Results in Engineering 20 (2023) 101428

Contents lists available at ScienceDirect

Results in Engineering

journal homepage: www.sciencedirect.com/journal/results-in-engineering

Research paper

Energy consumption prediction in water treatment plants using deep learning with data augmentation

Fouzi Harrou [a,*], Abdelkader Dairi [b], Abdelhakim Dorbane [c], Ying Sun [a]

[a] King Abdullah University of Science and Technology (KAUST), Computer, Electrical and Mathematical Sciences and Engineering (CEMSE) Division, Thuwal 23955-6900, Saudi Arabia

[b] Computer Science Department, University of Science and Technology of Oran-Mohamed Boudiaf (USTO-MB), El Mnaouar, BP 1505, Oran 31000, Algeria

[c] Smart Structures Laboratory (SSL), Department of Mechanical Engineering, Belhadj Bouchaib University of Ain Temouchent, Ain Temouchent, Algeria

ARTICLE INFO

Keywords:
Deep learning
Data augmentation
Features selection
Energy consumption
Wastewater treatment plants
Data-based methods

ABSTRACT

Wastewater treatment plants (WWTPs) are energy-intensive facilities that play a critical role in meeting stringent effluent quality regulations. Accurate prediction of energy consumption in WWTPs is essential for cost savings, process optimization, regulatory compliance, and reducing carbon footprint. This paper introduces an efficient approach for predicting energy consumption in WWTPs, leveraging deep learning models, data augmentation, and feature selection. Specifically, Spline Cubic interpolation enriches the dataset, while the Random Forest model identifies important features. The study investigates the impact of lagged data to capture temporal dependencies. Comparative analysis of five deep learning models on original and augmented datasets from Melbourne WWTP demonstrates substantial performance improvement with augmented data. Incorporating lagged energy consumption data further enhances accuracy, providing valuable insights for effective energy management. Notably, the Long Short-Term Memory (LSTM) and Bidirectional Gated Recurrent Unit (BiGRU) models achieve Mean Absolute Percentage Error (MAPE) values of 1.36% and 1.436%, outperforming state-of-the-art methods.

1. Introduction

Wastewater treatment plants (WWTPs) are vital infrastructure components responsible for treating and purifying wastewater before its discharge into the environment [1,2]. However, the operation of WWTPs entails substantial energy consumption due to the various processes involved, such as aeration, pumping, and sludge treatment [3,4]. Accurate prediction of energy consumption in WWTPs is of utmost importance as it enables cost savings, facilitates process optimization, ensures regulatory compliance, and supports sustainability goals by reducing carbon footprint.

Over the past decade, the use of machine learning techniques in optimizing the control of WWTPs and predicting their key features has gained significant attention [5–8]. One of the main motivations for utilizing machine learning methods is their flexibility, as they only require historical data for construction, unlike model-based approaches that often rely on physical knowledge and can be time-consuming to develop. By leveraging machine learning, WWTP operators can make accurate predictions and optimize the control strategies based on historical data, leading to improved energy efficiency and cost savings. A wide range of machine learning techniques have been explored and applied to address this prediction problem and provide valuable insights into energy usage in WWTPs [9–11]. For instance, Ramli and Abdul Hamid conducted a study on data-based modeling of a wastewater treatment plant using machine learning methods [12]. They compared techniques such as Artificial Neural Network (ANN), K-Nearest Neighbors (KNN), Support Vector Regression (SVR), and Linear Regression for energy consumption prediction. Energy consumption data was collected from the electrical bills of Tenaga National Berhad (TNB) over a period from March 2011 to February 2015. The study found that ANN produced the most accurate predictions compared to the other machine learning methods. Recently, in [13], Nnaji et al. used ANFIS (Adaptive Neuro-Fuzzy Inference System) and RSM (Response Surface Methodology) to predict coagulation-flocculation parameters for color and COD (Chemical Oxygen Demand) removal, as well as SVI (Sludge Volume Index) in dye wastewater treatment at 298 K. ANFIS produced higher averaged R2 values (0.9842 for color removal, 0.9752 for COD removal, and 0.9940 for SVI) and lower averaged RMSE values (0.2317

* Corresponding author.
E-mail addresses: fouzi.harrou@kaust.edu.sa (F. Harrou), abdelkader.dairi@univ-usto.dz (A. Dairi), abdelhakim.dorbane@univ-temouchent.edu.dz (A. Dorbane), ying.sun@kaust.edu.sa (Y. Sun).

https://doi.org/10.1016/j.rineng.2023.101428
Received 2 August 2023; Received in revised form 11 September 2023; Accepted 15 September 2023
Available online 26 September 2023

for color removal, 0.0069 for COD removal, and 0.1889 for SVI) compared to RSM. In [14], Zhang et al. presented an energy consumption model for WWTPs using the random forest (RF) algorithm. The study incorporated a non-numerical variable, discharge standard, and utilized 2387 data records from the China Urban Drainage Yearbook. The model aimed to predict energy consumption after upgrading discharge standards and contribute to understanding the water-energy nexus in WWTPs. The model achieved a correlation coefficient (R^2) of 0.702. However, the study did not explore the influence of local climate and treatment technologies on energy consumption. Using logistic regression, Boncescu et al. conducted a study on energy consumption in a Romanian wastewater treatment plant [15]. The study utilized data from 403 random samples collected over two years and included parameters such as flow rate, BOD (Biochemical Oxygen Demand), TSS (Total Suspended Solids), COD (Chemical Oxygen Demand), and overall energy consumption. The logistic regression model achieved accuracy ranging from 80% to 87% for different parameters compared to actual values. However, the model did not consider all the parameters influencing water quality in the wastewater treatment plant. In [16], Torregrossa et al. proposed a methodology for daily benchmarking in WWTPs to achieve energy savings. They utilized the Energy Online System (EOS) to calculate Key Performance Indicators (KPIs) considering pollutant load, enabling comparisons between different plants. The study evaluated the performance of SVR, ANN, and RF algorithms using data from the Solingen-Burg WWTP, serving a population of 120,000. Among the three algorithms, RF demonstrated the highest efficiency, with an R^2 value of 0.72 in validation and 0.71 in testing. However, the study had limitations in data availability, which may affect the generalizability of the findings. In another study, Torregrossa et al. presented a machine-learning approach for energy cost modeling in WWTPs [17]. Specifically, they focused on a database of 279 WWTPs in northwest Europe and employed NN and RF algorithms to develop cost functions and assess their efficiency. This methodology demonstrated improved performance compared to traditional techniques, establishing it as a valuable tool for energy cost modeling. However, the study acknowledged this methodology's complexity and time-consuming nature and the requirement for a significant amount of data for validation. Qiao and Zhou proposed a density peaks-based adaptive fuzzy neural network (DP-AFNN) for modeling energy consumption (EC) and effluent quality (EQ) in WWTPs [18]. They used the improved Levenberg-Marquardt method to optimize the parameters of the fuzzy neural network, and results showed that the DP-AFNN achieved better prediction accuracy, outperforming existing methods for modeling EC and EQ in WWTPs. Oulebsir et al. [19] proposed a methodology for optimizing EC in WWTPs using an ANN. The model achieved a coefficient of determination ranging from 90% to 92% during training and 74% to 82% during testing, demonstrating its effectiveness in predicting daily EC and offering potential for improved energy efficiency in WWTP operations. Adibimanesh et al. recently proposed an intelligent tool for optimizing EC in WWTPs, specifically focusing on sewage sludge incineration [20]. They used machine learning algorithms, including ANN, Parallel models (GBM, RF, and SVM), and Chained models. In the parallel model, the multi-output problem was divided into several single-output models, and the accuracy of GBM, RF, and SVM algorithms was tested to select the most accurate model. Simulation results from ASPEN PLUS, based on real data from a wastewater treatment plant in Gdynia (Poland), were used for validation. The ML models achieved R^2 values of 0.85, 0.91, and 0.94 for Parallel models, Chained models, and ANN, respectively, demonstrating the accuracy of the proposed tool.

In recent years, deep learning models have emerged as powerful tools for modeling complex relationships and capturing temporal dynamics in various domains. These models have shown promising results in time-series prediction tasks. Therefore, they hold significant potential for accurate energy consumption prediction in WWTPs. Bagherzadeh et al. conducted a study on predicting EC in the East Melbourne WWTP using machine learning techniques [21]. They examined the impact of wastewater, hydraulic, and climate-based parameters on daily EC by analyzing data collected over six years (2014-2019). The study employed various machine learning models for EC prediction, including ANN, RF, Gradient Boosting Machine (GBM), and Recurrent Neural Network (RNN), where the GBM demonstrates the best performance. In [22], Das et al. conducted a study to predict EC in WWTPs using four deep learning models, namely Artificial Neural Network (ANN), Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU). The results indicated that the GRU model outperformed the other models. In [23], a transfer learning approach is adopted to forecast EC in WWTPs based on deep learning models, including Long LSTMs, GRUs, and uni-dimensional Convolutional Neural Networks (CNNs). The CNN-based models demonstrated superior performance, outperforming LSTMs and GRUs, with the best model achieving an MAE of 630 kWh and an RMSE of 690 kWh. Yusuf et al. [24] compared the performance of Auto-Regressive Integrated Moving Average (ARIMA) and LSTM for electric load forecasting in water utility sites. LSTM showed significant improvements over ARIMA in terms of RMSE, with reductions of 75.54% for WWTPs and 17.33% for pumping stations. Similarly, LSTM achieved substantial decreases in MAPE, with reductions of 78.46% for water treatment plants and 9.9% for pumping stations, compared to ARIMA.

The limited availability of energy consumption time-series data in WWTPs poses significant challenges for accurate predictions. Traditional machine learning and deep learning models require substantial training data to achieve optimal performance. However, the scarcity of long-term records and the sparse nature of daily data can lead to overfitting or poor generalization of the models, especially deep learning models known for capturing complex temporal dependencies. Data augmentation techniques offer a promising solution to enhance the learning capabilities of these models. While data augmentation has shown effectiveness in fields like computer vision [25,26], its application in time-series data tasks, such as time-series classification, has been relatively less common [27,28]. Recent research has developed data augmentation methods tailored to time-series data, which have shown promise in various domains, including medical data, pollution prediction, and wearable sensor data [29–31], and these methods have shown promising results in various fields, including medical data [32], pollution prediction [27], and wearable sensor data [33].

In the context of WWTPs, the utilization of deep learning for time-series prediction faces the common challenge of limited large time-series datasets [34]. Data scarcity hampers the use of machine learning in designing, optimizing, and controlling WWTPs [35]. Collecting experimental data is costly and challenging, and the dynamic nature of wastewater treatment processes compounds these issues. The resulting insufficient training data can hinder the performance of deep learning models, which rely on large datasets for improved generalization [34]. Addressing data scarcity is crucial for enhancing the accuracy and reliability of time-series prediction models for WWTPs. This paper makes a threefold contribution. Firstly, it explores the application of various deep learning models, including RNN, LSTM, GRU, BiLSTM, and BiGRU, for energy consumption prediction in WWTPs. Through a comprehensive evaluation, we provide insights into their effectiveness in capturing temporal dependencies and improving prediction accuracy. Secondly, the paper proposes an efficient approach to predict energy consumption in WWTPs by employing data augmentation, specifically using cubic spline interpolation, to address the limited size of WWTP time-series datasets. By generating synthetic data points between observed values, the dataset is expanded, temporal resolution is enhanced, and the generalization capabilities of the predictive models are improved. The approach's efficacy is demonstrated by evaluating the performance of the deep learning models with and without data augmentation. Lastly, the paper investigates the impact of incorporating lagged energy consumption data into the prediction models. By considering historical energy consumption information, it aims to capture the temporal dynamics and dependencies influencing current energy consumption patterns. This

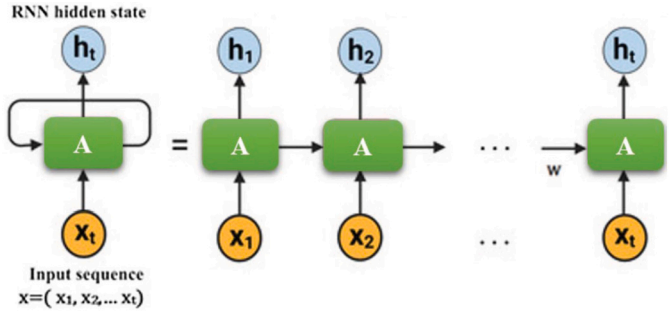

Fig. 1. A schematic illustration of an unrolled RNN architecture. This unrolling visually elucidates the temporal dynamics and connections within the network, enabling a more comprehensive understanding of its inner workings.

analysis provides valuable insights into the benefits of incorporating lagged data for improved prediction performance. These contributions are validated using real-world WWTP data from the Melbourne WWTP, containing multivariate variables collected over a five-year period.

The paper follows a structured approach to investigate using deep learning models with data augmentation for energy consumption prediction in WWTPs. Section 2 briefly introduces the deep learning models employed in the study and explains the data augmentation technique using cubic spline methods. Section 3 presents the proposed predictive framework that integrates deep learning models with data augmentation to enhance the accuracy of energy consumption predictions. Section 4 is dedicated to presenting the datasets used in the study and discussing the results obtained from the proposed framework. A comparison with state-of-the-art methods on the same dataset is also provided. Finally, in Section 5, the paper summarizes the findings and outlines future research directions in this domain.

2. Material and methods

2.1. RNN model

While regular feedforward neural networks have proven successful in various domains, they have limitations when modeling sequential data and capturing time dependencies. In these networks, data flow transformations occur solely in one direction through hidden layers, with the output influenced solely by the current input. This lack of memory hinders their ability to effectively model data sequencing and capture temporal dependencies within historical data.

RNN have proven to be effective models for capturing temporal dependencies in sequential data, making them well-suited for energy consumption prediction in WWTPs. RNNs process input sequences step-by-step while maintaining a hidden state that retains information from previous steps. This enables the model to leverage historical context and capture long-term dependencies in energy consumption patterns.

The basic concept of an RNN revolves around considering the influence of past information when generating the output [36]. Unlike traditional feedforward neural networks, which process inputs independently, RNNs introduce the concept of cells that retain the memory of previous inputs and utilize this information to influence the output. In a basic RNN model, the hidden state at each time step is updated based on the current input and the previous hidden state (Fig. 1). This allows the network to capture temporal dependencies and model sequential data effectively. The update equation for a basic RNN can be expressed as:

$$h_t = \sigma\big(W[h_{t-1}, x_t] + b\big). \tag{1}$$

where x_t represents the input at time step t, h_t denotes the hidden state at time step t, W represents the weight matrix, and b represents the bias vector. The σ function is usually a non-linear activation function such as the sigmoid or hyperbolic tangent function.

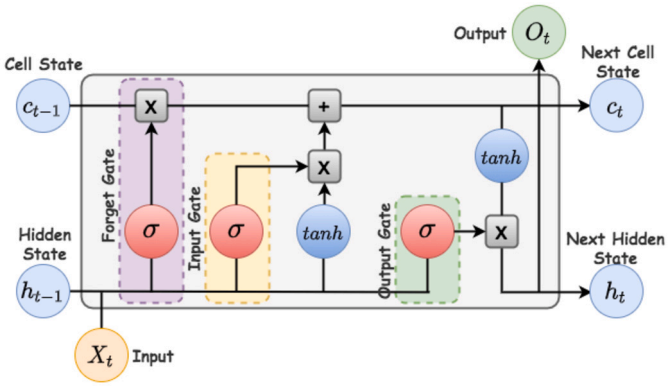

Fig. 2. Structure of an LSTM unit, highlighting its intricate architecture for capturing long-range dependencies in sequential data. It showcases the three LSTM gates (input, forget, and output), along with candidate memory cells and their content.

In the context of energy consumption prediction in WWTPs, RNNs can leverage their ability to retain historical information to capture patterns and trends in energy consumption data over time. The recurrent nature of RNNs allows them to model the sequential nature of the energy consumption data, considering both short-term fluctuations and long-term dependencies. By learning from past inputs, the RNN can capture the temporal dynamics inherent in the data, enabling accurate prediction of energy consumption in WWTPs. However, one limitation of basic RNNs is that they may suffer from the vanishing or exploding gradient problem, where the gradients used for updating the weights either become extremely small or very large, hindering effective learning over long sequences. This limitation led to more advanced RNN architectures, such as LSTM and GRU, which mitigate the gradient problem and capture long-term dependencies more effectively.

2.2. LSTM model

LSTM is a specialized type of RNN that addresses the challenge of capturing long-term dependencies in sequential data. The LSTM architecture includes memory cells and gating mechanisms that enable the model to selectively retain or forget information from previous time steps, facilitating the modeling of complex temporal dynamics. Indeed, LSTM has garnered considerable attention as a powerful tool for time-series forecasting [37–39]. For example, Ruma et al. [40] effectively employed an optimized LSTM model to predict water levels in the Bangladesh river network. Likewise, Ahmed et al. [41] harnessed deep learning time series forecasting networks and LSTM to forecast municipal solid waste (MSW) generation. Furthermore, Singh et al. [42] showed LSTM's good forecasting capabilities, particularly in wind power prediction.

In an LSTM, each memory cell is responsible for storing and propagating information across time steps. The key components of an LSTM cell are the input gate, forget gate, output gate, and memory cell. These gates and the memory cell work in concert to control the flow of information within the LSTM and determine which information should be remembered, forgotten, or outputted (see Fig. 2).

The equations governing the LSTM cell operations are as follows:

$$\begin{cases} i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \\ f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \\ o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \\ \tilde{C}t = \tanh(W_c \cdot [ht-1, x_t] + b_c), \\ C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t, \\ h_t = o_t \odot \tanh(C_t), \end{cases} \tag{2}$$

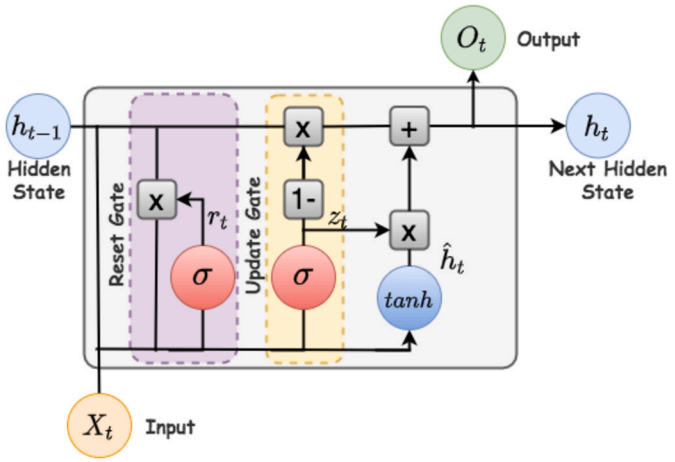

Fig. 3. Internal structure of a GRU highlighting its architectural design for capturing sequential dependencies. It illustrates the key components, including the update and reset gates, as well as the candidate hidden state.

The input gate (i_t) controls how much information from the current input should be stored in the memory cell. The forget gate (f_t) determines the amount of information from the previous cell state (C_{t-1}) that should be discarded. The output gate (o_t) regulates the amount of information from the current cell state (C_t) to be outputted. The candidate cell state ($\tilde{C}_t$) is a potential update to the current cell state, computed based on the current input and previous hidden state. The final cell state (C_t) is a combination of the previous cell state and the candidate cell state, weighted by the forget and input gates. Finally, the hidden state (h_t) is obtained by applying the output gate to the cell state after applying a hyperbolic tangent function.

Using memory cells and gating mechanisms, LSTMs can effectively capture and retain relevant information over longer sequences, overcoming the vanishing or exploding gradient problem often encountered in basic RNNs. This makes LSTMs particularly suitable for modeling complex temporal dependencies and capturing long-term patterns in sequential data, including energy consumption patterns in WWTPs.

2.3. GRU model

GRU is another variant of RNN that addresses the challenges of capturing and modeling long-term dependencies in sequential data [43]. The GRU architecture simplifies the structure of the LSTM while maintaining comparable performance, making it computationally efficient. In a GRU, each unit has two key components: the update gate and the reset gate. These gates determine how much information from the previous time step should be passed on and how much new information should be incorporated. The update gate controls the balance between retaining the previous hidden state and incorporating new information, while the reset gate influences the extent to which the past hidden state affects the current state.

Overall, the GRU model simplifies the LSTM architecture by combining the forget and input gates into a single gate. The overall structure of the GRU model is illustrated in Fig. 3. The equations governing the operations of a GRU unit are as follows:

$$\begin{cases} z_t = \sigma(W_z \cdot [h_{t-1}, x_t] + b_z), \\ r_t = \sigma(W_r \cdot [h_{t-1}, x_t] + b_r), \\ \tilde{h}t = \tanh(W_h \cdot [r_t \odot ht - 1, x_t] + b_h), \\ h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t, \end{cases} \tag{3}$$

Here, x_t represents the input at time step t, h_t denotes the hidden state at time step t, and W and b denote the weight matrices and bias vectors, respectively. The σ function represents the sigmoid activation function, and $\odot$ denotes element-wise multiplication.

The update gate (z_t) determines how much of the previous hidden state (h_{t-1}) should be retained, while the reset gate (r_t) controls the extent to which the past hidden state influences the current state. The candidate hidden state ($\tilde{h}_t$) represents the potential update to the current hidden state and is computed based on the reset gate, the previous hidden state, and the current input. Finally, the new hidden state (h_t) is a combination of the previous and candidate hidden states, weighted by the update gate. By leveraging the update and reset gates, the GRU model can effectively control the flow of information, selectively retaining relevant past information and incorporating new information as necessary. This allows the model to capture long-term dependencies while being computationally efficient.

2.4. BiLSTM model

BiLSTM is a variant of RNN that enhances the modeling capabilities of traditional LSTM by incorporating information from both past and future time steps [44]. This bidirectional processing allows the model to capture dependencies in both directions and effectively model temporal dynamics in sequential data [45]. In a BiLSTM, the input sequence is processed in two separate directions: one from the past to the future (forward direction) and the other from the future to the past (backward direction). This is achieved by having two sets of LSTM units, with one set processing the input sequence in the original order and the other set processing it in reverse. The forward LSTM units capture the information from the past to the current time step, while the backward LSTM units capture the information from the future to the current time step. The outputs of the forward and backward units are then combined to obtain the final representation for each time step. The equations governing the operations of a BiLSTM unit are presented in Equations (4) and (5).

Forward LSTM:

$$\begin{cases} i_t^f = \sigma(W_i^f \cdot [h_{t-1}^f, x_t] + b_i^f), \\ f_t^f = \sigma(W_f^f \cdot [h_{t-1}^f, x_t] + b_f^f), \\ o_t^f = \sigma(W_o^f \cdot [h_{t-1}^f, x_t] + b_o^f), \\ \tilde{C}t^f = \tanh(W_c^f \cdot [ht - 1^f, x_t] + b_c^f), \\ C_t^f = f_t^f \odot C_{t-1}^f + i_t^f \odot \tilde{C}_t^f, \\ h_t^f = o_t^f \odot \tanh(C_t^f), \end{cases} \tag{4}$$

Backward LSTM:

$$\begin{cases} i_t^b = \sigma(W_i^b \cdot [h_{t+1}^b, x_t] + b_i^b), \\ f_t^b = \sigma(W_f^b \cdot [h_{t+1}^b, x_t] + b_f^b), \\ o_t^b = \sigma(W_o^b \cdot [h_{t+1}^b, x_t] + b_o^b), \\ \tilde{C}t^b = \tanh(W_c^b \cdot [ht + 1^b, x_t] + b_c^b), \\ C_t^b = f_t^b \odot C_{t+1}^b + i_t^b \odot \tilde{C}_t^b, \\ h_t^b = o_t^b \odot \tanh(C_t^b), \end{cases} \tag{5}$$

Here, x_t represents the input at time step t, h_t^f and h_t^b denote the hidden states at time step t for the forward and backward units, respectively. C_t^f and C_t^b represent the cell states at time step t for the forward and backward units, respectively. The W and b terms denote the weight matrices and bias vectors, respectively. The σ function represents the sigmoid activation function, and $\odot$ denotes element-wise multiplication.

The final representation for each time step is obtained by concatenating the forward and backward hidden states: $h_t = [h_t^f, h_t^f]$. By incorporating information from both past and future time steps, the BiLSTM model can effectively capture contextual information from the entire

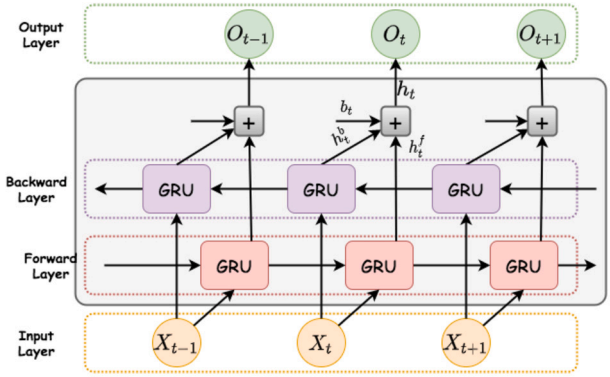

Fig. 4. Architectural diagram of the BiGRU model, illustrating its dual recurrent pathways for capturing both past and future sequential dependencies, enhancing its predictive capabilities.

sequence and capture dependencies that are not apparent when processing the data in a single direction. This makes the BiLSTM model well-suited for tasks involving sequential data, such as energy consumption prediction in WWTPs, where both past and future observations can impact the energy consumption patterns.

2.5. BiGRU model

Similar to BiLSTM, BiGRU processes the input sequence in both forward and backward directions simultaneously, allowing it to capture dependencies in both directions and model complex temporal dynamics [46,47]. In a BiGRU, the input sequence is split into two streams: one processed in the original order (forward stream) and the other in reverse order (backward stream). Each stream consists of GRU units, which perform the gating operations to control the flow of information (Fig. 4). The forward GRU units capture information from past to current time steps, while the backward GRU units capture information from future to current time steps. The outputs from both directions are combined to form the final representation for each time step.

The computation of the forward hidden state at time step t, referred to as h_t^f, and the backward hidden state at time step t, denoted as h_t^b, can be described by the following equations:

$$\begin{cases} h_t^f = GRU(X_t, h_{t-1}^f) \\ h_t^b = GRU(X_t, h_{t-1}^b) \end{cases} \tag{6}$$

Then, the final representation, h_t, for each time step is obtained by concatenating the hidden states from both directions:

$$h^t = w_t^f * h_t^f + w_t^b * h_t^b + b_t, \tag{7}$$

where the function $GRU(.)$ denotes the utilization of the GRU model to apply a nonlinear transformation to the input data. The weights w_t^f and w_t^b correspond to the parameters associated with the forward hidden state h_t^f and the backward hidden state h_t^b, respectively. Furthermore, the bias term b_t represents the additional parameter associated with the hidden layer state. These weights and bias terms play a vital role in the computation of the forward and backward hidden states within the GRU model.

Table 1 summarizes the key concepts, advantages, and limitations associated with the considered RNN models.

2.6. Data augmentation using interpolatiom

Time-series data augmentation plays a vital role in addressing the challenges posed by limited size datasets [48–50], particularly in WWTP time-series data. With the increasing demand for accurate predictions and optimization in WWTP operations, extensive and diverse data availability becomes crucial. However, acquiring large-scale WWTP time-series datasets are often constrained by practical limitations, resulting in small-sized datasets, especially when considering daily records. To overcome these limitations, data augmentation techniques provide a promising solution. Among various augmentation methods, interpolation techniques have shown effectiveness in generating synthetic data points that closely resemble the original time series. By leveraging interpolation methods, we can artificially expand the dataset and capture a broader range of temporal patterns and variations within the limited data.

Cubic spline interpolation is a mathematical technique used to estimate missing values and generate synthetic data points by constructing smooth and continuous curves between existing observations. It can be applied to each variable in the WWTP dataset, not limited to energy consumption alone. By applying cubic spline interpolation to the different variables, we can effectively increase the temporal resolution and capture finer details of the entire WWTP system, including biological, hydraulic, and climate factors. The interpolation process involves fitting a cubic polynomial between adjacent data points for each variable. Let's consider a specific variable with observed data points at time steps $t_1, t_2, \ldots, t_n$, denoted by $y_1, y_2, \ldots, y_n$. The goal is to estimate the variable's values at additional time steps within this range. The cubic spline interpolation function for estimating the value of the variable at a given time step t can be defined as:

$$y(t) = a_i + b_i(t - t_i) + c_i(t - t_i)^2 + d_i(t - t_i)^3, \tag{8}$$

where a_i, b_i, c_i, and d_i are coefficients specific to each interval $[t_i, t_{i+1}]$. These coefficients are determined by satisfying certain continuity and smoothness conditions at the data points.

To generate synthetic data points, we select a set of additional time steps within the range of the original time series, denoted as t_{aug}. Using the cubic spline interpolation function, we estimate the variable's values at these time steps by substituting the corresponding values of t into the equation. This process allows us to create a larger dataset with augmented data points that fill the gaps between the observed values for each WWTP variable.

3. Deep learning with data augmentation: a strategy for improved energy consumption prediction

Predicting energy consumption (EC) in WWTPs is crucial for designing and operating sustainable and energy-efficient facilities. However, the complexity of EC, influenced by various biological and environmental factors, poses challenges in developing accurate soft sensors. To address this, we propose an integrated approach that combines data augmentation using cubic spline interpolation with deep learning models for EC prediction in WWTPs. The adopted deep learning framework for energy consumption prediction in WWTPs used in this study is illustrated in Fig. 5.

The first step of our approach involves augmenting the dataset through cubic spline interpolation. This augmented data, which fills the gaps between observed values, provides a denser representation of the WWTP variables, capturing finer temporal details and variations. By merging the augmented data with the original dataset, we create a larger and more comprehensive training set. This augmented dataset offers a richer representation of the temporal dynamics and patterns in the WWTP variables.

Then, to ensure a comprehensive evaluation of the deep learning models trained on the augmented data, the augmented dataset is divided into training and testing sets. Specifically, 75% of the augmented data is allocated for training the deep learning algorithms. This large portion of the data enables the models to learn from a diverse range of temporal patterns and dependencies present in the WWTP variables.

We then employ deep learning models, including RNN, LSTM, GRU, BiLSTM, and BiGRU, designed to capture the temporal dependencies and patterns present in the data. These models are trained using gradient-based optimization algorithms, such as Adam, to minimize the

Table 1
Comprehensive Overview of RNN Models: A summary highlighting the key concepts, advantages, and limitations of RNN, LSTM, GRU, BiLSTM, and BiGRU models.

Model	Key Concept	Advantages	Limitations
RNN	Recurrent connections allow information flow over time	Simplicity, low computational cost	Difficulty capturing long-term dependencies (vanishing/exploding gradients)
LSTM	Memory cells and gating mechanisms handle long-term memory	Captures long-term dependencies, mitigates gradient problem	Higher computational complexity than Simple RNN
GRU	Combines forget and input gates, simplifying LSTM structure	Efficient, comparable performance to LSTM	May not capture long-term dependencies as effectively
BiLSTM	Processes input in both forward and backward directions	Captures dependencies in both directions, enhanced modeling	Higher computational complexity than unidirectional models
BiGRU	Processes input bidirectionally with simplified GRU structure	Efficient, captures dependencies in both directions	May have limitations in capturing very long-term dependencies

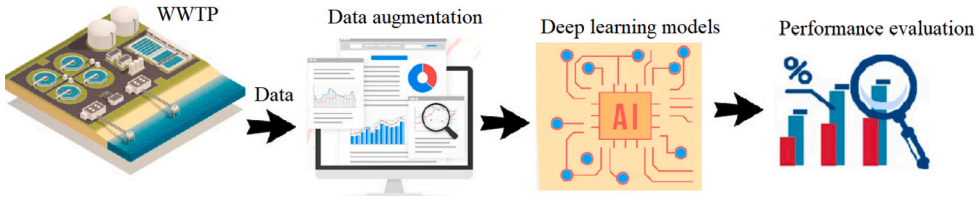

Fig. 5. Deep Learning with Data Augmentation for Energy Consumption Prediction: An overview of the main steps, including data collection, augmentation, development of deep learning models, energy consumption prediction, and performance evaluation using statistical metrics.

loss function between the predicted and actual target values. The training process iteratively updates the model parameters to improve the models' predictive performance. During the training phase, the models are optimized using the Adam optimization algorithm. Adam is an adaptive learning rate optimization algorithm that adjusts the learning rate for each parameter individually based on their past gradients. This adaptive mechanism facilitates efficient convergence and improved training performance of the deep learning models. The training process involves iteratively updating the model parameters by minimizing the loss function between the predicted and actual target values. This optimization aims to improve the models' predictive performance and their ability to accurately predict energy consumption in WWTPs. The models learn from the augmented training data, adjusting their internal representations and parameters to capture the temporal dependencies and patterns within the variables.

Following the training phase, the trained models are evaluated using the testing data, which constitutes the remaining 25% of the augmented dataset. The testing data consists of unseen data points that were not utilized during the training process. By assessing the models' performance on this independent dataset, we gain insights into their ability to generalize to unseen data and make accurate predictions. During the evaluation phase, the models take the testing data as input and generate predictions for the target variable, which is the energy consumption in this context. The predictions are compared against the actual energy consumption values in the testing data to assess the models' accuracy and performance. In this study, evaluation metrics, namely Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), R-squared (R2): R2, also known as the coefficient of determination, Mean Squared Logarithmic Error (MSLE), and Mean Absolute Percentage Error (MAPE) are computed to quantify the models' predictive capabilities and assess their effectiveness.

- RMSE is a widely used metric that measures the average deviation between the predicted values and the actual values. It provides an overall indication of the model's accuracy, with lower values indicating better performance.

$$\mathrm{RMSE}(y,\hat{y}) = \sqrt{\frac{1}{n}\sum_{i=0}^{n-1}(y_i-\hat{y}_i)^2}, \tag{9}$$

where y_t represents the actual or observed value, $\hat{y}_t$ represents the actual or observed value, and n represents the total number of data points.

- MAE is another commonly used metric that quantifies the average absolute difference between the predicted and actual values. It provides a measure of the average prediction error without considering the direction of the errors.

$$\mathrm{MAE}(y,\hat{y}) = \frac{1}{n}\sum_{i=0}^{n-1}\|y_i-\hat{y}_i\|_1, \tag{10}$$

- MAPE calculates the average percentage difference between the predicted and actual values. It is a relative measure that provides insights into the magnitude of errors compared to the actual values.

$$\mathrm{MAPE}(y,\hat{y}) = \frac{1}{n}\sum_{i=0}^{n-1}\frac{\|y_i-\hat{y}_i\|_1}{\max(\epsilon,\|y_i\|_1)}, \tag{11}$$

- R2 assesses the proportion of the variance in the dependent variable that the independent variables can explain. It indicates how well the model fits the data, with higher values indicating better fit.

$$R^2 = 1-\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}. \tag{12}$$

- MSLE is a metric that measures the average logarithmic difference between the predicted and actual values. It is particularly useful when dealing with skewed or exponential data distributions.

$$MSLE = \frac{1}{n}\sum_{i=1}^{n}(\log(1+y_i)-\log(1+\hat{y}_i))^2. \tag{13}$$

4. Results and discussion

4.1. Data description and analysis

This study capitalizes on using multivariate data sourced from the Melbourne water treatment plant and airport weather stations[1] to ad-

[1] https://data.mendeley.com/datasets/pprkvz3vbd/1.

Table 2
Abbreviations and Units of Analyzed Variables: This table provides the abbreviations and units for the variables analyzed in the study, encompassing hydraulic, wastewater, climate, and temporal parameters.

	Parameters (Abbreviation)	Unit
	Energy Consumption (EC)	MWh
Hydraulic	Average Inflow (Q_{in})	m^3/s
	Average Outflow (Q_{out})	m^3/s
Wastewater	Ammonia (NH4-N)	mg/L
	Biological Oxygen Demand (BOD)	mg/L
	Chemical Oxygen Demand (COD)	mg/L
	Total Nitrogen (TN)	mg/L
Climate	Average Temperature (Tavg)	°C
	Maximum temperature (Tmax)	°C
	Minimum temperature (Tmin)	°C
	Atmospheric pressure (AP)	(hPa)
	Average humidity (H)	%
	Total rainfall and / or snowmelt (Pr)	mm
	Average visibility (VIS)	Km
	Average wind speed (WSavg)	Km/h
	Maximum wind speed (WSmax)	Km/h
Time	Year (year)	-
	Month (month)	-
	Day (day)	-

vance energy consumption prediction methodologies. Spanning a timeframe of five years, from January 2014 to June 2019, the dataset encompasses a diverse array of nineteen variables as detailed in Table 2. These variables encompass crucial aspects such as power consumption, biological characteristics, hydraulic factors, and climate variables. Water quality and biological data are meticulously collected through sensor-based measurements, while the weather data originates from the conveniently located Melbourne airport weather station adjacent to the water treatment plant. By integrating this comprehensive dataset, researchers gain a holistic perspective of the intricate interrelationships that influence energy consumption at a wastewater treatment plant. Additionally, the dataset offers valuable time-domain information that is considered in this study to enhance prediction performance. Detailed information about the dataset can be found in the comprehensive work by Bagherzadeh et al. [21]. To ensure data reliability, a careful data cleaning process was undertaken to eliminate outliers. Data points exhibiting unusually high or low energy consumption were identified as outliers and subsequently removed from the dataset, resulting in the removal of approximately 5% of the data points [51].

Fig. 6 depicts Violin plots of the time-series data used in the WWTPs, revealing their non-Gaussian distribution. These plots offer more insights than traditional boxplots and histograms by combining the benefits of both visualization methods [52]. They provide summary statistics and depict data density across the entire range of the variable, offering a comprehensive understanding of the distribution. The width of the violin plot at each data point represents the density or frequency, facilitating comparisons between datasets. This feature is especially useful for comparing different variables or groups within the dataset. Additionally, violin plots excel at revealing asymmetries or multimodalities in the data, which might not be apparent in boxplots or histograms.

Fig. 7 displays the RadViz visualization, showcasing the impact of various factors on power consumption in the WWTP. RadViz is a data visualization technique that helps understand multiple variables' impact on a target variable. It works by projecting high-dimensional data onto a 2D plane while maintaining the relationships between the variables [53]. In a RadViz plot, each variable is represented by a point on a circle, and the target variable (in this case, power consumption in the WWTP) is represented at the circle's center. The position of each point on the circle is determined by the value of the corresponding variable, and the distance of the point from the center represents the contribution of that variable to the target variable.

From the Radviz plot in Fig. 7, it is evident that certain variables significantly impact the energy consumption of the wastewater treatment plant (WWTP). Notably, the inflow and outflow rates are positioned close to the corresponding dimensional anchors, indicating their strong influence on energy consumption. This suggests that the volume of wastewater entering and leaving the plant plays a crucial role in determining the energy requirements.

Furthermore, variables related to time, such as the day, also exhibit a noticeable effect on energy consumption. This could imply that specific days or time periods may lead to higher energy demands due to fluctuations in wastewater flow or operational requirements. Temperature variables, including average temperature and minimum and maximum temperatures, are also positioned close to their respective anchors. This indicates that temperature variations can significantly impact the energy consumption of the WWTP, as certain treatment processes might be more energy-intensive under different temperature conditions. In addition to the aforementioned factors, chemical composition variables, such as Total Nitrogen (TN), Chemical Oxygen Demand (COD), and Biological Oxygen Demand (BOD), appear to have a substantial influence on energy consumption. This suggests that the composition of the wastewater, particularly in terms of these pollutants, can affect the energy requirements for treatment processes. By analyzing the Radviz plot, we gain valuable insights into the relative importance of different variables on energy consumption in the WWTP. These observations can guide decision-making and resource allocation strategies to optimize energy usage and improve the overall efficiency of wastewater treatment operations. Researchers and stakeholders can utilize this information to focus on specific variables and devise targeted measures to reduce energy consumption, enhance sustainability, and achieve cost-effectiveness in the WWTP's energy management.

Fig. 8 illustrates the yearly distribution of key variables, including Energy Consumption, Average Inflow, and Average Outflow, throughout the study period from 2014 to 2018. The violin plots offer a comprehensive view of the data distribution, highlighting the variability and patterns in these variables over time. By examining these plots, we gain valuable insights into the seasonal trends and long-term changes in energy consumption and water flow rates within WWTPs. Upon analyzing the plots, a notable trend emerges, revealing a slight decrease in the annual distribution of energy consumption in 2018 compared to 2017. This reduction is evident in both the average values and standard deviations, suggesting a potential improvement in the operational efficiency or management practices of the WWTP during that period. The observed decrease in energy consumption aligns with the increasing emphasis on sustainable practices and energy-efficient technologies in wastewater treatment facilities. Furthermore, the violin plots provide evidence of a strong correlation between Energy Consumption, Average Inflow, and Average Outflow. This suggests that the inflow rate plays a significant role in influencing the energy consumption of the WWTP. In 2018, a noticeable decline in the inflow rate is observed compared to the previous year (2017). This decrease in the inflow rate could be attributed to factors such as changes in weather patterns, fluctuations in population or industrial activities, or alterations in water usage behavior. The reduction in the inflow rate likely contributed to the observed decrease in power consumption, as the WWTP would have had to process a lower volume of wastewater, thus requiring less energy for treatment. Overall, the presented violin plots offer a comprehensive overview of the energy consumption and water flow dynamics in the WWTP over the studied years.

The violin plots in Fig. 9 provide a comprehensive visual representation of the monthly energy consumption patterns throughout the study period. One striking observation is the substantial increase in variance during the hot months of October, November, and December. This heightened variability underscores the importance of considering

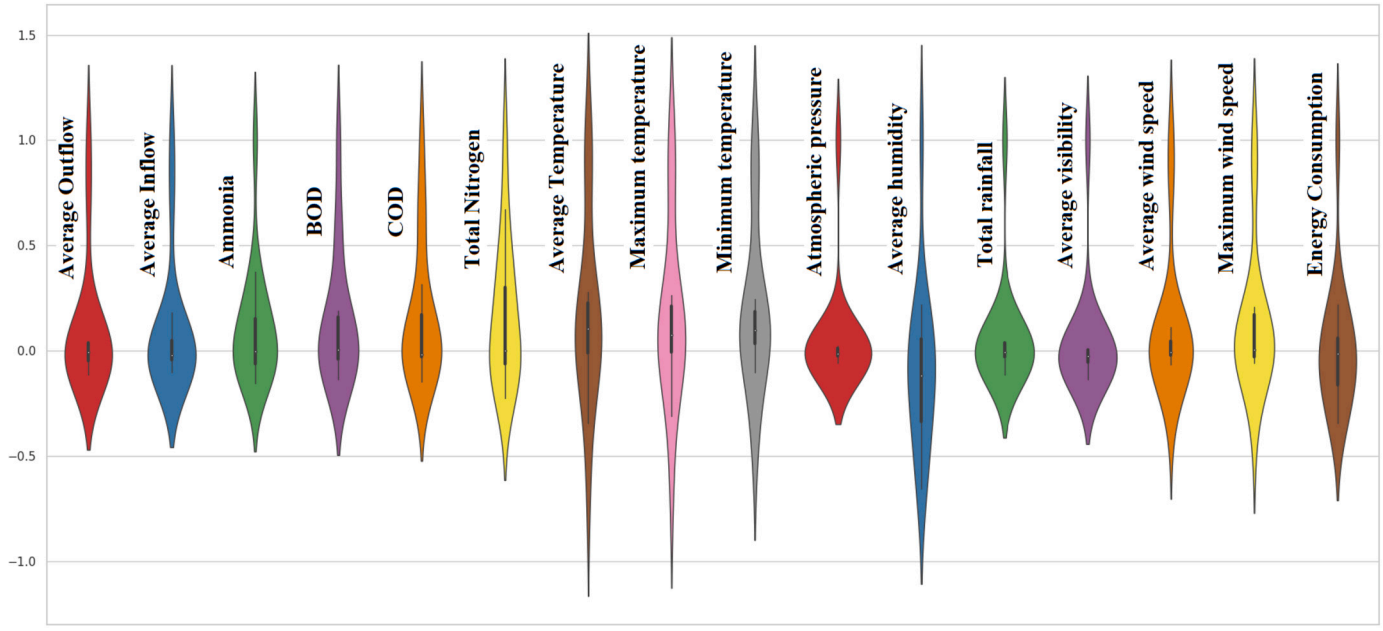

Fig. 6. Exploratory Violin Plots: Violin plots visually represent the distribution and spread of the time-series data used in this study.

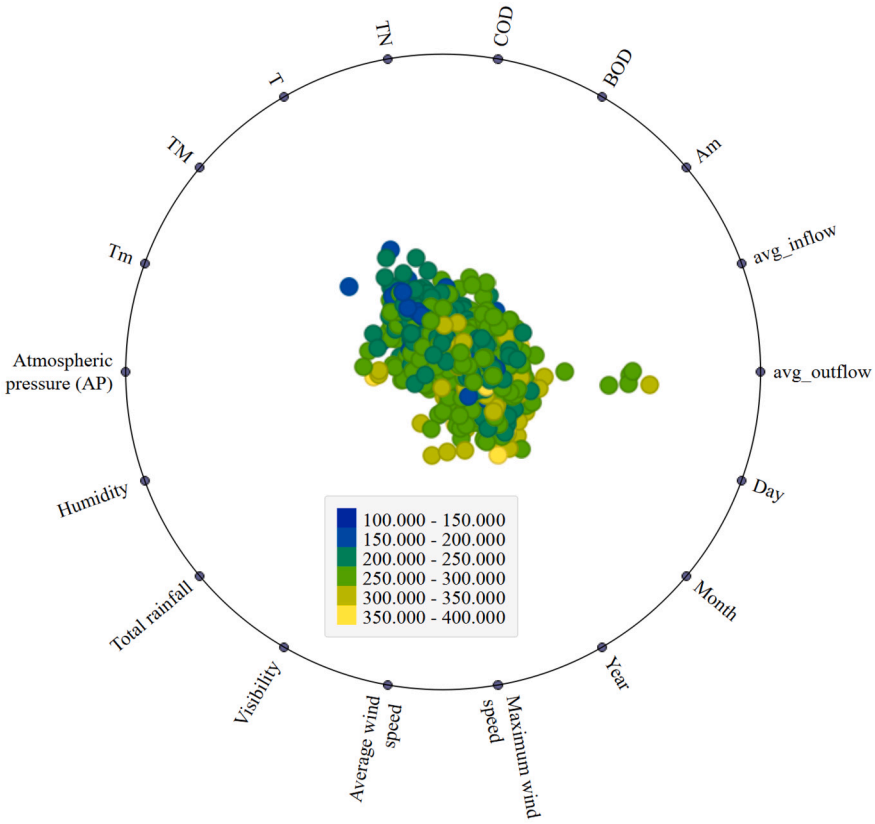

Fig. 7. RadViz Visualization of Factors Impacting WWTP Power Consumption: This visualization, based on RadViz, offers insights into how various factors influence power consumption within the WWTP.

external factors, such as weather conditions and tourism, when developing accurate energy consumption prediction models. The rise in energy consumption during the hot months can be attributed to the surge in water usage for cooling and recreational activities. As temperatures increase, the demand for water for various cooling purposes, such as air conditioning and irrigation, also rises. This heightened water demand translates into increased energy requirements for pumping, aeration, and other processes involved in treating and managing wastewater. The connection between higher temperatures and increased energy consumption emphasizes the need for WWTPs to be well-prepared for such seasonal variations to ensure continuous and reliable operation. Furthermore, the influx of tourists during the holiday season can exert additional pressure on local water and wastewater infrastructure. Tourist destinations experience a substantial rise in visitors during this

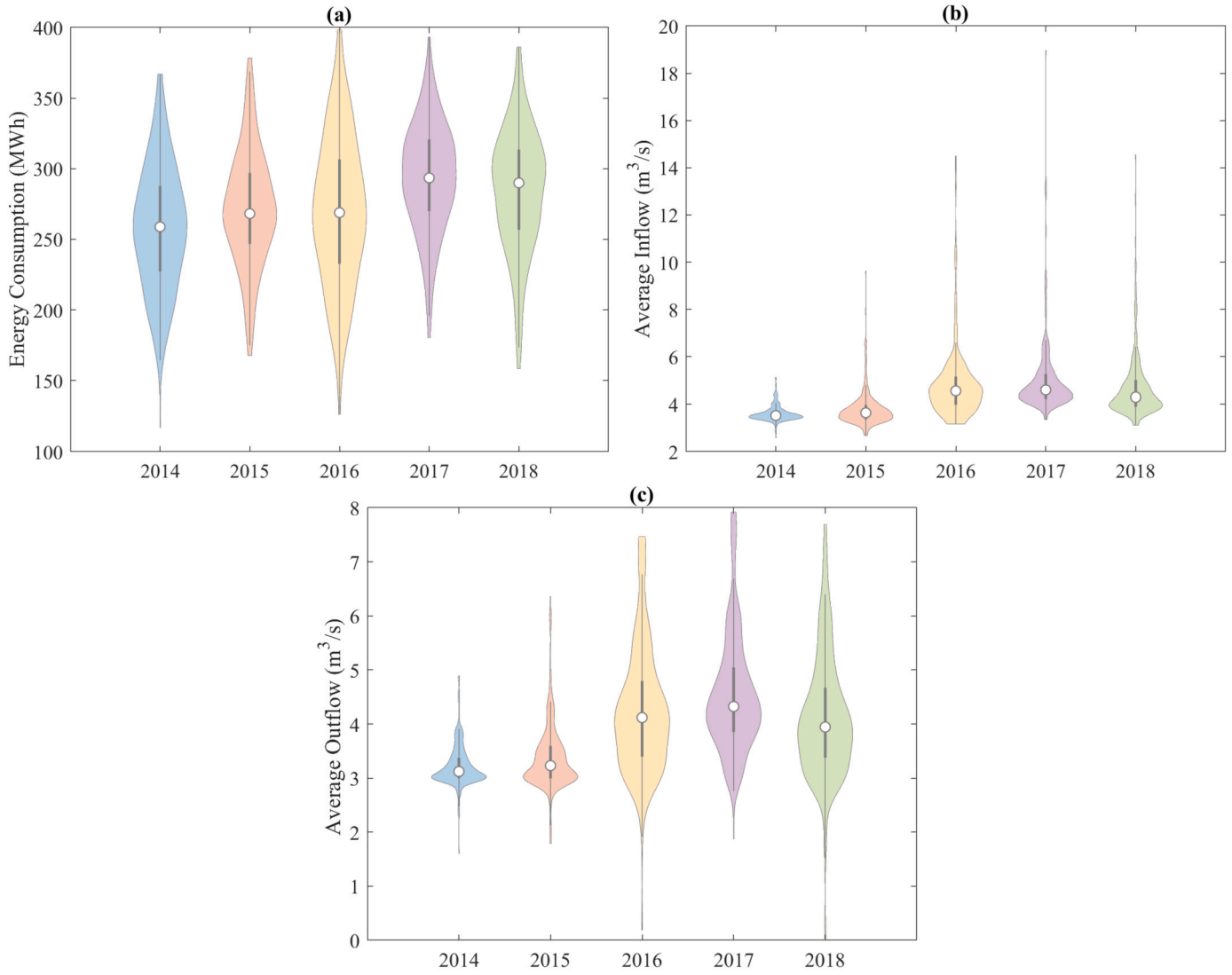

Fig. 8. Annual distribution of key variables in the monitored WWTP over the study period from 2014 to 2018, visualized using violin plots: (a) Energy Consumption, (b) Average Inflow, and (c) Average Outflow.

time, leading to a spike in water usage and wastewater generation. As a result, WWTPs may need to ramp up their capacity to cope with the augmented load, further contributing to higher energy consumption. The insights gained from the monthly energy consumption patterns can have practical implications for the management and optimization of WWTPs. By understanding the seasonal fluctuations in energy demand, operators can implement strategies to improve energy efficiency and reduce operational costs. For instance, implementing energy-saving measures during low-demand periods or adopting demand response strategies can help manage energy consumption during peak periods. In addition, the knowledge of seasonal energy consumption trends can aid in infrastructure planning and investment decisions. For areas experiencing significant tourism-related fluctuations in water usage, authorities can plan and allocate resources to expand WWTP capacity accordingly, ensuring that the infrastructure is well-equipped to handle peak demands without compromising operational efficiency. Overall, considering the seasonal and climatic influences on energy consumption in WWTPs is essential for developing robust and accurate energy consumption prediction models.

4.2. Predicting energy consumption using all input variables

The first experiment involves exploring the application of deep learning models, namely RNN, LSTM, GRU, BiLSTM, and BiGRU, for the prediction of energy consumption in WWTPs. All input variables, such as power consumption, biological characteristics, hydraulic conditions, and climate variables, are fed into these models. The output layer generates predictions for energy consumption based on the learned representations and patterns. For the experiment, we divide the data into training and testing subsets. The training data consists of 75% of the available data from January 1, 2014, to January 28, 2018, while the testing data covers the remaining 25% from January 29, 2018, to June 27, 2019. To address the challenge of limited data availability, we employ cubic spline interpolation as a data augmentation technique. This process increases the dataset size, expanding the original 1382 records by a factor of 10. By augmenting the data, we aim to enhance the models' learning capabilities and improve their predictive performance, even with limited training data.

In this study, the deep learning models are trained with specific hyperparameters to optimize their performance in predicting WWTP energy consumption. We use 100 epochs and a batch size of 250 during the training process. The models consist of 32 hidden units and use

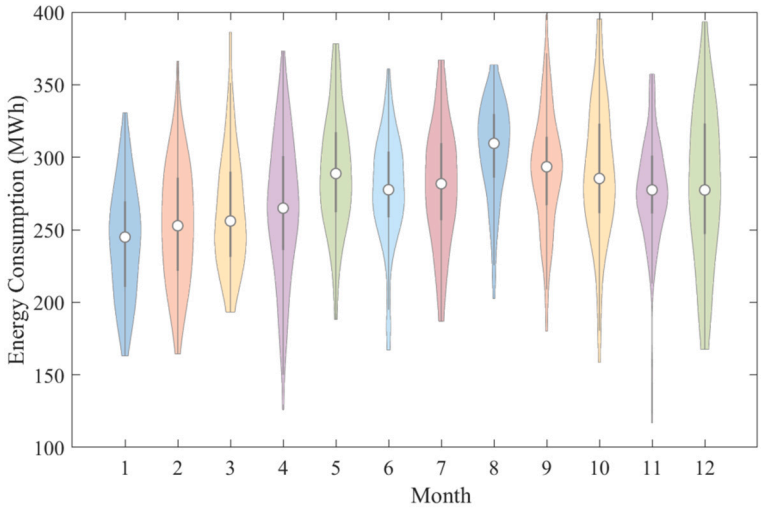

Fig. 9. Violin plots showing the monthly energy consumption distribution throughout the study period, visually highlighting density variations and potential patterns in the data beyond conventional statistical summaries.

Table 3
Comparison of prediction results between deep learning models (i.e., RNN, LSTM, GRU, BiLSTM, and BiGRU) using augmented and non-augmented data: evaluating RMSE, MAE, R^2, MSLE, and MAPE metrics.

Data	Model	RMSE	MAE	R^2	MSLE	MAPE
Non-Augmented data	RNN	45.753	35.247	-0.331	0.03	13.225
	GRU	50.049	38.643	-0.593	0.034	14.676
	LSTM	45.912	35.779	-0.34	0.029	13.504
	BiGRU	50.185	39.246	-0.601	0.035	14.678
	BiLSTM	49.009	37.544	-0.527	0.033	14.598
Augmented data	RNN	44.447	34.445	-0.182	0.028	13.097
	GRU	42.883	33.752	-0.1	0.026	12.885
	LSTM	42.861	33.622	-0.099	0.026	12.854
	BiGRU	47.448	38.407	-0.347	0.031	13.831
	BiLSTM	43.135	33.994	-0.113	0.026	12.985

the 'Relu' activation function. The cross-entropy loss function and the 'Adam' optimizer are employed to guide the learning process. These parameter values are carefully selected to ensure effective training and improve the predictive capabilities of the deep learning models. To evaluate the performance of the models. To further investigate the impact of data augmentation, we compare the performance of the deep learning models trained on both augmented and non-augmented datasets. The results are presented in Table 3, allowing us to assess the data augmentation technique's effectiveness in improving the models' predictive accuracy.

The results presented in Table 3 demonstrate that no single approach consistently outperforms all others in terms of all statistical scores. However, the MAPE values, which range between 12% and 14%, indicate that there is still room for improvement in predicting WWTP energy consumption. Upon analyzing the results in Table 3, it becomes evident that the use of data augmentation has a positive impact on the predictive performance of the deep learning models. The LSTM model trained on the augmented data achieves the best performance, with a MAPE value of 12.854%. This indicates that the data augmentation technique helps the model to capture and generalizes underlying patterns in the time-series data more effectively. Following closely, the GRU model trained on augmented data achieves a MAPE value of 12.885%, further supporting data augmentation's effectiveness. On the other hand, when training the models on non-augmented data, the LSTM model still performs relatively well, achieving a MAPE value of 13.504%. However, the GRU model shows a higher MAPE value of 14.676%, suggesting that the use of data augmentation significantly benefits the GRU model's predictive accuracy.

In conclusion, the results indicate that the application of data augmentation, especially with LSTM and GRU models, leads to improved predictions of WWTP energy consumption. While the models show promising performance, there is still scope for further enhancing the accuracy of the predictions and refining the models for better decision-making and resource allocation in WWTPs.

4.3. EC prediction using dynamic models

This section aims to construct parsimonious models by using only the relevant features as inputs for WWTP energy consumption prediction. Building upon our previous study [11], which demonstrated the benefits of incorporating important features and lagged energy consumption, we further investigate the performance of dynamic deep learning models in predicting energy consumption in WWTPs. By adopting the important features identified through XGBoost and RF algorithms, including the month, daily inflow rate (Q_{in}), average humidity, TN, BOD, and Ammonia, we focus on constructing models that capture the essential variables influencing energy consumption. Additionally, we incorporate lagged energy consumption data to account for the temporal dynamics. Here, we consider two scenarios: training the models using a reduced dataset containing the relevant features, with and without augmentation. Table 4 compares prediction results between deep learning models using augmented and non-augmented data with selected features and lagged energy consumption.

The models are evaluated using various performance metrics to assess their accuracy in predicting WWTP energy consumption. The comparison allows us to examine the impact of data augmentation on the models' predictive capabilities and determine the effectiveness of incorporating lagged energy consumption.

The initial observation highlights a substantial improvement in the performance of deep learning models (RNN, GRU, LSTM, BiGRU, and BiLSTM) when using non-augmented data with selected features and lagged energy consumption (Table 4), as compared to using all features without feature selection (Table 3). Including relevant features and temporal dynamics enhances the models' ability to capture the critical variables affecting energy consumption more accurately. This finding aligns with the concept of parsimony in model construction, where focusing on important features leads to improved predictive capabilities. By incorporating only the significant variables and considering the temporal dependencies in energy consumption, the models are better equipped to understand the underlying patterns and trends, resulting in improved predictions. This observation highlights the importance of feature selection and temporal consideration in optimizing

Table 4
Evaluating the considered deep learning approaches for accurate energy consumption prediction: augmented vs. non-augmented data, incorporating lagged energy consumption.

Data	Model	RMSE	MAE	R^2	MSLE	MAPE
Augmented data	RNN	7.498	5.924	0.964	0.001	2.182
	GRU	6.513	4.698	0.973	0.001	1.738
	LSTM	5.821	3.685	0.978	0.001	1.36
	BiGRU	5.921	3.912	0.978	0.001	1.436
	BiLSTM	6.273	4.165	0.975	0.001	1.533
Non-Augmented data	RNN	38.655	29.446	0.106	0.021	11.196
	GRU	38.683	29.771	0.105	0.021	10.88
	LSTM	37.759	28.698	0.147	0.02	10.738
	BiGRU	39.064	30.087	0.087	0.021	11
	BiLSTM	39.114	30.055	0.085	0.021	11.001

the predictive performance of deep learning models for WWTP energy consumption prediction.

The second observation from Table 4 highlights a significant enhancement in prediction accuracy when employing deep learning models based on augmented data compared to non-augmented data. Specifically, the deep learning models trained on augmented data achieved MAPE values ranging between 1% and 2.8%. In contrast, when using non-augmented data, the MAPE values ranged between 10% and 11%. This improvement in prediction accuracy can be attributed to the benefits of data augmentation through cubic spline interpolation. By augmenting the original dataset, we were able to generate additional data points that capture finer temporal details and variations in the WWTP variables. The augmented data provided a richer representation of the underlying patterns and dynamics, enabling the deep-learning models to learn from an expanded range of temporal dependencies. Comparatively, the deep learning models trained solely on non-augmented data had limited exposure to temporal patterns and dependencies due to the smaller dataset size. This resulted in higher MAPE values, indicating a relatively higher average percentage difference between the predicted and actual energy consumption values. The significant reduction in MAPE values observed when using deep learning models with augmented data highlights the effectiveness of data augmentation in capturing the complex temporal dynamics of WWTP energy consumption.

When considering the models trained on augmented data, we observe that all models achieved improved performance compared to those trained on non-augmented data. The R^2 values for all models were high, ranging from 0.964 to 0.978. These values indicate that a significant portion of the variance in the energy consumption data was captured by the models. The MSLE values for all models were close to zero, suggesting accurate predictions. Additionally, the MAPE values were relatively low for all models, ranging from 1.36% to 2.182%. The MAPE values achieved by the deep learning models trained on augmented data further highlight the effectiveness of data augmentation in improving prediction accuracy. Overall, the combination of data augmentation with cubic spline interpolation and the selection of relevant input features proved to be a successful strategy for achieving satisfactory prediction performance. This approach not only improved the accuracy of the predictions but also reduced the computational burden by utilizing a more concise set of features.

Fig. 10 demonstrates the prediction results obtained from the various deep learning models using the testing data. The plots reveal that the deep learning models exhibit the ability to accurately predict the future trend of energy consumption in the studied WWTP. This indicates that the models have successfully captured the underlying patterns and temporal dependencies in the time-series data, allowing them to generate reliable predictions.

Furthermore, Fig. 11 provides valuable insights into the prediction performance of the five deep learning models based on augmented data.

Table 5
Comparison of the predictive performance of the investigated deep learning models in this study with SOTA methods reported in the literature.

Methods	RMSE (MWh)	MAE (MWh)	MAPE(%)	R2
GBM [21]	33.9	26.9	–	0.18
RF [21]	34.8	27.7	–	0.14
ANN [21]	39.8	32.1	–	0
RNN [21]	37.3	29.3	–	0.01
KNN [11]	37.33	28.23	10.65	–
XGBoost [11]	37.14	28.5	10.81	–
LightGBM [11]	37.38	28.63	10.96	–
GPRRQ [11]	37.45	28.65	10.04	–
GSVR [11]	37.7	28.88	10.12	–
BT [11]	37.56	28.75	10.27	–
This study, RNN	7.498	5.924	2.182	0.964
This study, GRU	6.513	4.698	1.738	0.973
This study, LSTM	5.821	3.685	1.36	0.978
This study, BiGRU	5.921	3.912	1.436	0.978
This study, BiLSTM	6.273	4.165	1.533	0.975

The boxplot and empirical cumulative distribution function (ECDF) of the prediction errors offer a comprehensive view of the distribution of errors for each model. Upon analysis, it is evident that both the LSTM and BiGRU models demonstrate superior prediction performance compared to the other models. The narrower spread and lower variability in the prediction errors of LSTM and BiGRU highlight their robustness in generating more accurate predictions, indicating that they have effectively learned and captured the complex patterns in the WWTP energy consumption time-series data. These findings corroborate the effectiveness of data augmentation and the selection of relevant features, as discussed earlier, in improving the performance of deep learning models.

4.4. Comparison with previous studies

Table 5 presents a comparison of the performance of our proposed deep learning models (RNN, LSTM, GRU, BiLSTM, BiGRU) with the state-of-the-art (SOTA) methods for WWTP energy consumption prediction. It is worth noting that the '-' symbol indicates that the values were not provided in the original study for direct comparison. In [21], the prediction results of GBM, RF, ANN, and RNN models are obtained based on selected features. In [11], various machine learning techniques, including KNN, XGBoost, LightGBM, GPR, SVR, and Bagged Tree, have been investigated using selected relevant features and lagged 1 energy consumption data. Comparing the performance of these machine learning models with our proposed deep learning models, it is evident that our framework significantly outperforms the machine learning approaches. For instance, the best-performing model in [11], kNN, achieved an RMSE of 37.33 MWh, MAE of 28.23 MWh, and MAPE of 10.65%.

In terms of RMSE and MAE, our deep learning models (RNN, GRU, LSTM, BiGRU, and BiLSTM) outperform the previous methods by a significant margin. The lower RMSE and MAE values in our models indicate their ability to provide more accurate predictions of WWTP energy consumption. Specifically, the LSTM model achieved the best performance with an RMSE of 5.821 MWh and an MAE of 3.685 MWh, showcasing its effectiveness in capturing complex temporal dependencies and patterns in the data. The MAPE values further confirm the superiority of our deep learning models, as they exhibit substantially lower error percentages compared to the previous methods. The LSTM and BiGRU models achieved the lowest MAPE values of 1.36% and 1.436%, respectively, demonstrating their capability to predict WWTP energy consumption with high precision.

Overall, the results in Table 5 demonstrate that the proposed deep learning models, combined with data augmentation and feature selection, outperform the previous state-of-the-art methods in predicting

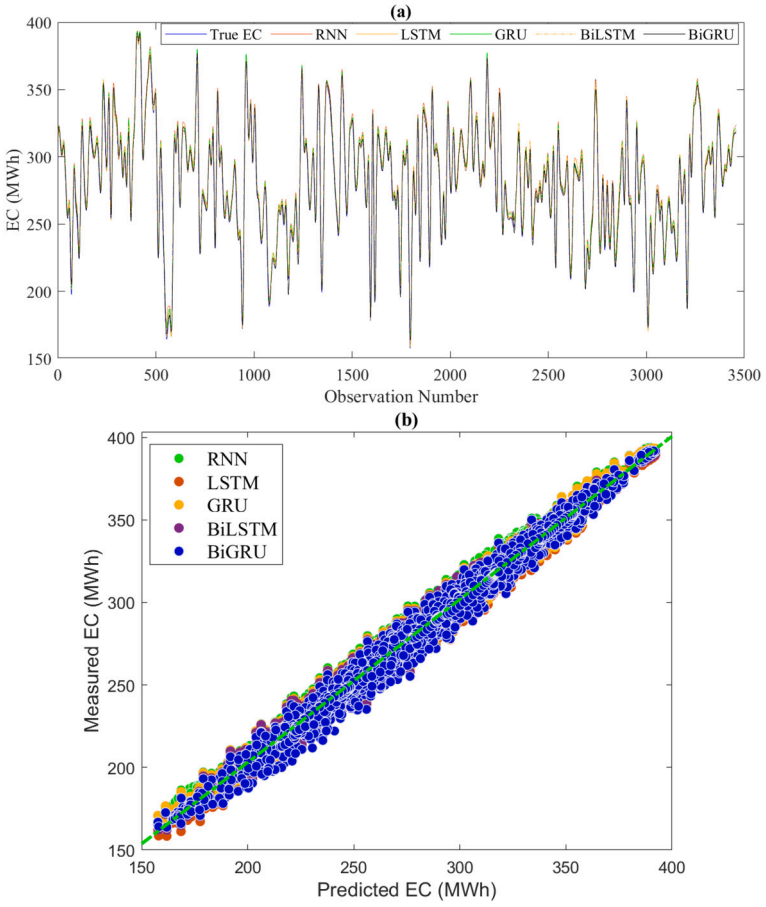

Fig. 10. (a) Line plots and (b) scatter plots illustrating the predicted energy consumption by the five deep learning models using augmented testing data. The green line in (b) represents perfect prediction.

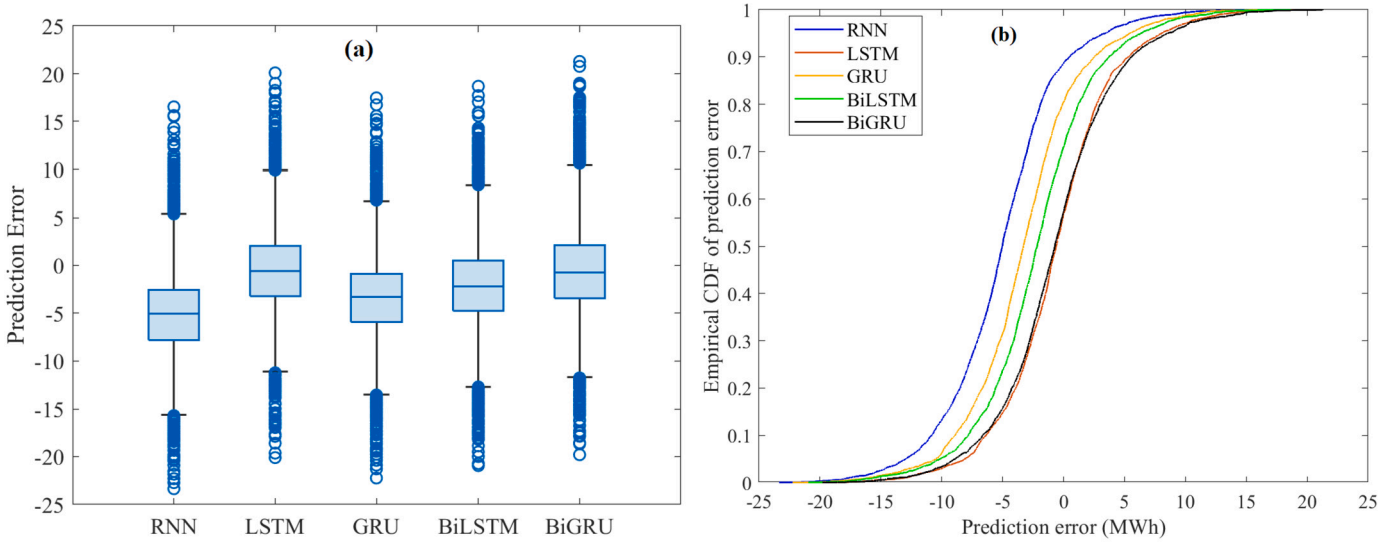

Fig. 11. (a) Boxplot illustrating the distribution of prediction errors; models with more compact boxplots around zero exhibit better predictive performance. (b) ECDF displaying the cumulative distribution of prediction errors obtained from the five considered deep learning models.

WWTP energy consumption. The remarkable reduction in RMSE, MAE, and MAPE values achieved by our LSTM model compared to the best-performing machine learning model highlights the advantage of leveraging the temporal dependencies and nonlinear relationships inherent in deep learning models. The significant improvements in accuracy and precision achieved by our models can have practical implications for WWTP management and decision-making, leading to more efficient resource allocation and enhanced sustainability.

5. Conclusion

This study proposed an efficient approach for energy consumption prediction in WWTPs, employing deep learning models, data augmentation, and feature selection. Integrating data augmentation using cubic spline interpolation provided a denser representation of WWTP variables, capturing finer temporal details. Additionally, RF and XGBoost algorithms were employed for feature selection, improving model efficiency. Furthermore, the study explored the inclusion of lagged power consumption data, enhancing the models' ability to capture temporal dependencies. The deep learning models outperformed conventional machine learning methods, highlighting the benefits of this approach for accurate energy prediction in WWTPs.

The models developed in this study serve as valuable tools for informed decision-making and proactive energy management in WWTP operations. Indeed, accurate energy consumption prediction has multifaceted benefits. Firstly, it enables substantial cost savings, a crucial factor for the efficient operation of WWTPs. By accurately forecasting energy needs, resource allocation can be optimized, reducing unnecessary expenditures and enhancing overall financial sustainability. Secondly, this approach could empower WWTPs to streamline their processes and improve operational efficiency. When energy consumption is predicted with precision, adjustments, and optimizations can be made in real-time, ensuring that the facility operates at its highest efficiency levels. This not only saves costs but also contributes to the WWTP's compliance with stringent effluent quality regulations. Additionally, by accurately forecasting energy needs, WWTPs can proactively manage their energy sources, potentially incorporating renewable energy solutions. This shift towards more sustainable practices aligns with global efforts to reduce carbon emissions and mitigate the environmental impact of industrial processes.

CRediT authorship contribution statement

Fouzi Harrou: Conceptualization, Data curation, Formal analysis, Methodology, Validation, Writing – original draft, Writing – review & editing. **Abdelkader Dairi:** Conceptualization, Data curation, Formal analysis, Methodology, Validation, Writing – original draft, Writing – review & editing. **Abdelhakim Dorbane:** Conceptualization, Formal analysis, Writing – original draft, Writing – review & editing. **Ying Sun:** Conceptualization, Methodology, Supervision, Writing – review & editing.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgement

This publication is based upon work supported by King Abdullah University of Science and Technology (KAUST) Research Funding (KRF) from the Climate and Livability Initiative (CLI) under Award No. ORA-2022-5339.

References

[1] A.G. Capodaglio, G. Olsson, Energy issues in sustainable urban wastewater management: use, demand reduction and recovery in the urban water cycle, Sustainability 12 (1) (2019) 266.

[2] Z. Sheikholeslami, D.Y. Kebria, F. Qaderi, Nanoparticle for degradation of btex in produced water; an experimental procedure, J. Mol. Liq. 264 (2018) 476–482.

[3] M. Grzegorzek, K. Wartalska, B. Kaźmierczak, Review of water treatment methods with a focus on energy consumption, Int. Commun. Heat Mass Transf. 143 (2023) 106674.

[4] S. Longo, B.M. d'Antoni, M. Bongards, A. Chaparro, A. Cronrath, F. Fatone, J.M. Lema, M. Mauricio-Iglesias, A. Soares, A. Hospido, Monitoring and diagnosis of energy consumption in wastewater treatment plants. a state of the art and proposals for improvement, Appl. Energy 179 (2016) 1251–1268.

[5] G. Kyritsakas, J. Boxall, V. Speight, Forecasting bacteriological presence in treated drinking water using machine learning, Front. Water 5 (2023) 1199632.

[6] L. Li, S. Rong, R. Wang, S. Yu, Recent advances in artificial intelligence and machine learning for nonlinear relationship analysis and process control in drinking water treatment: a review, Chem. Eng. J. 405 (2021) 126673.

[7] S. Abba, Q.B. Pham, A. Usman, N.T.T. Linh, D. Aliyu, Q. Nguyen, Q.-V. Bach, Emerging evolutionary algorithm integrated with kernel principal component analysis for modeling the performance of a water treatment plant, J. Water Process Eng. 33 (2020) 101081.

[8] H.C. Croll, K. Ikuma, S.K. Ong, S. Sarkar, Systematic performance evaluation of reinforcement learning algorithms applied to wastewater treatment control optimization, Environ. Sci. Technol. (2023).

[9] A. Azma, Y. Liu, M. Azma, M. Saadat, D. Zhang, J. Cho, S. Rezania, Hybrid machine learning models for prediction of daily dissolved oxygen, J. Water Process Eng. 54 (2023) 103957.

[10] Y. Alali, F. Harrou, Y. Sun, Predicting energy consumption in wastewater treatment plants through light gradient boosting machine: a comparative study, in: 2022 10th International Conference on Systems and Control (ICSC), IEEE, 2022, pp. 137–142.

[11] Y. Alali, F. Harrou, Y. Sun, Unlocking the potential of wastewater treatment: machine learning based energy consumption prediction, Water 15 (13) (2023) 2349.

[12] N.A. Ramli, M.A. Hamid, Data based modeling of a wastewater treatment plant by using machine learning methods, J. Eng. Technol. 6 (2018) 14–21.

[13] P.C. Nnaji, V.C. Anadebe, C. Agu, I.G. Ezemagu, J.C. Edeh, A.A. Ohanehi, O.D. Onukwuli, E.E. Eluno, Statistical computation and artificial neural algorithm modeling for the treatment of dye wastewater using mucuna sloanei as coagulant and study of the generated sludge, Results Eng. 19 (2023) 101216.

[14] S. Zhang, H. Wang, A.A. Keller, Novel machine learning-based energy consumption model of wastewater treatment plants, ACS EST Water 1 (12) (2021) 2531–2540.

[15] C. Boncescu, L. Robescu, D. Bondrea, M. Măcinic, Study of Energy Consumption in a Wastewater Treatment Plant Using Logistic Regression, IOP Conference Series: Earth and Environmental Science, vol. 664, IOP Publishing, 2021, p. 012054.

[16] D. Torregrossa, G. Schutz, A. Cornelissen, F. Hernández-Sancho, J. Hansen, Energy saving in wwtp: daily benchmarking under uncertainty and data availability limitations, Environ. Res. 148 (2016) 330–337.

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

[18] J. Qiao, H. Zhou, Modeling of energy consumption and effluent quality using density peaks-based adaptive fuzzy neural network, IEEE/CAA J. Autom. Sin. 5 (5) (2018) 968–976.

[19] R. Oulebsir, A. Lefkir, A. Safri, A. Bermad, Optimization of the energy consumption in activated sludge process using deep learning selective modeling, Biomass Bioenergy 132 (2020) 105420.

[20] B. Adibimanesh, S. Polesek-Karczewska, F. Bagherzadeh, P. Szczuko, T. Shafighfard, Energy consumption optimization in wastewater treatment plants: machine learning for monitoring incineration of sewage sludge, Sustain. Energy Technol. Assess. 56 (2023) 103040.

[21] F. Bagherzadeh, A.S. Nouri, M.-J. Mehrani, S. Thennadil, Prediction of energy consumption and evaluation of affecting factors in a full-scale wwtp using a machine learning approach, Process Saf. Environ. Prot. 154 (2021) 458–466.

[22] A. Das, P.K. Kumawat, N.D. Chaturvedi, A Study to Target Energy Consumption in Wastewater Treatment Plant Using Machine Learning Algorithms, Computer Aided Chemical Engineering, vol. 50, Elsevier, 2021, pp. 1511–1516.

[23] P. Oliveira, B. Fernandes, C. Analide, P. Novais, Forecasting energy consumption of wastewater treatment plants with a transfer learning approach for sustainable cities, Electronics 10 (10) (2021) 1149.

[24] J. Yusuf, R.B. Faruque, A.J. Hasan, S. Ula, Statistical and deep learning methods for electric load forecasting in multiple water utility sites, in: 2019 IEEE Green Energy and Smart Systems Conference (IGESSC), IEEE, 2019, pp. 1–5.

[25] C. Shorten, T.M. Khoshgoftaar, A survey on image data augmentation for deep learning, J. Big Data 6 (1) (2019) 1–48.

[26] L. Perez, J. Wang, The effectiveness of data augmentation in image classification using deep learning, arXiv preprint, arXiv:1712.04621, 2017.

[27] A. Flores, J. Valeriano-Zapana, V. Yana-Mamani, H. Tito-Chura, Pm2. 5 prediction with recurrent neural networks and data augmentation, in: 2021 IEEE Latin American Conference on Computational Intelligence (LA-CCI), IEEE, 2021, pp. 1–6.

[28] E. Talavera, G. Iglesias, Á. González-Prieto, A. Mozo, S. Gómez-Canaval, Data augmentation techniques in time series domain: a survey and taxonomy, arXiv preprint, arXiv:2206.13508, 2022.

[29] I.Y. Javeri, M. Toutiaee, I.B. Arpinar, J.A. Miller, T.W. Miller, Improving neural networks for time-series forecasting using data augmentation and automl, in: 2021 IEEE Seventh International Conference on Big Data Computing Service and Applications (BigDataService), IEEE, 2021, pp. 1–8.

[30] J. Yeomans, S. Thwaites, W.S. Robertson, D. Booth, B. Ng, D. Thewlis, Simulating time-series data for improved deep neural network performance, IEEE Access 7 (2019) 131248–131255.
[31] A.-A. Semenoglou, E. Spiliotis, V. Assimakopoulos, Data augmentation for univariate time series forecasting with neural networks, Pattern Recognit. 134 (2023) 109132.
[32] C. Esteban, S.L. Hyland, G. Rätsch, Real-valued (medical) time series generation with recurrent conditional gans, arXiv preprint, arXiv:1706.02633, 2017.
[33] T.T. Um, F.M. Pfister, D. Pichler, S. Endo, M. Lang, S. Hirche, U. Fietzek, D. Kulić, Data augmentation of wearable sensor data for Parkinson's disease monitoring using convolutional neural networks, in: Proceedings of the 19th ACM International Conference on Multimodal Interaction, 2017, pp. 216–220.
[34] S. Borzooei, Y. Amerlinck, S. Abolfathi, D. Panepinto, I. Nopens, E. Lorenzi, L. Meucci, M.C. Zanetti, Data scarcity in modelling and simulation of a large-scale wwtp: stop sign or a challenge, J. Water Process Eng. 28 (2019) 10–20.
[35] A. Dairi, F. Harrou, S. Khadraoui, Y. Sun, Integrated multiple directed attention-based deep learning for improved air pollution forecasting, IEEE Trans. Instrum. Meas. 70 (2021) 1–15.
[36] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780.
[37] F. Harrou, F. Kadri, Y. Sun, Forecasting of photovoltaic solar power production using lstm approach, in: Advanced Statistical Modeling, Forecasting, and Fault Detection in Renewable Energy Systems 3, 2020.
[38] F. Harrou, T. Cheng, Y. Sun, T. Leiknes, N. Ghaffour, A data-driven soft sensor to forecast energy consumption in wastewater treatment plants: a case study, IEEE Sens. J. 21 (4) (2020) 4908–4917.
[39] A.K. Shaikh, A. Nazir, N. Khalique, A.S. Shah, N. Adhikari, A new approach to seasonal energy consumption forecasting using temporal convolutional networks, Results Eng. 19 (2023) 101296.
[40] J.F. Ruma, M.S.G. Adnan, A. Dewan, R.M. Rahman, Particle swarm optimization based lstm networks for water level forecasting: a case study on Bangladesh river network, Results Eng. 17 (2023) 100951.
[41] A.K.A. Ahmed, A.M. Ibraheem, M.K. Abd-Ellah, Forecasting of municipal solid waste multi-classification by using time-series deep learning depending on the living standard, Results Eng. 16 (2022) 100655.
[42] U. Singh, M. Rizwan, Scada system dataset exploration and machine learning based forecast for wind turbines, Results Eng. 16 (2022) 100640.
[43] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, Learning phrase representations using rnn encoder-decoder for statistical machine translation, arXiv preprint, arXiv:1406.1078, 2014.
[44] A. Graves, J. Schmidhuber, Framewise phoneme classification with bidirectional lstm and other neural network architectures, Neural Netw. 18 (5–6) (2005) 602–610.
[45] A. Zeroual, F. Harrou, A. Dairi, Y. Sun, Deep learning methods for forecasting Covid-19 time-series data: a comparative study, Chaos Solitons Fractals 140 (2020) 110121.
[46] X. Lin, Z. Quan, Z.-J. Wang, H. Huang, X. Zeng, A novel molecular representation with bigru neural networks for learning atom, Brief. Bioinform. 21 (6) (2020) 2099–2111.
[47] B. Khaldi, F. Harrou, A. Dairi, Y. Sun, A deep recurrent neural network framework for swarm motion speed prediction, J. Electr. Eng. Technol. (2023) 1–15.
[48] A. Kummer, T. Ruppert, T. Medvegy, J. Abonyi, Machine learning-based software sensors for machine state monitoring-the role of smote-based data augmentation, Results Eng. 16 (2022) 100778.
[49] B.K. Iwana, S. Uchida, An empirical survey of data augmentation for time series classification with neural networks, PLoS ONE 16 (7) (2021) e0254841.
[50] Q. Wen, L. Sun, F. Yang, X. Song, J. Gao, X. Wang, H. Xu, Time series data augmentation for deep learning: a survey, arXiv preprint, arXiv:2002.12478, 2020.
[51] Y. Gu, Y. Li, X. Li, P. Luo, H. Wang, X. Wang, J. Wu, F. Li, Energy self-sufficient wastewater treatment plants: feasibilities and challenges, Energy Proc. 105 (2017) 3741–3751.
[52] J.L. Hintze, R.D. Nelson, Violin plots: a box plot-density trace synergism, Am. Stat. 52 (2) (1998) 181–184.
[53] Y. Abraham, B. Gerrits, M.-G. Ludwig, M. Rebhan, C. Gubser Keller, Exploring glucocorticoid receptor agonists mechanism of action through mass cytometry and radial visualizations, Cytometry, Part B Clin. Cytom. 92 (1) (2017) 42–56.