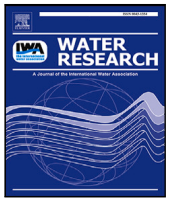
Water Research 245 (2023) 120518

Contents lists available at ScienceDirect

Water Research

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



Review

Deep learning in wastewater treatment: a critical review

Maira Alvi [a,*], Damien Batstone [b], Christian Kazadi Mbamba [b], Philip Keymer [b], Tim French [a], Andrew Ward [b], Jason Dwyer [c], Rachel Cardell-Oliver [a]

[a] Department of Computer Science & Software Engineering, The University of Western Australia, Australia
[b] Australian Centre for Water and Environmental Biotechnology, University of Queensland, Brisbane, Australia
[c] Urban Utilities, Brisbane, Australia

ARTICLE INFO

Keywords:
Deep learning
Review
Wastewater
Artificial Intelligence
Machine learning
Mechanistic modeling

ABSTRACT

Modeling wastewater processes supports tasks such as process prediction, soft sensing, data analysis and computer assisted design of wastewater systems. Wastewater treatment processes are large, complex processes, with multiple controlling mechanisms, a high degree of disturbance variability and non-linear (generally stable) behavior with multiple internal recycle loops. Semi-mechanistic biochemical models currently dominate research and application, with data-driven deep learning models emerging as an alternative and supplementary approach. But these modeling approaches have grown in separate communities of research and practice, and so there is limited appreciation of the strengths, weaknesses, contrasts and similarities between the methods. This review addresses that gap by providing a detailed guide to deep learning methods and their application to wastewater process modeling. The review is aimed at wastewater modeling experts who are familiar with established mechanistic modeling approach, and are curious about the opportunities and challenges afforded by deep learning methods. We conclude with a discussion and needs analysis on the value of different ways of modeling wastewater processes and open research problems.

1. Introduction

Modern wastewater treatment plants represent one of the greatest public health advances of the last century (Naik and Stenstrom, 2012). They have a substantial role in preserving ecosystems and safeguarding human health by removing pathogens, contaminants and nutrients prior to environmental discharge. Modern treatment facilities comprise multiple unit operations, relying on many physical, biological, and chemical mechanisms and are subject to substantial external dynamic disturbances, strong non-linearities, widely varying system time constants for different processes, and treatment targets (Burton et al., 2013). Treatment facilities are capital-intensive civil infrastructure that is non-flexible in purpose, highly exposed to localized climatic changes, with costs that are strongly dependent on the selection of effective processes and unit operations (Pinheiro et al., 2018). There is a strong focus in research and practice on machine assisted asset design and optimization (Yuan et al., 2019b). Furthermore, wastewater treatment processes are data rich, have strong research support, and generally an excellent history of transfer from research to engineering practice (Corominas et al., 2018b).

Wastewater treatment processes are large, complex processes, with multiple controlling mechanisms, a high degree of disturbance variability and non-linear, generally stable behavior with multiple internal recycle loops. Hence, machine-based analysis has focused on dynamic analysis, particularly plant-wide analysis rather than unit analysis with the ambition to expand upstream (sewer and catchment), and downstream (receiving environment). Dynamic process analysis will be the focus of this review. Use of semi-mechanistic biochemical models such as the Activated Sludge Model series (ASM) (Henze et al., 2000), and Anaerobic Digestion Model 1 (ADM1) (Batstone et al., 2002) and related commercial models, particularly in design and control system analysis is ubiquitous. However, these models focus on biochemical and chemical conversion only, with the hydraulics generally being mixed tanks in series, with some extensions for plug-flow and biofilm systems. Models for units dominated by physical and chemical processes are commonly empirical (Jeppsson et al., 2013), due to the complexity of underlying phenomena i.e., non-ideal separation and reaction. This is being challenged, with the semi-empirical Takacs secondary clarifier model being an outstanding example (Takács and Nolasco, 1991). Multi-phase and aquatic chemistry is being increasingly represented on a fundamental level, including the incorporation of advanced non-ideality (Batstone et al., 2012). However, this is resulting in an extremely high level of model complexity, substantial solver issues involving multi-unit non-linear, stiff differential–algebraic systems, often involving hundreds of state variables (Flores-Alsina et al., 2015).

* Corresponding author.
E-mail address: maira.alvi@research.uwa.edu.au (M. Alvi).

https://doi.org/10.1016/j.watres.2023.120518
Received 21 February 2023; Received in revised form 19 August 2023; Accepted 22 August 2023
Available online 25 August 2023
0043-1354/© 2023 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Nomenclature

ADM	Anaerobic Digestion Model
AETL	Auto-encoder Transfer Learning
AE	Auto-encoder
AI	Artificial Intelligence
ANN	Artificial Neural Network
ARIMA	Auto-regressive Integrated Moving Average
ASM	Activated Sludge Model
AS	Activated Sludge
Bi	Bidirectional
BM	Boltzmann Machine
BOD	Biochemical Oxygen Demand
BSM1	Benchmark Simulation Model 1
BSM2	Benchmark Simulation Model 2
CFD	Computational Fluid Dynamics
CNN	Convolutional Neural Network
COD	Chemical Oxygen Demand
CUSUM	Shewhart Cumulative Sum
DAE	Differential Algebraic Equation
DBN	Deep Belief Network
DL	Deep Learning
DO	Dissolved Oxygen
DRL	Deep Reinforcement Learning
D	Dimension
EWMA	Exponentially Weighted Moving Average
FFNN	Feed Forward Neural Network
GAN	Generative Adversarial Network
GDBN	Growing Deep Belief Network
GHG	Greenhouse Gas
GPU	Graphical Processing Unit
GRNN	Gated Recurrent Neural Network
GRU	Gated Recurrent Unit
I/O	Input/Output
IWA	International Water Association
LSTM	Long Short Term Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MLP	Multi-Layer Perceptron
MLSS	Mixed Liquor Suspended Solids
ML	Machine Learning
MPC	Model Predictive Control
N_2O	Nitrous Oxide
NH_4-N	Ammonium
NN	Neural Network
NO_3–N	Nitrate
ODE	Ordinary Differential Equation
PCA	Principal Component Analysis
PI	Proportional Integral
PLS	Principal Least Square
PO_4	Phosphate
R^2	Coefficient of Determination
RBM	Restricted Boltzmann Machine
ReLU	Rectified Linear Unit
ResNet	Residual Neural Network
RL	Reinforcement Learning
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
SDAE	Stacked Denoising Auto-encoder
SHAP	SHapley Additive exPlanation
SMOTE	Synthetic Minority Over-Sampling Technique
SPC	Statistical Process Control
SSM	State Space Model
SS	Suspended Solids
SVM	Support Vector Machine
TL	Transfer Learning
TN	Total Nitrogen
TP	Total Phosphorous
TSS	Total Suspended Solids
VAE	Variational Auto-encoder
WWTP	Wastewater Treatment Plant

The biochemical and chemical focus of mechanistic modeling in the wastewater sector means that they are generally excellent when simulating processes dominated by these mechanisms, but weaker when other phenomena (e.g., hydraulics) dominate (Samstag et al., 2016). Fundamentally, there are behaviors in wastewater treatment systems that cannot be represented with reasonable computational resources using mechanistic models. This issue expands greatly if the scope of the model expands upstream (to the catchment), or downstream (to the receiving environment) (Jeppsson et al., 2013). Indeed, parameter and system analysis often use simplified models (Ocampo-Martinez, 2010; Olsson et al., 2005) that allow the several thousand repeated simulations required.

Mechanistic models are also challenged by the rapid emergence of new processes focusing on efficiency and resource recovery utilizing different fundamental biochemistry such as phototrophic (algae and purple bacteria), chemotrophic, and chemical processes. While modeling is used to analyze these processes (Puyol et al., 2017; Wágner et al., 2016), selection of an appropriate model requires domain expertise, and is a balance of model and operator capability, complexity, suitability to the application, model maturity, identifiability, and data availability. This is beyond the scope of the paper, and is discussed further in Rieger et al. (2012).

Data-driven models, in contrast to mechanistic models, seldom assume structural knowledge of the underlying physical process of a system. These models search for empirical relationships between process state variables (Solomatine and Ostfeld, 2008) and have been widely used in the wastewater industry for predictive control, fault detection, variable prediction, and process management (Newhart et al., 2019). The basis of these models are data relationships rather than underlying knowledge of the process. These form a broad range of models including regressive (Alwan, 1992), linear state-space models (SSM) (Durbin and Koopman, 2012), dimensional reduction (or simplification techniques) such as principal component analysis (PCA) (Rosén and Lennox, 2001), and most recently, machine learning (ML) and deep learning (DL) (Dargan et al., 2020).

Classic empirical models are normally effective at representation and analysis of data, but fairly poor at prediction (including model based control), particularly where the prediction domain lies outside the training domain, or where the I/O relationships are non-linear (including most wastewater processes) (Puyol et al., 2017; Wágner et al., 2016). ML models have similar prediction difficulties when expanding beyond their training domain, however emerging ML techniques offer the potential for better prediction capabilities, and effective representation of non-linear relationships (Corominas et al., 2018a). ML offers a remedy to tackle a range of problems that do not assume prior knowledge including statistical stationarity, and linearity. Instead, it directly explores the input features to learn a non-parametric model that can theoretically express any linear or non-linear process (Meir,

2000). ML is a subset of Artificial Intelligence (AI) that builds mathematical models from features without being explicitly programmed and addresses the broader problem of programming computers to perform tasks that normally require human intelligence. The field of machine learning is complex and somewhat inaccessible to non-(ML)-experts.

Previous reviews of data-driven models for wastewater processes considered the translation from development to practice (Corominas et al., 2018a) and the use of data-driven models in general focusing on linear I/O (Newhart et al., 2019; Corominas et al., 2018a). Recent reviews of deep learning consider the application of DL to: urban water supply and sewage infrastructure (Fu et al., 2022) and AI applied to drinking water process systems (Alam et al., 2022) and membrane based treatment systems (Jawad et al., 2021). This review of deep learning focuses on a different application area: wastewater treatment design, optimization, and control, where as discussed above, the default approach is mechanistic modeling. This review surveys applications of deep learning for wastewater process modeling, and offers insights into how deep learning complements existing approaches, and where and how deep learning offers new opportunities for modeling complex wastewater processes including hybrid approaches that combine both methods.

2. Comparison of modeling approaches

2.1. Mechanistic models

What is mechanistic modeling? Mechanistic process modeling utilizes prior knowledge of the system to build a mathematical representation of the system. It imposes basic dynamic mass balances in multiple units, generally with a dilute assumption (water as non-reactive carrier), and the majority of units represented as single, or multiple 0-D completely mixed reactors, though 1-D models with advection/diffusion/sedimentation may be included where appropriate (e.g., biofilms, sedimentation basins, contact basins). The principles of 0-D conservation models are highly developed, and for each state variable, consist of accumulation, advection, diffusion (if in contact with a mass transfer boundary), and source terms (Hangos and Cameron, 2001). The general mass balance equation is applicable across a wide range of industries and physics problems.

Models are focused on biochemistry. Since the conservation equation is based on well established principles, the majority of the intellectual effort specific to wastewater is development of the source (reaction) terms. This has focused on biochemical conversion processes, with non-chemical reactions generally prioritized to a lesser extent and focused on specific issues such as liquid–gas transfer. A broader inorganic chemistry framework is now being established (Batstone et al., 2012). Standardized mechanistic models such as the ASM series (Henze et al., 2000) and ADM1 (Batstone et al., 2002) focused on the presentation of a biochemical conversion framework involving multiple functional groups to represent key units effectively. These models have been modified and implemented in various forms in commercial software and are highly mature. Over the last 20 years, a key focus has been the expansion and synthesis of unit models and inclusion of ancillary models to a plant wide mechanistic framework, assisted by initiatives such as the Benchmark Simulation Model 2 (BSM2) (Jeppsson et al., 2007).

The resulting non-linear differential equations are numerically solved to emulate the non-linear behavior of wastewater treatment plants (WWTPs). In many cases, implicit algebraic equations are also present, requiring a differential–algebraic equation (DAE) solution approach (Solon et al., 2017). The complexity of the biochemical models and the process itself can mean that a plant wide model consists of thousands of differential equations, and model reduction is common for specific tasks. However, improved computing power and targeted solver approaches means that computing power is no longer a critical barrier for mechanistic process modeling. As mechanistic models are conservation based and represent the underlying mechanisms (though they use a simplified, semi-empirical approach to individual processes such as biochemical conversion), they have excellent predictive capabilities when the dominant mechanisms are included.

Issues with mechanistic models. The main barrier, particularly when modeling these complex systems, is the requirement of a high degree of application specific information (e.g., wastewater characterization, parameter identification, input and disturbance acquisition) and expert domain knowledge (Ocampo-Martinez, 2010; Olsson et al., 2005). There is an intrinsic barrier to emergent processes which utilize a fundamentally different biology (e.g., photosynthetic, fermentative, etc.), as they do not benefit from 50 years of specific model development. Finally, mechanistic models are limited to the mechanisms which have been included in the underlying model and parameterize elements which may be time dependent (e.g., biochemical parameters) either due to matrix effects (e.g., temperature, inhibitors, nature of the substrate), or unknown mechanisms. The application of data-driven models is motivated where the system is too large, too different, or too noisy to be effectively represented by a mechanistic approach.

2.2. Empirical models

Empirical models are data-driven models that establish a relationship between inputs and outputs without an expert-provided specification of the system behavior. Domain expertise is not required to apply empirical models, but achieving a useful outcome generally requires expertise to identify applications or erroneous outcomes. All data-driven models are types of machine learning, since they represent a system using generalizable rules, without expert domain knowledge of the system. Data-driven models can be parametric (i.e., the relationship between inputs and outputs is numerically associated by parameters), or non-parametric (i.e., the relationship is associated mainly through model structure). The ubiquitous deployment of cost-effective sensors has generated large volumes of WWTP data. However, extracting useful information from this data is still a challenge. Traditional approaches used to transform and enrich data, rely on manual interpretation which is impractical for vast data sets (Fayyad et al., 1996). While not I/O empirical models, large data processing techniques such as Statistical Process Control (SPC) charts, i.e. Shewhart, cumulative sum (CUSUM), and exponentially weighted moving average (EWMA) charts (Venkatasubramanian et al., 2003; Gustafsson, 2007) have been applied to these data processing problems. Unfortunately, these methods do not yield a reasonable degree of confidence, due to the inherent non-linearities of WWTPs.

Linear and simple non-linear empirical models. Classical empirical modeling represents the relationship between inputs and outputs in time space or state space, generally using simple linear relationships. Time space models are discrete, while state space models may be continuous or discrete in time. In general, simple ML techniques assume that the time series is stationary which is often not the case for many real-world time series (Le Guen and Thome, 2019).

Time-space models. The simplest time-space models are: regression models: outputs are a linear or simple non-linear function of current inputs; regressive or moving average (MA) regression: outputs are a function of inputs and past inputs (McKenzie, 1984); or auto-regressive moving average (ARMA): outputs are a function of inputs and past inputs and outputs (Alwan, 1992). The function may be linear or non-linear in parameters e.g., where an ARIMA (Reagan, 1984) model is used, but the majority of wastewater applications use ARMA models, and are hence linear in parameters. Statistical analysis and optimized approaches to linear (in parameters) and non-linear problems is provided by the Box–Jenkins methodology (Box et al., 1974). PCA (Rosén

and Lennox, 2001) is commonly used for dimensional reduction in correlated inputs, or principal least squares (PLS) as a way to co-analyze relationships between multiple inputs and outputs. PCA for input reduction is particularly useful, since many concentration inputs are highly correlated in wastewater systems and will be co-represented in a single major principal component. Time dependence is represented by the moving average and auto-regressive elements and any non-linearity in time is represented (generally inadequately) by the parameters in the moving average function i.e., an exponential decay kinetic is represented by 2–3 parameters. Finally, they are also highly exposed to mismatches between model discrete frequency and the fundamental system time constants.

State space models. State space models (Durbin and Koopman, 2012) address many of the issues of time-space models by representing the system as a set of linear ODEs in inputs and outputs. The system states are generally mapped to outputs using translation functions. This means that they can provide excellent representation of a simple, linear time-dependent system, and have enhanced predictive capability. However, wastewater processes are generally non-linear (biochemical equations have an order between 0 and 1), and are non-stationary over longer periods due to multiple processes with different time constants. Due to these issues, state space models have reasonable predictive power only in the short term, and do not extract fundamental longer-term system characteristics since variance is generally dominated by short term behavior.

Issues with empirical models. Due to the high number of potential inputs, and level of correlation between these inputs, feature engineering is required to prepare data for building empirical models. In a process context, feature engineering involves the selection and pre-processing of inputs to identify the minimal and most important inputs. This may involve numerical reduction (e.g., via PCA), or numerical selection (by forward or backwards feature selection). Feature selection requires domain expertise and is a time intensive process, as it is often difficult to describe features that appear obvious to humans.

However, feature engineering governs the performance of empirical models (Bengio et al., 2013). Consequently, the design of a pre-processing pipeline for training an empirical model is the critical path to implementation of a successful model.

2.3. Deep learning models

What is deep learning? Deep learning is a technique for creating a model of a complex non-linear system by learning from examples of the system's inputs and outputs. It extracts essential features or representations directly from the raw input data by organizing its computational processing units, called artificial neurons, into a hierarchy of layers. This organization as layers provides the notion of the computational 'depth' of the network, inspiring the name 'deep learning'. Once *trained* for a particular task with measured data, a deep learning model can be used for generating outputs for the same task from previously unseen input data.

How does DL work? DL is a technique based on Artificial Neural Networks (ANNs), which are loosely inspired by biological processing units of a human brain (Goodfellow et al., 2016). The simplest processing unit of these algorithms is called a neuron or perceptron. Fig. 1 illustrates the basic operation of a single neuron/perceptron. When these units are arranged in layers in a hierarchical manner they form a deep neural network, where deep refers to the number of layers. The layers are connected by activation functions, which essentially switch connections on or off by continuous and discontinuous functions. The weights of the connections between the neurons determine the model response. Training a deep learning network involves adjusting weights across a range of inputs and outputs in order to optimize the neuron weights so as to maximize the accuracy of outputs predicted by the DL model for given inputs as shown in Fig. 2.

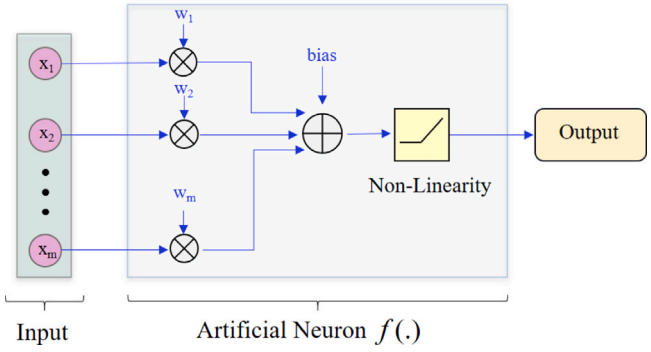


Fig. 1. Illustration of an artificial neuron or perceptron in a standard ANN. A weighted sum is computed via linear combination of inputs features by weights of parameters alongside a scalar bias term. Afterwards, the output is computed by application of non-linear function called an 'activation function' (e.g., rectified linear unit (ReLU)) to the weighted sum.

Unlike mechanistic models, where the function of the model is largely defined by the specific state equations, the function of a neural network is defined by its *architecture* i.e., the specific arrangement of neurons and activation functions. A detailed review of architectures relevant to wastewater processes is provided in this section. A wide range of architectures are available, and these are summarized graphically in Fu et al. (2022) and Van Veen and Leijnen (2016) The Neural Network GNN[1] for reference. Design choices for DL architectures include the type and number of neurons, activation function type, number of layers, and specific inter-connections between neurons. The Multi-Layer Perceptron (MLP) (also called feed forward, or deep feed forward) is a feed forward model where each neuron in each layer is connected to all neurons in the following layer and vice versa. The input layer connects to the inputs while the output layer is the final layer which produces the output. In addition to these layers, a network may contain a range of internal layers called hidden layers (hidden neurons).

The *feed forward* model translates input to output data in a forward direction through activation functions and weights. As noted above, while the structure of the neural network largely defines its function, a number of specific neuron types may be used for specific applications. For a basic neural network, inter-layer connection is *dense*, that is all neurons in a given layer are connected to all neurons in a previous layer by activation functions and weights, but an architecture can be specifically built to be sparse, or specific hidden layer neurons (e.g., convolution or deconvolution) may inherently utilize 2:1 or 1:2 interlayer mapping. Input neurons generally consist of processed input features. The back-fed variation receives information back from a hidden layer or another input cell and is used in Restricted Boltzmann Machines and Deep Belief Networks. Random noise may also be included in input cells, often to train the network to avoid this noise.

Hidden neurons contain the internal system states, computed from inputs and other hidden layers, and are mapped to outputs via activation functions. In recurrent neurons, information from a hidden layer is fed back to that layer as input information. These are used in recurrent neural networks (RNN). Previous time states may be stored (memory recurrent neuron), forming long, short term memory (LSTM) networks and this information may be filtered (gated) in gated recurrent neural networks (GRNN). Hidden layer information may be inter-layer, forward, or backward (bidirectional RNNs). Convolutional neurons apply a convolution operation to reduce dimensionality. Deconvolution may be applied to increase dimensionality (generally considering time series data). Output neurons generate the desired output from a number of

[1] https://www.asimovinstitute.org/neural-network-GNN

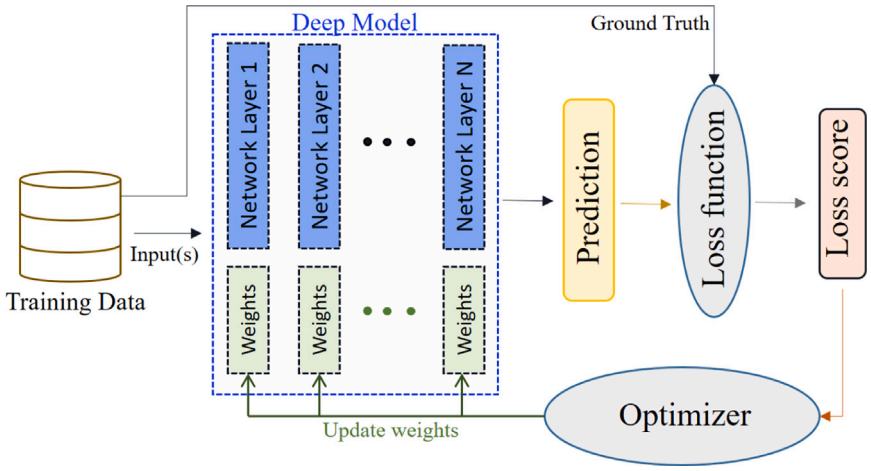


Fig. 2. Illustration of a training loop in a deep neural network. Initially, the model parameters (weights) are initialized to random values. Then the weights are adjusted iteratively over training samples by an optimizer with the aim of minimizing the loss score of the difference between the predicted and actual outputs.

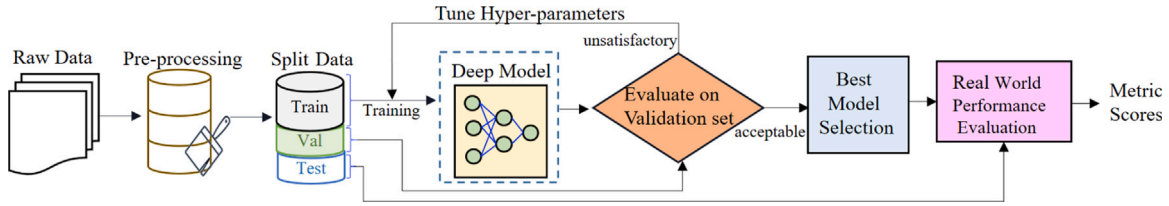


Fig. 3. Illustration of a workflow to build a Deep learning model. The raw data is pre-processed and afterwards split to train, validation, and test sets. The training data is used to train a model while validation data evaluate model performance and aid to tune hyper-parameters. Finally, the best model is evaluated over unseen test data to estimate real-world performance.

hidden layers, which normally includes dimensional reduction. An alternative is an auto-encoder output layer, which allows for an increase in dimensionality via a decoding function (generally with the aim of reflecting a processed input layer).

A *deep learning network* automatically learns appropriate *features* for a problem as well as learning the desired output. This property distinguishes DL from other ML techniques which rely on expert-informed feature engineering of the data to achieve accurate learning. A deep network learns features and outputs by adjusting its parameters using a *cost function* to assess how well or poorly a DL algorithm performs with respect to the actual training output and the predicted results, through a technique called *back-propagation*. In this technique, the model first processes an input to produce an output, which is then compared with the desired output and an error score is computed. The error score information is then passed back for the next round of training which updates neuron weights in an effort to reduce the error score (see Fig. 2). Searching for a minimal error score can be achieved with the aid of *gradient descent* optimization strategies (Rumelhart et al., 1986). Computation of gradients for every training sample often demands large computation time, therefore for practical purposes, a variant, *Stochastic Gradient Descent*, that only utilizes a subset of samples for gradient computation, is used. Generally, multiple traversals of training data, known as epochs, lead to acceptable training convergence to an accurate deep learning model. However, *vanishing gradient* and *exploding gradient* problems may occur when gradient descent fails to converge.

2.3.1. Deep learning workflow

A typical workflow to build a deep model with a reasonable estimate of performance over unseen data consists of several steps (Chollet, 2021). It includes data pre-processing, data splitting, training, and model evaluation over the unseen data set as illustrated in Fig. 3. The four key steps are discussed below in detail.

(1) Data pre-processing. Raw data is seldom suitable to be directly consumed by the training loop of a deep model. Therefore, it is often necessary to pre-process raw data. This may demand review and interpretation from field domain experts. At a simplified level, it serves two primary goals i.e., data cleaning and feature scaling. During data cleaning, the treatment of outliers is an important decision. They are often filtered via data imputation. However, filtering can potentially impact the reliability of data. Feature scaling refers to the normalization / standardization operations that encourage a similar spread of different feature distributions. It is known that features with relatively different scales can pose a challenge to the optimization algorithms and may result in slower convergence (Bengio, 2012).

(2) Data splitting. Given a clean data set, the next steps include partitioning the data to train, validation, and test splits. These data splits enable us to train and find the model hyper-parameters and estimate the model's generalization performance. This division requires careful treatment such that each split represents the same underlying data distribution. Therefore, it is a relatively complicated task to split a time series data set when the data is scarce. A naive split in time (also known as fixed partitioning) may result in very different distributions of the train, test, and validation split that can potentially result in sub-optimal performance (Alvi et al., 2022b).

(3) Training loop. This is the core part of the deep learning pipeline. It is broadly comprised of two parts, i.e., network design and model parameters. Network design refers to different design choices of neural network architectures - hyper-parameters. This includes the choice of neuron type, number of layers and number of neurons in each layer, learning rate, and activation functions. The hyper-parameters can be either manually set via trial and error or searched via Bayesian Optimization and evolutionary algorithms. Once the network design is completed, the model weights are optimized to perform the desired task accurately. The process of iteratively changing model weights is called

training. It is guided by the quality of model performance, estimated using different evaluation metrics. For simplicity, we deliberately defer the discussion of evaluation metrics to the next step. The model weights are learned directly from the training data with the aid of back-propagation-based optimization techniques. During the training phase, a validation set is also utilized to avoid over-fitting the model to the training set, and further, it also assists in the selection of hyper-parameters. Once the best-performing model is identified based on the train and validation splits, it is ready for performance evaluation over an unseen test data set.

(4) Model evaluation. Evaluation of a model refers to quantification of model performance. In developing a deep learning model, it is iteratively performed during all phases, i.e., training, validation, and testing. During training, the model is evaluated over train and validation splits to make updates in the hyper-parameters of the model. Evaluation over a test set yields test error or generalization error. This estimates the expected performance of the model over unseen data set i.e., generalization performance. It is noteworthy that the generalizability of a model is evaluated on the assumption that training set and the test set are drawn from the same probability distribution. Describing the desired behavior mathematically or numerically is a challenging process, and often, simplified and intuitive definitions are used. In the case of regression, root means square error (RMSE), mean absolute error (MAE), and the coefficient of determination (R^2) may be selected depending on model goal. This contrasts with mechanistic modeling, where R^2 is commonly most suitable as objective function, due to the desire to conduct inferential statistics on model and parameter quality (Dochain and Vanrolleghem, 2001). Similarly, for classification models, accuracy, recall, precision, and F1-score are used. Practically, instead of relying on one single metric, a combination is used to identify the best model.

2.3.2. Machine learning methods

ML problems where both input and expected output data is available for learning are called *supervised* learning problems. In contrast, in *unsupervised* learning the network learns patterns and structures in the data but there is no ground truth of the 'correct' output.

Supervised learning. In supervised learning, the computational model is fed with two sets of data: an input vector, and the observed output, indicating the ground-truth (observed) data. Examples of supervised learning in wastewater treatment include predicting outlet ammonia based on inputs like flow, ammonia, and COD. Another example is classifying wastewater influent into domestic, municipal, or industrial categories based on various quality variables. The learning process involves solving an optimization problem using gradient descent algorithms.

Unsupervised learning & semi-supervised learning. In contrast to supervised techniques, unsupervised learning techniques do not require training labels. Instead, they endeavor to discover latent patterns or subgroups in the input data. Thus, they are naturally suitable for exploratory data analysis (Han et al., 2011). Some common applications include clustering, association rules, dimensionality reduction and anomaly detection (Pang et al., 2021). Similarly, in the *semi-supervised learning*, the algorithm learns from data where only some samples are labeled.

Deep reinforcement learning (DRL). In *reinforcement learning* (RL), a reward/penalty approach is used to learn the best strategy (policy) for taking a certain action given a particular state. It is often formulated as a Markov decision process whereby an agent perceives the information in a state space, acts on the environment that results in a new state (transition) and receives a scalar reward as a feedback on that action. This theory's combination with deep learning architectures, as approximators for the value or policy function, gives rise to *deep reinforcement learning* (Seo et al., 2021). Deep RL approaches are generally categorized as model-free and model-based methods. In model-free, an agent learns a policy through direct interactions with the environment and makes decisions by trial and error. While, in model-based deep RL, an internal model of the environment is utilized that allows an agent to plan a trajectory of actions, leading to the desired goals. For WWTPs, model-based methods can be potentially utilized for robust control purposes where the agent represents the model predictive control (MPC) (Ocampo-Martinez, 2010).

Model and hyper-parameters. In designing a machine learning solution, there are many design decisions. Among these, the choice of appropriate model architecture is important. This often depends on the modality of the data. For instance, a recurrent neural-inspired architecture can process sequential data systems more appropriately, while a convolutional neural network is more suitable for images. For a model design, there are generally two kinds of parameters, namely, model parameters and hyper-parameters. The model parameters refers to variables such as model internal parameter weights, which are learned automatically from the data using optimization strategies. Whereas hyper-parameters refers to design variables that are external to the neural network and includes variables such as learning rate, decay rate, momentum, optimizer, number of hidden layers, and size of different layers. These variables are often fine-tuned through manual iterations, but occasionally a computationally intensive brute-force search is employed (Snoek et al., 2012).

2.3.3. Key deep learning structures

As noted above, to a large degree, neuron types and activation functions are simple and somewhat generic, and models are largely differentiated in function and application by their structure. This section provides a brief overview of different types of relevant neural networks structures. For further details and potential use cases for wastewater systems refer to the supplementary material.

RNN (Rumelhart et al., 1986) is a specialized type of architecture to cater sequential or temporal data. They are distinguished from other types of neurons by inclusion of 'memory' that enables previous inputs to influence the outputs in future. Some popular RNN architecture includes LSTMs, Gated Recurrent Units (GRUs) and their bidirectional flavors. In contrast with RNNs, recently, a new type of architecture called *'Transformers'* (Vaswani et al., 2017) has demonstrated state-of-the-art performance over sequential data. The model has an attention based mechanism that enables better generalization than the RNNs alongside a faster acceleration over the graphical processing units.

Convolutional neural networks (CNNs) is another popular architecture that is widely used to handle structured data sets like images. At its core, it performs a convolutional operation i.e, sliding spatial filters across the input to produce a feature maps. The learned feature maps are then used for a variety tasks including classification, segmentation and detection.

A rich class of deep learning algorithms includes schemes that can learn to produce new samples of some modality such as images or text. Some of the popular architectures include *Generative Adversarial Networks (GANs)*, *Auto-encoders (AEs)* and *Diffusion Models*. At a high level GANs (Goodfellow et al., 2014) consist of a pair of neural networks that are engaged in a competitive interplay, such that one aims to produce high-quality data that other tries to discriminate it from real data. In contrast, Diffusion models (Ho et al., 2020) are inspired from Markov chains that learn by iteratively adding noise to the data and then gradually removing it in a reverse process. Similarly, the 'Auto Encoders' work by learning an identify function. This is often achieved by constraining the model to learn a smaller dimensional hidden distribution.

Table 1
Comparison of modeling approaches.

Mechanistic	Based on biochemical dynamic mass balances
Advantages	Mature field with standardized models and software support. Better prediction on unfamiliar scenarios. Can be applied in scaled approach (time and space).
Disadvantages	Requires knowledge of underlying mechanism (expert knowledge). Complex equations and can be computationally intensive. Many parameters (commonly cannot be exhaustively fit). Requires numerical solver. Requires specific monitoring and analytical campaign.
Empirical	Learn relationships between inputs and outputs without explicit modeling of the underlying system.
Advantages	Computationally very cheap. Does not require expert knowledge. Not restricted to large data sets. Limited parameter count.
Disadvantages	Need to engineer suitable features of the data for effective learning. Can only represent a limited group of I/O relationships. Black box model: provides no insight into processes.
Deep Learning	Use multiple layers to learn complex relationships between inputs and outputs.
Advantages	Not dependent on selection of specific features and outputs (can use available process data). I/O can be categorical or continuous.
Disadvantages	Black box model: provides no insight into processes. Large number of parameters and hyper-parameters to tune, requiring expert knowledge. Learning requires large data sets. Learning is computationally expensive.

2.4. Hybrid modeling

Both mechanistic and DL models have advantages and disadvantages as discussed earlier in this review. Mechanistic models are useful for process design and simulation of the dynamic behavior of the treatment plants but as noted above, development requires expert knowledge and a set of assumptions in underlying mechanisms. Both modeling approaches require field data, and potentially experimentation, with (for example) mechanistic modeling requiring targeted sampling (Corominas et al., 2018b), and machine learning being dependent on relevant data. Both can benefit from appropriate experimental design (Dochain and Vanrolleghem, 2001). Mechanistic models cannot capture all behavior and require targeted data collection. In contrast, DL models cannot be used for process scenario analysis where the model has not received relevant data and DL models have higher requirements for relevant data.

A hybrid approach offers opportunities that leverage the advantages of both methodologies across a range of applications. In limited wastewater data sets, mechanistic models can be used to augment the data required for DL modeling. Schneider et al. (2022) conducted a review of generalized empirical-mechanistic hybrid modeling, identifying the general classifications (serial, parallel, and surrogate), and the rapid emergence of ML techniques amongst empirical models. A recent study has shown that using the ASM1 (Henze et al., 2000) to augment the data sets led to an improved performance of DL models in comparison to models that were trained with raw plant data (Li et al., 2022). This increase in performance could also be explained by the fact that the mechanistic model acted as a filter that removed inherent noise in the data. Heo et al. (2021) demonstrated the combined use of mechanistic modeling (i.e., BSM2) to generate objective function outputs, together with a DL approach to optimize and map the output space. Another important aspect of hybrid modeling could be to apply the model developed with data from the mechanistic modeling through transfer learning. There is a need to investigate the potential of DL complementing engineering modeling approaches for designing, managing, monitoring and predicting wastewater treatment processes, particularly for real-time optimization for the recovery of clean water, energy, and nutrients. Table 1 presents a comparison of the three major modeling approaches.

3. Deep learning applications in wastewater treatment processes

Over the last few decades, mechanistic modeling has been the dominant approach for describing physico-chemical and biological processes taking place in wastewater treatment systems (Mannina et al., 2016). Mechanistic models such as the International Water Association (IWA) ASM series and ADM1, while very efficient in design and simulation, do not readily support analysis of complex systems in data rich-environments but instead rely on controlling mechanisms. However, in recent years, data-driven modeling approaches, in particular DL, have been rapidly-growing within the wastewater industry and are employed for a wide range of purposes. Representative applications of DL include predicting performance, process control and automation, soft sensing, fault-detection, diagnosis and missing data imputation. WWTP studies have used various deep learning algorithms: RNN, LSTM, Bi-LSTM, CNN, GRU,Bi-GRU, deep VAE (deep variational autoencoders), DBN, DRL, and GANs and hybrid methods. The technical details of the algorithms of these techniques are discussed in Section 2. This section focuses on the applications of deep learning techniques employed in wastewater treatment systems.

3.1. Process simulation

In the wastewater industry, there is an increasing need to improve the performance of treatment processes in response to stringent safety and environmental regulations. DL methods are emerging as efficient tools for predicting key performance and effluent quality variables. DL offers the potential to model wastewater processes without the need for underlying mechanistic principles and account for their non-linear nature, thus driving its adoption in activated sludge and anaerobic digestion systems. Studies on DL approaches have focused on predicting the performance and effluent quality in WWTPs. These include predicting variables such as total nitrogen (TN), total chemical oxygen demand (COD), soluble COD, biochemical oxygen demand (BOD) (Cheng et al., 2020), ammonium (NH_4-N) (Alvi et al., 2022b) and nitrate (NO_3-N) (Guo et al., 2015), total phosphorous (TP) (Yaqub et al., 2020), sludge bulking (Bagheri et al., 2015) and removal efficiency of total suspended solids (El-Rawy et al., 2021), effluent quality and biogas production (Kazadi Mbamba and Batstone, 2023).

The most widely studied DL approaches in wastewater treatment prediction are supervised learning techniques such as LSTM, RNN,

GRU, CNN, and Bi-LSTM (Cheng et al., 2020; Guo et al., 2020; Kang et al., 2020; Pisa et al., 2019; Qiao et al., 2012; Wang et al., 2019). Of these, GRU has been shown to be more efficient in terms of convergence, while LSTM networks have proven to have higher predictive capabilities for wastewater treatment systems (Cheng et al., 2020). Other studies have combined deep learning models. In this approach, the first method, such as CNN, is used to automatically extract the local features of each independent timestamp in the data and encode them, whereas the second method is used to represent the global sequential features based on the local feature encoding. Combined methods have been shown to outperform single methods. For instance, the prediction results of a hybrid CNN-LSTM prediction model showed higher accuracy and better prediction performance than stand-alone CNN or LSTM models (Guo et al., 2020; Wang et al., 2019). Similarly, Alvi et al. (2022b) proposed 'GRUconv' a combined model that exploits the strengths of GRU to extract global long-term temporal features and CNN to facilitate learning of localized short-term trends the sensor time series. The 'GRUconv' model outperforms state-of-the-art DL models.

DL studies have also have explored other aspects of wastewater treatment such as optimization of methane production (Qdais et al., 2010), quantification of long-term N_2O emissions from full-scale wastewater treatment plants using LSTM (Hwangbo et al., 2021), forecasting of influent flow with historical flow and meteorological data (Kang et al., 2020; Oliveira et al., 2020), and energy consumption optimization in activated sludge systems (Oulebsir et al., 2020; Wang et al., 2020a). These studies have indicated that deep learning models are essential for identifying both opportunities for improvement as well as causes of poor performance within wastewater systems. For example, the study by Oliveira et al. (2020) demonstrated that a hybrid deep learning model combining LSTM and a convolutional neural network (CNN) was effective in accurately forecasting influent flow rates, enabling optimal treatment process control and reducing operational costs. Similarly, the study by Hwangbo et al. (2021) developed a model for predicting N_2O emissions from wastewater treatment plants, highlighting the importance of incorporating long-term data for more accurate predictions. These studies demonstrate the potential of DL methods to address various challenges in wastewater treatment and improve its overall efficiency and sustainability.

In recent years, transfer learning techniques have been to overcome the issue related to data scarcity in WWTPs. Promising results have been achieved by using a pre-trained DL model to simulate a different but related system. For example, the usefulness of transfer learning based on GDBN was demonstrated for predicting total phosphorus in the effluent using a real-world wastewater case system (Wang et al., 2020a). Similarly, a pre-trained uni-variate CNN model was used for forecasting energy consumption of WWTPs with satisfactory results (Oliveira et al., 2021). Transfer learning has shown great potential in leveraging knowledge from existing data to enhance the performance of DL models in new and under-studied domains such as WWTPs.

DL has also been applied to classification tasks in microscopy analysis of activated sludge (AS) flocs to provide insight about the AS properties that impact on the performance of AS systems. Satoh et al. (2021) developed a CNN-based approach, using a pre-trained image classification neural network (i.e., inception architecture for computer vision), to automatically classify AS flocs in full-scale WWTPs based on morphological properties. The CNN model was trained and tested with approximately 13,000 images of AS flocs and achieved a 95 percent accuracy in recognizing aggregated or dispersed flocs and the presence or absence of filamentous bacteria. This classification model has potential as an early warning system to identify settleability deterioration and the abundance of filamentous bacteria in the aeration tanks of a full-scale AS system. However, the model architecture is highly complex with more than 22 deep layers and requires more computational resources, though it has similar predictive capabilities as the regression DL models.

Unsupervised methods have gained traction in the prediction of key performance variables in wastewater treatment systems. Niu et al. (2020) employed a genetic-deep belief network model to forecast the COD and SS in the effluent of pulp and paper mill wastewater treatment. Their study also used a genetic algorithm to reduce the input variable dimensionality and simplify the network architecture. In addition, Asadi and McPhedran (2021) utilized GAN to estimating greenhouse gas (GHG) emission rates and augment data in domestic wastewater treatment. Another unsupervised method, based on partial least squares and deep belief network (PLS-DBN), was employed by Han and Zhang (2017) for multi-step ahead permeability forecasting in membrane bioreactors, allowing early detection of fouling and pollution risks. PLS-DBN exhibits strong learning and feature-extracting capabilities, enabling its application for nonlinear system modeling and identification, with notable improvements in water quality prediction compared to single-layered neural network compared to single layered neural networks (Qiao et al., 2018).

Overall, DL techniques have proven to be powerful tools for process simulation tasks such as predicting key performance variables and identifying opportunities for improvement in wastewater treatment systems. Supervised learning techniques such as LSTM, RNN, GRU, CNN, and Bi-LSTM have been widely used for regression tasks with time-series data sets. Transfer learning and unsupervised learning methods have also been applied to overcome data scarcity issues and for prediction of other key performance variables. Moreover, DL has also been used for classification tasks in microscopy analysis of the morphology of activated sludge flocs. These studies have shown that DL models have the potential to identify opportunities for improvement as well as causes of poor performance within wastewater systems. As the field of DL continues to advance, it is expected that more sophisticated models and techniques will emerge to address the challenges and limitations of current methods, making DL an even more indispensable tool for sustainable wastewater management.

3.2. Process control and optimization

The field of process monitoring and control in wastewater treatment has been well-established for many years, relying on basic controllers like PID and digital, as well as mechanistic models for testing (Olsson and Newell, 1999). Common parameters targeted for control include NH_4-N, NO_3-N, PO_4-P, pH, temperature, and methane concentration. However, with the emergence of new paradigms in resource recovery and the advancement of data mining techniques based on DL, this field is rapidly expanding. These developments are leading to new opportunities for process monitoring and control in wastewater treatment.

Both supervised and unsupervised DL methods are increasingly being used in process control and automation. For instance, Filipe et al. (2019) proposed a control method that combined deep reinforcement learning using Proximal Policy Optimization for optimizing energy consumption in wastewater pumping stations. Similarly, Seo et al. (2021) developed a deep reinforcement learning-based control scheme for reducing the energy consumption and energy cost of pumping systems in WWTPs. This control scheme accounted for the constraints of an on/off pumping system through a designed reward function and demonstrated good performance and robustness in closed-loop systems treating industrial wastewater, as observed by Zhuang et al. (2018). Additionally, Qiao et al. (2012) demonstrated that an RNN-based multivariate control system applied to BSM1 for controlling the dissolved oxygen (*DO*) concentration, NO concentration and mixed liquor suspended solids (MLSS) concentration showed a better performance compared to a traditional PID controller. These studies demonstrate the potential of DL in improving process control and automation in wastewater treatment.

DL techniques can be applied in WWTPs for nutrient removal and energy conservation using MPC. MPC is a widely used industrial control

strategy that optimizes control actions of a system in real-time. The MPC approach typically involves a control target, a system model, and an optimization process that determines the optimal inputs required to achieve the desired plant performance. In the past, linear space models have been used to develop MPCs for wastewater treatment systems, even though such systems are highly nonlinear and dynamic (Corominas et al., 2018a). However, recent advances have enabled the development of nonlinear dynamic process models for optimal control of these systems.

DL methods are also becoming popular for nonlinear system identification and modeling, and hence have been applied to MPC in wastewater treatment. By designing an MPC based on DL, it is possible to optimize the plant responses over a specified time horizon, while simultaneously simplifying the objective function to reduce the computational requirements of the optimization. Despite recent advances in DL-based MPC (Kumar et al., 2018), this approach is still in its early stages of application in the wastewater industry. A DL-based MPC has been proposed for a continuous stirred tank reactor (Wang et al., 2020a) . The proposed MPC consisted of a growing deep belief network (GDBN) and an optimal controller. Due to its higher learning speed, GDBN was used for system identification, providing a predictive model of the controlled system. This study also used a quadratic gradient descent optimization and demonstrated better tracking control performances than other state-of-the-art MPC methods without DL dynamics. Practical implementation of DL-based MPC for wastewater systems requires improvements in system identification and reductions in the computational requirements to solve the optimization problem online or in real-time. As the field continues to advance, it is expected that DL-based MPC will become increasingly prevalent in the water resource recovery industry for optimizing wastewater treatment processes and reducing energy consumption.

Transfer learning is still in its early stages of implementation, but there is a growing interest in applying it to process control in wastewater treatment systems. Transfer learning allows control strategies developed for one system to be transferred toanother system with ease. For instance, a LSTM-based proportional-integral (PI) controller to improve the conventional PI controller was implemented to improve the conventional PI controller strategies for maintaining the concentration of *DO* at 2 mg/L in an aerobic tank of a simulated WWTP based on the Benchmark Simulation Model 1 (BSM1) system (Alex et al., 2008). Once developed, the controller was transferred to DO control in the remaining aerobic tanks, without substantial modification of the hyperparameter values and neural architecture of the pre-trained LSTM network (Pisa et al., 2021). Results show that the adoption of this transfer learning-based technique can allow for the development of new control loops with less effort, without requiring substantial knowledge of the processes being controlled, and with improved control performance compared to conventional PI control structures. However, further research is needed to investigate the potential of transfer learning in process control of wastewater treatment systems and to determine the limitations of this approach.

3.3. State estimation and soft sensing

One of the major challenges in optimizing the operation of biological wastewater treatment processes is the lack of online sensors to provide direct measurement of key parameters, such as the microbial concentrations in activated sludge which are vital for managing the process. State estimation and soft sensing are two approaches used in process control to estimate such variables that are difficult or impossible to measure directly (Yan et al., 2021). The purpose of state estimators or soft-sensors is to provide a double check, or even a replacement for physical sensors (Haimi et al., 2013).

3.3.1. State estimation

State estimation involves using mathematical models and measurements of other variables to estimate the values of unmeasured variables. It is based on the principle that the behavior of a system can be predicted based on its current state and the inputs that it receives. The estimated values of the unmeasured variables are used to improve the control of the system. This is important because the complete state of a wastewater treatment system is only partially observable and state estimation is useful to determine the underlying dynamic behavior of such a system. State estimation allows for improved process monitoring, fault detection and diagnosis and development of enhanced control strategies.

The methods for state estimation used today are normally based on using first principles techniques (Kadlec et al., 2009). These include state observers, Kalman Filter, the extended Kalman filter, unscented Kalman filter, particle filtering and moving horizon estimator (Jones et al., 1989; Fortuna et al., 2007; Welch et al., 1995). Such models describe the physical, chemical and biological processes and compute the values of states of interest on the basis of these mathematical formulations. For example, Jones et al. (1989) implemented a state estimation using extended Kalman filter for online tracking of unmeasured process states through a simulation applied to a high rate anaerobic digestion system. In this study, the algorithm of state estimation coupled available process measurements, and an adequate mechanistic dynamic model. Prediction and modeling of the major sources of stochastic disturbances was able to track several key unmeasured variables. However, the use of first principles models is prohibitive due to the cost of developing high-fidelity models, which include detailed description of all physicochemical and biological processes affecting the dynamics of treatment systems. However, the burden of developing high-fidelity mechanistic models limits the use of first principles models due to high-degree complexity and non-linearity in wastewater treatment systems. Alternatively, a surrogate model which approximates the dynamics of wastewater treatment systems may be used to design extended Kalman filters for state estimation (Yin and Liu, 2018), at the expense of model prediction accuracy. Since DL models do not require prior knowledge of the system dynamics, their capabilities for state estimation and system identification have also been investigated for control applications in robotics (Talebi et al., 2009), but there is currently no application to wastewater treatment processes.

3.3.2. Soft-sensing

On the other hand, soft sensing involves the use of data-driven models such as neural networks or support vector machines to estimate the values of unmeasured variables based on the available measurements. Unlike state estimation, soft sensing does not rely on mathematical models of the system. Instead, it uses historical data to build a model that can predict the values of unmeasured variables.

DL is increasingly being used for soft-sensing of key variables for process monitoring to ensure operational excellence in wastewater treatment. One of the advantages of DL in developing soft sensors is that the model continually learns with new data. Soft sensors or software/virtual sensors are models that can infer the values of process variables that are otherwise difficult and costly to measure with sensors.

A growing number of researchers have developed soft sensors using DL methods and their studies have shown that such methods are capable of predicting major trends in the concentrations of key variables with a high degree of accuracy. For example, soft sensors have been developed for key variables such as NO_3-N, TN, NH_4-N, PO_4-P, TP, BOD, COD and SS (Cheng et al., 2020; Pisa et al., 2019; Qiu et al., 2016; Wang et al., 2021). More recently, a NH_4 soft-sensor was developed based on 'GRUconv' model by coupling sequential modeling GRU to capture global trends and CNN kernels to facilitate learning of local behaviors (Alvi et al., 2022b). Effectiveness of this hybrid network was demonstrated using real-world data from a two-stage (high-rate

anaerobic followed by high-rate algal treatment) pilot wastewater treatment plant with excellent outcomes. This study also demonstrated that DL-based soft-sensors are cost-effective in comparison to traditional sensors. In Alvi et al. (2022a), two soft sensors were developed to approximate NH_4–N, and NO_2–N in real time using limited data. Transfer learning was leveraged to address the data scarcity issue in this paper, which proposed a method called $AETL$ (autoencoder transfer learning). In this method an LSTM autoencoder is used to systematically augment the target domain data and generate synthetic samples that closely follow the target domain distribution. The synthesized data is then used to train a source model, which is fine-tuned with real-world data from the target domain to create a more generalizable DL model.

Both state estimation and soft sensing have emerged as promising techniques for estimating difficult-to-measure process variables in wastewater treatment. The studies above demonstrate the potential of DL algorithms for improving state estimation and soft sensing by providing accurate estimates of process variables, allowing for better process control and optimization.

3.4. Data pre-processing

Pattern discovery through DL in in wastewater industry data presents significant challenges that need to be addressed to improve decision-making and process performance. These challenges include high-dimensionality of the data, process uncertainty and dynamics, as well as variations sampling time among variables (Corominas et al., 2018a). In addition, raw process data are often suffer from errors and have missing values caused by disturbances and sensor faults (Alvi et al., 2022b). These issues can result in incorrect control response, misinterpretation of data, and can negatively impact the performance of machine learning models.

To overcome these challenges, a well-designed machine learning methodology should incorporate additional steps to ensure the accuracy and reliability of the data. Quality assurance becomes critical in identifying inconsistencies and anomalies within the data requiring activities such as missing data imputation and noise removal for effective data cleaning.

DL methods are increasingly being advocated for fault detection, diagnosis and missing data imputation in WWTPs. One example is the use of LSTM-based methods, which have demonstrated superior fault detection capabilities compared to traditional sensors. In a study by (Mamandipoor et al., 2020), LSTM-based methods achieved a fault detection rate exceeding 92%, enabling timely identification of collective faults. Transformer-based models have also emerged as a powerful approach for fault detection in wastewater treatment processes. Peng and Fanchao (2022) proposed a transformer-based model that incorporates position encoding, residual connections, and multi-head attention mechanisms. The empirical evaluation of their approach on the BSM1 model showcased significantly lower false alarm and missed alarm rates compared to state-of-the-art methods like PCA and support vector machine (SVM) approaches.

Unsupervised DL approaches such as DBN combined with one-class SVM have proven effective for fault detection. Additionally, deep variational autoencoders (VAE) based on residual neural networks (ResNet) were applied for missing data imputation and sensor self-validation (Ba-Alawi et al., 2022). The integration of ResNet-VAE frameworks offers advantages in extracting complex features from membrane bioreactor data and overcoming vanishing gradient issues. The VAE-ResNet approach exhibited improved performance in detecting and reconstructing faulty sensors, as well as imputing missing values. with mean absolute percentage errors (MAPE) varying between 3.98% to 10.44%. However, it should be noted that while ResNet has excelled in classification and computer vision tasks, its computational resource requirements, primarily due to its deep structure, raise concerns about its feasibility in simple regression tasks such as those found in wastewater systems. Another proposed approach for anomaly detection in WWTPs is the coupling of RNN and Restricted Boltzmann Machines (RNN-RBM) with classifiers, as highlighted by Dairi et al. (2019).

These DL techniques offer promising avenues for enhancing fault detection, diagnosis, and missing data imputation in the wastewater industry. Continued research and development in this area can lead to more accurate and efficient decision-making processes within WWTPs.

3.5. CFD modeling and deep learning

Computational fluid dynamics (CFD) is a powerful tool for simulating fluid flows in the water industry, but its high computational cost limits its for iterative applications such as iterative CFD-assisted design. To address this issue, a hybrid approach that combines traditional CFD with machine learning techniques has been proposed. By mapping expected outputs and providing design candidates at relatively low computational costs, this approach has been successfully applied in other fields such as mechanical and aeronautical engineering (Hammond et al., 2022). However its applications in the water industry is still limited. In recent years, deep learning, a subset of machine learning, has gained popularity in the water industry due to its ability to learn complex relationships from data. The combination of CFD and deep learning has the potential to revolutionize the design and upscaling of bioreactors in the water industry, allowing for more efficient and cost-effective solutions. While this approach is still in its infancy, it holds great promise for the future of water treatment and resource recovery.

4. Discussion and needs analysis

This section identifies opportunities for DL in WWTPs and possible directions for this emerging field to benefit wastewater designers, operators and researchers. We then discuss key challenges for DL in the particular context of wastewater treatment-related tasks.

4.1. Value of deep learning for wastewater treatment

As noted in the previous section, DL has been applied in the wastewater sector for prediction, predictive control, soft sensing, and fault diagnostics (Alvi et al., 2022b; Cheng et al., 2020; Pisa et al., 2019; Mamandipoor et al., 2020). Table 2 summarizes cited references, showing the methods that have been applied and the WWTP problems addressed. Despite their prominence and potential success, it is important to note that these models are application and case specific. The characteristics of WWTPs can vary depending on geographic region and underlying treatment processes. As a consequence of this heterogeneous behavior, for each application a new model needs to be developed. The latest advances in deep learning technologies provide new techniques to transfer learning from existing trained models to data from same/other plants. For instance, transferring knowledge from one trained LSTM-based PI controller to another different control loop in a WWTP (Pisa et al., 2021).

There are many aspects of deep learning that could be fruitful for the wastewater industry, such as end-to-end learning with the capability of handling complex and multi-modal data. For example, in a typical WWTP, data is generated from multiple sources such as cameras, microscopic images, sensor probes, laboratory test and even structured text records for maintenance and plant management. A DL model can gain knowledge and actionable insights from complex, high-dimensional and heterogeneous biochemical data. Multi-modal data fusion can provide complementary information due to its dependence on variable acquisition parameters (Wang et al., 2020b). This capability means that previously un-utilized information from a WWTP may prove to be useful for learning accurate decision support systems. Multi-modal data fusion examples can be seen in core computer science literature (Ngiam et al., 2011), but not in the wastewater domain.

Table 2
Overview of deep learning approaches and their applications in WWTPs. Highlighted cells note examples of published studies where an architecture is applied to a certain problem type. Applications of State Estimation does not yet have WWTP case studies. Refer to the table of nomenclature for acronyms.

	Application in WWTPs			
Architecture	Process Simulation	Process Control	Soft Sensing	Data Quality
FFNN	–	–	Alvi 2022a	–
RNN	–	Qiao 2012	Chang 2021	Dairi 2019
(Bi)LSTM	Cheng 2020 Wang 2019 Oliveira 2021 Kang 2020 Hwangbo 2021 Yaqub 2020 El-Rawy 2021	Pisa 2019 Pisa 2021 Guo 2020 Kumar 2018	Pisa 2019 Alvi 2022b Cheng 2020	Mamandipoor 2020
(Bi)GRU	Cheng 2020 Oliveira 2021	–	Alvi 2022b Cheng 2020	–
CNN	Satoh 2021 Wang 2019 Oliveira 2021	Guo 2020	Alvi 2022b	–
Hybrid Network	Wang 2019 Bagheri 2015	Guo 2020	Alvi 2022b	–
Transformer	–	–	–	Peng 2022
AE	–	–	Alvi2022a Qui 2016	Ba-Alwi 2022 Alvi2022a
GAN	–	–	–	Asadi 2021
DBN	Niu 2022 Hou 2017	Wang 2020	–	–
RBM	–	–	–	Dairi 2019
DRL	–	Seo 2021 Filipe 2019 Zhuang 2018	–	–
TL	Wang 2020 Oliveira 2021	Pisa 2021	Alvi 2022a	Alvi 2022a

4.2. Open problems for applying deep learning to wastewater treatment

Section 3 has discussed and summarized many promising applications that can benefit from advances in deep learning as listed in Table 2. However, there are various challenges that have hindered widespread adoption of these techniques in the wastewater industry. This subsection will discuss some of the challenges and possible ways to address them.

Lack of training data. The wastewater industry lacks large, public data sets. This is due to the high cost in terms of time, resources, and equipment required to collect the necessary measurements, need to curate data, and obligations of authorities in terms of public safety. However, we believe that DL will emerge as an important technique for wastewater applications in the future because large data sets are beginning to become available, given the availability of low-cost sensors and wider uptake of IoT sensing systems. Some public data sets have been recently introduced on data hubs such as Arcgis.[2] However, many existing public data sets suffer from low frequency (daily) and lack important metadata such as climate information. Available data sets are derived from diverse types of wastewater treatment processes (Alvi et al., 2022b; Kazadi Mbamba and Batstone, 2023), which reduces their suitability for training specialized deep learning models. Data logged from wastewater treatment processes may also lack annotations (classification labels and context information) that are necessary to train deep algorithms. Unlike domains such as image processing, data annotation in wastewater requires domain expertise, and cannot be performed using crowd-sourcing annotation tools such as Amazon Turk.

Training models with a limited data set is a challenging problem as the data set can suffer from inherent problems. One well-known issue is the problem of imbalanced data distribution (Longadge and Dongre, 2013; Chawla, 2009; Alvi et al., 2022b). Other problems include the presence of non-recurring seasonal patterns and noise or anomalies in the data. For instance, in a WWTP, a sensor may perform anomalously due to physical problems, such as clogging, or a natural external perturbation like rainfall. High rainfall can cause biological washout, or may be diverted, perturbing main treatment plant performance. Although, these events are rare, because of their significance to operations it is important to predict the effect of such events. This emphasizes the need for metadata and event logging to provide context. Addressing this problem demands the development of algorithms that can generalize in presence of skewed data.

Alternatively, a skewed data set can be augmented so it is more suitable for DL. Alvi et al. (2022b) has proposed an algorithm that divides the original limited skewed data, such that each data split has a similar statistical distribution. That in turn enables training of deep models with limited data and effectively provides fair evaluation of the generalizability of the models. Other techniques that remedy imbalanced data sets include SMOTE: Synthetic Minority Over-Sampling Technique (Chawla et al., 2002) and cost-sensitive learning (Domingos, 1999; Zadrozny et al., 2003). In cost-sensitive learning, a cost matrix is defined that has different penalties for misclassification with respect to the majority and minority class samples.

Similarly, another body of research has tackled the limited data issues by hybrid modeling based augmentation methods. In these techniques, synthetic data is generated from mechanistic models and is used alongside real-world data to improve performance of deep learning models (Li et al., 2022). Recently, instead of mechanistic techniques, deep generative models have been also explored for data augmentation. The generative models are trained to learn real distribution of data and afterwards leveraged to generate additional data samples (Asadi and McPhedran, 2021; Alvi et al., 2022a). In Alvi et al. (2022a) autoencoder is employed for data augmentation within the transfer learning paradigm. It is evident from these studies that these approaches not only alleviate the data scarcity issue but also enhance the generalizability of predictive models. Relevant core machine learning literature

[2] https://hub.arcgis.com/search?q=wastewater

also demonstrates that generative models produces high quality samples than traditional techniques and generally augmented data sets improves performance of deep models. Transfer learning techniques also have potential to induce effective predictive models with limited data. However, both generative and TL approaches are relatively unexplored and need more attention for possible adoption in the wastewater industry.

Model generalization. Model generalization is the ability of a learned model to perform well on unseen data. It relies on the assumption that the training data and the unseen test data are independently sampled from the same underlying data distribution (Ben-David et al., 2010). In the context of WWTPs, this assumption may not hold because of *concept drift* and *data drift*. For instance, the minimal control over the feed to the WWTPs often results in high system variabilities, such as changing hydraulic loads, weather conditions, and complex biochemical phenomena. Consequently, the system can behave differently than expected. In addition, the treatment process may evolve in unforeseen ways, such as changes in operating conditions, i.e., chemical dosing, temperature, etc. These intricacies often create out-of-distribution scenarios for a deployed model, and consequently, the models suffer from performance loss (Lu et al., 2018).

The other factor that hinders the generalization of deep models is *open set recognition problems* (i.e., a model trained on limited training conditions). For instance, a model trained over data samples from a treatment plant collected in dry weather conditions is most likely to fail when it is tested over samples from wet weather conditions. These factors can introduce additional distributional differences between the train and test data. In Alvi et al. (2022b), it was found that deep learning models were unable to predict certain days, and these were identified as wet or rainy days. Further analysis revealed that the models lacked sufficient training data on rainy days, which contributed to their poor performance in these cases. Thus, routine capture and update of deep model parameters/design is an important design and deployment practice. These factors are domain-specific, but the rich body of literature in data augmentation, generative models like GAN, diffusion models, and autoencoders can be explored for possible benefits.

Model interpretation. Access to inner mechanisms and understanding how outputs are determined are fundamental to model transparency and reliability. It also inculcates confidence for deployment in practical settings. The classical machine learning models including decision trees, k-means clustering, and support vector machines enable experienced researchers to examine the process of decision making. However, interpreting the results of a deep learning model is a daunting challenge and an inherent difficulty related to 'black box' terminologies. This challenges trust in the models, particularly in dynamic environments like WWTPs. For instance, an incorrect or biased decision made by predictive model in the wastewater industry can cause significant economic and environmental harm.

Explainable AI aims to provide human understandable explanations for the outputs of machine learning models. As this area is in its nascent phase various research challenges exist. These include (i) the definition of model explainability, (ii) the formulation of explainable algorithms to assess model behavior and develop solutions to improve them, and finally, (iii) the design of measures to evaluate the performance of models in their explanation. Although there is a growing body of articles in other application areas (Mokhtari et al., 2019; Lundberg and Lee, 2017), this area has yet to be explored in the wastewater research community. There are only a handful of works. For instance, Alvi et al. (2022b) explored feature importance via its absence to understand its role in model predictions. More formal evaluations are needed using techniques such as SHAP (SHapley Additive exPlanations) (Lundberg and Lee, 2017) to foster trust in the model's output and enable informed decision making.

Preprocessing & hyper-parameters. Data pre-processing is an essential step in developing deep learning models for prediction, fault detection, and process control. However, most studies have not addressed the issue of input variable selection, which is critical to optimal model performance (Kazadi Mbamba and Batstone, 2023). A comprehensive methodology for data pre-processing in the application of deep learning in wastewater systems is required. In addition, deep learning algorithms involve setting of many hyper-parameters. These include definition of the architecture, type of neurons, and number of layers that need to be set before training and testing. Presently, there is no standard systematic methodology for obtaining the best hyper-parameters. However, we see an increasing interest in this direction (Kazadi Mbamba and Batstone, 2023). More attention to this area may lead to improved model performance, resource utilization, and generalization across different data sets and tasks.

Model robustness. A deep model's robustness to different types of noise is a desirable property for practical deployment. In real world plants, there are several perturbation sources including: internal sensor variance, external perturbations, and natural fluctuations in weather conditions. It is important that model behavior remains stable under these circumstances. Deep models developed for wastewater treatment facilities that do not have an explicit notion of robustness during their training phase are possibly prone to different sources of noise. This is especially true for Gaussian and adversarial noise: a carefully designed signal that can completely alter the behavior of a model (Yuan et al., 2019a). An illustrative example in Alvi et al. (2022b) demonstrates that simple changes in the input feature space have a drastic impact on a model's performance. Specifically, the model's ability to predict ammonium decreased by $\approx$ 25% when only a single input feature pH was altered by its mean value. In addition, adding a little Gaussian noise to the input signal also deteriorates a model's performance. Due to the lack of robustness, Ba-Alawi et al. (2021) utilized a stacked denoising autoencoder (SDAE) to reconcile faulty/noisy sensor data from a real WWTP. Similarly, other noise injection training approaches (Yin et al., 2015) and denoising diffusion models (Ho et al., 2020) can be explored to enhance the robustness of deep models.

5. Conclusion

This article identifies broad applications of deep learning in wastewater systems, including system representation, process control, soft sensing, and data quality and aggregation. Further opportunities have been identified in computer assisted design, exploration of design spaces, and in hybrid modeling (mechanistic-empirical models). A number of practical challenges have been identified, including data quality and quantity, the non-transparent nature of deep learning models, and model robustness. In particular, DL requires a large volume of quality samples that describe the empirical distribution of processes of interest. While open data sets are becoming more available, the quality, range of process variables, and particularly explanatory metadata to accompany this is critical. DL also demands the development of advanced algorithms to ascertain their robustness and methods are needed to interpret the decision flow within neural networks. Additional research directions include a robust and repeatable framework for feature selection and ability to transfer models between systems using both transfer learning and use of mechanistic models for the same.

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 paper.

Data availability

No data was used for the research described in the article.

Acknowledgments

This work was supported in part by the Cooperative Research Centre Projects (CRC-P) *Transforming wastewater treatment in regional Australia with robust technology for multiple benefits* through the Australian Government, Urban Utilities Queensland, under Grant CRCPSIX000079; and by a Ph.D. Scholarship from The University of Western Australia.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.watres.2023.120518.

References

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., et al., 2008. Benchmark Simulation Model No. 1 (BSM1). Report by the IWA Taskgroup on Benchmarking of Control Strategies for WWTPs, pp. 19–20.

Alvi, M., Cardell-Oliver, R., French, T., 2022a. Utilizing autoencoders to improve transfer learning when sensor data is sparse. In: Proceedings of the 9th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation. Association for Computing Machinery, New York, NY, USA, pp. 500–503. http://dx.doi.org/10.1145/3563357.3567407.

Alvi, M., French, T., Cardell-Oliver, R., Keymer, P., Ward, A., 2022b. Cost effective soft sensing for wastewater treatment facilities. IEEE Access 10, 55694–55708. http://dx.doi.org/10.1109/ACCESS.2022.3177201.

Alwan, L.C., 1992. Effects of autocorrelation on control chart performance. Commun. Stat.-Theory Methods 21, 1025–1049.

Asadi, M., McPhedran, K.N., 2021. Greenhouse gas emission estimation from municipal wastewater using a hybrid approach of generative adversarial network and data-driven modelling. Sci. Total Environ. 800, 149508.

Ba-Alawi, A.H., Loy-Benitez, J., Kim, S., Yoo, C., 2022. Missing data imputation and sensor self-validation towards a sustainable operation of wastewater treatment plants via deep variational residual autoencoders. Chemosphere 288, 132647.

Ba-Alawi, A.H., Vilela, P., Loy-Benitez, J., Heo, S., Yoo, C., 2021. Intelligent sensor validation for sustainable influent quality monitoring in wastewater treatment plants using stacked denoising autoencoders. J. Water Process Eng. 43, 102206.

Bagheri, M., Mirbagheri, S.A., Bagheri, Z., Kamarkhani, A.M., 2015. Modeling and optimization of activated sludge bulking for a real wastewater treatment plant using hybrid artificial neural networks-genetic algorithm approach. Process Saf. Environ. Prot. 95, 12–25.

Batstone, D.J., Amerlinck, Y., Ekama, G., Goel, R., Grau, P., Johnson, B., Kaya, I., Steyer, J.P., Tait, S., Takács, I., Vanrolleghem, P.A., Brouckaert, C.J., Volcke, E., 2012. Towards a generalized physicochemical framework. Water Sci. Technol. 66, 1147–1161. http://dx.doi.org/10.2166/wst.2012.300.

Batstone, D.J., Keller, J., Angelidaki, I., Kalyuzhnyi, S., Pavlostathis, S., Rozzi, A., Sanders, W., Siegrist, H., Vavilin, V., 2002. The IWA anaerobic digestion model no 1 (ADM1). Water Sci. Technol. 45, 65–73.

Ben-David, S., Blitzer, J., Crammer, K., Kulesza, A., Pereira, F., Vaughan, J.W., 2010. A theory of learning from different domains. Mach. Learn. 79, 151–175.

Bengio, Y., 2012. Practical recommendations for gradient-based training of deep architectures. In: Neural Networks: Tricks of the Trade, second ed. pp. 437–478.

Bengio, Y., Courville, A., Vincent, P., 2013. Representation learning: A review and new perspectives. IEEE Trans. Pattern Anal. Mach. Intell. 35, 1798–1828.

Box, G.E., Jenkins, G.M., MacGregor, J.F., 1974. Some recent advances in forecasting and control. J. R. Stat. Soc. Ser. C. Appl. Stat. 23, 158–179.

Burton, F., Tchobanoglous, G., Tsuchihashi, R., Stensel, H., Metcalf and Eddy Inc, 2013. Wastewater Engineering: Treatment and Resource Recovery. McGraw-Hill Education.

Chawla, N.V., 2009. Data mining for imbalanced datasets: An overview. In: Data Mining and Knowledge Discovery Handbook. pp. 875–886.

Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. Smote: Synthetic minority over-sampling technique. J. Artif. Intell. Res. 16, 321–357.

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.

Chollet, F., 2021. Deep Learning with Python. Simon and Schuster.

Corominas, L., Garrido-Baserba, M., Villez, K., Olsson, G., Cortés, U., Poch, M., 2018b. Transforming data into knowledge for improved wastewater treatment operation: A critical review of techniques. Environ. Model. Softw. 106, 89–103. http://dx.doi.org/10.1016/j.envsoft.2017.11.023.

Corominas, L., Garrido-Baserba, M., Villez, K., Olsson, G., Cortés, M., 2018a. Transforming data into knowledge for improved wastewater treatment operation: A critical review of techniques. Environ. Model. Softw. 106, 89–103.

Dairi, A., Cheng, T., Harrou, F., Sun, Y., Leiknes, T., 2019. Deep learning approach for sustainable wwtp operation: A case study on data-driven influent conditions monitoring. Sustainable Cities Soc. 50, 101670.

Dargan, S., Kumar, M., Ayyagari, M.R., Kumar, G., 2020. A survey of deep learning and its applications: A new paradigm to machine learning. Arch. Comput. Methods Eng. 27, 1071–1092.

Dochain, D., Vanrolleghem, P.A., 2001. Dynamical Modelling & Estimation in Wastewater Treatment Processes. IWA publishing.

Domingos, P., 1999. Metacost: A general method for making classifiers cost-sensitive. In: Proceedings of the fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 155–164.

Durbin, J., Koopman, S.J., 2012. Time Series Analysis by State Space Methods. Oxford University Press.

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 Process Eng. 44, 102380.

Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., 1996. From data mining to knowledge discovery in databases. AI Mag. 17, 37.

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.

Flores-Alsina, X., Mbamba, C.K., Solon, K., Vrecko, D., Tait, S., Batstone, D.J., Jeppsson, U., Gernaey, K.V., 2015. A plant-wide aqueous phase chemistry module describing pH variations and ion speciation/pairing in wastewater treatment process models. Water Res. 85, 255–265. http://dx.doi.org/10.1016/j.watres.2015.07.014.

Fortuna, L., Graziani, S., Rizzo, A., Xibilia, M.G., et al., 2007. Soft Sensors for Monitoring and Control of Industrial Processes, Vol. 22. Springer.

Fu, G., Jin, Y., Sun, S., Yuan, Z., Butler, D., 2022. The role of deep learning in urban water management: A critical review. Water Res. 118973.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets. Adv. Neural Inf. Process. Syst. 27.

Guo, Z., Du, B., Wang, J., Shen, Y., Li, Q., Feng, D., Gao, X., Wang, H., 2020. Data-driven prediction and control of wastewater treatment process through the combination of convolutional neural network and recurrent neural network. RSC Adv. 10, 13410–13419.

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

Gustafsson, F., 2007. Statistical signal processing approaches to fault detection. Annu. Rev. Control 31, 41–54.

Haimi, H., Mulas, M., Corona, F., Vahala, R., 2013. Data-derived soft-sensors for biological wastewater treatment plants: An overview. Environ. Model. Softw. 47, 88–107.

Hammond, J., Pepper, N., Montomoli, F., Michelassi, V., 2022. Machine learning methods in CFD for turbomachinery: A review. Int. J. Turbomach. Propul. Power 7, 16.

Han, J., Pei, J., Kamber, M., 2011. Data Mining: Concepts and Techniques. Elsevier.

Han, H.G., Zhang, S., 2017. An early warning system for mbr based on multi-step prediction and deep belief network classifier. In: 2017 Chinese Automation Congress. CAC, IEEE, pp. 7442–7447.

Hangos, K., Cameron, I. (Eds.), 2001. Process Modelling and Model Analysis, Vol. 4. Science Direct.

Henze, M., Gujer, W., Mino, T., van Loosdrecht, M.C., 2000. Activated sludge models ASM1, ASM2, ASM2d and ASM3. IWA publishing.

Heo, S., Nam, K., Tariq, S., Lim, J.Y., Park, J., Yoo, C., 2021. A hybrid machine learning–based multi-objective supervisory control strategy of a full-scale wastewater treatment for cost-effective and sustainable operation under varying influent conditions. J. Clean. Prod. 291, 125853.

Ho, J., Jain, A., Abbeel, P., 2020. Denoising diffusion probabilistic models. Adv. Neural Inf. Process. Syst. 33, 6840–6851.

Hwangbo, S., Al, R., Chen, X., Sin, G., 2021. Integrated model for understanding N2O emissions from wastewater treatment plants: A deep learning approach. Environ. Sci. Technol. 55, 2143–2151.

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

Jeppsson, U., Alex, J., Batstone, D.J., Benedetti, L., Comas, J., Copp, J.B., Corominas, L., Flores-Alsina, X., Gernaey, K.V., Nopens, I., Pons, M.N., Rodrıguez-Roda, I., Rosen, C., Steyer, J.P., Vanrolleghem, P.A., Volcke, E.I.P., Vrecko, D., 2013. Benchmark simulation models, QUO vadis? Water Sci. Technology 68, 1–15. http://dx.doi.org/10.2166/wst.2013.246.

Jeppsson, U., Pons, M.N., Nopens, I., Alex, J., Copp, J., Gernaey, K., Rosén, J.P., Vanrolleghem, P., 2007. Benchmark simulation model no 2: General protocol and exploratory case studies. Water Sci. Technol. 56, 67–78.

Jones, R., MacGregor, J., Murphy, K., 1989. State estimation in wastewater engineering: Application to an anaerobic process. Environ. Monit. Assess. 13, 271–282.

Kadlec, P., Gabrys, B., Strandt, S., 2009. Data-driven soft sensors in the process industry. Comput. Chem. Eng. 33, 795–814.

Kang, H., Yang, S., Huang, J., Oh, J., 2020. Time series prediction of wastewater flow rate by bidirectional lstm deep learning. Int. J. Control Autom. Syst. 18, 3023–3030.
Kazadi Mbamba, C., Batstone, D., 2023. Optimization of deep learning models with genetic algorithms for forecasting performance in water industry. Comput. Chem. Eng. 108276.
Kumar, S.S.P., Tulsyan, A., Gopaluni, B., Loewen, P., 2018. A deep learning architecture for predictive control. IFAC-PapersOnLine 51, 512–517.
Le Guen, V., Thome, N., 2019. Shape and time distortion loss for training deep time series forecasting models. In: Advances in Neural Information Processing Systems. Curran Associates, Inc.
Li, K., Duan, H., Liu, L., Qiu, R., Akker, B.van den., Ni, B.J., Chen, T., Yin, H., Yuan, Z., Ye, L., 2022. An integrated first principal and deep learning approach for modeling nitrous oxide emissions from wastewater treatment plants. Environ. Sci. Technol. 56, 2816–2826.
Longadge, R., Dongre, S., 2013. Class imbalance problem in data mining review. arXiv preprint arXiv:1305.1707.
Lu, J., Liu, A., Dong, F., Gu, F., Gama, J., Zhang, G., 2018. Learning under concept drift: A review. IEEE Trans. Knowl. Data Eng. 31, 2346–2363.
Lundberg, S.M., Lee, S.I., 2017. A unified approach to interpreting model predictions. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), Advances in Neural Information Processing Systems, Vol. 30. Curran Associates, Inc., pp. 4765–4774, URL http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf.
Mamandipoor, B., Majd, M., Sheikhalishahi, S., Modena, C., Osmani, V., 2020. Monitoring and detecting faults in wastewater treatment plants using deep learning. Environ. Monit. Assess. 192, 1–12.
Mannina, G., Ekama, G., Caniani, D., Cosenza, A., Esposito, G., Gori, R., Garrido-Baserba, M., Rosso, D., Olsson, G., 2016. Greenhouse gases from wastewater treatment–A review of modelling tools. Sci. Total Environ. 551, 254–270.
McKenzie, E., 1984. General exponential smoothing and the equivalent arma process. J. Forecast. 3, 333–344.
Meir, R., 2000. Nonparametric time series prediction through adaptive model selection. Mach. Learn. 39, 5–34.
Mokhtari, K.E., Higdon, B.P., Başar, A., 2019. Interpreting financial time series with shap values. In: Proceedings of the 29th Annual International Conference on Computer Science and Software Engineering. pp. 166–172.
Naik, K.S., Stenstrom, M.K., 2012. Evidence of the influence of wastewater treatment on improved public health. Water Sci. Technol. 66, 644–652. http://dx.doi.org/10.2166/wst.2012.144, arXiv:https://iwaponline.com/wst/article-pdf/66/3/644/443099/644.pdf.
Newhart, K.B., Holloway, R.W., Hering, A.S., Cath, T.Y., 2019. Data-driven performance analyses of wastewater treatment plants: A review. Water Res. 157, 498–513.
Ngiam, J., Khosla, A., Kim, M., Nam, J., Lee, H., Ng, A.Y., 2011. Multimodal deep learning. In: ICML.
Niu, G., Yi, X., Chen, C., Li, X., Han, D., Yan, B., Huang, M., Ying, G., 2020. A novel effluent quality predicting model based on genetic-deep belief network algorithm for cleaner production in a full-scale paper-making wastewater treatment. J. Clean. Prod. 265, 121787.
Ocampo-Martinez, C., 2010. Model Predictive Control of Wastewater Systems. Springer Science & Business Media.
Oliveira, P., Fernandes, B., Aguiar, F., Pereira, M.A., Analide, C., Novais, P., 2020. A deep learning approach to forecast the influent flow in wastewater treatment plants. In: International Conference on Intelligent Data Engineering and Automated Learning. Springer, pp. 362–373.
Oliveira, P., Fernandes, B., Analide, C., Novais, P., 2021. Forecasting energy consumption of wastewater treatment plants with a transfer learning approach for sustainable cities. Electronics 10, 1149.
Olsson, G., Newell, B., 1999. Wastewater Treatment Systems. IWA publishing.
Olsson, G., Nielsen, M., Yuan, Z., Lynggaard-Jensen, A., Steyer, J.P., 2005. Instrumentation, Control and Automation in Wastewater Systems. IWA publishing.
Oulebsir, R., Lefkir, A., Safri, A., Bermad, A., 2020. Optimization of the energy consumption in activated sludge process using deep learning selective modeling. Biomass Bioenergy 132, 105420.
Pang, G., Shen, C., Cao, L., Hengel, A.V.D., 2021. Deep learning for anomaly detection: A review. ACM Comput. Surv. 54, 1–38.
Peng, C., Fanchao, M., 2022. Fault detection of urban wastewater treatment process based on combination of deep information and transformer network. IEEE Trans. Neural Netw. Learn. Syst..
Pinheiro, A., Cabral, M., Antunes, S., Brôco, N., Covas, D., 2018. Estimating capital costs of wastewater treatment plants at the strategical level. Urban Water J. 15, 732–740. http://dx.doi.org/10.1080/1573062X.2018.1547409.
Pisa, I., Morell, A., Vilanova, R., Vicario, J.L., 2021. Transfer learning in wastewater treatment plant control design: From conventional to long short-term memory-based controllers. Sensors 21, 6315.
Pisa, I., Santin, I., Morell, A., Vicario, J.L., Vilanova, R., 2019. LSTM-based wastewater treatment plants operation strategies for effluent quality improvement. IEEE Access 7, 159773–159786.
Puyol, D., Barry, E., Hülsen, T., Batstone, D., 2017. A mechanistic model for anaerobic phototrophs in domestic wastewater applications: Photo-anaerobic model (PANM). Water Res. 116, 241–253. http://dx.doi.org/10.1016/j.watres.2017.03.022.
Qdais, H.A., Hani, K.B., Shatnawi, N., 2010. Modeling and optimization of biogas production from a waste digester using artificial neural network and genetic algorithm. Resour. Conserv. Recy. 54, 359–363.
Qiao, J., Huang, X., Han, H., 2012. Recurrent neural network-based control for wastewater treatment process. In: International Symposium on Neural Networks. Springer, pp. 496–506.
Qiao, J., Wang, G., Li, W., Li, X., 2018. A deep belief network with plsr for nonlinear system modeling. Neural Netw. 104, 68–79.
Qiu, Y., Liu, Y., Huang, D., 2016. Date-driven soft-sensor design for biological wastewater treatment using deep neural networks and genetic algorithms. J. Chem. Eng. Jpn. 49, 925–936.
Reagan, K.M., 1984. An Evaluation of ARIMA(Box-Jenkins) Models for Forecasting Wastewater Treatment Process Variables (Master's thesis). UCLA.
Rieger, L., Gillot, S., Langergraber, G., Ohtsuki, T., Shaw, A., Takacs, I., Winkler, S., 2012. Guidelines for using Activated Sludge Models. IWA Publishing.
Rosén, C., Lennox, J., 2001. Multivariate and multiscale monitoring of wastewater treatment operation. Water Res. 35, 3402–3410.
Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. Nature 323, 533–536.
Samstag, R.W., Ducoste, J.J., Griborio, A., Nopens, I., Batstone, D.J., Wicks, J.D., Saunders, S., Wicklein, E.A., Kenny, G., Laurent, J., 2016. Cfd for wastewater treatment: An overview. Water Sci. Technol. 74, 549–563. http://dx.doi.org/10.2166/wst.2016.249.
Satoh, H., Kashimoto, Y., Takahashi, N., Tsujimura, T., 2021. Deep learning-based morphology classification of activated sludge flocs in wastewater treatment plants. Environ. Sci.: Water Res. Technol. 7, 298–305.
Schneider, M.Y., Quaghebeur, W., Borzooei, S., Froemelt, A., Li, F., Saagi, R., Wade, M.J., Zhu, J.J., Torfs, E., 2022. Hybrid modelling of water resource recovery facilities: Status and opportunities. Water Sci. Technol. 85, 2503–2524.
Seo, G., Yoon, S., Kim, M., Mun, C., Hwang, E., 2021. Deep reinforcement learning-based smart joint control scheme for on/off pumping systems in wastewater treatment plants. IEEE Access 9, 95360–95371.
Snoek, J., Larochelle, H., Adams, R.P., 2012. Practical bayesian optimization of machine learning algorithms. Adv. Neural Inf. Process. Syst. 25.
Solomatine, D.P., Ostfeld, A., 2008. Data-driven modelling: Some past experiences and new approaches. J. Hydroinform. 10, 3–22.
Solon, K., Flores-Alsina, X., Mbamba, C.K., Ikumi, D., Volcke, E., Vaneeckhaute, C., Ekama, G., Vanrolleghem, P., Batstone, D., Gernaey, K., Jeppsson, U., 2017. Plant-wide modelling of phosphorus transformations in wastewater treatment systems: Impacts of control and operational strategies. Water Res. 113, 97–110. http://dx.doi.org/10.1016/j.watres.2017.02.007.
Takács, G., Nolasco, D., 1991. A dynamic model of the clarification-thickening process. Water Res. 25, 1263–1271. http://dx.doi.org/10.1016/0043-1354(91)90066-Y, URL https://www.sciencedirect.com/science/article/pii/004313549190066Y.
Talebi, H.A., Abdollahi, F., Patel, R.V., Khorasani, K., 2009. Neural Network-Based State Estimation of Nonlinear Systems: Application to Fault Detection and Isolation, Vol. 395. Springer.
Van Veen, F., Leijnen, S., 2016. The Neural Network Zoo. The Asimov Institute.
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. In: Advances in Neural Information Processing Systems. pp. 5998–6008.
Venkatasubramanian, V., Rengaswamy, R., Kavuri, S.N., Yin, K., 2003. A review of process fault detection and diagnosis: Part III: Process history based methods. Comput. Chem. Eng. 27, 327–346.
Wágner, D.S., Valverde-Pérez, B., Saebø, M., de la Sotilla, M.B., Wagenen, J.V., Smets, B.F., Plósz, B.G., 2016. Towards a consensus-based biokinetic model for green microalgae ? the ASM-A. Water Res. 103, 485–499. http://dx.doi.org/10.1016/j.watres.2016.07.026.
Wang, Y., Huang, W., Sun, F., Xu, T., Rong, Y., Huang, J., 2020b. Deep multimodal fusion by channel exchanging. In: Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., Lin, H. (Eds.), Advances in Neural Information Processing Systems. Curran Associates, Inc., pp. 4835–4845, URL https://proceedings.neurips.cc/paper/2020/file/339a18def9898dd60a634b2ad8fbbd58-Paper.pdf.
Wang, G., Jia, Q.S., Qiao, J., Bi, J., Zhou, M., 2020a. Deep learning-based model predictive control for continuous stirred-tank reactor system. IEEE Trans. Neural Netw. Learn. Syst..
Wang, G., Jia, Q.S., Zhou, M., Bi, J., Qiao, J., Abusorrah, A., 2021. Artificial neural networks for water quality soft-sensing in wastewater treatment: A review. Artif. Intell. Rev. 1–23.
Wang, Z., Man, Y., Hu, Y., Li, J., Hong, M., Cui, P., 2019. A deep learning based dynamic cod prediction model for urban sewage. Environ. Sci.: Water Res. Technol. 5, 2210–2218.
Welch, G., Bishop, G., et al., 1995. An Introduction to the Kalman Filter. Chapel Hill, NC, USA.
Yan, P., Gai, M., Wang, Y., Gao, X., 2021. Review of soft sensors in anaerobic digestion process. Processes 9, 1434.
Yaqub, M., Asif, H., Kim, S., Lee, W., 2020. Modeling of a full-scale sewage treatment plant to predict the nutrient removal efficiency using a long short-term memory (LSTM) neural network. J. Water Process Eng. 37, 101388.
Yin, X., Liu, J., 2018. State estimation of wastewater treatment plants based on model approximation. Comput. Chem. Eng. 111, 79–91.

Yin, S., Liu, C., Zhang, Z., Lin, Y., Wang, D., Tejedor, J., Zheng, T.F., Li, Y., 2015. Noisy training for deep neural networks in speech recognition. EURASIP J. Audio Speech Music Process. 2015, 1–14.

Yuan, X., He, P., Zhu, Q., Li, X., 2019a. Adversarial examples: Attacks and defenses for deep learning. IEEE Trans. Neural Netw. Learn. Syst. 30, 2805–2824.

Yuan, Z., Olsson, G., Cardell-Oliver, R., Schagen, K.van., Marchi, A., Deletic, A., Urich, C., Rauch, W., Liu, Y., Jiang, G., 2019b. Sweating the assets–the role of instrumentation, control and automation in urban water systems. Water Res. 155, 381–402.

Zadrozny, B., Langford, J., Abe, N., 2003. Cost-sensitive learning by cost-proportionate example weighting. In: Third IEEE International Conference on Data Mining. IEEE, pp. 435–442.

Zhuang, Z., Sun, Z., Cheng, Y., Yao, R., Zhang, W., 2018. Modeling and optimization of paper-making wastewater treatment based on reinforcement learning. In: 2018 37th Chinese Control Conference. CCC, IEEE, pp. 8342–8346.