" will be addressed by this strategy. Researchers and business experts may make substantial strides toward improved management and long-term conservation of our water resources by collaborating.

High-frequency Keywords and Groups

Keywords are crucial for emphasising the main goal of papers and making mapping easier for readers. We must preprocess the abstract to remove noise and unnecessary information before we can extract information from it in our database. Typically, this procedure entails cleaning up data by removing any non-informative language, digits, punctuation, and blank words. To guarantee uniform spelling, the material must also be standardised and divided into words or phrases. The foundation of this standardisation is the conversion of capital characters to lowercase. The keywords are examined and assessed using R tools to show their frequency and associations based only on "the textual processing of the abstracts in our database". A minimum correlation of 0.25 and a minimum of 25 keywords were required. The keywords are shown in Figure 09, which highlights the connections between various fields by compiling a list of terms associated with artificial intelligence and water.

Figure 09's overall interpretation led us to the conclusion that machine learning and modelling approaches have been essential to the recent rise in "the use of AI in the wastewater industry". Reusing wastewater for irrigation is one area where AI has made a big difference. AI-directed or AI-supervised technologies can enhance irrigation, guaranteeing crops receive the proper amount of water and nutrients while consuming less water. Better water management and resource conservation have benefited from this.

Water resource management has also been made easier by AI-hosted and AI-driven technology, guaranteeing adherence to quality requirements. Water quality has improved thanks to the employment of AI in quality management, making it suitable for agricultural and human use. Furthermore, resource recovery, sustainability, and environmental protection have all benefited greatly from AI technology.

Wastewater infrastructure management is another area where AI has had a significant impact. By predicting the lifespan of infrastructure components, AI-guided technologies enable asset managers to schedule maintenance and replacements before they break. Because of the decreased expenses and downtime, wastewater management is now more economical and effective. AI technology have been utilised to lessen the effects of climate change, which is a major waste management concern. AI-model tools may lower greenhouse gas emissions, encourage energy efficiency, and improve operations. By enabling the effective use of water resources, these techniques can also aid in managing water scarcity.

Typological Analysis of Models in the Wastewater Fields

Maintaining public health and environmental sustainability depends heavily on wastewater treatment. The use of cutting-edge technology, such machine learning, has drawn a lot of attention as waste management becomes more complicated. Depending on the difficulties encountered, these models are used differently in each study. This variation in model utilisation is heavily influenced by “the typology of the data”. These models may be broadly categorised into four groups: “Clustering, Dimensionality Reduction, Regression, and Classification”.

Input data is assigned to predetermined classes or categories in the first category, classification models. Applying classification models to wastewater treatment is the subject of several studies in the literature. For instance, Choi et al. identified smell sources in metropolitan areas using DT and RF models. Wodicka et al. used SVM and Ribalta et al. used a case study to forecast wastewater quality fluctuation at wastewater treatment facility inlets.

In the second class of regression models, continuous numerical values are predicted using input components. “Six machine learning models”, including an ANN, were utilised to compare the research that employed the Zhao et al. regression models. Based on anoxic water "oxidation-reduction potential (ORP) data, the optimal model for predicting the readily biodegradable chemical oxygen demand (RB-COD) and the slowly biodegradable chemical oxygen demand (SB-COD) in municipal wastewater" is selected using SVR, an ANN-based AdaBoost, an SVR-based AdaBoost, DT, and RF. Granata et al. used "the M5P regression tree, bagging, and random forest algorithms" in another case study to handle complex wastewater engineering problems, such as forecasting lateral flow in a spillway and predicting energy loss.

Models for regression include time series. This kind of model, which is frequently used in the wastewater industry, detects and forecasts patterns and relationships in sequential data across time. Time series analysis frequently employs "autoregressive integrated moving average (ARIMA), long short-term memory (LSTM) networks, and recurrent neural networks (RNNs)". In order to assess the predictive power of linear and nonlinear models for complex processes, “Nourani et al. used the ARIMA linear model to predict the parameters of chemical oxygen demand (COD-eff) and biological oxygen demand (BOD-eff)”. Kang et al. forecast “wastewater flow rates for a municipal wastewater treatment” facility using bidirectional “long short-term memory (LSTM)”. Another study by Dairi et al. suggests a method for identifying abnormalities in wastewater treatment plants by combining clustering algorithms with deep learning techniques (recurrent neural networks).

Dimensionality reduction is the third category. Many models are used to minimise the quantity of features in a dataset while retaining important information in order to escape "the curse of dimensionality and boost computing performance". One popular technique for reducing dimensionality in wastewater treatment is principal component analysis (PCA). In order to assess the absorption effectiveness of various metal oxide nanoparticle (MONP) composite hydrogels for wastewater dye treatment, Murshid et al. used PCA. Other machine learning models' inputs were also chosen using PCA.

“Clustering models”, which are “unsupervised learning methods” for grouping related data according to shared traits, make up the fourth category. The goal of clustering is to identify underlying structures or hidden connections in the data without any previous labels or classification. The literature describes a number of applications of "clustering models in wastewater treatment plants". In a separate research, Navato et al. evaluated four models—"K-means, GMM, Hierarchical, and DBSCAN clustering"—to identify anomalies in the complex chemical environment of wastewater.

A typological classification of publications is carried out in order to comprehend “the scientific direction of using machine learning models” in wastewater treatment. The results are summarised in Figure No. 10.

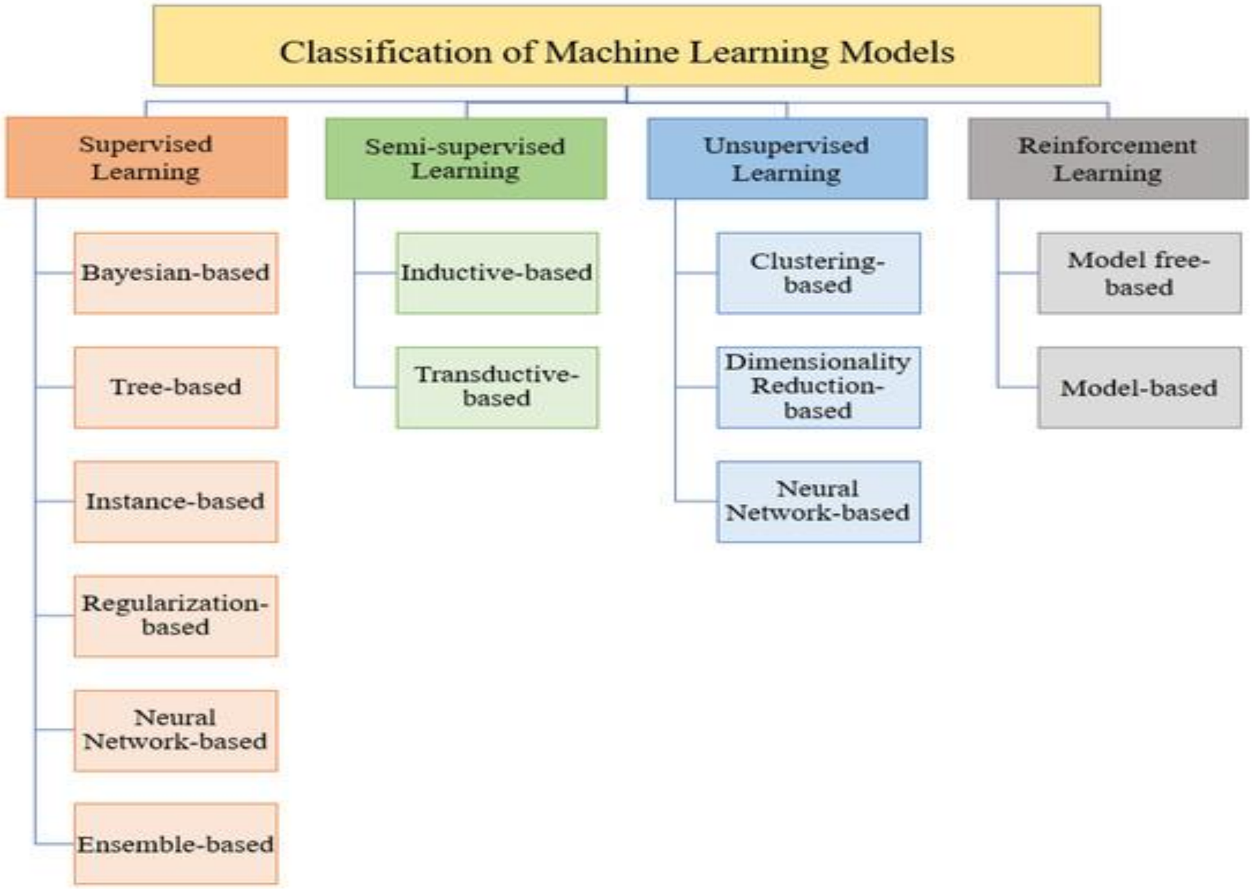

Figure no. 7: "Classification of machine learning models".

Based on the availability of labelled data, machine learning models may be broadly divided into two major categories. Supervised learning models are those that need labelled data for training, while unsupervised learning models do not. In Figure 9, 52 of the 67 models are classified as "supervised learning models, while the remaining 15 are classified as unsupervised learning models".

Examples of machine learning models that might be enhanced by classifying them according to their characteristics include "regression, clustering, classification, and dimensionality reduction models". Classification models forecast categorical values, whereas regression models forecast continuous values. Clustering methods separate the data into subsets or clusters, whereas dimensionality reduction techniques reduce the amount of features while preserving the structure of the data. Of the "67 models, 22 are regression models, 11 are classification models, 10 are clustering models, 6 are dimensionality reduction techniques", and 18 may combine regression and classification models.

Choosing the optimal machine learning approach depending on the type of data and the issues at hand requires an understanding of these typologies. By using "regression, classification, dimensionality reduction, or time-series analysis techniques", researchers and practitioners can efficiently analyse data, produce precise predictions, and obtain valuable insights.

AI IN MICROBIAL SELECTION AND OPTIMISATION

AI uses artificial algorithms to carry out challenging jobs. Artificial Neural Networks (ANNs), Random Forests, and Support Vector Machines (SVMs) are promising machine learning methods. In order to facilitate effective degradation processes, methods like clustering and classification aid in the identification and characterisation of various microbial populations. In order to advise effective bioremediation solutions, ANNs and ensemble approaches may anticipate "microbial community dynamics" and explore the relationship between environmental conditions and the microbial community, as seen in figure 1.

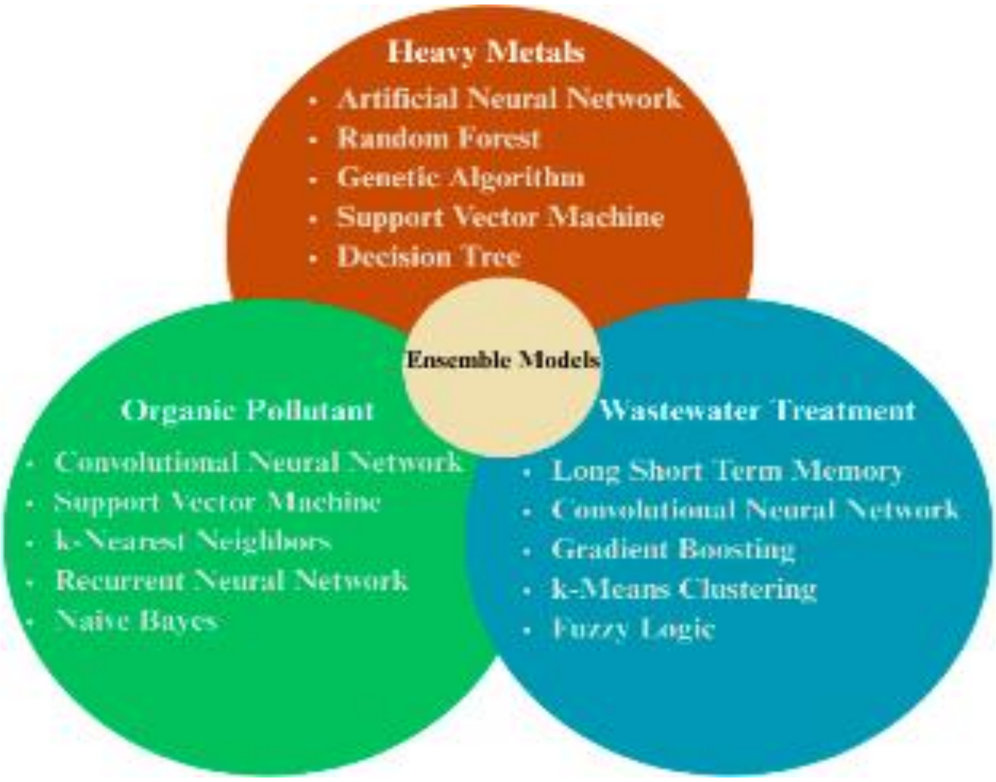

Figure no. 8 :- "Machine learning models used in heavy metals, organic pollutants and wastewater treatment with ensemble models at the intersection".

According to a new study, random forests are crucial for forecasting shifts in the bacterial microbiota in a variety of polluted settings. The authors gathered six variables for analysis and used the "Web of Science to assemble data". To predict the effects of "microplastics on bacteria that break down antibiotics", three algorithms were put into practice. With AUC values of 90% and 93%, "the random forest model" outperformed the others. When combined with deep learning algorithms, AI-driven sensors allow for real-time monitoring and reveal information about the catalytic activity of enzymes. Biosensors and learning tools can help predict long-term treatment effects, according to another study. Using in "silico analysis (ligand-binding site modelling, molecular docking, and molecular dynamics simulations)", researchers evaluated the interaction between our commercial dyes and Bacillus megaterium H2 Az reductases. The binding findings that were obtained indicated that the complexes were stabilised. Additionally, a hypothesis regarding the use of nano-biosensors for pesticide detection in the field and their modification and exploitation for bioremediation was presented by a group of researchers. Factors influencing the results of bioremediation are modelled using random forests. In contrast, the complex architecture of artificial neural networks and their composites renders them less interpretable, despite their high accuracy. In practical deployment, advanced techniques such as SHAP and LIME are employed.

One study employed "22 metagenomic and genomic datasets of microbial communities" in conjunction with AI and ML algorithms to enhance the breakdown of environmental contaminants and toxins and to uncover the genetic and functional possibilities of these communities. Some research have shown that machine learning (ML) can anticipate degradation routes by examining sequences. Simultaneously, AI-based functional profiling helps improve bioremediation processes by identifying important taxa and enzymes that cause degradation. Genes, metabolites, and proteins are examples of biomarkers that can show the existence and activity of particular microbial populations that can break down contaminants. Two distinct study groups working at various times have shown that algorithms like "SVMs, ANNs, and RF" may uncover potential biomarkers by finding "correlations between microbial activity and pollution concentrations". Figure 2 illustrates these ML integration techniques for microbiome analysis.

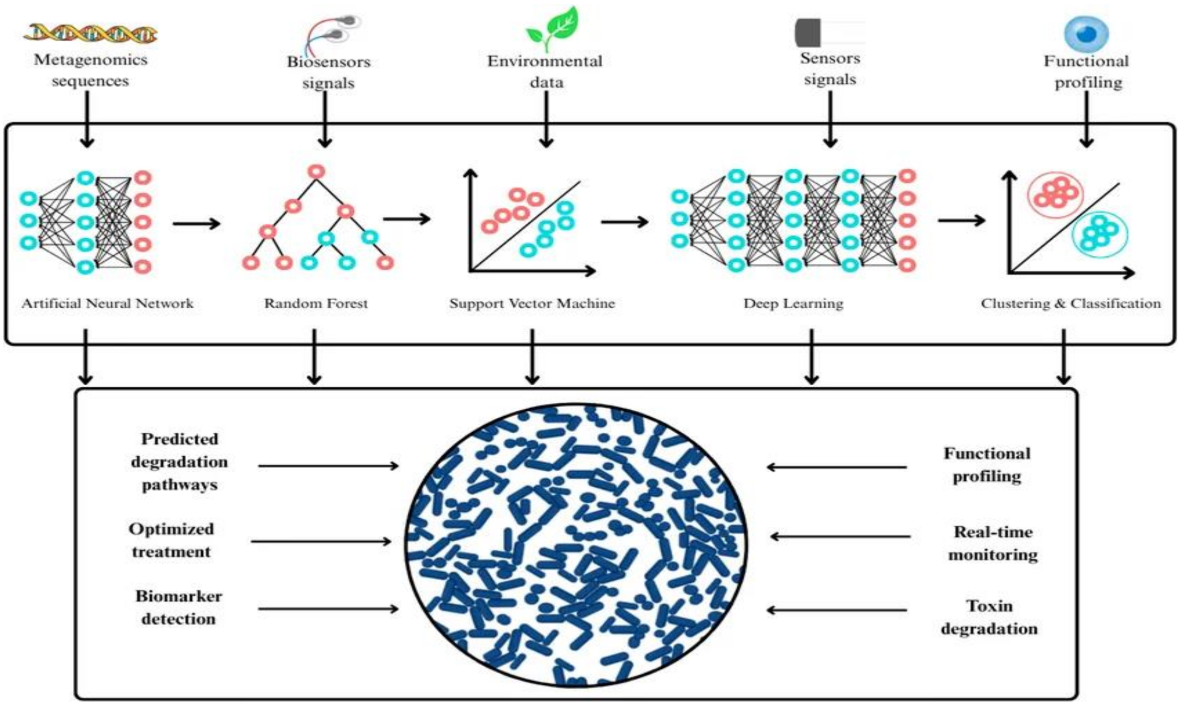

Figure no. 9ResearchResearch: "Integration of machine techniques in microbiome analysis for environmental and clinical applications".

CHALLENGES AND PROSPECTS

Moving toward a more effective, economical, and sustainable bioremediation solution is the main objective. In order to train AI and ML models, it is necessary to have "high-quality, comprehensive datasets, as incomplete or biased data" can result in inaccurate predictions. Another significant concern for AI and ML is the availability of databases. The current databases may not account for the complete "diversity of microbial communities and their metabolites".

OUTLOOK

Protein modelling, a crucial stage in metagenomics, is being carried out using "bioinformatics tools like AlphaFold, I-TASSER, and SWISS-MODEL". Metagenomics can now study "the genetic material of microbial communities", resolving issues and finding novel "enzymes, genes, and metabolic pathways" thanks to recent advancements in sequencing technology.

The advancement of omics depends on these online resources. "Artificial intelligence (AI) and machine learning (ML)" have demonstrated the capacity to resolve intricate challenges within bioremediation processes. In order to predict, model, and optimise water treatment processes to break down pollutants and toxins, machine learning algorithms have demonstrated improved performance. Data management, reproducibility, and collaboration are common issues and constraints found in many of the reviewed studies. Overcoming these obstacles will depend critically on the integration of "bioinformation, metagenomics, and machine learning". We can eventually help develop more successful bioremediation techniques and environmental solutions by filling up the gaps between these multidisciplinary collaborations.

CONCLUSION AND FUTURE WORK

The following significant conclusions may be drawn from this study (AI) on WWTPs: Forecasting is a component of almost all AI applications in WWTPs. WWTP effluents were among the results from AI models. Filamentous sludge, sludge bulking, "energy consumption, mass flow, COD operation in real time, TSS, BOD, and other factors". Simple, constructed, "hybrid models, supervised, and unsupervised learning algorithms" differ from AI models. SVM and IoT algorithms, including CNN, LSTM, hybrid models, and others, have been investigated for use in WWTPs. AI-based sensors were more significant and active as a practical support for the WWTP's operation; intelligent control increases the WWTP's efficiency and enhances it.

Although the R2 values and errors produced satisfactory results, the comparison of algorithms across different research was relative; T0T sensors were also essential in gathering a large number of data points for the development of "data-driven AI models". The datasets provide trustworthy data that helped the WWTPs function, particularly in the areas of anomaly and defect detection. In the wastewater industry, AI algorithms and methods generally performed better than traditional methods. More research is needed on the application of AI to image processing or analysis, such as identifying biofilms and sludge in WWTPs and forecasting effluent quality. The use of AI technology for picture analysis and pattern identification of different wastewater metrics will be crucial for improving WWTP performance and optimising maintenance operations. The following recommendations for the future are based on recent applied AI research on WWTPs: large-scale dataset sharing and cooperation must be enabled, and data solutions must be carefully designed and developed with a knowledge of "the data and processes involved in WWTPs". Future research on "WWTPs" should use more cutting-edge AI methods. Data-driven model detection is now feasible in "WWTPs with the availability of T0T and contemporary on-time sensors".

REFERENCES

[1] Salome, MS., Gilberto, M Pedro, O, Bruno, F, Eugenio, CF, Madalena, AM, Federico, LM, Alcina, P., & Paulo, N., ACS Es&t water, 4(3),784-804 (2023)

[2] Sarkar, IH., SN Computer Science, 3(2), 158 (2022).

[3] Tortajada, C., NPJ Clean Water, 3 (1), 22 (2020).

[4] Gil de Zúñiga, H., Goyanes, M., & Durotoye, T. Political Communication, 41(2), 317-334 (2024).

[5] Doorn, N., Science of the Total Environment, 755, 142561 (2021).

[6] M. S. Aliyar and O. O. Babalola, Bioremediation of Environmental Waste: The role of the Micro-organism, Front. Agron, 2023, 5, 1-15.

[7] P.K. Gupta, B. Yadav, A. Kumar and S.K. Himanshu, Machine learning and artificial intelligence application in constructed wetlands for industrial effluent treatment: Advances and Challenges in Assessment and Bioremediation Modelling, In Bioremediation for Environmental Sustainability, Elsevier, 2021, pp. 403-414.

[8] Y. Yang, A. Jerger, S. Feng, Z. Wang, C. Brasfield, M.S. Cheung, et al., Improved Enzyme Functional Annotation Prediction Using Contrastive Learning with Structural Inference, Common. Biol., 2024, 7 (1), 1995.

[9] H.A. Oyewusi, R.A. Wahab, K.A. Akinyele, G.M. Albadrani, M. Q. AI- Ghadi, M.M. Abdel, M. M. Abdel- Daim et al, Bioinformatics analysis and molecular Dynamics Simulation of Az reductases (AzrBmH2) from Bacillus Megaterium H2 for the Decolourisation of Commercial Dyes, Environ. Sci. Eur., 2024, 36 (1), 31.

[10] P.P. Biswas, W.H. Chen, S.S. Lam, Y.K. Park, J. S. Chang and A.T. Hoang, A comprehensive Study of Artificial Neural Network for Sensitivity analysis and Hazardous Elements Sorption Prediction via Bone Char for Wastewater Treatment, J. Hazard. Mater., 2024, 465, 133154.

[11] C. Cortes and V. Vanik, Support- Vector Networks, Mach. Learn., 1995, 20 (3), 75-80.

[12] R. E. Uhrin, Introduction to Artificial Neural Networks, in Proceedings of IECON 95- 21st Annual Conference on IEEE Industrial Electronics, IEEE, pp. 33-38.

[13] N. B. Chang, G. Mohiuddin, A.J. Crawford, K. Bai and K. R. Jin, Diagnosis of the Artificial Intelligence- Based Prediction of Flow Regime in a Constructed Wetland for Stormwater Pollution Control, Ecol. Inform., 2015, 28, 42-60.

[14] S. Sahu, A. Kaur, G. Singh and A.S. Kumar, Harnessing the Potential of Microalgae- Bacteria Interaction for Eco-Friendly Wastewater Treatment: A review on new Strategies Involving Machine Learning and Artificial Intelligence, J. Environ. Manage., 2023, 346, 119004.

[15] T. Uttaradit, N. Mukjang, N. Chaisoung, W. Pathom- Aree, J. Pekkoh, C. Pumas, et al., Putative Protein Discovery from Microalgal Genomes as a Synthetic Biology Protein Library for Heavy Metal Bio- Removal, Biology, Biology, 2022, 11 (8), 1226.

[16] S. Khanna and A. Kumar, Bioinformatics Toward Improving Bioremediation, In Biotechnological Innovations for Environmental Bioremediation, Springer Nature Singapore, Singapore, 2022, pp. 631-669.

[17] D. R. Dudhagara, R. K. Rajpara, J.K. Bhatt, H.B. Gosai and B. P. Dave, Bioengineering for polycyclic Aromatic Hydrocarbon Degradation by Mycobacterium Littorale: Statistical and Artificial Neural Networks (ANN) approach, Chemo. Intell. Lab. Syst., 2016, 159, 155- 163.

[18] Z.M. Yaseen, the next generation of soil and water bodies heavy metals prediction and detection: - New expert system-based edge cloud server and

federated learning technology, *Environ.Pollut., 2022, 313,120081.*

[19] Y. Mussallem, A. A. El Hadj, M. Laidi, S. Hanini and M. Hentabli, computational intelligence techniques for modelling of dynamic adsorption of organic pollutants on activated carbon, *Neural Comput. Appl*. 2021, 33 (19) 12493-12512.

[20] Ravi, N., Johnson, D.P. Artificial Intelligence-Based monitoring system for onsite septic systems failure. Process, Saf., Environ. Prot. 2021, 148, 1090-1096.

[21] Wodecki, B., Drewnowski, J., Bialek, A., Lazuka, E., Szulzyk- Cieplak, J. Prediction of wastewater Quality at a wastewater Treatment Plant Inlet Using a System Based on Machine Learning Methods. Processes 2022, 10, 85.

[22] Rane, N, Choudhary, S. R, Manu, P, Cheung, C, Yunusa- Kaltungo, A., & Zayed, T., Ye, L., Environmental Science & Technology, 56 for enhanced wastewater treatment systems (2023).

[23] Li, K, Duan, H, Liu, L, Qiu, R, Van Den Akker, B, Ni, Bj, Chen, T, Yin, H, Yuan, Z., & Ye, L., Environmental Science & Technology, 56 (4), 2816- 2836 (2022).

[24] Zokor, Y, Gzar, H., & Zakur, Y., In E3S Web of Conferences. EDP Sciences, 448, 03038 2023).

[25] Gholizadeh, M., Saeedi, R, Bagheri, A., & Peezy, M., Environmental Research, 246, 118146 (2024).

[26] Murshid, N., Mouhtady, O., Abu- Samha, M., Obied, E, Kharboutly, Y., & Chouk, H. Halwani. J.: Younes, K. Metal Oxide Hydrogel Composites for Remediation of Dye- Contaminated Wastewater: Principal Component Analysis. 2022.

[27] Sharma, S., Aggarwal, A., Choudhary, 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, 22-24 December 2019: IEEE: Piscataway, NJ, USA, 2018: pp 119.

[28] Zidan, K; Sbahi, S; Hejjai, 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 neighbour algorithm. Environ. Sci. Pollute. Res. 2022, 29, 75716-75729.

[29] Banks, A.; Vincent, J; Anyakoha, C.A review of particle on genetic algorithms: Past, Present and Future. Multimer. Tools Appl. 2021, 80, 8091-8123.

[30] Jain, N. K.; Nangia, U.; Jain, J. Review of Particle Swarm Optimisation. J. Inst. Eng. Ser. B 2018, 99, 408-412.

[31] Bencke, L.; Cochineal, C; Muñoz, R. Automated Classification of Social Network Messages into Smart Cities Dimension. Future Generation. Compute. Syst, 2020, 109- 218-237.

[32] 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. 2019, 155, 1298-1232.

[33] 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- 1876.

[34] World Health Organisation: WHO, Antimicrobial Resistance, World Health Organisation: WHO, 2023, 2023, November 21, Internet, https://www.who.int/fact-sheet/details/antimicrobial-resistance.

[35] E. Raper, T. Stephenson, D.R. Anderson, R. Fisher, A. Patel, et al, Bioremediation and bioelectricity from Himalayan Rock Soil in a sediment-microbial fuel cell using carbon-rich substrates, Fuel, 2023, 341, 127019.

[36] R. Bhattacharya, D. Bose, J. Yadav, B. Sharma, E. Sangili, A. Patel, et al, Bioremediation and Bioelectricity from Himalayan rock Soil in sediment- microbial fuel cell using carbon-rich substrates, Fuel, 2024, 341, 134907.

[37] R. Bhattacharya, D. Bose, P. Gantri, A. Rizvi, G. Halder and A. Sarkar, Bioelectricity production and bioremediation potential of within-somatic plant microbial fuel cells with food waste as enrichment, Energy Nexus, 2024, 15, 100315.

[38] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahedoudy, B. Shuai, et al, Recent advances in convolutional neural networks, Pattern Recognition, 2019, 77, 355-378.

[39] G. Alam, I. Ehsanullah, M. Naushad and M. Sillanpaa, Application of artificial Intelligence in water treatment for optimisation and automation of adsorption processes: Recent advances and Prospects, Chem. Eng. J. 2022, 435, 130011.

[40] Barrere- Animas, A. Y. Oyedele, L.O. Bilal, M.; Akimoto, T. D.; Delgado, J. M.D.; Akanbi, L.A.

Rainfall prediction: A comparative analysis of modern machine learning algorithms for series forecasting. Machine Learning Algorithms. 2018,8, 100205.

[41] 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 T01 (ICCBI-2020), Pandian, A.P., Palanisamy, R., Natalias, K., Eds.: Springer International Publishing: Berlin/Heidelberg, Germany, 2019; Volume 48, pp. 196-206.

[42] Schmitt, M. Artificial Intelligence in Business Analysis, Capturing Value with Machine Learning Application in Financial Services, University of Strathclyde: Glasgow, UK, 2021.

[43] Masruri, A, Luo, W, Zaidi, N. A., Robles-Kelly, A. Machine Learning for Financial Risk Management: A Survey. IEEE Access 2021, 8, 203203-203223.

[44] 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 978-3-030-023294/978-3-030-02330-0.

[45] 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- 423.

[46] Jain, N.K. Nangia, U.; Jain, J. A Review of Particle Swarm Optimisation. J. Inst. Eng. Ser. B2019, 99, 407-411. Hor

[47] Xin, D; Gonnella, N. C.; He, X; Horspool, K., Solvate Prediction for Pharmaceutical Organic Molecules with Machine Learning. Crysts. Growth Des. 2019, 19, 1903- 1911.

[48] Robles-Velasco, A.; Cortes, P.; Munuzuri, J.; Oneita, L. Estimation of a logistic regression model by a genetic algorithm to predict pipe failure in sewer Networks or Spectra. 2021, 43, 759-776.Research