Engineering Applications of Artificial Intelligence 144 (2025) 110132

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Insights into the application of explainable artificial intelligence for biological wastewater treatment plants: Updates and perspectives

Abdul Gaffar Sheik [a,b], Arvind Kumar [a], Chandra Sainadh Srungavarapu [c], Mohammad Azari [d], Seshagiri Rao Ambati [e], Faizal Bux [a,*], Ameer Khan Patan [f]

[a] Institute for Water and Wastewater Technology, Durban University of Technology, Durban, 4001, South Africa
[b] School of Engineering, The University of British Columbia Okanagan, 3333 University Way, Kelowna, BC, V1V 1V7, Canada
[c] Department of Chemical Engineering, National Institute of Technology, Warangal, 506004, India
[d] Institute for Water and Environment, Department of Water Quality Management, Karlsruhe Institute of Technology, Karlsruhe, 76131, Germany
[e] Department of Chemical Engineering, Indian Institute of Petroleum and Energy, Vishakhapatnam, 530 003, Andhra Pradesh, India
[f] Laboratory of Food Process Engineering, Wageningen University and Research, Bornse Weilanden 9, 6708, WG Wageningen, the Netherlands

ARTICLE INFO

Keywords:
Explainable artificial intelligence
Trustworthiness of artificial intelligence
Machine-learning
Process modeling and control
Smart water engineering
Wastewater treatment plants

ABSTRACT

Explainable artificial intelligence (XAI) is an interactive platform that assists users in comprehending the decisions and predictions made by machine learning (ML) models. This allows users to enhance their knowledge of ML models and their functioning, which not only helps in mitigating bias and errors but also aids in improving user decision-making confidence. XAI, due to its ability to increase the model output interpretation, has gained significant attention in biological wastewater treatment plants (WWTPs). This is owing, in particular, to the fact that it facilitates the experts in steering knowledge about the predictions and decisions made by ML, thus guaranteeing that the model decisions are fair and unbiased. ML has made amazing advances in recent years, thanks to its exponential growth in possessing the power to process massive volumes of data, allowing it to be widely embraced in WWTPs. This review seeks to illustrate the potential of XAI for WWTP applications such as process modeling and control, soft sensing, fusion of data, and the internet of things, and fill the knowledge gap by thoroughly introducing XAI techniques and their use in smart wastewater engineering. Overall, the features of XAI can aid in establishing reliable and efficient water resource management, which is quintessential to achieving environmental sustainability. It is envisioned that the prospects offered would spark new lines of study, helping to reduce the current skepticism and apprehension about ML adoption and integration in WWTP.

1. Introduction

The accessibility of clean water is regarded as being of utmost importance around the world because of the growing world population and the consequent rise in the influence of pollutants on the climate, as well as the conversion of land into freshwater habitats (Sagan et al., 2020). Contamination of the waters is caused mostly by both inorganic and organic residue, sediments, radioactive chemicals, effluents, sewers, and toxic metals (Dubey et al., 2015). Indeed, prompt sewage treatment to clean up polluted water is necessary in order to meet emission regulations (Yu et al., 2018). Recent breakthroughs in modeling techniques to support the related systems have led to dramatic gains in biological wastewater treatment plants (WWTPs). The WWTPs are large and complex systems facilitating the treatment of wastewater generated from industries through numerous interconnected processes such as chemical breakdown treatments, filtration, clarification, and biological processes. They have the capacity to treat harmful substances present in water so that the water can either be safely returned to the environment or reused. This clearly marks the importance of maintaining and managing these treatment plants to ensure the efficient and effective treatment of wastewater. Additionally, efforts have been made to reform the waste management procedures so that they are more lucrative and environmentally friendly through the use of advanced technologies. For the WWTP process to be effective in terms of cost and operations, modeling and optimizing are of utmost importance, which are usually carried out by “regression” and “time series” analysis (Shojaeimehr et al., 2014). The advantages of these methods lie in their relative ease of use and practicality in application. However, it is important to consider their very limited predictive abilities under certain circumstances,

[*] Corresponding author.
E-mail address: faizalb@dut.ac.za (F. Bux).

https://doi.org/10.1016/j.engappai.2025.110132
Received 15 May 2024; Received in revised form 4 December 2024; Accepted 20 January 2025
Available online 30 January 2025

Nomenclature

AI	Artificial intelligence
ASM	Activated sludge model
ADM	Anaerobic digestion model
BSM	Benchmark simulation model
BOD	Biochemical Oxygen Demand
BiLSTM	Bidirectional long short-term memory
COD	Chemical oxygen demand
Chl-a	Chlorophyll *a*
DM	Decision making
DL	Deep learning
DT	Decision tree
DNN	Deep neural network
EQ	Effluent quality
GAO	Glycogen-accumulating organisms
IoT	Internet of Things
LR	Linear regression
LSTM	Long short-term memory
ML	Machine learning
MNN	Mechanical Neural Network
MAPE	Mean absolute percentage error
MSE	Mean square error
MAE	Mean absolute error
RMSE	Root mean square error
PAO	Phosphorus-accumulating organisms
RF	Random forest
SVM	Support vector machine
SVRL	Regression with the linear kernel
SVI	Sludge Volume Index
RL	Reinforcement learning
RQ	Research questions
R2	Coefficient of determination
RNN	Recurrent Neural Networks
RBF	Radial Basis Function
TSS	Total suspended solids
TN	Total nitrogen
TP	Total phosphorous
TKN	Total Kjeldahl Nitrogen
WWTP	Wastewater treatment plant
WQ	Water quality
WW	Wastewater
WS	Water Sector
XAI	Explainable artificial intelligence
XGB	Xtreme gradient boost

especially when non-linear patterns and a lot of noisy data are available (Rajaee et al., 2019). The underlying fact might be due to the complex relationships between the variables and high variability in the data, making it difficult to capture the data behavioural patterns. As a result of this fact, this significant interest leads the researchers to pursue alternative approaches in which the processes are integrated with machine learning (ML) and deep learning (DL) models for capturing data behavioural patterns, thereby providing the platform for more accurate predictions (Liu et al., 2023). The latest advancement of ML has resulted in significant advancements in the water sector (WS) such as capturing big data, pattern recognition, intelligent search, and creating human-computer interfaces. These features of ML/DL technology will have a significant impact on addressing the complexities that are generated in the wastewater (WW), and water industries (Singh et al., 2022). With the application of ML/DL models, one can foresee that the water and WW industries will have the potential to improve their efficient water resource handling techniques. In the context of WWTPs, designing an efficient process monitoring system ensures these plants function smoothly even under disturbances that occur due to fluctuating flow and load conditions. This is ensured when wastewater after treatment meets strict emission standards. Also, the available historical data generated from the process and ML techniques can be used to create effective WWTP process monitoring systems (Khurshid and Pani, 2023; Ismail et al., 2021). Safeer et al. (2022) provide a comprehensive overview of AI technologies used to determine source water quality (WQ), coagulation/flocculation, disinfection, membrane filtration, desalination, modeling WWTPs, membrane fouling prediction, heavy metal removal, and biological oxygen demand (BOD)/chemical oxygen demand (COD) monitoring. In one of the reviews, it has also emphasized that despite the success in control, optimization, and modeling achieved with the AI methods incorporated with the Internet of Things (IoT) and smart sensors, there have been consistent and widespread major problems and challenges in treatment and monitoring in WWTP (Lowe et al., 2022). The abbreviations used in this manuscript are detailed in the nomenclature section.

In recent days, to leverage the advantage of AI over process industries, a significant step has been made by data scientists by creating better classifications and modifications in ML models as an enroute towards DL with hybrid combination techniques. In a few instances, these DL techniques appear to be more effective, noise-resistant, and accurate than conventional models. The underlying problem with such models, despite their obvious effectiveness, is that it may not be clear how or why they make particular conclusions or how they handle actual data. This makes water professionals not to trust over the conclusions generated from the DL models unless there are proper or reliable interpretable explanations. This leads to factors for reducing the usefulness of DL models created by modifying and updating the existing ML models. This challenge is particularly noticeable in circumstances when a high level of reliability is required, which is typical in the WS. With the expanded range of computing WW systems and the fact that it requires better predictability for a greater variety of datasets, DL has recently achieved significant advancements. The input data in this DL procedure will train independently using convolutional neural networks (CNN), long short-term memory (LSTM), bidirectional LSTM, support vector regression (SVR), deep feed-forward networks (DFFN), etc. According to Singh et al., 2022); Sheik et al., 2024c. Three components known as data pre-processing, feature extraction and recognition, and model optimization make up DL models (Ning et al., 2020). The ML algorithm was employed to force the user to decide regarding their water processes, but individuals were unaware of the ML's output or the process by which it arrived at its conclusion. This problem has prompted the development of new methods and ideas over the past few years to make ML/DL models more comprehensible and hence increase the quality of their output. The term "explainable artificial intelligence" (XAI) is used to refer to this idea in the scientific community (Gunning et al., 2019). In order to maintain the superior performance and precision of ML/DL models, XAI helps scientists, developers, experts in the field, and users better understand exactly how the models work inside. This way it creates the possibility for humans to retain the intellectual oversight on the methods adopted by these models to achieve the desired output. The feature of understanding the working process of ML models helps users to implement reliable decision-making (DM) on the process systems.

In the literature, several XAI methods are usually used with ML/DL models (Gupta et al., 2022; Ba-Alawi et al., 2023a, 2023b, 2023c). The practical use of XAI in the WW and water domain has only started. Fig. 1 (A) describes the modeling road path from 1960 to 2020 in WWTP. It depicts the modeling roadmap from mathematical modeling, computational fluid dynamics, and the application of advanced controllers to

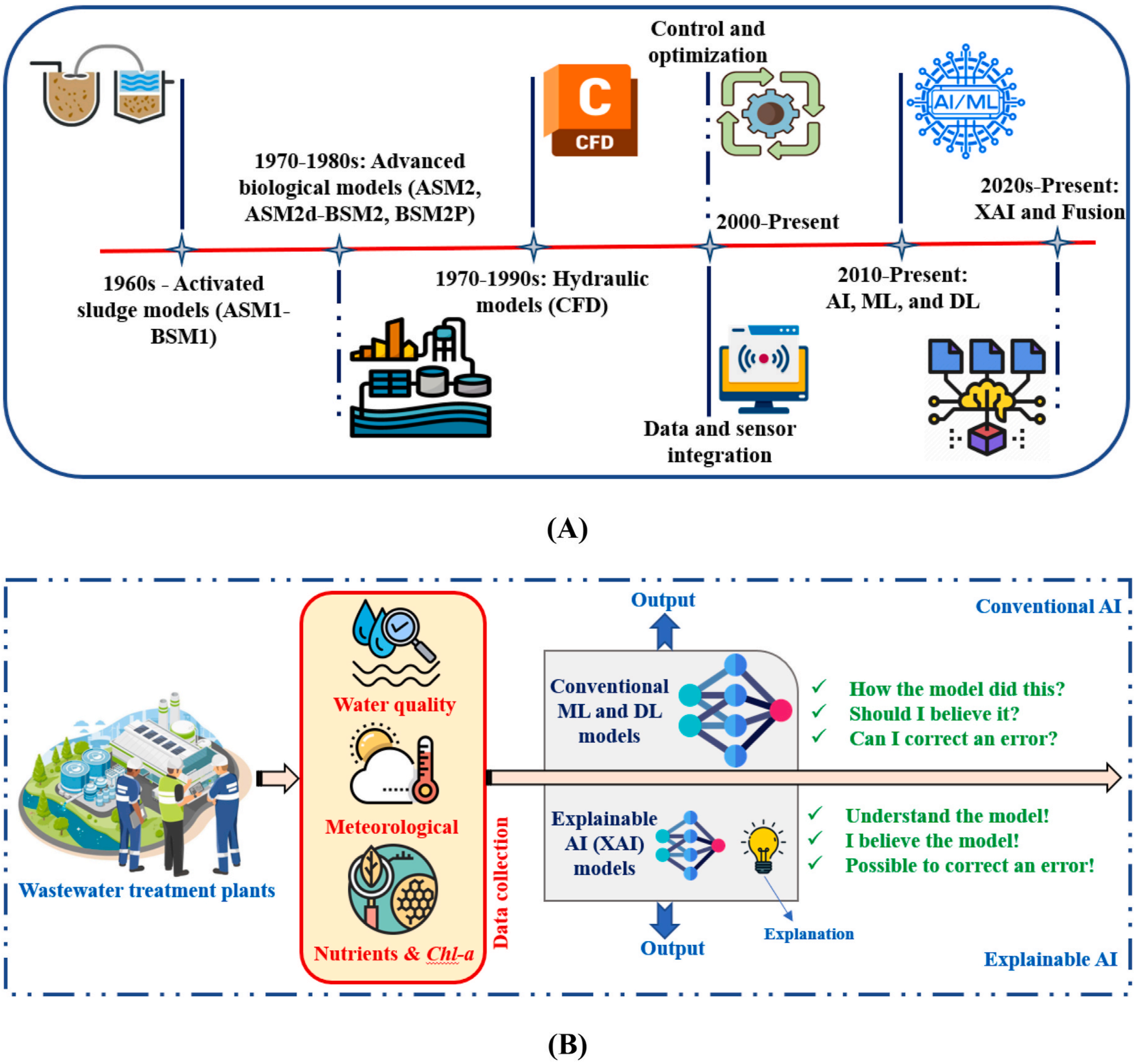
Control and optimization
1970–1980s: Advanced biological models (ASM2, ASM2d-BSM2, BSM2P)
CFD
2000-Present
2020s-Present: XAI and Fusion
AI/ML
1960s - Activated sludge models (ASM1-BSM1)
1970–1990s: Hydraulic models (CFD)
2010-Present: AI, ML, and DL
Data and sensor integration
(A)
Output
Conventional AI
Water quality
Meteorological
Nutrients & Chl-a
Data collection
Wastewater treatment plants
Conventional ML and DL models
✓ How the model did this?
✓ Should I believe it?
✓ Can I correct an error?
Explainable AI (XAI) models
✓ Understand the model!
✓ I believe the model!
✓ Possible to correct an error!
Output
Explanation
Explainable AI
(B)

Fig. 1. **(A)** The technological advancement path for WWTP, and **(B)** Concepts of XAI for WWTP applications.

data-driven modeling (ML, DL, and XAI). Fig. 1(B) illustrates the comparison of conventional AI and XAI applications on WWTP which depicts the concepts of traditional AI and XAI and how these models will help decision-making. When compared to mechanistic models (e.g., BSM1, BSM2) (Sheik et al., 2023), the advantages of ML/DL-based WWTP process modeling include (a) shorter execution time, (b) no requirement for multi-disciplinary knowledge related to biokinetics (enzymatic reactions), microbiome (types of microorganisms), heat/mass transfer, and (c) avoidance of model recalibration if trained on large datasets. Although a wide range of regression and classification models have been developed to predict biogas yield, process stability parameters, and effluent quality indicators (El-Rawy et al., 2021; Ly et al., 2022), the researchers are yet skeptical due to the black-box nature of ML techniques. There are two kinds of ML techniques: (I) black-box ML and (II) explainable ML which is comparable to white-box, with the latter seeking to provide a deeper comprehension of the functional reliance of the output on the input. It should be noted that the ML scientific community supports the use of explainable (or interpretable) ML in all scenarios. Several recent studies on industrial process modeling have proved the benefits of ML combined with numerous explainability metrics such as feature importance testing, partial dependence analysis, and so on (Wang et al., 2022b; Zhang et al., 2023a, 2023b, 2023c; Wang et al., 2021; Park et al., 2022a, 2022b). The investigation's current inquiry question is, "What is XAI in WS in the context of quality assessment, bias risk, and data fusion?" As a result, the authors suggest a systematic assessment of the available literature which attempts to provide information on XAI in WS and to assist scholars in identifying present gaps and solutions. Furthermore, this work delivers a cutting-edge contribution by creating a comprehensive map of XAI in the water treatment sector to create a coherent taxonomy system, aiding users with a thorough understanding of XAI in WWTPs. The bibliometric evaluation presented in section 2 was utilized to reorganize and summarize the findings of earlier studies, as well as the general knowledge picture, by offering a mapping analysis for the research stream of XAI usage in the WS (Zhang et al., 2023).

The purpose of this research is to assist in spotlighting the significance of XAI in the wastewater treatment division and aiding to contribute towards a more sustainable and environmentally friendly approach to managing water resources in the WWTP sector. Previous studies of XAI models for WW processes focused on the translation from advancement to practice (Zahra et al., 2023a, 2023b), and the application of data-driven models in general (Singh et al., 2022). Recent deep learning reviews look at the use of XAI in urban water supply and sewage infrastructure (Liu et al., 2023), as well as in drinking water process systems (Alam et al., 2022) and membrane-based treatment systems (Jawad et al., 2021). However, the methodologies and applications of

XAI devised in this innovative field of study on WWTP have not been thoroughly examined regarding the positive and negative aspects. To the best of our knowledge, there are currently limited comprehensive articles on XAI-based models and their applications on WWTPs. Overall, in this review, we explore the potential of XAI techniques to enhance the understanding of the ML model's working mechanisms and to improve the DM process in the context of achieving optimal design and control over WWTPs. The scope of the review paper covers.

- An up-to-date comprehensive review of the explainable AI models and their application into WWTPs.
- Research questions on applying XAI in WWTPs and in-depth analysis on literature search.
- The fusion of XAI techniques in WWTPs with a focus on process modeling, data handling, control systems, soft sensing, and the Internet of Things.
- Bridging gaps and unlocking potentials of XAI in the usage of WWTPs.
- Discusses several challenges and reveals future trends for XAI research in WWTPs.

2. Bibliographic analysis

Following a keyword search on Google Scholar (GS), PUBMED (PM), Scopus (SP), Science Direct (SD), and Web of Science (WofS) databases, the statistics of publications have been collected and reported by the authors on the topics XAI and WWTP, and ML and WWTP. XAI and Wastewater treatment: GS (18200), SD (2627), PM (D'Alterio et al., 2020), SP (Ba-Alawi et al., 2023a), and WofS (Alvi et al., 2022). ML and Wastewater treatment: GS (25000), SD (7680), SP (990), PM (433), and WofS (638). The statistical precis of XAI on WWTP is illustrated in Fig. 2 (A). In summary, the number of publications for the provided keywords varied across the different databases. However, there are several articles from the provided references that discuss topics relevant to the optimization of WWTP, the use of AI, ML/DL in process systems, and XAI application in process systems. In general, the methodology adopted for bibliographic analysis served the purpose of this review. The re006Cative and generic keywords used in this study were: ''AI, XAI, and WWTP." Furthermore, the Scopus database, which is one of the most trustworthy scientific databases has been chosen as this study's primary data source. Information was gathered between 2014 and 2024, and each keyword has been checked individually. A maximum of less than 50 ″new'' OR ″highly cited'' articles were retrieved and translated to CSV files for each round of search. VOS viewer, a freely available and freely accessible bibliometric tool, is used to evaluate information. For specific topics like XAI in WWTP, there are less than 650 publications in the period of 2043 to 2024. The initial data mapping result is shown in Fig. 2(B).

Following the loading of data into VOSviewer, data filtration was initiated to remove unrelated repetitive keywords (such as paper, study, etc.), and various combinations of the terms (WW and WS) were merged and accounted for as one unique term. The research on XAI and WS can be divided into three groups (colors). Large circles indicate the significance and repetition of keywords like WW, different models, and sensors (XAI, WQ, etc.). Furthermore, the closer the distance between two items the more prominent interactions between two items implying that there is overlap among two keywords. Overall, the bibliometric analysis tool provided valuable insights into the current state of research on ML and XAI in WWTP.

3. Methodology

3.1. Review of XAI in wastewater treatment

As the application of ML/DL spreads and gets intertwined in process systems design, an emphasis is now put on comprehending about the algorithmic DM process. Subsequently, this allows numerous other fields, organizations, and researchers to focus on the need to change from emphasizing model correctness to explainability. This represents a paradigm shift in understanding the degree of accuracy achieved by ML systems in predicting the process outcomes (Angelov et al., 2021). Explainability is a critical tool in ensuring the DM process of ML models and the outcomes that are produced by these models are transparent and fair. This aids in comprehending the limitations and uncertainties of ML models in predicting the process outcomes, thereby fostering researchers the need to develop more reliable, ethical, and logical ML models for process industries such as WWTP- "A smart infrastructure." The choice of methodology to use XAI for a particular problem depends on a trade-off between model performance and explainability (Arrieta et al., 2020; Sokol et al., 2022). However, the existing literature emphasizes that there is not much attention given to explaining the working methodology of ML models which are integrated with the process systems, such as WWTP.

In general, there are two categories of ML models namely transparent and opaque models as shown in Fig. 3(A). The transparent models sometimes perform poorly and they either underestimate or overestimate the state variables of process systems. To avoid this uncertainty, models such as the RF, SVM, CNN, MNN, and RNN are sometimes employed in place of them. Models can sometimes be highly opaque representing a black box model making users difficult to comprehend the working methodology of a model (Rudin, 2019). The black box model typically requires users to engage in post-hoc explainability efforts to attempt and create possible explanations related to the working procedures of opaque AI models (Hasenstab et al., 2023). A greater understanding of the internal working mechanisms and features of an opaque model will be attained by building proper explainable strategies. Notably, there are distinct differences between model-agnostic and model-specific based on the nature of the techniques used to explain machine learning models in post hoc explainability as shown in Fig. 3 (A). A model-agnostic model works for all models. In this, the techniques of XAI are to be broadly applicable in a manner adaptable enough to function only on the basis of connecting a model's input to its output, independent of the inherent architecture of the model (Dieber and Kirrane, 2020). In contrast, the model-specific models only work for a specific single or a group of models (Speith, 2022). Also, in this the XAI techniques frequently capitalize on understanding a particular model and seek to increase transparency aiming to shed light on how the model arrives at its prediction of process state variables, making it easier for users to trust and comprehend the decisions made by the model (Bach et al., 2015). Tritscher et al. (2020) suggest that through simplification of a model at first allows users to understand the underlying working mechanism of the model, which thereby helps ultimately in identifying the desired data predicting patterns. Starting with a simplified model, researchers thereafter can progressively plan to enhance and extend the model through iterations by systematically integrating the additional variables and complexities. This methodological approach in turn guarantees that the resulting model comprehensively captures all the phenomenological complexities associated with the working systems. This way it helps the enhanced iterative ML models in maintaining the interpretability and thereby resulting in a potent tool for understanding and making well-informed decisions. The application of ML/DL and RL models are largely used in terms of publications and practical usages in the WWTP field according to Alvi et al., 2023); Singh et al., 2022; Croll et al., 2023. If a model can be understood independently, it is considered transparent. In contrast, transparency serves as the opposite of a "black box" (Adadi and Berrada, 2018). According to Gilpin et al. (2018), interpretability and explainability are capable of delivering interpretations and explanations in a human-understandable way. In the case of opaque models, they are difficult to interpret necessitating the post hoc explainability techniques in interpreting the opaque models after training without degrading their predictive performance (Lipton, 2018; Speith, 2022).

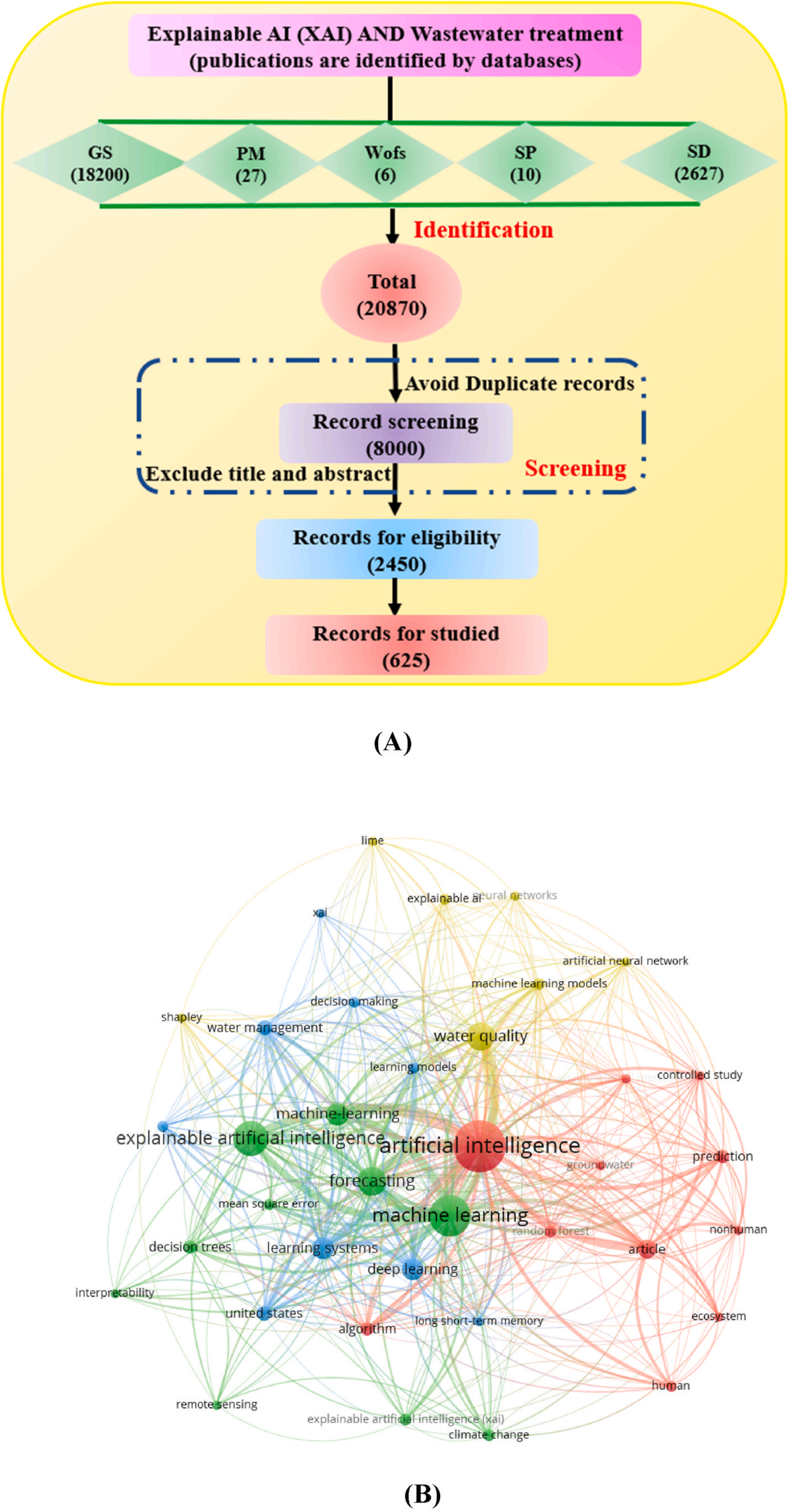

Fig. 2. (A)A precis of the procedure for identifying, selecting, and including relevant contributions (B) Co-occurrence network map of author keywords or ML with XAI.

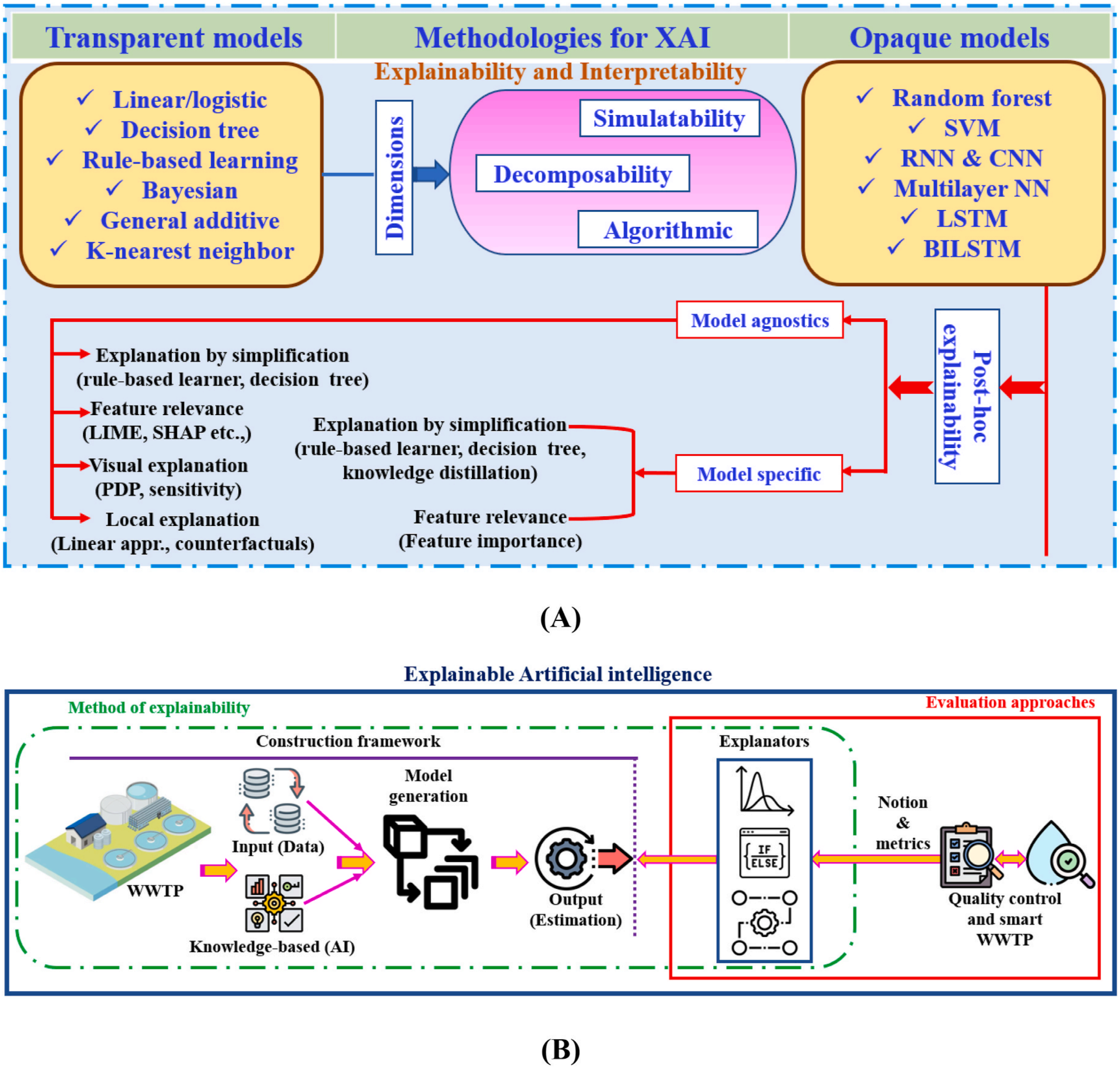

Fig. 3. **(A)** Various artificial intelligence models that are involved in WWTP **(B)** Diagrammatic view of XAI with the interaction between methods for explanations and their evaluation approaches in WWTP.

Therefore, this review presents the post hoc explainability techniques paving the way to interpret the working procedure of models rather than directly reporting the application of ML/DL and RL models onto WWTP. The ontology of the XAI taxonomy with post hoc method categories are shown in Fig. 3(A). With a high flow of publications published each year, XAI is expanding tremendously as a field of study for future application into complex processes such as industrial and financing sectors (Yang et al., 2022). It's difficult to evaluate recent XAI improvements in the WWTP sector. However, the tenets, models, and post-hoc justifications serve as a foundation for comprehending the particular features and specifications of XAI in application into WWTPs. Very few research publications presented the idea of XAI in WWTP (Wang et al., 2022a, 2022b, 2022c; Wang et al., 2021). However, the lack of comprehensive studies makes it difficult to assess the true impact and effectiveness of XAI in the WWTP sector. This review aims to provide more practical insights and advantages associated with the application of XAI in WWTP operations.

3.1.1. Post hoc explainability

Even though the semantic implications of these clauses are similar, they require various levels of AI before humans will accept them. The high-level ontology and taxonomy of XAI can be found in the details below for further information. A transparent model aids in displaying transparency in the decisions made by the models in predicting the states of the process systems. The k-nearest neighbors (kNN), decision trees, rule-based learning, Bayesian networks, and so forth are examples of typical transparent models (Adadi and Berrada, 2018). These models frequently produce transparent decisions, but openness alone does not ensure that a model will be easily understood. One must have the ability to comprehend the working scenario and the DM process of the models. This way it aids users in improving the DM process of the models by modifying the existing model's programming scripts to predict the outcomes of any complex process systems such as WWTP. The conceptual framework at the basis of the proposed goal is represented in Fig. 3 (B) whereby methods for explainability are built from automatically induced models using explanators, and these can be evaluated by employing notions and metrics of WWTP operations. In Fig. 3(B) the explanation methods generate outputs based on AI model predictions. These outputs flow into the evaluation approaches, where they are assessed using key metrics. Feedback arrows from evaluation approaches go back to the methods, showing the need for continuous improvement based on evaluation results. User feedback is intertwined with the evaluation loop, providing insights into whether explanations meet practical WWTP needs. The explanation types that are essential in post hoc explainability can be categorized as below.

- Feature-relevant explanation: This is a concept that can be closely related to a simplified explanation of the model's feature influencing the prediction of the process outcomes. After all potential combinations have been considered, this kind of XAI technique aims to assess a model's feature according to its average expected marginal contribution to the model's choice for the prediction (Chen et al., 2019).
- Visual explanation: According to Chattopadhay et al. (2018), this kind of XAI technique is centered on visualization. Therefore, the interpretation of the prediction or judgment over the input data can be facilitated by utilizing the family of data visualization tools.
- Local explanation: According to Selvaraju et al. (2017), local explanations provide insight into how the model functions in a limited region surrounding a particular instance of interest. The region of interest can either be a case of identifying the model's sensitivity to certain inputs and highlighting any biases or limitations it may have or to understanding how deterrent the features of the model are in predicting the process outcomes. For instance, the local explanations are valuable in knowing the DM process of a model as they shed light on the specific features that contribute to a prediction of process outcomes. This knowledge can help identify areas where the model may not accurately predict outcomes, enabling researchers to refine and improve the model's performance in sensitive regions to make them function more robust.

The terms "explainability" and "interpretability" which are commonly used in the ML field are sometimes deemed inadequate since they do not address every potential issue related to comprehending "black box" models (Burkart and Huber, 2020). Explainability refers to the ability of the model to provide insights into the reasons behind its decisions or predictions, making them understandable to users. It is essential in the process industries to generate human-understandable explanations for model outputs. The explainability focuses more on post-hoc explanations of the model's behaviour, even for complex models like DL networks. For explaining black-box models techniques like feature importance, surrogate models, or visualizations (e.g., SHAP values, LIME) are used. On the other hand, interpretability refers to the degree to which a human can understand the cause of a decision made by an AI model. A model is interpretable if its internal mechanics (like its parameters or decision rules) are clear and understandable. It focuses on how easily a person can comprehend the internal mechanics of the model in predicting the outcomes of WWTP operations. It is about the clarity of the relationship between inputs and outputs. Highly interpretable models (like decision trees) provide simple, intuitive insight into how decisions are made. In the domain of XAI technology, both concepts aim to make AI decisions more transparent when they are applied to different types of models and scenarios.

Explainability along with interpretability is needed most of the time to triumph over user's trust and obtain significant insights into the motivations, decisions, and causes behind "black box" techniques. It's not always the case that explainable models translate well by default. Adadi and Berrada (2018) divide the XAI taxonomy in the literature currently in use by (a) scope (Chen et al., 2019) (b) usage (Tritscher et al., 2020) (c) methodology (Dieber and Kirrane, 2020). According to Phillips et al. (2020), there are "Four Principles of XAI" that explain the scenarios of XAI which are increasingly importance in any process application. These principles outline the essential requirements that an AI must meet to qualify as an XAI in WWTP, and they are as follows.

- ✓ Explanation: An AI system must provide justification, proof, or both for every action it makes in the process of WWTP.
- ✓ Meaningful: The AI system's explanations must be intelligible and significant to its consumers. Since several user groups may have varied wants and backgrounds, the AI system's explanation must be customized to each group's unique traits and requirements when it tends to apply for the WWTP.
- ✓ Accuracy: In accordance with this concept, the AI system's explanation must correctly interpret and predict the workings process of the WWTP system.
- ✓ Knowledge limits: AI systems must be able to recognize the situations that are to be generated in which they were not intended to obtain in any WWTP process. Failing to do so makes their responses could not be trustworthy to the users anymore.

3.2. Research questions of XAI in WWTP

In the context of XAI-based WWTP's parameter prediction, the goal-question-metric (GQM) approach is employed (Rini and Berghout, 1999) to gain a critical understanding of the performance of the selected models through XAI tools in predicting the outcomes of the process. In the context of WWTP operations, the XAI approaches experts mainly from two domains, a manager water quality test, and an operational process manager. Engaging XAI experts with the domain experts helps not only to grasp the data supplied very easily but also aids in ascertaining the concerns regarding explainability, thereby allowing AI experts to enhance the working functionality of the XAI tools. The "working functionality" of XAI tools refers to the specific methods, techniques, and processes to provide transparency and explainability in AI models. These functionalities are designed to help users, particularly non-experts, understand how AI models make decisions or predictions. The key working functionalities of XAI tools include model-agnostic explanations, feature importance, bias detection, and fairness analysis, local & global explanations, etc. The close collaboration between the AI expert and process domain experts helps ensure that the explanations generated by the XAI tools are relevant and meaningful, thereby enhancing the credibility and practicality of the XAI tools in WWTP applications. To assess how disruptive the XAI tool explanations are when contrasted with domain experts' previous WWTP understandings, one can formulate and explore the following research questions for investigations (RQ).

RQ:1- How accurate is the XAI-enabled ML/DL system in predicting WWTPs?

Response: The prediction accuracy of the process outcomes after training the data set generated from the processing system gives insight to any domain expert about the working efficiency of the XAI. However, one has to be aware that the efficiency in predicting the state variables of WWTPs depends on various factors like the quality and the quantity of data generated, the degree of complexities involved in the process, the explainable techniques available to make the DM process of ML and DL models more transparent. In the scenario of evaluating the working performance of XAI, domain expertise is crucial for restructuring and redesigning the models to increase the level of accuracy of XAI in providing meaningful explanations of the ML/DL models used for the prediction of the process outcomes without deviating from the underlying principles of WWTPs. The "working performance" of an XAI model refers to how well the model operates regarding its core objectives, which are to provide accurate and interpretable results. In the context of XAI, "working performance" can be broken down into two key dimensions 1) Predictive Performance and 2) Explanatory Performance. For XAI models, both predictive and explanatory performance must be balanced. A highly accurate model that cannot explain its results is less useful in situations where accountability or transparency is required, like in healthcare, finance, or legal applications. So, the working performance of an XAI model is a measure of how effectively it performs its tasks while providing understandable and reliable explanations for its actions. To interpret and understand the predictions of ML/DL models one can choose to work with techniques such as LIME, PDP, and SHAP (Parsa et al., 2020; Alvi et al., 2023).

RQ:2- When the ML/DL solution with XAI support forecasts WWTP events, how much noise is present?

Response: The noise occurrence in predictions is often influenced by factors such as the quality of the data and the model complexity towards data sensitivity. To train the model, one has to first plan the mitigative steps to reduce the noise in the data. Noisy and inconsistent data usually lead to poor performance of the ML/DL models in predicting the desired process outcomes. Further, the presence of complexity and sensitivity in the models not only captures the unwanted noise but also overfits or underfits the data, thereby not making reasonable predictions. To avoid model dysfunctionality due to the dynamic variation in the WWTP process, it is highly recommended to create and adopt a holistic approach in regularly updating the model with new data and monitoring its performance. This way, it helps the users enhance and tune the model to make the model more adaptive and reliable to evolving conditions and, as a result, reduce the impact of noise. It is also recommended to know that for post-hoc predictions, techniques such as LIME, PDP, FI, and SHAP can be employed to gain greater insights into understanding the model's performance under different conditions. This way, it helps users perceive the potential ways to fine-tune the model, making it more reliable and adaptable (Chen et al., 2019; Wang et al., 2022a).

RQ:3- To what extent do the XAI tools produce simple explanations?

Response: One can assess the simplicity of these products by using both quantitative measurements and expert interviews. To measure simplicity quantitatively, one should calculate the entropy of explanations generated by XAI tools. Entropy quantifies uncertainty and can indicate the distinctiveness and variability of explanations. Additionally, conducting interviews with domain specialists will provide valuable insights into explanations from various XAI tools, facilitating to gathering of expert opinions and assessing the reliability and adaptability of XAI. The synergistic approach of combining expert opinions with quantitative measures can help determine if the explanations produced by XAI are easy to comprehend, enhancing the DM process for better predictions (Dwivedi et al., 2023; Páez et al., 2019).

RQ:4- How reliable are the justifications produced by the XAI tools?

Response: Investigating the soundness of an explanation can be challenging because it depends on the interests of the user's history, as Gilpin et al. (2018) have emphasized. Nevertheless, XAI tools such as LRP (Love et al., 2023), SHAP (Wang et al., 2022a, 2022b, 2022c), and DeepLIFT (Zahra et al., 2023a, 2023b) all sometimes mask the most important data values, limiting the XAI tool's ability to generate simpler explanations. In general, the reliability of the justifications provided by XAI tools depends on several factors. At first, it depends on the algorithm of the model used by the XAI tool. Some models help provide justifications in a more transparent and interpretable way while the other are quite complex and challenging to interpret. Second, it is clear that the quality and quantity of data have an impact on the training of XAI tools. Effective training of the XAI tool with diverse data aids in producing reliable justifications. However, if the training data is biased or incomplete, the explanations may lack reliability. Further, the role of the domain expert also plays a crucial part in making the justifications or explanations provided by the XAI tool more reliable. Overall, adopting approaches to test and validate XAI tools rigorously by experts, the justifications are considered to be more reliable.

RQ:5- To what extent do the explanations provided by the XAI tools generate new insights?

Response: Performance is crucial to our WWTP beneficiaries. To justify its use, an ML/DL system needs to outperform its current system and be equally comprehensible. If a model deviates too far from their predictions, it could be hard to put users trust in XAI tools. The methodology in generating clear explanations about the DM process of ML models has the potential to lead to more accurate assessments of the efficiency and performance of XAI tools, allowing experts to better manage and optimize WWTP operations. Overall, this could result in cost savings, reduced environmental impact, and improved overall performance for WWTPs (Alvi et al., 2023; Sheik et al., 2023; Dwivedi et al., 2023). When it comes to high-stakes applications such as WWTP, the most significant metric for evaluating ML/DL model performances lies in accurately detecting events, ensuring that users have confidence in making informed decisions and implementing them in the process. Hence, in this review, the authors tried to introduce the concept of XAI which helps users comprehend the working methodology of ML models in predicting the outcomes of any process operation, making them feel more confident and empowered in their interactions with such technology. This can lead to increased adoption and acceptance of AI in various industries.

In the context of XAI, several techniques are available for developing the taxonomy of explainability, aimed at improving interpretability and providing transparency and comprehensibility to the behaviour and decisions of AI architectures being employed for human users. The techniques also aid in identifying any biases or limitations within the model. Although certain XAI approaches are specifically developed to tackle specific issues, at times it can be difficult to understand their fundamental intuitions. Presented here is a clear and concise significance of many prevalent XAI techniques and architectures.

- Shapely Additive Explanation (SHAP), for example, demonstrates how each of the components of a model's prediction can be broken down into contributions from each input feature.
- Local Interpretable Model-Agnostic Explanations (LIME) focuses on providing local explanations for individual predictions rather than a global understanding of the entire model. It approximates complex models with simpler and interpretable models for better understanding.
- Partial Dependence Plots (PDP) directly illustrate the relationship between a feature and the target. The PDP can help identify the impact of a particular feature on the target variable, providing valuable insights into the behaviour of the model. They can also be used to detect interactions between features and non-linear relationships.
- The attention mechanism (AM) is a method employed in ML and AI to enhance the efficacy of models by directing attention toward pertinent information. This feature enables models to choose to focus on specific portions of the input data which is critical to the process of study, allocating variable levels of significance or weight to individual components.
- Decision-making in rule-based systems is characterized by transparency and clarity. The presence of explicit rules in systems enables human users to track the source of each choice, which is based on the manifestation of particular conditions.
- A counterfactual explanation aims to address inquiries such as "What modifications to the input features would have led to an alternative prediction?" This facilitates the comprehension of the decision-making process of the model by users.

Besides the above-discussed XAI methods, techniques such as Integrated Gradients (IG), Layer-Wise Relevance Propagation (LWRP), and many more can be employed, depending on the complexity associated with the data, to interpret and understand the inner workings of complex models. All of these techniques are useful in gaining insights into how complex ML models make predictions and the important features driving those predictions (Shao et al., 2023). Overall, these techniques adhere to the idea that “simpler explanations are preferred over more likely complicated ones,” which is also the basic ideology to be followed in addressing the queries raised in RQs.

4. XAI in wastewater treatment plants

Mechanistic modeling has dominated the field for the past several decades when it comes to explaining both biological and chemical reactions occurring in WWTPs (Mannina et al., 2016). For the design and simulations to be more effective, mechanistic models like the International Water Association (IWA) ASM series and ADM1 rely on governing mechanisms rather than conveniently supporting the systematic study of complex systems in data-rich contexts. However, the WW sector has been using data-driven modeling techniques for a wide range of applications and has seen a sharp increase in use in recent years. Anomaly detection, performance prediction, process management and automation, soft sensing, diagnostics, and missing data imputation are many examples of applications that use XAI. The use of XAI models in WWTP is the main topic of this section.

4.1. Process modeling and simulation in WWTPs

In the WW sector, XAI techniques are emerging as effective tools for forecasting the performance of ML models in predicting the EQ parameters. XAI can be a prospective tool that can be used in WW processes without the requirement for underlying mechanistic concepts, which is driving its adoption in activated sludge (AS) and anaerobic digestion (AD) systems.

4.1.1. XAI techniques for modeling and simulation of WWTPs

For generating insights into the performance of the ML model’s prediction over WQ parameters in WWTP, XAI methods such as SHAP and PDP have received the greatest research attention. More particularly, generating explanations about which particular input feature of the WWTP has an influence over predicting the critical variables such as sludge volume index (SVI), sludge quantity, TN, TP, COD, TSS, etc. (Alvi et al., 2022; Wongburi and Park, 2022; Ba-Alawi et al., 2023c; Shao et al., 2023). For process modeling tasks like forecasting important performance factors and locating areas where WWTPs may be improved, DL models in combination with XAI techniques have proven to be more promising and effective tools. Similarly, for regression tasks using time series data sets, XAI utilizes techniques such as SHAP, FI, and PDP to explain the predictions of ML models. Overall, XAI is an effective technique for revealing how specific model traits affect model predictions. Each feature is given a value that indicates how much it contributes to the model’s output. DL models, such as DNN, LSTM, and BiLSTM (Alvi et al., 2023; Farhi et al., 2021; Zhang et al., 2023a, 2023b, 2023c), help improve the efficiency, accuracy, and effectiveness of various processes involved in the treatment of WW. However, their complexity often makes it challenging to understand why they make specific predictions, which can be a problem in critical applications where interpretability and transparency are required. Although XAI models have the potential to provide the explanations for ML model’s ability to predict important variables of the processes, there is a noticeable lack of research articles explicitly focusing on their use in the predictive control of EQ. In general, XAI allows one to effectively explore how specific model attributes affect model predictions. It assigns a number to each feature, indicating how much impact it has on the model’s output. By understanding the impact of each attribute, decision-makers can make informed adjustments to optimize the process and ensure desired outcomes. As the field of XAI advances, more sophisticated models and approaches are expected to emerge to address the shortcomings and challenges of current methods.

4.1.2. Real-time WWTP modeling and simulation, and assessment through XAI techniques - case studies

The idea of implementing XAI techniques on a real-time WWTP has been explored by many researchers to comprehend the influence of input parameters over the predictive output of the plant. Two different case studies have been considered in this study to know how the XAI technique like SHAP provided the explanations on how the sludge volume index and the sludge quantity generated are influenced by the input parameters of a WWTP. The readers of this manuscript are encouraged to explore the literature presented in Table 1 as it presents the literature available on applications of ML and XAI models on the data obtained from real-time WWTPs. Table 1 also highlights the user’s variable of interest to predict and control, and the models and techniques employed in doing so which helped in generating explanations to understand the impact of input parameters on predictive outputs generated by WWTP.

Case study 1: Performance evaluation of ML (RNN) with XAI (SHAP) model in predicting sludge volume index (SVI) on a real-time WWTP data.

One of the most crucial operational variables in an activated sludge process is the Sludge Volume Index (SVI). SVI is difficult to anticipate because of the nonlinearity of the data and the unpredictability of the operating conditions. Wongburi and Park (2022) explored Recurrent Neural Network (RNN) with XAI (SHAP), using complex time-series data obtained from Nine Springs WWTP in Madison, Wisconsin. The schematic of this case study is shown in Fig. 4. The data was used to predict SVI using ML (RNN) and interpret the prediction result using XAI (SHAP). Initially, the data was collected from 1996 to 2020, which was then divided into three datasets to check the efficacy of the model over datasets created over different periods. The first dataset is the actual data collected from 1996 to 2020; the second data set was created from 2010 to 2020 because of the presence of significant errors in the data obtained in 2000; and the third dataset was created from 2010 to 2020 by removing the out-of-range (50–150 mL/g) SVI values. As a first step, the data was collected, analyzed, and cleaned using the Python and data analytics approaches. Following data cleaning, the RNN model was applied to the different datasets created to predict SVI values accurately. The XAI techniques were then applied to interpret the model’s predictions and provide insights into the factors influencing SVI values. In data-based process assessment, it is always important to know which input parameter is influencing more on the output parameter. In this study, the input parameters such as flow rate, influent BOD, Total Suspended Solids (TSS), Total Kjeldahl Nitrogen (TKN), Ammoniacal Nitrogen (NH_3-N), Total Phosphorus (TP), and organic loading were selected as influencing factors on the output parameter Sludge Volume Index (SVI). After training the RNN model using all the datasets, it was found that for the first dataset, the prediction gave an RMSE value of 4.161 and an MAE value of 3.284. For the second dataset, the prediction model performed better, which resulted in lower RMSE (3.360) and MAE (2.156) values in comparison to the first dataset. Similar types of trends were observed for the third dataset. The results of the study demonstrated that the RNN design is effective in handling typical fluctuations that occur in the activated sludge systems, but selecting the relevant data using data analysis is one of the key steps in making the model perform better. Finally, the prediction result was explained using the Shapley interpretation to check which parameter is influencing much on SVI. It was found that the organic loading, which is related to influent BOD and flow rate, primarily affects SVI. The insights obtained through the results of this study suggest that improving the aeration of the system can lead to better control over SVI.

Case study 2: Application of ML models with XAI (SHAP) in predicting sludge production on a real-time WWTP data.

Sludge is produced from urban sewage in China due to its extensive WWTP investment. China had 2827 WWTP with 60.16 billion cubic meters of processing capacity in 31 provinces, municipalities, and autonomous areas in 2021. About 14.229 million metric tons of dry sludge are produced, making it difficult either to use or treat the sludge

Table 1
Applications of ML and XAI models in WWTP.

S. No	Variables	ML & XAI models	Comment	Performance matrix	Reference
1	TSS and OP	RF (ML), DNN (ML), VIM (XAI), PDP (XAI)	DNN models were used to build and validate RF models, and then VIM and PDP studies were carried out. VIM determined the factors that had the greatest impact on the effluent parameters (in this case, TSSe and PO4e), whereas PDP clarified their effects on TSSe and PO4e.	R2-RF, and DNN (TSS) = 0.92 R2-RF and DNN (PO) = 0.886 and 0.872	Wang et al., 2021
2	PAO and GAO	LR (ML), SVRL (ML), SVR (ML), RBF (ML), RF (ML), and SHAP (XAI)	New insights into how PAOs and environmental factors interact may be revealed by ML-enabled analysis, which has immediate implications for the sustainable design and functioning of full-scale EBPR systems.	R2 = 0.4–0.7	Oh and Kim, 2021
3	Temperature, pH, EC, DO, *Chl-a*, TUR	SHAP (XAI), FI (XAI), XGB (ML), PDP (XAI), and VIF (Statistical ML)	This finding showed that SHAP analysis, an XAI method, gives valuable information that permits a decrease in the necessary number of independent variables for creating a ML model, hence reducing the labor and expenses associated with field data collecting.	RMSE-1.872, RSR-0.630, and NSE-0.603	Park et al., 2022a, 2022b
4	BOD, TN, TP, TKN, TSS, NH, SVI, flow rate (FR)	RNN (ML), SHAP (XAI)	The ability to predict SVI will help WWTPs create corrective actions to keep SVI steady. The wastewater treatment industry will benefit from improved operational performance, system management, and process dependability thanks to the SVI prediction model and XAI technique.	RMSE-3.360 and MAE-2.156 for SVI	Wongburi and Park (2022)
5	BOD, pH, TSS, TKN, FR, Temp	XGBoost (ML), k-NN (ML), SHAP (XAI), and eight ML models	The findings of this study have shown that the use of ML techniques can assist preserve chemical resources by improving chemical dosage management in wastewater treatment.	R^2 -0.605) for valve XGBoost, RF of R2-0.436. RMSE -8.056, and 4.466	Xu et al., 2022
6	COD, TN, FR	Explainable deep multi-task learning UNet (DMTL-UNet) (DL), XGBoost (ML), KSHAP (XAI)	A promising strategy for increasing the effectiveness and precision of sensor diagnosis and reconstruction in WWTPs is the suggested DMTL-UNet concept	R2-0.9175 and MSE-0.08408, F-score −99.08 % RMSE-31.1175	Ba-Alawi et al., 2023a
7	TSS, OP	RF (ML), XGboost (ML) and LightGBM (ML), SHAP (XAI)	The model comparison should be done from a variety of angles to make sure that all of the underlying details are exposed and looked at. It was found that SHAP to be really useful in this investigation.	RMSE-0.020 RME- 0.0050 R2-0.882	Wang et al., 2022a, 2022b, 2022c
8	COD, TSS, TN	multisensor fusion-based automated data reconciliation and imputation (MFS-ARI: Data fusion), KSHAP (XAI)	To evaluate the effects of missing, inaccurate, reconciled, and imputed data on the MBR performance operation utilizing R2AU-Net, the ASM-SMP-ARS integrated MBR model was used. In light of this, the suggested MSF-ARI based on R2AU-Net might, under suitable environmental discharge circumstances, reduce energy consumption by 37.44% and the appearance of early fouling by 10 days.	RMSE = 1.96 MAE = 0.31	Ba-Alawi et al., 2023b
9	TN and TP	Convolutional autoencoder (CAE) integrated with deep fully connected layers (DFC) (Neural Network), SHAP (XAI)	analysis-based on XAI, the relationships between variables and how they affected the output of the CAE-DFC model were clearly understood thanks to kernel SHAP.	R2 for TN and TP are 0.9607, and 0.9137	Ba-Alawi et al., 2023c
10	Influent phosphorus and chemical dosage data for phosphorus removal, 42 variables	OLS (Statistical ML), SVM (ML), DT (ML), RF (ML), ANN (ML), and SHAP (XAI)	Incomplete data sets can be used in this study as an illustration of how AI might be applied to process improvement and potential cost reduction.	R2 -0.496, accuracy of 79.7%	Xu et al., 2023a, 2023b
11	COD, BOD, SS, TN, and TP	Nine ML algorithms (KRR (ML) DT (ML), SVR (ML), kNN (ML), FCNNs (ML), RF (ML), XGBoost (ML)), SHAP (XAI)	The novel aspect of this work is how machine learning algorithms were used to estimate the formation of sludge in wastewater treatment facilities.	RMSE, MAE, MAPE, and R2 values of 4.4815, 2.1169, 1.7032, 0.0415, and 0.8218, respectively	Shao et al., 2023

effectively. This marked the importance of predicting the sludge production data and identifying the key factors that influence sludge production. Shao et al. (2023) used nine different ML models to predict the sludge production data obtained from Liaoning WWTP in China. The schematic of this case study is shown in Fig. 5. It came to know that for the data collected, XGBoost predicts better than other ML and ensemble learning models when metrics are compared. Its RMSE, MAE, MAPE, and R2 are 4.4815, 2.1169, 1.7032, 0.0415, and 0.8218. Ensemble learning fits highly nonlinear data better than the RF model, which has proven to be the second-best algorithm only to XGBoost. Traditional base learners like DTs, lasso regression, and kernel ridge regression forecast the prediction of sludge generation very poorly. On the other hand, complicated models like FCNNs have many parameters demanding a greater time in training the model. Despite equal prediction accuracy, NNs are not considered to be cost-effective when compared to XGBoost and RFs. It was also inferred in this study that on small and medium-scale datasets, complex ensemble learning models outperform base learners in prediction accuracy. To infer the details on influencing parameters over sludge production, the SHAP methodology is employed in interpreting the predictions of the XGBoost model. The method helps in interpreting the influence of each variable on predictions, which demonstrates the model's sensitivity to certain attributes. According to the SHAP plot, the influent wastewater volume (Q) and environmental temperature (T) were found to have the most substantial influences on the prediction of sludge generation. These results correspond with the input feature contributions of the XGBoost model. It

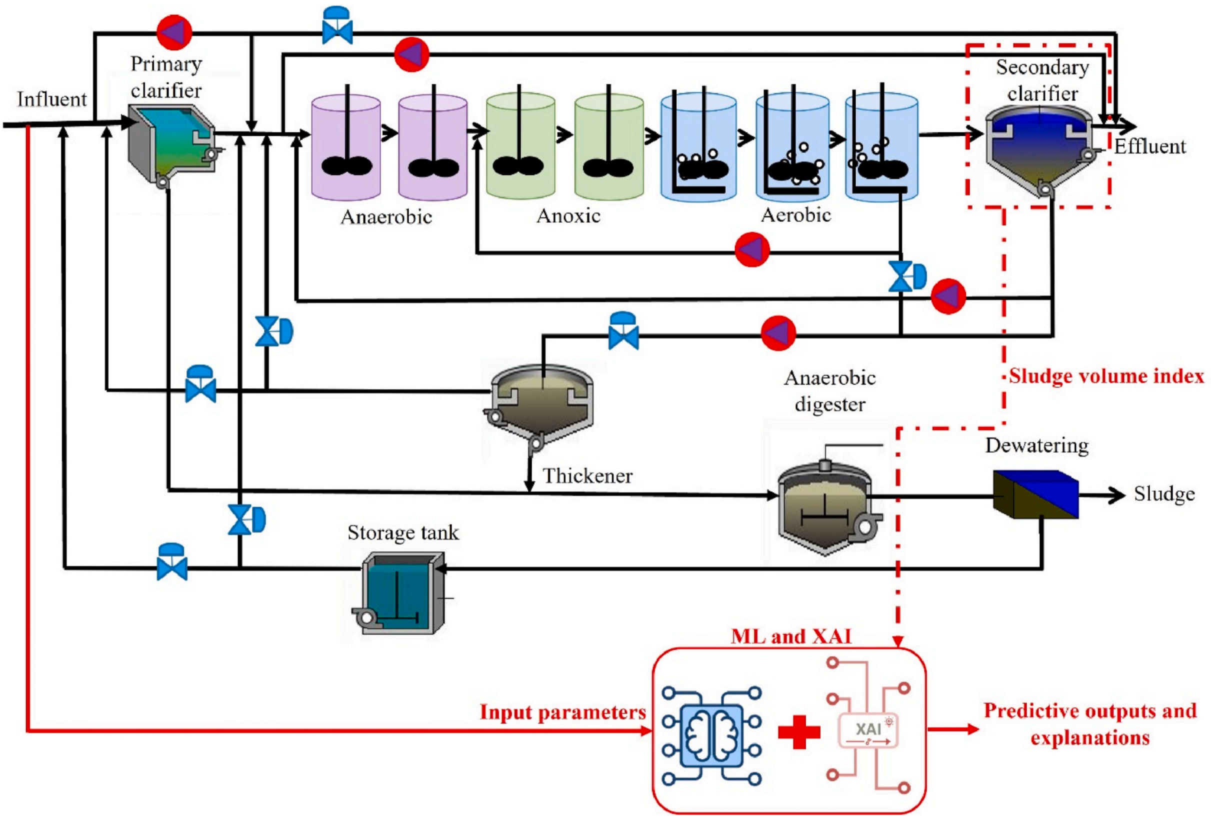

Fig. 4. The schematic of ML and SHAP applied to WWTP to predict and generate an explanation for the sludge volume index.

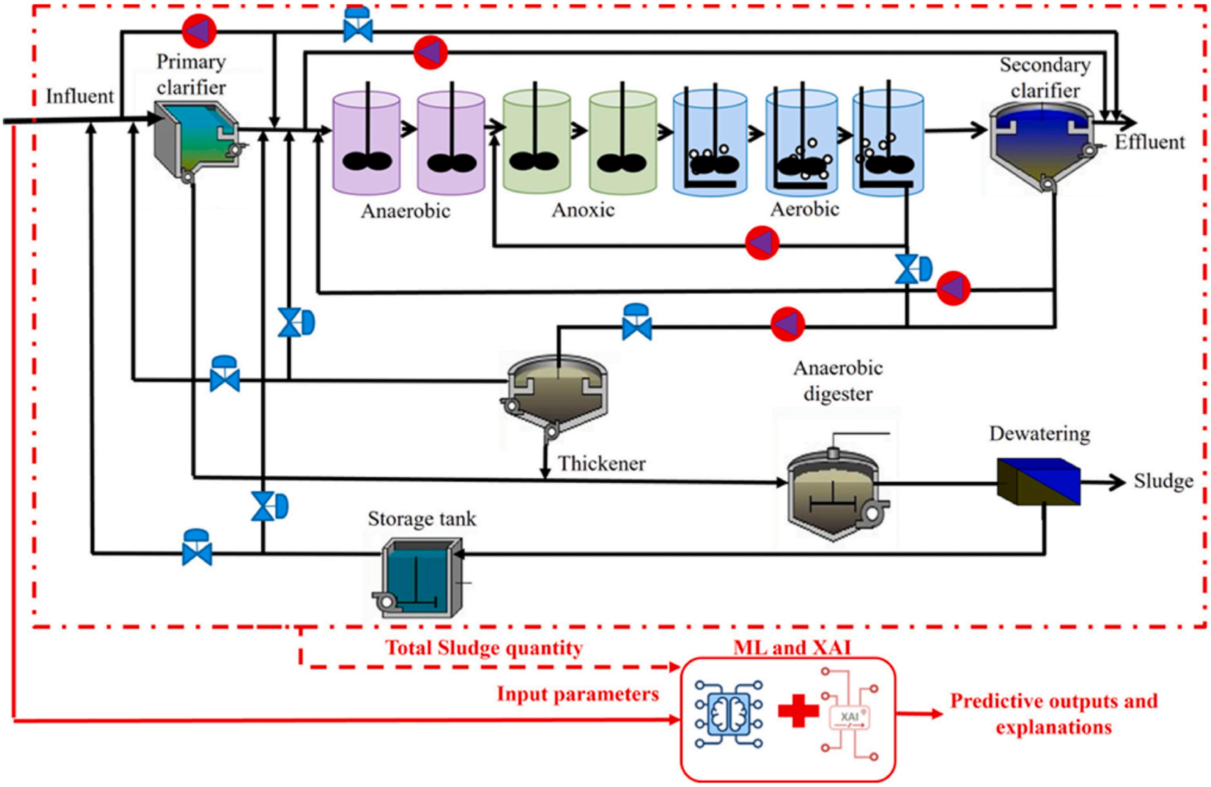

Fig. 5. The schematic of ML and SHAP models applied to WWTP to predict and generate an explanation for sludge quantity produced.

can also be noted that the significance of total nitrogen has risen in the SHAP study of water quality indicators, but the significance of suspended solids has diminished. The mismatch arises because SHAP values emphasize the importance of each feature within individual samples, but the model's feature contribution indicates the feature's weight, which is optimized through overall prediction bias, leading to divergent outcomes between the two approaches. Overall, the insights obtained through the results of this study suggest that sludge production is greatly impacted by the input features, such as daily wastewater treatment volume and temperature. Keeping control over these parameters can help to reduce sludge production and improve the overall efficiency of the treatment process.

4.2. *Process design and control in WWTPs*

4.2.1. *Adoption of XAI-based controllers in WWTPs*

Since XAI techniques are becoming more and more popular for modeling and identifying nonlinear systems, they have also been used for controllers in WWTP. Plant responses can be optimized over a certain time horizon by creating a control based on XAI. At the same time, the objective function can be made simpler to lower the optimization's computational requirements. Even with recent developments in XAI-based controllers (Machlev et al., 2022; Utama et al., 2022), the WW industry is still only beginning to use this strategy due to several factors. The reason could be a lack of knowledge and understanding of XAI techniques among WWTP operators and engineers. In addition, there are concerns about the reliability and robustness of XAI models in real-world applications. This represents challenges in the implementation of XAI-based controllers, which may necessitate significant investment in terms of training and infrastructure improvements. However, with the continuous advancements and success stories in XAI applications, it is expected that more WWTPs will embrace this strategy in the near future. In the context of creating smart process control, it is crucial to include an optimal controller and an emerging XAI network (XAIN). The XAIN helps identify the system and creates predictive models for the controllers, improving control effectiveness. One must notice that for XAI-based controllers to be used practically in WWTPs, system identification must be improved, and the amount of computing power needed for system optimization must be decreased (Kumar et al., 2018). The incorporation of advanced ML models such as DL and reinforcement learning can be a potential solution to improve system identification in XAI-based controllers for WWTPs.

4.2.2. *Reinforcement learning and deep learning in WWTP optimization and control*

Mohammadi et al. (2024) developed a simulator for the Deep reinforcement learning (DRL) environment using six models to determine the phosphorus removal process, reaching 97% accuracy. Without complicated system modeling or parameter estimates, DRL algorithm simulation scenarios are created using SCADA data with a suitable historical horizon to improve process control. The study of Croll et al. (2023), found that deep Q-learning, proximal policy optimization, and synchronous advantage actor criticism performed poorly in most circumstances. However, the twin delayed deep deterministic policy gradient (TD3) method consistently optimized control while meeting WWTP treatment requirements. TD3 control optimization reduced aeration and pumping energy requirements by 14.3% compared to BSM1 benchmark control. The important intermediate parameters prediction problem is solved using the deep neural network (DNN) model to guide control decisions. This study advances data-driven IoT system management and control, especially in circumstances with limited monitoring data resources (Shen et al., 2024). On influent pollution concentration tops and bottoms, RL responds differently. RL agents are more influenced by fines on tops (because of high effluent pollutant concentrations) and energy usage on bottoms. Finally, on weekends and in rainy and stormy weather, the RL agent cuts usage more than Proportional–integral–Derivative controls (PID). The adaptable RL agent can adapt to changing conditions better than PIDs (Hernández-del-Olmo et al., 2023), making it more efficient in managing and controlling the process under varying circumstances. A comprehensive and sophisticated explanation system, SHAP, was implemented to compare models and provide an in-depth analysis of the best model. XGboost is the optimal model for both Total Suspended Solids (TSS) and Orthophosphates (PO_4) tasks, whereas RF is the least optimal model due to overfitting and polarised fitting (Wang et al., 2022a, 2022b, 2022c). Using Proximal Policy Optimization, Filipe et al. (2019) proposed a method to optimize WW pumping station energy consumption by using deep-reinforcement learning. These models have shown greater potential in handling complex and non-linear systems by automatically extracting meaningful features from the data. Additionally, advancements in hardware technology, such as the development of specialized processors for accomplishing AI tasks, can greatly reduce the computing power required for real-time optimization, making XAI-based controllers more practical and efficient for WWTP applications.

4.2.3. *Emerging trends: transfer learning and smart automation*

The recent developments made in XAI have sparked anticipation that as the field develops, XAI-based controllers will be used more frequently in the water resource recovery sector to optimize WWTP procedures and tap energy usage. Even with recent developments in DL-based MPC with XAI, the WW sector is still only beginning to use this strategy. In addition to this, the use of transfer learning to process control in WWTP is still in its infancy, although it has gained significant attention in its growing stage. Thanks to transfer learning, with the transfer learning methodology, one can easily transfer control techniques developed for one system to another system, reducing the time spent in creating and implementing new control techniques in a newer system. For instance, to enhance conventional PI and PID controller strategies, an LSTM-based proportional-integral (PI) controller was developed based on the Benchmark Simulation Model 1 (BSM1) system for maintaining a DO concentration of 2 mg/L in an aerobic tank of a simulated WWTP (Alex et al., 2008; Sheik et al., 2023). This LSTM-based PI controller showed improved control performance compared to traditional PID controllers. This approach can be extended to LSTM-XAI-based controllers for better prediction and control of process variables. Later, XAI tools such as SHAP, LIME, etc. can be used for interpreting the controller's predictions. Once the benchmark strategy is developed, the pre-trained LSTM-XAI controller network can now be transferred to control DO in the remaining aerobic tanks without substantial modifications to hyperparameter values or neural architecture. This way the methodology developed using the source model, including explanations and insights can be transferred to a target model for obtaining interpretability and predictions with the knowledge acquired by the source model. Besides DO, the approach can also be used for other important variables of WWTP such as ammonia, TN, and TSS in the aerobic, anoxic, and anaerobic tanks. In general, this is the area with the most scope towards future digitalization in terms of smart automation in WWTP, enabling practical control and monitoring in WWTP. However, deeper research is required to investigate the possible application of transfer learning in WWTP process control and to identify the downsides of these techniques.

4.3. *Soft sensing*

Conversely, soft sensing uses data-driven models. Data-driven models such as SVM, ANN, RF, and Principal component analysis (PCA) are used to infer the values of variables that cannot be measured, difficult to measure, and expensive to measure in real-time using the measurements that are already available (Shyu et al., 2023). The type of model used in inferring measurements is data specific.

- For instance, SVM is used when there is a highly nonlinear relationship between the input variables and the target outputs. This is also used to handle the data comprising many input variables (pH, temperature, pressure, flow rate, component concentrations), leading the system to high dimensional space situations.
- On the other hand, ANN is also used when highly complex non-linear relationships exist between the input variables and the target outputs, which is the case in many industries. The other important feature of ANN is that it has self-learning capabilities, which help to learn the complex patterns between the input variables and target outputs, making them ideal for soft sensing.
- RF besides handling the nonlinearity between the variables, also provides information related to input variables, which are highly important for predictions of the target outputs. RF is also considered one of the most robust methods used for predictions, particularly in terms of missing data and even when there is noise associated with the data.
- In general, the high dimensional data with many input features makes it hard for the model to work and generate interpretations. The PCA model is employed in condensing the high dimensional data into a smaller set of newer variables. These newest variables are called principal components, and they are created without much altering the variance information from the original data sets.

The features of the above-mentioned data-driven models are important for soft sensing to estimate the variables that are hard to measure. Models that can infer the values of process variables that are normally challenging and expensive to quantify with hard sensors are known as soft sensors. Soft sensors create a model that can forecast the values of unknown variables or unmeasurable factors using previous data. The model's ability to continuously learn from fresh data is one of the benefits of XAI in the creation of soft sensors. This particular feature of soft sensing helps in maintaining the real-time monitoring of the system in a more systematic way, which thereby aids in controlling the process more efficiently. XAI is being utilized more and more to soft-sense important factors for tracking processes in WWTP to guarantee operational excellence (Alvi et al., 2022; Wang et al., 2022c). In WWTP, state estimation and soft sensing have both shown promise as methods for predicting process variables that are challenging to measure (Xu et al., 2023a; Chang et al., 2023). The aforementioned literature presents how XAI models can be used to enhance soft sensing and state estimation by giving precise estimates of process variables, which can lead to improved process control and optimization. It is also important to note that there hasn't been much research done in this area of work to predict difficulties, abnormal events, and operational glitches in WWTP using XAI (Ching et al., 2021).

4.4. Fusion of data and information

- **Importance:** The fusion of data and information is an essential requirement for several cutting-edge technologies, including the Internet of Things (IoT), computer vision, and remote sensing. However, fusion is a somewhat nebulous notion that can take numerous shapes (Murray, 2021). For example, in digital vision, feature fusion is the combination of features (Cheng et al., 2020l; Murray, 2021; Alvi et al., 2023; Sheik et al., 2024b). More accurate conclusions may typically be drawn by correlating and combining data from several sources than by analyzing a single dataset alone (Ding et al., 2019; Ly et al., 2022). Therefore, information and data fusion not only enhance the explainability of ML/DL models but also help in reducing process disruptions by improving the DM process (Zaghloul et al., 2022; Singh et al., 2022; Liu et al., 2023).
- **Challenges:** Data fusion can happen at three levels i.e., knowledge, models, and data (Arrieta et al., 2020). Smirnov, and Levashova (2019); Jiang et al. (2021), and Ba-Alawi et al., 2023a, 2023b, 2023c provide a thorough analysis of the reasons behind and methods by which fusion takes place to solve concerns associated with the IoT, privacy, and data security in WWTP. It is noteworthy that there is no relationship between data fusion and ML/DL models at the data level, making explainability difficult to explain the working methodology of ML/DL models in predicting the process variables. Though they perform better, there is still considerable confusion regarding the differences between information fusion and predictive modeling when using ML/DL models. The trade-off between explainability and performance is evident once more.
- **Scope of Improvement:** To obtain high-level features, the initial step involves the fusion of data with the initial layers of DL. This process tightly links the fusion and the tasks to be completed, making features correlated. To handle this correlation, various XAI strategies like LIME and SHAP have been advised by researchers (Ba-Alawi et al., 2023a; Wongburi and Park (2022); Park et al., 2021). The significance of these techniques was already discussed in the process modeling and simulation section of this review paper. These techniques help clarify how data sources are combined in a DL model, improving its usability. However, it is still unclear if the input features of a model may be inferred if a prior feature was known to be employed in that model. Further research is needed to determine the extent of the relationship between input features and prior features in a model. To gain a better understanding of what is happening in a model, it is recommended to enable XAI with the necessary ML/DL models.
- **Addressing Data Privacy and Security:** Empirical research attempting to solve data privacy issues has been lacking in the WWTP industry, posing a significant obstacle to XAI's usage, necessitating the need to overcome it. To overcome this challenge, federated learning, and differential privacy are identified as effective methods for addressing data privacy and security while promoting ML use (Xu et al., 2022, 2023b; Park et al., 2022a, 2022b). Finally, to make progress in utilizing ML for enhancing project performance (e.g., productivity, quality, and safety), it's crucial to make advancements toward appropriate data fusion strategies to enhance the understanding and interpretability of a model's decisions or predictions for better explainability. This will ensure accurate integration and interpretation of multiple data sources, leading to improved DM and problem-solving.

4.5. Using XAI on the internet of things

In the literature, there is a debate over the necessity of IoT in XAI. Doshi-Velez (2017) states that in certain situations, the integration of IoT in XAI might not be required and that the system can be trusted if the following conditions are met: (a) The need for IoT in XAI for better explainability is not a greater priority than the cost of implementing it in WWTP. (b) the impact of inaccurate results in the field of WWTP application is not too great, and (c) the problem has already been thoroughly studied and applied to real-world WWTP scenarios. However, the literature has emphasized the necessity of XAI-IoT in complex systems such as WWTP. Efficient IoT integration in WWTP is crucial for users to effectively handle ML results, regardless of whether this is due to business needs, moral dilemmas, or legal issues in the water industry (Confalonieri et al., 2021; Karthikeyan et al., 2022).

In the context of IoT, this section looks at the requirements and advantages of implementing IoT-XAI techniques in WWTP. XAI is very important to unravel the model behaviour in predicting the critical variables of a WWTP, particularly when it comes to implementing the decisions that are developed using the model predictions. Uncertainty and inefficiency in model predictions might lead to a situation where the decision system is impacted and may have significant consequences for the overall performance and effectiveness of the WWTP. The uncertainty and inefficient model predictions can be due to data abnormalities, which may occur for many reasons. For instance, in AI-driven processes that are completely data-driven, there might be situations of misrepresentation of data due to sensor failures or deviation of any equipment from normal operation, creating anomalies in the process system. Due to data anomalies, unexpected deviations could appear in the ML algorithm's training set. This could lead to inaccurate predictions or decisions, which can have serious consequences in complex processes such as WWTP. Integration of IoT with XAI can help address these failures by leveraging already-existing past data. The IoT in general, which has the ability to collect, process, and analyze real-time data, can overcome the impact of data anomalies in predicting process outcomes. At any given instant in a sensor failure situation, it can make use of past data to process it into ML models for the purpose of predicting process outcomes without any delay. Further, with the history of the past, IoT can also detect patterns and anomalies that may indicate potential sensor failures or degradation in process performance. With the IoT platforms, one can utilize predictive maintenance techniques, where models are trained on historical data to predict when sensors are likely to fail or require maintenance. This enables more accurate predictions and proactive DM, ultimately creating the synergy between IoT and AI to improve the accuracy and reliability of predictions and decisions. However, it has been observed that very few research articles specifically emphasize the use of IoT tools with XAI models in the predictive control of WWTP, even though such tools have a lot of potential for prediction and control. Fig. 6(A) depicts the flow path toward the usage of XAI for prediction and control in WWTP and Fig. 6(B) IoT integration in XAI for seeking different applications in WWTP.

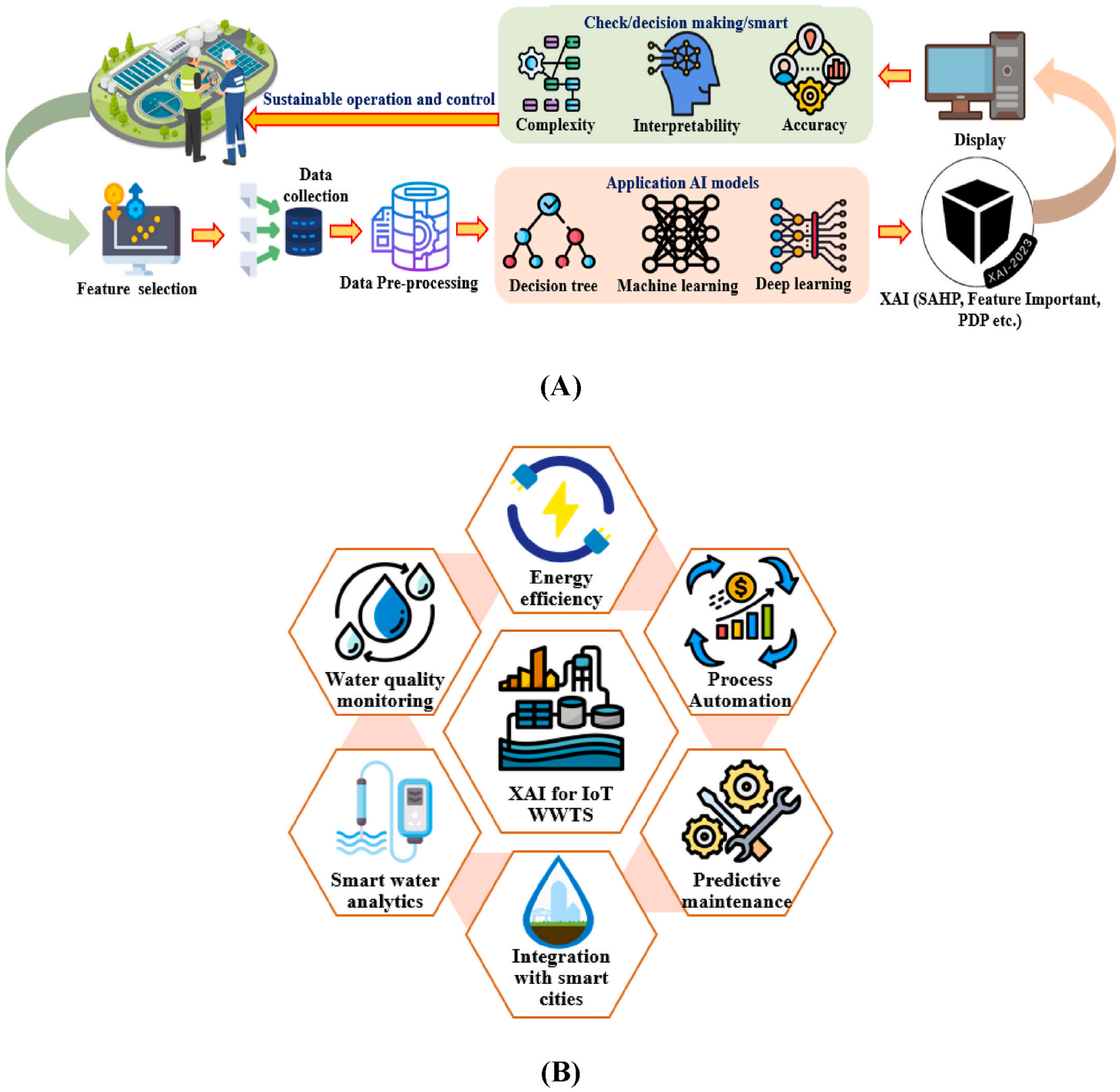

Fig. 6. (A) Flow path of usage of XAI for prediction and control in WWTP **(B)** IoT in WWTP for seeking different applications.

4.6. Future research directions, scope, and challenges

Future research in Explainable Artificial Intelligence (XAI) for wastewater treatment should prioritize the development of domain-specific interpretable models that can effectively integrate with future technologies such as the Internet of Things (IoT), digital twins, and edge computing. Furthermore, the development of techniques for interpretable reinforcement learning, energy optimization, and human-centered interfaces will guarantee that Explainable Artificial Intelligence (XAI) not only improves performance but also promotes confidence and transparency among plant operators and regulatory authorities. The following aspects can be the prospective research approaches that can be thought of in the context of XAI technology.

- Create digital twin models that can replicate the WWTP and include XAI to offer valuable insights into their recommendations. An AI-powered digital twin might replicate various operational situations and provide explanations for why specific modifications (such as decreasing aeration levels to monitor DO or increasing the chemical dosage to reach the optimal pH value of the process) would enhance efficiency.
- Exploring techniques to integrate sensor data in real-time while providing clear explanations of AI model results. One example is elucidating the reasons towards the need for modifications in chemical dosage when there is a rapid shift in pH sensor readings and the interaction between several IoT sensors in facilitating such adjustments.
- Focus on developing interpretable RL methods specifically for wastewater treatment. For example, incorporating explainable reward functions where the optimization objectives (e.g., minimizing energy use or chemical waste) are aligned with human-understandable metrics, like cost savings or environmental impact. Integrating RL into process control and addressing potential biases in ML models are essential for building robust, reliable, and fair AI systems in WWTP applications. Many WWTPs already use well-established control methods like PID controllers or Model Predictive Control (MPC). One should develop integrating RL into these systems without causing disruptions for better control and efficiency.

It is undeniably difficult to stay up to date with the most recent advancements in XAI research due to its development, which is happening at a rapid pace. However, it is crucial for researchers and practitioners to continually educate themselves to keep pace with new XAI developments and make meaningful contributions to applying them to WWTP, paving the way for creating research opportunities towards building a smart and sustainable water industry.

Additionally, XAI offers a crucial step in developing process fairness and considering bias during the algorithmic DM process (Mougen et al., 2021). Furthermore, XAI enhances transparency and trust by offering explanations for algorithmic decisions. This ultimately leads to better user understanding and acceptance of ML systems and increased public trust in the field. For this to happen, the following areas need to be addressed and prioritized.

(a) developing frameworks to bridge the gap between WWTP experts and XAI developers for smooth design and implementation.
(b) establishing a framework for independent auditing and validation of models.
(c) promoting transparency and explainability in algorithmic DM processes.
(d) fostering trust and public confidence in employing algorithmic systems by emphasizing the importance of their usage in leveraging process system behaviour.
(e) emphasizing the importance of the process of data fusion to enhance explainability and improve DM capabilities, which is crucial in establishing efficient and effective algorithmic systems. These areas are crucial for understanding the potential of XAI in improving water treatment and management.

The literature that is currently available highlights the conflicting scenarios that arise in the field of developing and implementing XAI in WWTPs (Belle, and Papantonis, 2021a, 2021b; Khalil et al., 2023; Yang et al., 2022; Shao et al., 2023). While they are being developed, there is often a lack of consensus regarding explicit objectives for explainability and techniques to evaluate the quality of explanations. One reason might be attributed to the fact that there is a lack of collaboration between the varying professions and expertise levels of individuals involved in XAI research and development. The other reason could be the necessity for different models tailored to specific datasets that pertain to individual process systems, depending on their complexity levels. This makes everyone agree that the process of assessing explanation approaches is not rigorous enough (Doshi-Velez and Kim, 2017), a critical issue that needs to be addressed in order to advance XAI. Often the evaluation criteria for successful implementation of XAI in WWTPs is primarily based on both the opinions of computer scientists and WWTP process managers. For example, computer scientists mostly act as developers of XAI, and WWTP process managers act as experts in evaluating the working efficacy of XAI based on the explanations they provide. Thus, in the context of implementing XAI in WWTP, the collaboration between different professions is essential to meet the unique needs, expectations, and demands for the successful integration of XAI into WWTPs (Langer et al., 2021; Love et al., 2023). Fig. 7 depicts the usage of the XAI application for WWTP with a flowchart and future directions. Also, when we speak in the context of process control, XAI can benefit in identifying the key variables (such as WQ parameters, dissolved oxygen, and flow rates) that significantly influence decision-making models. XAI models like SHAP or LIME can visually represent model decisions. AI systems could use these visualizations to provide operators in control rooms with a clear understanding of the rationale behind specific control actions. Additionally, with IoT sensor data, AI models frequently identify anomalies. However, XAI can explain these anomalies by illuminating the patterns that gave rise to the detection. This is essential for early fault diagnosis or preventing false positives. Data from IoT devices may be noisy or unreliable. The model's response to this noise can be explained by XAI, which also reveals which sensor data is most trustworthy for making decisions. Understanding how soft sensors estimate the values of unmeasured variables will help XAI improve soft sensor credibility. Also, when WWTP operators use soft sensors with lower confidence in their estimations, XAI may be able to help them understand prediction uncertainty and make better judgments.

The concept of XAI gained prominence in the context of WWTP when process managers expressed dissatisfaction with the lack of established standards for process assessment. This need for better process assessment to enhance plant operation has sparked lively debates between AI developers and WWTP process experts, propelling the field of XAI ahead. As a whole, meeting process manager demands is the primary driver behind XAI's growing appeal for enhancing WWTP operations. The literature clearly shows how keen researchers are to use newly developed models to tackle complex process system problems. Table 2 reports the summary of XAI taxonomies and methods. Table 3 reports the comparison of different modeling strategies in WWTP. A lot of evidence suggests that humans have over-trusted ML systems in the past, and there is still a long way to go before they can trust ML systems completely in the present. XAI has significant limitations, which are detailed below, in addition to the advantages and possibilities it offers in the field of complex decision systems (Watson, 2020; Xu et al., 2023a, 2023b). Some basic concepts in XAI are unclear or contradictory because there is no standard terminology existing in the field of XAI, which leads to confusion and differing interpretations between the developers and the users. For example, while everyone agrees that explanations by models should be precise, it's uncertain whether the focus should be on

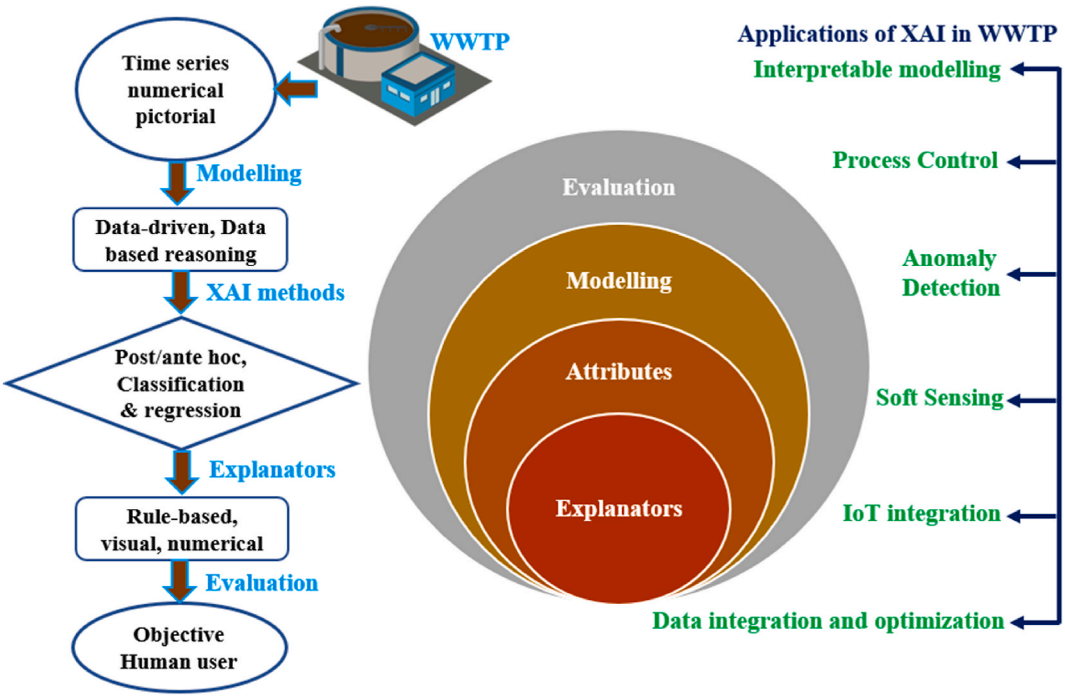
WWTP
Time series numerical pictorial
Modelling
Data-driven, Data based reasoning
XAI methods
Post/ante hoc, Classification & regression
Explanators
Rule-based, visual, numerical
Evaluation
Objective Human user
Evaluation
Modelling
Attributes
Explanators
Applications of XAI in WWTP
Interpretable modelling
Process Control
Anomaly Detection
Soft Sensing
IoT integration
Data integration and optimization

Fig. 7. Usage of XAI application for WWTP with flowchart and future direction.

the data generation process or the target model. Some argue that the focus should be on the data-generation process, while others argue that the focus should be on the target model. This lack of standard terminology will not only hamper the advancement of XAI application techniques but also cause philosophical debates that often lead to confusion and unproductive conversations. Hence, it is crucial to establish a common understanding and terminology in the field of XAI. Also, the lack of a measurable indicator to verify the accuracy of interpretability in the ML/DL models used always poses challenges. To overcome challenges in evaluating the accuracy of the model's interpretability, the systems field's expertise should be taken into consideration. Involving process expertise can help AI developers choose the appropriate ML/DL models specific to the data to produce the desired outcomes. This can lead to a more reliable method for choosing ML/DL models that are specific to user needs, leading to an accurate evaluation of the model's interpretability and performance in real-world scenarios.

In the WWTP sector, interpretability becomes a significant challenge when employing complex models like deep neural networks (DNNs) and reinforcement learning (RL) to analyze or predict critical process variables. These models are powerful but often difficult to understand, creating a need for reliable interpretability to ensure transparency and build trust among the process operators. DNNs, for instance, are frequently criticized as "black boxes" because they make predictions in ways that are not easily understood by domain experts. Similarly, RL models, which rely on trial-and-error learning, are also challenging to explain, as their decision-making processes are less intuitive. The lack of interpretability makes it difficult to deploy these models in real-time operations within WWTPs, where quick and clear decision-making is essential. To overcome these limitations, researchers must focus on improving the interpretability of complex models in the WWTP context. One approach is to develop techniques like saliency maps, which visually highlight the most influential parts of the input data that drive the model's decisions. Another promising direction is integrating interpretable models, such as decision trees, with complex models like DNNs and RL. This hybrid approach can provide both the accuracy of advanced models, and the transparency needed for practical deployment, offering a balance between performance and clarity. Adopting these techniques will not only enhance trust but also enable real-time deployment of these advanced models in critical WWTPs. Although there are challenges and opportunities related to integrating XAI with different technologies, there has to be considerable ideas related to the scalability of XAI in WWTP, real-time deployment, and handling noisy data. These can be outlined with the following research questions, which can also be an enroute to further scope and challenges with a brief explanation about how they can impact the successful implementation of XAI in wastewater treatment plants.

RQ: How can the scalability of XAI in WWTPs be planned to manage the complexity and size involved with WWTP operations?

Response: WWTPs are large and complex systems involving numerous interconnected processes such as chemical breakdown treatments, filtration, and biological processes. Each of these processes generates a vast amount of data, which is quite complex and difficult to interpret in real-time. Scaling XAI techniques to handle the complexities involved in the WWTPs is crucial. WWTPs, which are larger in size and incorporated with complex control systems, demand XAI solutions to evolve in a way that can handle massive datasets without compromising on providing clear and interpretable insights to operators. This can help them make informed decisions to optimize performance and efficiency. On the other hand, it is also necessary to investigate how XAI techniques or methods perform in real-time scenarios to ensure their effectiveness and reliability in the dynamic environment of WWTPs. This could further assist in highlighting the areas of optimization in a typical large-scale WWTP. Although there are huge complexities associated with the scalability of XAI, a promising direction is to break down the complexities by developing hierarchical XAI frameworks, where explanations are provided in-depth for each subsystem involved in the WWTP. These frameworks help in explaining the decisions at both the individual process level (e.g., why a chemical dosage was made to adjust pH) and at the system-wide level (e.g., how different modifications across various processes affect overall plant performance). This approach of multi-level interpretability can help plant operators understand both localized and broader system dynamics associated with the WWTP. Also, handling large data sets in real-time to provide interpretable outputs is quite complex. It is advisable to explore the potential of cloud-based or

Table 2
Summary of XAI taxonomies and methods.

XAI approaches	Model patterns	Remarks	References
Functioning-based model	Linear Model, Decision Trees, Local Perturbations, Leveraging Framework, Meta Explanation, Architecture Modification, Partial Dependence Plots (PDP), Shapley Values, Local Interpretable Model-Agnostic Explanations	In XAI, the use of function-based models aims to shed light on the process and reasons behind which a complicated model makes specific predictions. This can make the model's behaviour and judgments easier for users, key players, and policymakers to gain insight and trust. Choosing which function-based model to apply will rely on the particular issue at hand as well as the intricacy of the black-box model that needs to be described.	Karamichailidou et al., 2022; Han et al., 2019; Castillo et al., 2016; Xu et al., 2023a, 2023b, Confalonieri et al., 2021; Angelov et al., 2021; Love et al., 2023

Rule-based model	Rule Extraction, explicit If-Then Rules, Transparency, and Interpretability, Fidelity to the Original Model, Manual or Automatic Rule Generation, Interpretable Variables, Consistency, and Fairness, Hybrid Models	Rule-based models are particularly useful when it comes to providing clear and human-readable explanations for AI choices. Nonetheless, they may not capture all of the intricacies and intricacy of specific activities, and alternative XAI techniques may be better applicable in some circumstances. The XAI approach of choice is determined by the individual application and its proportions of transparency, interpretability, and prediction performance.	Irani and Kamal, 2014;
			Li and Gong, 2019;
			Dupuitet al., 2007;
			Balla et al., 2022;
			Love et al., 2023

Explainable AI methods			
Methods	**Model prediction types and concepts**	**Remarks**	**References**
SHapely Additive explanation (SHAP)	Summary Plot, Individual Instance Plot, Force Plot, Waterfall Plot, Dependency Plot, Interaction Value Plot, Feature Attribution Heatmap, Time Series Shapley Explanations, Text Shapley Explanations, Kernel Shap, and Deep Shap	The precise application case and the requirement for global or local interpretability determine the SHAP explanation type to be used. To fully comprehend their models, practitioners frequently combine these several explanation kinds. To increase openness, equity, and trust in AI systems, SHAP is a flexible framework that can be used with a variety of ML models and data formats.	Parsa et al., 2020; Wang et al., 2022a, 2022b, 2022c; Love et al., 2023

Fuzzy Classifier	Adaptive Learning Capabilities, Handling Uncertainty, Interpretability, Fuzzy Partitioning, Fuzzy Aggregation, Linguistic Variables, Fuzzy Inference System, Fuzzy Rules, Membership Functions, and Defuzzifier.	Applications in wastewater such as expert systems, and control systems, all frequently make use of fuzzy classifiers since they require human judgment or domain knowledge to make decisions. As a result of the presence of hazy or ambiguous information, they perform best in situations when clear, rule-based, or probabilistic classifiers may not be effective but helpful tools for developing interpretable AI systems. Fuzzy classifiers are an important tool for creating interpretable AI systems because they provide transparency and comprehensibility in the realm of XAI. This is because their rules and linguistic variables may be utilized to describe how a decision was made.	D'Alterio et al., 2020;
			Duarte et al., 2023;
			Love et al., 2023

Gradient-weighted Class Activation Mapping (Grad-CAM)	Vanilla Grad-CAM, Grad-CAM++, smooth Grad-CAM, Grad-CAM with Box, Layer-wise Grad-CAM, Multi-Class Grad-CAM	Each form of Grad-CAM has unique benefits and applications, so choosing one to employ depends on the particular issue at hand as well as one's tolerance for visual detail and noise. DL can be better understood and trusted by using these techniques, which act as useful explanation and interpretation tools.	Gireesh et al., 2023; Akkajit et al., 2023; Love et al., 2023
Layer-wise Relevance Propagation (LRP)	Epsilon-LRP, Alpha-Beta LRP, Deeplift, Layer-wise Scaling LRP, PatternNet and Pattern Attribution, Sequential LRP, Layer-wise Relevance Visualization	The DL architecture and level of interpretability required will determine which LRP variation is best. In applications of WWTP such as process, energy, and autonomous systems, where model openness and	Montavon et al., 2019; Love et al., 2023

(continued on next page)

Table 2 (*continued*)

XAI approaches	Model patterns	Remarks	References
		accountability are essential, LRP approaches are very helpful for explaining deep learning models. Users can learn more about a model's information processing methods and the features that have the most effects on how it makes decisions by examining these LRP versions.	
Local Interpretable Model-agnostic Explanations (LIME)	Tabular Data LIME, Time Series LIME, Regression LIME, Multimodal LIME, Time Series Forecasting LIME, Structured Data LIME	The type of LIME to employ will depend on the specific scenario, the kind of data, and the ML model being utilized. LIME is a flexible XAI tool. It offers a method for producing locally precise and comprehensible explanations for model predictions, making it simpler for users to comprehend, believe in, and perhaps even troubleshoot complex models.	Davagdorj et al., 2021; Love et al., 2023; Zahra et al., 2023a, 2023b

distributed XAI models that could streamline the fast processing of large data sets while maintaining high accuracy and reliability in the interpretation of results, enabling operators to quickly react to changes in system behaviour.

RQ: What are the methods for real-time deploying of XAI in WWTPs?

Response: Methods such as edge computing, hybrid systems, and real-time visualizations can be used effectively for the real-time deployment of XAI in WWTPs. In edge computing, it can be assured to deploy XAI directly on the local hardware where the data is generated. Through edge computing the data generated from the sensors and other monitoring devices of WWTP is processed locally. This allows XAI models to generate immediate predictions and explanations about how the input variables are related to desired process outcomes. It is also well known that some of the XAI models are quite complex and often take more time to process real-time data. This marks the need for developing an approach that can quickly analyze and respond to real-time data. For the smooth deployment of this technology, it is better to choose very simplified or streamlined XAI models that are particularly designed for quick decision-making without compromising too much on accuracy or interpretability. These models focus on providing explanations that are concise and relevant to the operator's needs. For example, if we consider the parameter pH, the streamlined XAI models give a simple explanation to increase the chemical dosage considering the pH level has dropped below the required range for optimal performance. This way, by focusing on the key variables like pH level and flow rates, the system gives a clear explanation about how these variables influenced the decision. This will help the operator take quick action without needing to know every tiny detail about how the model arrived at such a decision.

Table 3
Comparison of modeling strategies in WWTP.

Modeling	Advantages	Limitations	
Mechanistic	•Established field with computer assistance and established models •Improved forecasting in novel situations •Adaptable to a scaled perspective in both space and time	•Needs an in-depth understanding of the underlying mechanism. •Mathematically demanding and can involve complex equations. •Numerous factors, and solvers.	Meirlaen et al., 2001; Sheik et al., 2023b; Monje et al., 2022; Ramin et al., 2022
Empirical	•Extremely low-cost computationally •Doesn't call for specialized knowledge •Not limited to substantial datasets	•It is necessary to develop appropriate data features for efficient learning. •Can only symbolize a small subset of I/O connections. •Offers no understanding of how things work.	Raduly et al., 2004; Poorasgari and Örmeci, 2022; Langeveld et al., 2017
Machine learning	•Cost reductions and better efficiency may result from this. •This facilitates data-driven decision-making, which can result in choices that are more precise and knowledgeable. •It is useful for jobs like predictive maintenance, picture and speech recognition, and fraud detection.	•Incomplete, skewed, or noisy data might produce poor conclusions and erroneous forecasts. •Overfitting is reduced through the use of regularization procedures. •Effective generalization might vary based on the data and algorithm used.	Guo et al., 2015; Singh et al., 2022; Sheik et al., 2024a; Torregrossa et al., 2018; Ly et al., 2022
Deep learning	•Not reliant on choosing particular characteristics and outputs (may make use of process data that is readily available). •I/O may be continuous or categorical.	•Offers no understanding of how things work. •There are many parameters and hyper-parameters to tune, which calls for expertise. •Learning necessitates big datasets.	Zhang et al., 2023a, 2023b, 2023c; Alvi et al., 2023; Li et al., 2022
Reinforcement learning	•It can be used in domains like resource allocation, recommendation systems, and stock trading. •RL is used to solve optimum control-related problems. •For multi-objective optimization issues, it is helpful.	•In real-world scenarios, this can be too expensive or impracticable. •Policies that are not ideal can result from inadequate exploration. •Selecting the right reinforcement learning algorithm for a given task might be difficult.	Aponte-Rengifo et al., 2023; Yang et al., 2021; Zhou et al., 2022
Explainable AI	•By revealing how AI systems make decisions, XAI helps to build trust in those systems. •Developers and data scientists may find and fix	•The interpretability and efficacy of AI models are frequently trade-offs. •For certain usage circumstances,	Ba-Alawi et al., 2023a; Xu et al., 2023a, 2023b; Duarte et al., 2023; Bourahla and Bourahla, 2022

(continued on next page)

Table 3 (*continued*)

Modeling	Advantages	Limitations
	mistakes in AI models and datasets with the aid of XAI. •XAI can direct maintenance or repair procedures and assist in determining the root causes of problems. •Resulting in more thoughtful choices and behaviours.	incomplete explanations may be deceptive or insufficient. •This may result in a lack of clarity and consistency while adopting and assessing XAI solutions. •AI methods might have trouble providing coherent explanations for deep neural networks.

Further, if XAI models are able to provide real-time explanations, the operators who are working in a highly dynamic and complex environment must be in a state to comprehend the explanations. Real-time visualization tools help operators to interpret and comprehend the model outputs by converting explanations into easily understandable intuitive insights. For example, in a typical WWTP control room, a visualization dashboard shows either the increase or decrease in DO level alongside a real-time explanation of why the AI model recommends increasing or decreasing aeration levels in a treatment tank. In this way, it is helpful for the operator to see both the decision and the explanation in an easily understandable format, thereby aiding operators confidently implement the informed decisions obtained from XAI models. All the above methods contribute to real-time smooth deployment of XAI in WWTPs in an efficient and effective manner.

RQ: How can XAI models handle noisy and incomplete data in WWTPs?

Response: There are also situations where the interpretations of the system are inaccurate. This can happen due to various reasons, such as noise, incompleteness, or bias in the training data, which might result in inaccurate interpretations and subpar model performance. To avoid data discrepancies due to noise and incompleteness associated with the data, it is recommended to create more robust XAI models (Random Forest and Gradient Boosting (RFGB), Robust Regression Models, Shapley Additive Explanations (SHAP) with Noise-Resilient Models, etc.) that can handle data uncertainty and unreliability situations without losing accuracy. Models of this type could help us provide explanations about which part of the data is impacted by noise, thus allowing operators to rely on the model's recommendations. On the other hand, it is recommended to develop appropriate data preprocessing techniques that help in cleaning the data or input missing data using soft sensing techniques based on the previous trends (historical data) available. This would lead to reliable predictions and explanations that can be trusted by decision-makers in WWTPs. The level of uncertainty in the predictions due to data discrepancies occurring owing to noise and incompleteness can be explained using uncertainty-aware XAI models (Bayesian Neural Networks (BNNs) and Gaussian Processes (GP)). These models help not only in providing insights about the level of uncertainty associated with their predictions but also offer explanations degree of confidence that should be placed in the prediction made. This could help operators understand how much trust they can place in the prediction before making an informed decision, especially when dealing with noisy or incomplete data.

Inadequate data preprocessing procedures, such as missing values or inaccurate feature scaling, can also lead to inaccuracies in the functioning of the model's logic. This could therefore complicate the accuracy of reasoning of conclusions (Ba-Alawi et al., 2023b). In cases when certain categories or results are not adequately represented in the data, the explanations may excessively emphasize more common outcomes, therefore reducing the interpretability or generalisability of the model (Belle, and Papantonis, 2021a, 2021b). Models of great complexity such as deep neural networks or ensemble techniques (e.g., RF, gradient boosting) are generally challenging to interpret. Comprehending the internal mechanisms of these models necessitates the use of specialized methodologies such as LIME and SHAP to elucidate specific choices (Gilpin et al., 2018). More straightforward models such as linear regression and decision trees are easily understandable, but they may not possess the same level of predictive capability in comparison to other complex models. Presenting local explanations, which explain a single prediction, as opposed to global explanations, which explain the overall behaviour of the model, can be challenging. Local interpretations may logically conflict with global explanations, hence introducing complexities in the understanding of the model behaviour (Love et al., 2023). This demands the need for mastery of the subject matter to be critical when integrating AI systems into wastewater treatment plants (WWTPs) to ensure that operators can accurately interpret AI-driven insights and make informed decisions. Without a strong grasp of the subject matter, it may be difficult to fully leverage the capabilities of AI systems in WWTP operations. Besides this, there can also be challenges related to how XAI can be scaled for real-time monitoring in large-scale plants, or how data fusion techniques can handle the growing influx of IoT data while maintaining interpretability.

Real-time WWTPs need to adjust the dynamic circumstances that occur in the plant such as production changes and machine malfunctions, and accordingly, the explainable artificial intelligence models must update their explanations. While scaling the technology of XAI in real-time monitoring large-scale WWTPs, there can arise several technical challenges pertaining to developing efficient algorithms for real-time use, establishing distributed computation, data prioritization before it is being processed to working algorithms, and deploying scalable XAI models. Ideally, large-scale WWTPs generate enormous amounts of timely data, and it is crucial that AI algorithms efficiently process and analyze this data to provide meaningful explanations. The complexities associated with the existing XAI algorithms may not be able to keep pace with the rapid data influx, resulting in slow or inadequate responses. This necessitates either the development of efficient algorithms or the improvising the existing algorithms that can swiftly analyze incoming data and provide immediate explanations for DM. On the other hand, to enhance the speed and efficiency of XAI algorithms in processing the vast volumes of data that are continuously generated by numerous sensors and equipment, it is necessary to implement distributed computing systems in large-scale WWTPs. This can offer solutions by spreading the computational workload across multiple machines or processing units. It is inevitably evident that data prioritization and unified model infrastructure are important for the efficient scalability of XAI in large plants. The advice is kept over emphasizing prioritizing critical data points for explanations and interpretability that reflects the anomalies or system failures rather than analysing all data in detail. At the same time, an underlying unified model infrastructure must be designed to scale as the WWTPs expand with new equipment, additional processing stages, increased processing capacity, or when processes become more complex promptly. The approach can aid XAI models to scale with these changes, ensuring that new data sources are incorporated seamlessly, making it easier to monitor new systems without overloading the existing infrastructure. This combination of data prioritization and scalable infrastructure allows XAI to deliver timely relevant insights, even in growing large-scale WWTP operations.

The knowing of how data fusion techniques can handle the growing influx of IoT data while maintaining interpretability is also important in the context of establishing XAI in WWTPs. IoT devices produce a wide variety of data types, including structured, unstructured, time series, spatial, and categorical data. Fusing such disparate types of data while maintaining coherence and interpretability is complex. Advanced data fusion techniques, such as DL, are often complex and behave as black-box models, which complicates interpretability. This underlies choosing data fusion techniques such as hierarchical fusion, sensor-level fusion, context-aware fusion, and feature-level fusion that contribute to both scalability and interpretability. These fusion techniques offer a more transparent way to combine different types of data and extract meaningful insights. Hierarchical fusion involves combining data at different levels of abstraction, allowing for a more holistic view of the information. Sensor-level fusion integrates data from multiple sensors and tries to reduce redundancy and noise before presenting the data to AI models. Context-aware fusion takes into account the surrounding environment to enhance the overall understanding of the data. Feature level fusion combines specific features from different datasets to create a more comprehensive representation of the information. By utilizing these techniques, it can be ensured that despite the growing influx of IoT data, the insights remain interpretable and relevant.

Large WWTPs typically consist of complex processes in which it necessitates controlling the critical process variables to ensure smooth and efficient operation without any process disruptions. Integrating reinforcement learning can help mitigate the risks associated with unexpected process disturbances by allowing the system to adapt and learn from past experiences. This adaptive learning approach enables large plants to make real-time adjustments based on changing conditions, ultimately improving overall performance and reducing downtime. For efficient performance through reinforcement learning, it also demands minimizing the severity of the biases in the data. The data imperfections can allow ML models to develop biases, making the models overestimate the water quality parameters and potentially leading to inappropriate control or actions. The biased ML models might incorrectly favour certain processes specific to WWTP or fail to detect less frequent but important anomalies, ultimately leading to uncontrolled inefficiencies in the WWTP. Therefore, it is crucial to thoroughly analyze and clean the data before implementing it into the machine learning algorithms. Regularly auditing the data, ensuring diversity in data collection, and applying unbiased techniques can reduce these biases. By ensuring that the data is accurate and free from biases, the system can effectively learn and adapt to inevitable disturbances that occur in the process, leading to more reliable and efficient operations. In summary, integrating RL into process control and addressing potential biases in ML models are essential for building robust, reliable, and fair AI systems in WWTP applications. In conclusion, by overcoming the aforementioned challenges with the suggested scope of improvements, and by exploring future research directions, XAI holds great potential for implementation in WWTPs.

5. Conclusion

The advanced ML models are excelling in energy, water, and power system applications. However, consumers and water specialists could have difficulty understanding such algorithms if they don't completely comprehend the working procedure and the rationale behind the prediction of outputs. Consequently, the objective of XAI is to make ML/DL models more credible and understandable. In order to achieve this, XAI focuses on developing techniques and tools that can provide explanations for the decisions made by ML and DL models. Over the past few years, XAI has gained considerable attention, facilitating researchers to increasingly incorporate its application into projects within the WWTP and WS. This research review highlights intriguing patterns in the field's recent work and could provide insight into the WWTP scenarios in which XAI approaches are applied. Even though XAI can look easier to apply in reality, there are still issues to consider when using it to its fullest potential to improve user's confidence. Besides having challenges in implementing XAI, the authors think that XAI approaches have a great deal of promise to explain the choices made by ML models when they are employed in the field of WWTP. Moreover, XAI has the capacity to meet the requirements of environmental quality researchers seeking comprehensive process assessment. In addition to various techniques available for explainability within XAI, the literature highlights SHAP and LIME as the predominant methods utilized. These methods are preferred for their efficacy in clarifying the processes and mechanisms by which models function in predicting process results. The difficulties and restrictions associated with embracing and applying XAI techniques in the realm of WWTP are significant additional topics addressed in this work. Furthermore, future research objectives and possible applications pertaining to WWTP and XAI were presented. Among these are WWTP monitoring, process efficiency, and effective process management and control. The advantages XAI possesses, like providing transparency in explanations for improving informed DM process, might speed up its integration in a variety of industrial applications. Furthermore, IoT integration with XAI presents promising avenues for addressing issues associated with predictive maintenance and anomaly detection, particularly in scenarios involving sensor failures. By combining IoT sensor data with XAI capabilities, it becomes feasible to predict and mitigate potential equipment malfunctions or irregularities more effectively, thereby enhancing system reliability and operational efficiency. In summary, this work offers numerous instances and prospects of how XAI might be helpful in the field of WWTP.

CRediT authorship contribution statement

Abdul Gaffar Sheik: Conceptualization, Methodology, Data curation, Writing – original draft, Writing – review & editing. **Arvind Kumar:** Writing – review & editing. **Chandra Sainadh Srungavarapu:** Writing – review & editing. **Mohammad Azari:** Writing – review & editing. **Seshagiri Rao Ambati:** Writing – review & editing. **Faizal Bux:** Supervision, Investigation, Writing – review & editing. **Ameer Khan Patan:** Supervision, Writing – original draft, Conceptualization, Methodology, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this review article.

Data availability

Data will be made available on request.

References

Arrieta, A.B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Herrera, F., 2020. Explainable Artificial Intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Inf. Fusion 58, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012.

Adadi, A., Berrada, M., 2018. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE Access 6, 52138–52160.

Alam, G., Ihsanullah, I., Naushad, M., Sillanpää, M., 2022. Applications of artificial intelligence in water treatment for optimization and automation of adsorption processes: recent advances and prospects. Chem. Eng. J. 427, 130011.

Alex, J., Benedetti, L., Copp, J., Gernaey, K., Jeppsson, U., Nopens, I., Pons, M., Rieger, L., Rosen, C., Steyer, J., 2008. Benchmark simulation model No.1 (BSM1). Report by the IWA Taskgroup on Benchmarking of Control Strategies for WWTPs, pp. 19–20.

Alvi, M., Batstone, D., Mbamba, C.K., Keymer, P., French, T., Ward, A., Cardell-Oliver, R., 2023. Deep learning in wastewater treatment: a critical review. Water Res., 120518

Alvi, M., French, T., Cardell-Oliver, R., Keymer, P., Ward, A., 2022. Cost effective soft sensing for wastewater treatment facilities. IEEE Access 10, 55694–55708.

Akkajit, P., Sukkuea, A., Thongnonghin, B., 2023. Comparative analysis of five convolutional neural networks and transfer learning classification approach for microplastics in wastewater treatment plants. Ecol. Inf., 102328

Aponte-Rengifo, O., Francisco, M., Vilanova, R., Vega, P., Revollar, S., 2023. Intelligent control of wastewater treatment plants based on model-free deep reinforcement learning. Proceso 11 (8), 2269.

Angelov, P.P., Soares, E.A., Jiang, R., Arnold, N.I., Atkinson, P.M., 2021. Explainable artificial intelligence: an analytical review. Wiley Interd. Rev.: Data Min. Knowl. Discov. 11 (5), 1424. https://doi.org/10.1002/widm.1424.

Ba-Alawi, A.H., Al-masni, M.A., Yoo, C., 2023a. Simultaneous sensor fault diagnosis and reconstruction for intelligent monitoring in wastewater treatment plants: an explainable deep multi-task learning model. J. Water Proc. Eng. 55, 104119.

Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K.R., Samek, W., 2015. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. PLoS One 10, e0130140.

Balla, K.M., Bendtsen, J.D., Schou, C., Kallesøe, C.S., Ocampo-Martinez, C., 2022. A learning-based approach towards the data-driven predictive control of combined wastewater networks–An experimental study. Water Res. 221, 118782.

Belle, V., Papantonis, I., 2021a. Principles and practice of explainable machine learning. Front. in big Data 39. https://doi.org/10.3389/fdata.2021.688969.

Ba-Alawi, A.H., Nam, K., Heo, S., Woo, T., Aamer, H., Yoo, C., 2023b. Explainable multisensor fusion-based automatic reconciliation and imputation of faulty and missing data in membrane bioreactor plants for fouling alleviation and energy saving. Chem. Eng. J. 452, 139220.

Ba-Alawi, A.H., Heo, S., Aamer, H., Chang, R., Woo, T., Kim, M., Yoo, C., 2023c. Development of transparent high-frequency soft sensor of total nitrogen and total phosphorus concentrations in rivers using stacked convolutional auto-encoder and explainable AI. J. Water Proc. Eng. 53, 103661.

Bourahla, M.Z., Bourahla, M., 2022. Sewer systems control using internet of things and eXplainable artificial intelligence. Arti. Intell. Doct. Symp 207–220.

Belle, V., Papantonis, I., 2021b. Principles and practice of explainable machine learning. Front. in Big Data, 688969. https://doi.org/10.3389/fdata.2021.688969.

Croll, H.C., Ikuma, K., Ong, S.K., Sarkar, S., 2023. Reinforcement learning applied to wastewater treatment process control optimization: approaches, challenges, and path forward. Crit. Rev. Environ. Sci. Technol. 1–20.

Chen, H., Lundberg, S., Lee, S.I., 2019. Explaining Models by Propagating Shapley Values of Local Components arXiv preprint arXiv: 1911.11888.

Chattopadhay, A., Sarkar, A., Howlader, P., Balasubramanian, V.N., 2018. Grad-CAM++: generalized gradient-based visual explanations for deep convolutional networks. In: 2018 IEEE Winter Conf. On Appli. of Comp. Vis, WACV.

Cheng, T., Harrou, F., Kadri, F., Sun, Y., Leiknes, T., 2020. Forecasting of wastewater treatment plant key features using deep learning-based models: a case study. IEEE Access 8, 184475–184485.

Chang, P., Zhang, S., Wang, Z., 2023. Soft sensor of the key effluent index in the municipal wastewater treatment process based on transformer. IEEE Trans. Ind. Inf. 20 (3), 4021–4028.

Ching, P.M., So, R.H., Morck, T., 2021. Advances in soft sensors for wastewater treatment plants: a systematic review. J. Water Proc. Eng. 44, 102367.

Burkart, N., Huber, M.F., 2020. A survey on the explainability of supervised machine learning. ArXiv, preprint arXiv:2011.07876 70, 245–317.

Castillo, A., Cheali, P., Gómez, V., Comas, J., Poch, M., Sin, G., 2016. An integrated knowledge-based and optimization tool for the sustainable selection of wastewater treatment process concepts. Environ. Model. Software 84, 177–192.

Confalonieri, R., Coba, L., Wagner, B., Besold, T.R., 2021. A historical perspective of explainable Artificial Intelligence. Wiley Interd. Rev: Data Min. Knowl. Discov. 11 (1), 1391.

D'Alterio, P., Garibaldi, J.M., John, R.I., 2020. Constrained interval type-2 fuzzy classification systems for explainable AI (XAI). In: 2020 IEEE Inter. Conf. On Fuzzy Sys. FUZZ-IEEE, pp. 1–8.

Dwivedi, R., Dave, D., Naik, H., Singhal, S., Omer, R., Patel, P., Ranjan, R., 2023. Explainable AI (XAI): core ideas, techniques, and solutions. ACM Comput. Surv. 55 (9), 1–33.

Davagdorj, K., Li, M., Ryu, K.H., 2021. Local interpretable model-agnostic explanations of predictive models for hypertension. Advances in Intelligent Information Hiding and Multimedia Signal Processing: Proc. of the 16th Inter. Conf. on IIHMSP 5–7, 426–433, 2.

Doshi-Velez, F., Kim, B., 2017. Towards a Rigorous Science of Interpretable Machine Learning. Available at: 10.48550/arXiv.1702.08608.

Dubey, R., Bajpai, J., Bajpai, A.K., 2015. Green synthesis of graphene sand composite (GSC) as novel adsorbent for efficient removal of Cr (VI) ions from aqueous solution. J. Water Proc. Eng. 5, 83–94.

Dupuit, E., Pouet, M.F., Thomas, O., Bourgois, J., 2007. Decision support methodology using rule-based reasoning coupled to non-parametric measurement for industrial wastewater network management. Environ. Model. Software 22 (8), 1153–1163.

Duarte, M.S., Martins, G., Oliveira, P., Fernandes, B., Ferreira, E.C., Alves, M.M., Novais, P., 2023. A review of computational modeling in wastewater treatment processes. ACS ES&T Water 4 (3), 784–804.

Dieber, J., Kirrane, S., 2020. Why Model Why? Assessing the Strengths and Limitations of Lime arXiv preprint, arXiv:2012.00093.

Ding, W., Jing, X., Yan, Z., Yang, L.T., 2019. A survey on data fusion in internet of things: towards secure and privacy-preserving fusion. Infor. Fusion 51, 129–144. https://doi.org/10.1016/j.inffus.2018.12.001.

El-Rawy, M., Abd-Ellah, M.K., Fathi, H., Ahmed, A.K.A., 2021. Forecasting effluent and performance of wastewater treatment plant using different machine learning techniques. J. Water Proc. Eng. 44, 102380.

Farhi, N., Kohen, E., Mamane, H., Shavitt, Y., 2021. Prediction of wastewater treatment quality using LSTM neural network. Environ. Technol. Innov. 23, 101632.

Filipe, J., Bessa, R.J., Reis, M., Alves, R., Póvoa, P., 2019. Data-driven predictive energy optimization in a wastewater pumping station. Appl. Energy 252, 113423.

Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., Yang, G.Z., 2019. XAI—explainable artificial intelligence. Sci. Robot. 4 (37), 7120.

Gupta, R., Zhang, L., Hou, J., Zhang, Z., Liu, H., You, S., Li, W., 2022. Review of explainable machine learning for anaerobic digestion. Bioresour. Technol. 128468.

Gireesh, E.D., Skinner, H., Seo, J., Ching, P., Hyeong, L.K., Baumgartner, J., Gurupur, V., 2023. Deep neural networks and gradient-weighted class activation mapping to classify and analyze EEG. Intell. Decis. Technol. 17 (1), 43–53.

Gilpin, L.H., Bau, D., Yuan, B.Z., Bajwa, A., Specter, M., Kagal, L., 2018. Explaining explanations: an overview of interpretability of machine learning. In: 2018 IEEE 5th Inter. Conf. On Data Sci and Adv. Analy. IEEE, pp. 80–89.

Guo, H., Jeong, K., Lim, J., Jo, J., Kim, Y.M., Park, J.P., Cho, K.H., 2015. Prediction of effluent concentration in a wastewater treatment plant using machine learning models. J. Environ. Sci. 32, 90–101.

Han, H., Zhang, L., Liu, H.X., Yang, C., Qiao, J., 2019. Intelligent optimal control system with flexible objective functions and its applications in wastewater treatment process. IEEE Tra. on Sys. Man. and Cyber. Sys 51 (6), 3464–3476.

Hasenstab, K.A., Huynh, J., Masoudi, S., Cunha, G.M., Pazzani, M., Hsiao, A., 2023. Feature interpretation using generative adversarial networks (figan): a framework for visualizing a CNN's learned features. IEEE Access 11, 5144–5160. https://doi.org/10.1109/ACCESS.2023.3236575.

Hernández-del-Olmo, F., Gaudioso, E., Duro, N., Dormido, R., Gorrotxategi, M., 2023. Advanced control by reinforcement learning for wastewater treatment plants: a comparison with traditional approaches. Appl. Sci. 13 (8), 4752.

Irani, Z., Kamal, M.M., 2014. Intelligent systems research in the construction industry. Expert Syst. Appl. 41 (4), 934–950. https://doi.org/10.1016/j.eswa.2013.06.061.

Ismail, W., Niknejad, N., Bahari, M., Hendradi, R., Zaizi, N.J.M., Zulkifli, M.Z., 2021. Water treatment and artificial intelligence techniques: systematic literature review research. Environ. Sci. Pollut. Res. 1–19.

Jawad, J., Hawari, Zaidi, S.J., 2021. Artificial neural network modeling of wastewater treatment and desalination using membrane processes: a review. Chem. Eng. J. 419, 129540.

Jiang, Y., Li, C., Sun, L., Guo, D., Zhang, Y., Wang, W., 2021. A deep learning algorithm for multi-source data fusion to predict water quality of urban sewer networks. J. Clean. Prod. 318, 128533.

Karamichailidou, D., Alexandridis, A., Anagnostopoulos, G., Syriopoulos, G., Sekkas, O., 2022. Modeling biogas production from anaerobic wastewater treatment plants using radial basis function networks and differential evolution. Comput. Chem. Eng. 157, 107629.

Karthikeyan, M., Vijayachitra, S., Pyingkodi, M., Arunkumar, S., Kaviya, K.N., Madhumitha, R., 2022. Study on water quality and wastewater treatment using IOT. In: 6th Inter. Conf. On Comp. Method. and Commu. (ICCMC), pp. 431–438.

Khalil, M., AlSayed, A., Liu, Y., Vanrolleghem, P.A., 2023. Machine learning for modeling N2O emissions from wastewater treatment plants: aligning model performance, complexity, and interpretability. Water Res. 245, 120667.

Khurshid, A., Pani, A.K., 2023. Machine learning approaches for data-driven process monitoring of biological wastewater treatment plant: a review of research works on benchmark simulation model no. 1 (bsm1). Environ. Monit. Assess. 195 (8), 916.

Kumar, S.S.P., Tulsyan, A., Gopaluni, B., Loewen, P., 2018. A deep learning architecture for predictive control. IFAC-PapersOnLine 51, 512–517.

Langeveld, J., Van Daal, P., Schilperoort, R., Nopens, I., Flameling, T., Weijers, S., 2017. Empirical sewer water quality model for generating influent data for WWTP modelling. Water (The Hague) 9 (7), 491.

Langer, M., Oster, D., Speith, T., Hermanns, H., Kästner, L., Schmidt, E., Baum, K., 2021. What do we want from Explainable Artificial Intelligence (XAI)?–A stakeholder perspective on XAI and a conceptual model guiding interdisciplinary XAI research. Artif. Intell. 296, 103473. https://doi.org/10.1016/j.artint.2021.103473.

Liu, Y., Ramin, P., Flores-Alsina, X., Gernaey, K.V., 2023. Transforming data into actionable knowledge for fault detection, diagnosis and prognosis in urban wastewater systems with AI techniques: a mini-review. Process Saf. Environ. Protect. 172, 501–512.

Lipton, Z.C., 2018. The mythos of model interpretability: in machine learning, the concept of interpretability is both important and slippery. Quest 16 (3), 31–57. https://doi.org/10.1145/3236386.3241340.

Lowe, M., Qin, R., Mao, X., 2022. A review on machine learning, artificial intelligence, and smart technology in water treatment and monitoring. Water 14 (9), 1384.

Love, P.E., Fang, W., Matthews, J., Porter, S., Luo, H., Ding, L., 2023. Explainable artificial intelligence (XAI): precepts, models, and opportunities for research in construction. Adv. Eng. Inf. 57, 102024.

Ly, Q.V., Truong, V.H., Ji, B., Nguyen, X.C., Cho, K.H., Ngo, H.H., Zhang, Z., 2022. Exploring potential machine learning application based on big data for prediction of wastewater quality from different full-scale wastewater treatment plants. Sci. of the Tot. Environ. Times 832, 154930.

Li, X., Gong, G., 2019. Predictive control of slurry pressure balance in shield tunneling using diagonal recurrent neural network and evolved particle swarm optimization. Autom. ConStruct. 107, 102928.

Li, K., Duan, H., Liu, L., Qiu, R., van den Akker, B., Ni, B.J., Ye, L., 2022. An integrated first principal and deep learning approach for modeling nitrous oxide emissions from wastewater treatment plants. Environ. Sci. Technol. 56 (4), 2816–2826.

Machlev, R., Heistrene, L., Perl, M., Levy, K.Y., Belikov, J., Mannor, S., Levron, Y., 2022. Explainable Artificial Intelligence (XAI) techniques for energy and power systems: review, challenges and opportunities. Ener. and AI 9, 100169.

Mannina, G., Ekama, G., Caniani, D., Cosenza, A., Esposito, G., Gori, R., GarridoBaserba, M., Rosso, D., Olsson, G., 2016. Greenhouse gases from wastewater treatment–A review of modelling tools. Sci. Total Environ. 551, 254–270.

Meirlaen, J., Huyghebaert, B., Sforzi, F., Benedetti, L., Vanrolleghem, P., 2001. Fast, simultaneous simulation of the integrated urban wastewater system using mechanistic surrogate models. Water Sci. Technol. 43 (7), 301–309.

Mohammadi, E., Stokholm-Bjerregaard, M., Hansen, A.A., Nielsen, P.H., Ortiz-Arroyo, D., Durdevic, P., 2024. Deep learning based simulators for the phosphorus removal process control in wastewater treatment via deep reinforcement learning algorithms. Eng. Appl. Artif. Intell. 133, 107992.

Monje, V., Owsianiak, M., Junicke, H., Kjellberg, K., Gernaey, K.V., Flores-Alsina, X., 2022. Economic, technical, and environmental evaluation of retrofitting scenarios in a full-scale industrial wastewater treatment system. Water Res. 223, 118997.

Montavon, G., Binder, A., Lapuschkin, S., Samek, W., Müller, K.R., 2019. Layer-wise relevance propagation: an overview. Explainable AI: inter. Expla. and Visual. Deep Lear 193–209.

Mougen, C., Kanellos, G., Gottron, T., 2021. Desiderata for Explainable AI in Statistical Production Systems of the European Central Bank. https://doi.org/10.48550/arXiv.2107.08045. Available at:

Murray, B.J., 2021. Explainable Data Fusion. Doctoral Thesis, May, University of Missouri, MI. Available at: https://mospace.umsystem.edu/xmlui/handle/10355/85805.

Oh, S., Kim, Y., 2021. Machine learning application reveal dynamic interaction of polyphosphate-accumulating organism in full-scale wastewater treatment plant. J. of Wat. Process Eng. 44, 102417.

Park, J., Lee, W.H., Kim, K.T., Park, C.Y., Lee, S., Heo, T.Y., 2022a. Interpretation of ensemble learning to predict water quality using explainable artificial intelligence. Sci. Total Environ. 832, 155070.

Park, J., Ahn, J., Kim, J., Yoon, Y., Park, J., 2022b. Prediction and interpretation of water quality recovery after a disturbance in a water treatment system using artificial intelligence. Water (The Hague) 14 (15), 2423.

Páez, A., 2019. The pragmatic turn in explainable artificial intelligence (XAI). Minds Mach. 29 (3), 441–459.

Parsa, A.B., Movahedi, A., Taghipour, H., Derrible, S., Mohammadian, A.K., 2020. Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis. Accid. Anal. Prev. 136, 105405.

Poorasgari, E., Örmeci, B., 2022. Development of non-linear empirical models to estimate the abundance of carbapenem resistance genes during anaerobic digestion of wastewater sludge at mesophilic and thermophilic temperatures. Chem. Eng. J. 450, 138290.

Phillips, P.J., Hahn, C.A., Fontana, P.C., Broniatowski, D.A., Przybocki, M.A., 2020. Four Principles of Explainable Artificial Intelligence.

Rajaee, T., Ebrahimi, H., Nourani, V., 2019. A review of the artificial intelligence methods in groundwater level modeling. J. Hydrobiol. 572, 336–351.

Ramin, E., Flores-Alsina, X., Gaszynski, C., Harding, T., Ikumi, D., Brouckaert, C., Gernaey, K.V., 2022. Plant-wide assessment of alternative activated sludge configurations for biological nutrient removal under uncertain influent characteristics. Sci. Total Environ. 822, 153678.

Raduly, B., Capodaglio, A.G., Vaccari, D.A., 2004. Simplification of wastewater treatment plant models using empirical modelling techniques. Young Res 51.

Rudin, C., 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat. Mach. Intell. 1, 206–215.

Rini, V.S., Berghout, E., 1999. The Goal/Question/Metric Method: A Practical Guide for Quality Improvement of Software Development. McGraw-Hill.

Sagan, V., Peterson, K.T., Maimaitijiang, M., Sidike, P., Sloan, J., Greeling, B.A., Adams, C., 2020. Monitoring inland water quality using remote sensing: potential and limitations of spectral indices, bio-optical simulations, machine learning, and cloud computing. Earth Sci. Rev. 205, 103187.

Safeer, S., Pandey, R.P., Rehman, B., Safdar, T., Ahmad, I., Hasan, S.W., Ullah, A., 2022. A review of artificial intelligence in water purification and wastewater treatment: recent advancements. J. Water Proc. Eng. 49, 102974.

Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., 2017. Grad-CAM: visual explanations from deep networks via gradient-based localization. In: Proc. Of the IEEE Inter. Conf. on Comp. Vis, pp. 618–626.

Shojaeimehr, T., Rahimpour, F., Khadivi, M.A., Sadeghi, M., 2014. A modeling study by response surface methodology (RSM) and artificial neural network (ANN) on Cu2+ adsorption optimization using light expended clay aggregate (LECA). J. Ind. Eng. Chem. (Seoul, Repub. Korea) 20 (3), 870–880.

Speith, T., 2022. A review of taxonomies of explainable artificial intelligence (XAI) methods. In: Proc. Of the 2022 ACM Conf. on Fair. Accou. and Transp, pp. 2239–2250. https://doi.org/10.1145/3531146.3534639.

Singh, N.K., Yadav, M., Singh, V., Padhiyar, H., Kumar, V., Bhatia, S.K., Show, P.L., 2022. Artificial intelligence and machine learning-based monitoring and design of biological wastewater treatment systems. Bio Technol., 128486

Shao, S., Fu, D., Yang, T., Mu, H., Gao, Q., Zhang, Y., 2023. Analysis of machine learning models for wastewater treatment plant sludge output prediction. Sustainability 15 (18), 13380.

Sheik, A.G., Kumar, A., Patnaik, R., Kumari, S., Bux, F., 2024a. Machine learning-based design and monitoring of algae blooms: recent trends and future perspectives–A short review. Crit. Rev. Environ. Sci. Technol. 1–24.

Sheik, A.G., Kumar, A., Ansari, F.A., Raj, V., Nicolas, P., Patan, A.K., Kumari, S., Bux, F., 2024b. Reinvigorating algal cultivation for biomass production with digital twin technology-a smart sustainable infrastructure. Algal Res., 103779

Sheik, A.G., Tejaswini, E.S.S., Seepana, M.M., Ambati, S.R., 2023. Control of anaerobic-anoxic-aerobic (A2/O) processes in wastewater treatment: a detailed review. Environ. Techn. Rev 12 (1), 420–440.

Sheik, A.G., Malla, M.A., Srungavarapu, C.S., Patan, A.K., Kumari, S., Bux, F., 2024c. Prediction of wastewater quality parameters using adaptive and machine learning models: a South African case study. J. Water Proc. Eng. 67, 106185.

Shen, Y., Zhu, X., Guo, Z., Yu, K., Alfarraj, O., Leung, V.C., Rodrigues, J.J., 2024. A deep learning-based data management scheme for intelligent control of wastewater treatment processes under resource-constrained IoT systems. IEEE Internet Things 11 (15), 25757–25770.

Shyu, H.Y., Castro, C.J., Bair, R.A., Lu, Q., Yeh, D.H., 2023. Development of a soft sensor using machine learning models for predicting the water quality of an onsite wastewater treatment system. ACS Environ. Au 3 (5), 308–318.

Sokol, K., Flach, P., 2022. Explainability is in the beholder's mind: establishing the foundations of explainable artificial intelligence. arXiv preprint arXiv:2112.14466.

Smirnov, A., Levashova, T., 2019. Knowledge fusion patterns, A survey. Inf. Fusion 52, 31–40. https://doi.org/10.1016/j.inffus.2018.11.007.

Torregrossa, D., Leopold, U., Hernández-Sancho, F., Hansen, J., 2018. Machine learning for energy cost modelling in wastewater treatment plants. J. Environ. Manag. 223, 1061–1067.

Tritscher, J., Ring, M., Schlr, D., Hettinger, L., Hotho, A., 2020. Evaluation of post-hoc XAI approaches through synthetic tabular data. In: Intern. Symp. On Method. for Intell. Sys, pp. 422–430.

Utama, C., Karg, B., Meske, C., Lucia, S., 2022. Explainable artificial intelligence for deep learning-based model predictive controllers. In: 2022 26th Inter. Conf. On Sys. The. Cont. and Comp. (ICSTCC), pp. 464–471.

Wang, D., Thunéll, S., Lindberg, U., Jiang, L., Trygg, J., Tysklind, M., 2022a. Towards better process management in wastewater treatment plants: process analytics based on SHAP values for tree-based machine learning methods. J. Environ. Manag. 301, 113941.

Wang, N., Wang, Y., Er, M.J., 2022b. Review on deep learning techniques for marine object recognition: architectures and models. Control Eng. Pract. 118, 104458.

Wang, G., Jia, Q.S., Zhou, M., Bi, J., Qiao, J., Abusorrah, A., 2022c. Artificial neural networks for water quality soft-sensing in wastewater treatment: a review. Artif. Intell. Rev. 55 (1), 565–587.

Wongburi, P., Park, J.K., 2022. Prediction of sludge volume index in a wastewater treatment plant using recurrent neural network. Sustain. Times 14 (10), 6276.

Wang, D., Thunéll, S., Lindberg, U., Jiang, L., Trygg, J., Tysklind, M., Souihi, N., 2021. A machine learning framework to improve effluent quality control in wastewater treatment plants. Sci. Total Environ. 784, 147138.

Watson, D., 2020. Conceptual Challenges for Interpretable Machine Learning. Available at: SSRN 3668444.

Xu, B., Pooi, C.K., Tan, K.M., Huang, S., Shi, X., Ng, H.Y., 2023a. A novel long short-term memory artificial neural network (LSTM)-based soft-sensor to monitor and forecast wastewater treatment performance. J. Water Proc. Eng. 54, 104041.

Xu, Y., Wang, Z., Nairat, S., Zhou, J., He, Z., 2023b. Artificial intelligence-assisted prediction of effluent phosphorus in a full-scale wastewater treatment plant with missing phosphorus input and removal data. ACS ES&T Water 4 (3), 880–889.

Xu, Y., Zeng, X., Bernard, S., He, Z., 2022. Data-driven prediction of neutralizer pH and valve position towards precise control of chemical dosage in a wastewater treatment plant. J. Clean. Prod. 348, 131360.

Yu, T., Yang, S., Bai, Y., Gao, X., Li, C., 2018. Inlet water quality forecasting of wastewater treatment based on kernel principal component analysis and an extreme learning machine. Water (The Hague) 10 (7), 873.

Yang, G., Ye, Q., Xia, J., 2022. Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion: a mini-review, two showcases and beyond. Inf. Fusion 77, 29–52. https://doi.org/10.1016/j.inffus.2021.07.016.

Yang, Q., Cao, W., Meng, W., Si, J., 2021. Reinforcement-learning-based tracking control of waste water treatment process under realistic system conditions and control performance requirements. IEEE Trans. on Sys. Man, and Cyber. Sys 52 (8), 5284–5294.

Zhang, Y., Li, C., Duan, H., Yan, K., Wang, J., Wang, W., 2023a. Deep learning-based data-driven model for detecting time-delay water quality indicators of wastewater treatment plant influent. Chem. Eng. J. 467, 143483.

Zahra, Q., Gul, J., Shah, A.R., Yasir, M., Karim, A.M., 2023a. Antibiotic resistance genes prevalence prediction and interpretation in beaches affected by urban wastewater discharge. One Heal, 100642.

Zahra, Q., Gul, J., Shah, A.R., Yasir, M., Karim, A.M., 2023b. Antibiotic resistance genes prevalence prediction and interpretation in beaches affected by urban wastewater discharge. One Heal, 100642.

Zhang, Y., Li, C., Duan, H., Yan, K., Wang, J., Wang, W., 2023b. Deep learning-based data-driven model for detecting time-delay water quality indicators of wastewater treatment plant influent. Chem. Eng. J. 467, 143483.

Zhou, P., Wang, X., Chai, T., 2022. Multiobjective operation optimization of wastewater treatment process based on reinforcement self-learning and knowledge guidance. IEEE Trans. Cybern. 53 (11), 6896–6909.

Zhang, W., Xie, J., Liu, X., Zhang, L., Geng, P., 2023c. CNN-BiLSTM sewage treatment dissolved oxygen concentration prediction model based on attention mechanism. Inter. Conf. on Elect. Inform. Eng. and Data Proc. (EIEDP) 12700, 311–318.

Zaghloul, M.S., Achari, G., 2022. Application of machine learning techniques to model a full-scale wastewater treatment plant with biological nutrient removal. J. Environ. Chem. Eng. 10 (3), 107430.