Journal of Environmental Management 392 (2025) 126648

Contents lists available at ScienceDirect

Journal of Environmental Management

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

Research article

An artificial intelligence modeling framework based on microbial community structure prediction enhances the pollutant removal efficiency of the algae-bacteria granular sludge system

Zhe Liu [a,b,*], Jie Lei [a], Rushuo Yang [a], Linshan Cheng [a], Ying Du [a], Yuhang Zhang [a], Jiaxuan Wang [c], Yongjun Liu [a,b]

[a] School of Environmental and Municipal Engineering, Xi'an University of Architecture and Technology, Yan Ta Road. No.13, Xi'an, 710055, China
[b] Key Lab of Northwest Water Resource, Environment and Ecology, Ministry of Education, Xi'an University of Architecture and Technology, Xi'an, 710055, China
[c] School of Architecture and Civil Engineering, Xi'an University of Science and Technology, Yan Ta Road, No. 58, Xi'an, 710054, China

ARTICLE INFO

Keywords:
Algae-bacteria granular sludge
Microbial community structure prediction model
Machine learning
Non-dominated sorting genetic algorithm
Optimal control model

ABSTRACT

Algae-bacteria granular sludge (ABGS) technology is a new energy-saving and low-carbon water treatment technology based on the algae-bacteria symbiotic system. However, due to its complex internal microbial system, the regulation mechanism of ABGS is unclear. To address this issue, the present study constructed a two-stage optimal control model for the ABGS system, which includes prediction of microbial community structure and planning of pollutant removal efficiency. This model enabled intelligent optimization of the system's pollutant removal efficiency through the regulation of operational parameters. In the first stage, seven machine learning (ML) algorithms were compared to predict the succession process of microbial community structure under the different conditions ($R^2 > 0.94$). In the second stage, six ML algorithms were compared to predict the pollutant removal efficiency of the ABGS system, combining regulatory indicators and microbial community structure ($R^2 > 0.94$). Finally, the non-dominated sorting genetic algorithm was used to integrate the prediction models of the two stages, and the microbial community structure was selectively shaped to enhance the removal efficiency of any two of the carbon, nitrogen, and phosphorus pollutants in the ABGS system (removal rate >90 %). The results of this study provided a universally applicable and quantitative intelligent guidance model for the performance optimization of ABGS technology and other biological systems.

1. Introduction

Algae-bacteria granular sludge (ABGS), as an efficient biological sludge consortium composed of bacteria and algae symbiosis, is increasingly becoming the focus of research in the field of wastewater treatment (Bao et al., 2024). This sludge combines the excellent ability of bacteria in organic matter degradation with the characteristic of algae producing oxygen through photosynthesis, exhibiting good settlement performance and resistance to shock loads. Through the synergistic effect between bacteria and algae, ABGS can efficiently remove nutrients such as nitrogen and phosphorus from wastewater and reduce aeration energy consumption to a certain extent, thereby achieving energy saving and emission reduction in the wastewater treatment process (Yang et al., 2025; Zhang et al., 2020). However, due to the complexity of the ABGS system and the current reliance on empirical operation, the efficiency of wastewater treatment is difficult to maintain stability and high performance. Therefore, there is an urgent require to develop a technical method that can provide quantitative guidance for the operational parameters of the ABGS system (Bao et al., 2024).

With the rapid development of artificial intelligence (AI) technology in recent years, the application of AI to solve complex process problems has become a trend (Zhu et al., 2023). Correspondingly, machine learning (ML) is also increasingly being used in environmental research to process datasets and decipher the complex relationships between system variables (Zhu et al., 2023). For example, Liu et al. used ML algorithms to predict the cultivation process of aerobic granular sludge systems (Liu et al., 2024b), achieving a coefficient of determination (R^2) of 0.98 for model performance evaluation. Additionally, prediction

* Corresponding author. School of Environmental and Municipal Engineering, Xi'an University of Architecture and Technology, Yan Ta Road. No.13, Xi'an, 710055, China.
E-mail address: zheliu@xauat.edu.cn (Z. Liu).

https://doi.org/10.1016/j.jenvman.2025.126648
Received 24 March 2025; Received in revised form 9 July 2025; Accepted 16 July 2025
Available online 26 July 2025

models for sulfate reduction intensity (Zhu et al., 2023) and effluent quality in constructed wetlands (Jiang et al., 2024) have also shown good performance ($R^2 > 0.9$), demonstrating the significant advantages of ML in solving complex biological process problems. Furthermore, for in-depth research on ABGS systems with complex biological relationships, microbial community analysis is essential, as it can reveal the ecological interactions between microorganisms and provide deeper insights into the biological mechanisms of the system (Guo et al., 2023). However, the "high-throughput sequencing" method, which is most commonly used in microbial community research, is mainly applied to key time points in ABGS system studies due to its high cost (Liu and Salles, 2024), resulting in highly discretized microbial community data. Given the continuous nature of microbial community succession processes, a discretized approach would inevitably lead to the loss of crucial information. Therefore, to achieve continuous research on microbial community succession, it is necessary to model the process using machine learning based on high-throughput technology.

Due to the complexity of microbial community structure and the susceptibility of its succession process to randomness, current advancements in microbial modeling research primarily focus on predicting the behavior of individual microorganisms (Smith et al., 2024). For example, Wu et al. predicted the colonization potential of specific exogenous microorganisms by analyzing the species composition of microbial communities (Wu et al., 2024). Although these single-microorganism prediction models can reveal certain characteristics of microbial communities, they have limitations in explaining the response of microbial communities to the macro-performance of complex biological systems (Liu and Salles, 2024). Therefore, it is particularly necessary to develop a microbial community structure prediction model that comprehensively reflects the composition of multiple microorganisms, in order to more accurately assess the overall performance of the ABGS system. Moreover, there are cases where microbial communities have been combined with AI technology. For instance, Pan et al., through ML training, achieved prediction of microbial enrichment based on carbon source types ($R^2 > 0.9$) (Pan et al., 2024b); Lesnik et al. predicted the stability of bio-batteries based on the genetic characteristics of microbial communities ($R^2 > 0.9$) (Lesnik et al., 2020). These cases demonstrate that ML is an effective method for accurately constructing the response relationship between complex microbial communities and the performance of water treatment systems. Based on existing research, it can be mainly categorized into two types: one type predicts the performance of complex systems through microbial indicators, while the other predicts the impact of external environmental indicators on microbial communities. These two types of research complement each other, and combining them can not only analyze the internal microbial situation during the operation of the reaction system but also improve the "black box" nature of AI models to a certain extent. Furthermore, this combination lays the foundation for research aimed at improving biological systems based on microbial community structure prediction.

In summary, this study aims to predict the pollutant removal efficiency of the algae-bacteria granular sludge (ABGS) system as the ultimate goal. Based on the prediction of microbial community structure succession in external environments, a two-stage ABGS system performance prediction model is established and trained and fitted using ML. Building upon this two-stage ABGS system performance prediction model, the study further integrates the Non-dominated Sorting Genetic Algorithm (NSGA) (Hossain et al., 2022) from intelligent algorithms to construct a two-stage optimal control model for improving the pollutant removal efficiency of the ABGS system based on microbial community structure prediction. This model enabled engineers to deeply analyze the regulatory mechanisms of the ABGS system from a microbiological perspective, effectively making the black-box model more "transparent". It provided a widely applicable intelligent quantitative guidance model for complex bioengineering systems.

2. Materials and methods

2.1. Experimental setup and data collection

In this study, nine identical lab-scale Photo-sequencing Batch Reactors (PSBR) (Text S1) were utilized to investigate the impact of different operational conditions on the cultivation of ABGS. Referring to commonly used operational indicators in experiments (Vincent et al., 2023), the study designed the reactors as follows: R1-R3 were regulated by the Organic Loading Rate (OLR) (Chen et al., 2022) indicator, R4-R6 by the Carbon-to-Nitrogen ratio (C/N) (Bao et al., 2024) indicator, and R7-R9 by the Organic Nitrogen content (ON) (Zhang et al., 2020) indicator. These are collectively referred to as the regulatory indicators. For the initial inoculation of ABGS, the reactors with the three different regulatory indicators were inoculated with sludge from different batches.

Influent, effluent, and biomass samples were collected daily to measure the influent parameters of the reactors at different time points, including chemical oxygen demand (COD), ammonium nitrogen (NH_4^+-N), total nitrogen (TN), and total phosphorus (TP), as well as the microbial community structure of bacteria and algae in the ABGS. In this study, 16S rRNA high-throughput sequencing technology was primarily used for bacterial community structure analysis, while 18S rRNA high-throughput sequencing was employed to study eukaryotic microorganisms, including algal communities. For detailed methods on high-throughput sequencing, please refer to Text S3 (Guo et al., 2023). Other parameters were measured using standard methods (Zhao et al., 2017). To facilitate model development, the pollutant removal efficiency of ABGS was defined in terms of carbon (C), nitrogen (N), and phosphorus (P) removal, represented by the removal rates of COD, TN, and TP, respectively (COD-OUT, TN-OUT, TP-OUT (%)) (Zhang et al., 2020), collectively referred to as effluent indicators. To improve model visualization, certain parameters were defined and described, with detailed explanations provided in Table S1. The data for the initial one-month cultivation state of the nine reactors are shown in Table S9. Subsequently, Pearson Correlation Coefficient (PCC) (Edelmann et al., 2021) analysis (Text S4) was conducted to examine the nonlinear relationships among these parameters.

2.2. Data pre-processing

The study ultimately identified 15 indicators, with their data ranges and detailed explanations provided in Table S1–S2. In this study, data preprocessing was systematically conducted on the operational datasets collected from nine reactors over a 240-day period. The workflow comprised two sequential phases: (1) classification and reorganization of input parameters and output metrics across all experimental groups; (2) implementation of rigorous data processing procedures, including outlier removal, missing data imputation, dataset randomization, and standardization of all feature variables to uniform dimensional units (Yang et al., 2023).

When the dataset is divided into training and evaluation datasets, randomization is employed to eliminate the effects of staged operations and the use of multiple reactors (Zhu et al., 2023). The statistical properties of the training and evaluation datasets must be as similar as possible to ensure the accurate assessment of the model (Yang et al., 2023).

After randomization, each influencing factor has different dimensions, so normalization was applied to each dimension. In this study, the Min-Max method was used for data normalization (Yang et al., 2023), as shown in Eq. (1).

$$x_{new} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

Where x_{min} represents the minimum value of the variable, x_{max} repre-

sents the maximum value of the variable, and x_{new} represents the normalized variable.

2.3. Model structure and the training process

To investigate the biological mechanisms of pollutant removal efficiency in the ABGS system and to achieve overall system optimization, a two-stage optimization and control model for ABGS system performance was developed in this study. The first stage is a predictive model for the succession of microbial community structure. Given the temporal coherence of complex systems, a time variable (cultivation time) was incorporated into the model to reflect the system's state at different time points (Chen et al., 2024b). Considering the continuity of microbial community succession (Liu and Salles, 2024), a time-series enhancement strategy (Chen et al., 2024b) was adopted. The initial state of the microbial community structure was used as an input variable to build a time-series predictive model to predict the community structure at a specific future time. To predict microbial community structures at future time points, this study introduced a timestamp metric derived from time-series modeling (Chen et al., 2025), specifically defined as the "cultivation duration" reflecting the microbial community development period. This temporal indicator serves as a critical input parameter in capturing chronological progression patterns within biological treatment systems. In summary, the input variables for the microbial community succession prediction model include the initial structure, cultivation time of the microbial community (time), influent indicators, and regulatory indicators, while the output variable is the future microbial community structure. To encompass the microbial community structure indicators as comprehensively as possible, the model divides it into two dimensions: diversity indicators and abundance indicators, for which separate predictive models are established. Microbial community diversity indicators include: the number of bacterial species (Ba) and their percentage of the total microbial species (PCT-Ba). Microbial community abundance indicators include: the abundance of the top 15 microbial species (Ab) and the percentage of non-bacterial species among the top 15 species relative to the total abundance (PCT-Uba). For clarity, these microbial community structure indicators were collectively referred to as microbial indicators. The detailed structure of this model was presented in Eq. (2).

$$M_{t_0+\Delta t} = \frac{dM}{dt}\Delta t + M_{t_0}, \frac{dM}{dt} = f\,(D, W, t) \tag{2}$$

Where $M_{t_0+\Delta t}$ represents the microbial indicators at time point $t_0+\Delta t$, M_{t_0} represents the initial microbial indicators, Δt represents the cultivation time of the microbial community under known environmental conditions, $\frac{dM}{dt}$ represents the growth rate of the microbial community, D represents the regulatory indicators of the ABGS system, and W represents the influent indicators of the ABGS system.

The second stage of the model was a prediction model for the pollutant removal efficiency of the ABGS system. In this model, the input indicators included the output indicators from the first-stage model and the regulatory indicators, while the output indicators were the effluent indicators. During the construction of a two-stage prediction model, the inherent black-box nature of machine learning models makes it challenging to determine the optimal model selection prior to training (Zhu et al., 2023). So, several ML algorithms were compared and selected to build the prediction models, including Neural Networks (NN) (Sun et al., 2024a), Gaussian Process Regression (GPR) (Jin and Xu, 2024), Efficient Linear Regression (ELR) (Badawi et al., 2024), Tree Regression (TR) (Swarnam et al., 2024), Tree Ensemble (TE) (Asif et al., 2024), and Support Vector Regression (SVR) (Keshun et al., 2024). During the model training phase, the dataset was partitioned into an 80 % training set and a 20 % testing set, with the testing set remaining completely isolated from the training process (Zhu et al., 2024a). To enhance the model's generalization capability, a 5-fold cross-validation strategy was implemented. The specific workflow proceeded as follows: First, the training set was randomly shuffled and subsequently divided into five mutually exclusive subsets of equal size through stratified sampling. Each subset was then sequentially designated as the validation set for performance evaluation, while the remaining four subsets were utilized for model training. Ultimately, the average of the five validation outcomes served as the comprehensive performance evaluation metric for the model on the training set (Zhu et al., 2023). During the model training process, an L_2 regularization strategy was introduced to further enhance the model's generalization capability. This method effectively controls model complexity and prevents overfitting on training data by incorporating a regularization term into the loss function of the machine learning model. The added term corresponds to the sum of squared parameters within the model's weight vector (Pan et al., 2024a). To determine the optimal hyperparameters for different ML models, Bayesian optimization was used as an outer-circulatory mechanism (Dao et al., 2024) to fine-tune the hyperparameters of each model. Details of the models and the hyperparameter tuning ranges are provided in Text S5. To evaluate the performance of the trained ML models, three evaluation metrics were selected: Mean Square Error (MSE), R^2, Mean Absolute Error (MAE),and Root Mean Square Error (RMSE) (Gomez et al., 2024) (see Text S7). When constructing the Ab prediction model, since its output indicators are the 15 most abundant species, a multi-task neural network (MTNN) (Yin et al., 2023) was used for training (see Text S6). During the model construction process, the influence of randomness on microbial community assembly was evaluated using a neutral community model (Dini-Andreote et al., 2015) to ensure the model's reliability. A detailed explanation of the neutral community model is provided in Text S8. For the resulting second-stage effluent indicators prediction model, the SHapley Additive exPlanations (SHAP) (Sun et al., 2024b) method was used for analysis (Text S11), aiming to reveal the key input indicator weights that influence the pollutant removal efficiency of the ABGS system (Wang et al., 2024b). The training process for the model is described in the attached table at the end of the supporting material: EMBRACE Checklist (Zhu et al., 2024b).

Finally, for the constructed two-stage prediction model, given the superiority of intelligent algorithms in finding optimal solutions (Hossain et al., 2022). And consider the advantages of NSGA algorithm in solving multi-objective constrained problems. This study employed the NSGA from intelligent algorithms to integrate the two-stage prediction model, thereby constructing a two-stage optimal control model for the ABGS system. In this model, the COD-OUT, TN-OUT, and TP-OUT prediction models were set as the first, second, and third objectives, respectively. This study proposes an optimized strategy for ABGS systems by holistically integrating the initial state at current timesteps and temporal costs of regulation. The core methodology involves defining the solution space through adjustable regulation indicators while setting scenario-specific constants for deterministic parameters—including influent characteristics, temporal spans, and microbial metrics—based on practical operational conditions. Through this optimization framework, the optimal combination of regulation parameters is systematically identified to maximize removal efficiencies for three target pollutants in ABGS systems. The model structure is shown in Fig. 1.

2.4. Model feature analysis

For the first-stage microbial community structure succession prediction model, this study first employed molecular ecological network analysis methods (Liu et al., 2024a) (see Text S9) to construct a correlation network of microbial community succession processes under the influence of different regulatory indicators. Subsequently, Linear Discriminant Analysis Effect Size (LEfSe) (Han et al., 2024) (see in Text S10) was used to analyze the differences in various microbial community structures, aiming to reveal the potential mechanisms of microbial community succession and provide a theoretical basis for the predictability of the model (Jia et al., 2024).

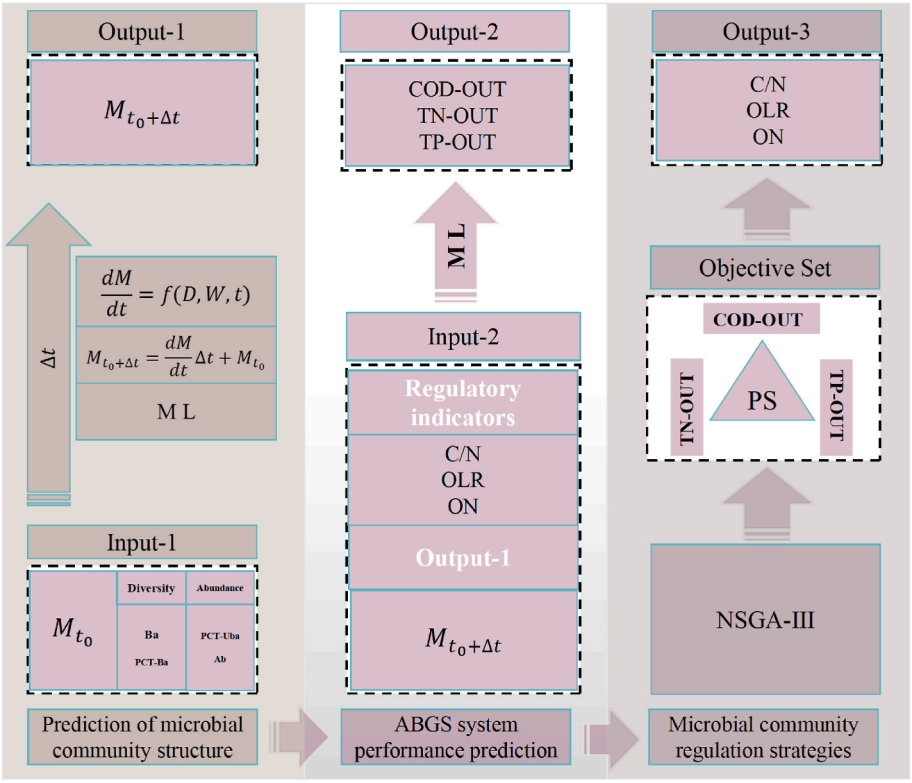

Fig. 1. Structure of the two-stage ABGS system optimal control model. Note: M represents the microbial community structure indicators, which is composed of microbial abundance indicators (PCT-Uba and Ab) and diversity indicators (Ba and PCT-Ba).

For the second-stage prediction model, the analysis of pollutant removal efficiency in the ABGS system aimed to determine the extent to which input indicators affect pollutant removal efficiency (Pederzoli et al., 2024). This study employed the SHAP method to analyze features, allowing for the quantitative assessment and ranking of the importance of input indicators (Wang et al., 2024b). For the key regulatory indicators, the study further utilized the one-dimensional Partial Dependence Plot (PDP) analysis method (see Text S12) (Divine et al., 2024) to explore the response relationships between regulatory indicators, microbial indicators, and the pollutant removal efficiency of the ABGS system. The goal is to provide theoretical guidance for the construction of the optimal control model (Wang et al., 2024a).

2.5. Integration of non-dominated sorting genetic algorithms with pollutant removal efficiency prediction models

To identify the optimal regulatory strategy for pollutant removal efficiency in the ABGS system, this study integrated the NSGA algorithm to determine the optimal control scheme (Chen et al., 2024b). Detailed explanations of the NSGA algorithm and its parameter configurations can be found in Text S13.

Due to this, the pollutant removal efficiency of the ABGS system in this study was represented by the removal rates of carbon, nitrogen, and phosphorus. The optimal solution obtained through the NSGA algorithm was not a single solution (Hossain et al., 2022), but rather a set of Pareto optimal solutions (PS) that satisfied a range of conditions. To evaluate the quality of each solution within the PS, this study considered the stringent requirement in practical engineering applications that all three pollutant removal rates must satisfy criteria (Zhang et al., 2020). Specifically, within the PS, the smaller the differences among the three objective function values, the better the characteristic of simultaneously achieving optimal solutions. Therefore, this study employed the degree of difference among the three objective variable values in the PS to assess the quality of the solutions, defining it as the equilibrium value (Eq). A smaller Eq indicates a smaller difference among the three objective values. Given that the removal efficiency data for carbon, nitrogen, and phosphorus pollutants all belong to a unidimensional dataset (ranging from 0 % to 100 %) and exhibit linear computational frameworks (removal efficiency formulas are linear equations), the design of the Eq calculation can appropriately reference Euclidean distance principles (Ghaziasgar et al., 2025). This approach ensures mathematical reliability in quantifying the geometric proximity of pollutant removal efficiencies to an idealized equilibrium state, where simultaneous maximization of all three removal rates is prioritized. For details, see Eq. (3).

$$\mathrm{Eq} = \mathrm{k}\sqrt{(\mathrm{COD\text{-}TN})^2 + (\mathrm{COD\text{-}TP})^2 + (\mathrm{TN\text{-}TP})^2} \quad (3)$$

Where: Eq represents the equilibrium value of the three objective functions in the PS; COD denotes the removal rate of COD in the PS; TP indicates the removal rate of TP in the PS; TN refers to the removal rate of TN in the PS; k is the equilibrium value coefficient, which was set to 4 in this study for the purpose of simplifying the analysis of Eq.

3. Results and discussion

3.1. Analysis of pollutant removal efficacy and microbial community structure succession in ABGS system

According to Fig. S2, this study revealed the distribution of influent indicators, regulatory indicators, and effluent indicators in nine ABGS reactors at different operational stages. A correlation analysis and clustering analysis were conducted using PCC (see Fig. 2a, b) (Chen et al., 2024b), with specific correlation values provided in Table S10–S11. The results of the clustering analysis indicate that OLR is closely related to COD-IN, C/N is related to TN-OUT, ON is associated with TP-OUT, and TN-IN is linked to NH_3-IN, suggesting a strong correlation among these four groups of indicators. To improve model

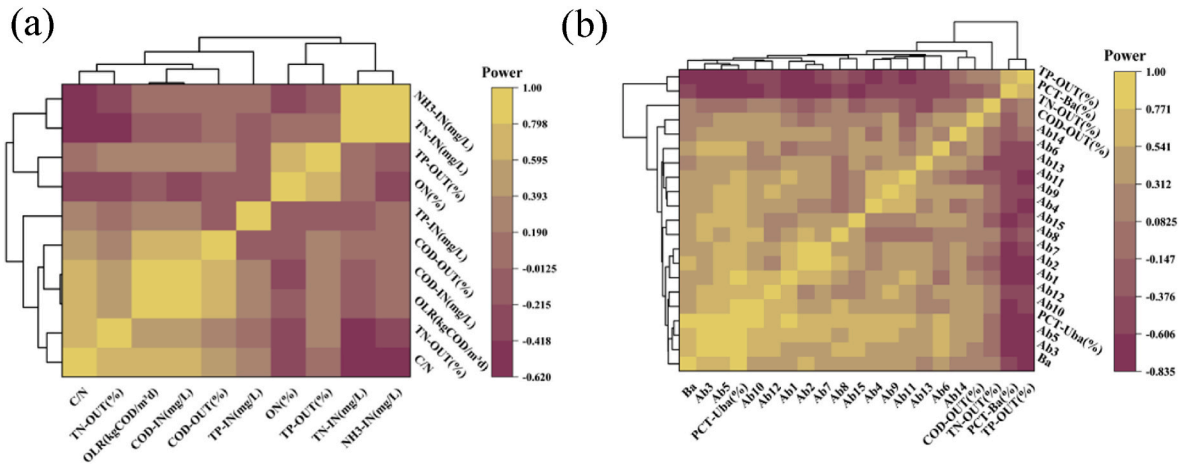

Fig. 2. (a)correlation and clustering analyses between influent indicators, effluent indicators and regulatory indicators. (b)correlation and cluster analysis between effluent indicators and microbial indicators.

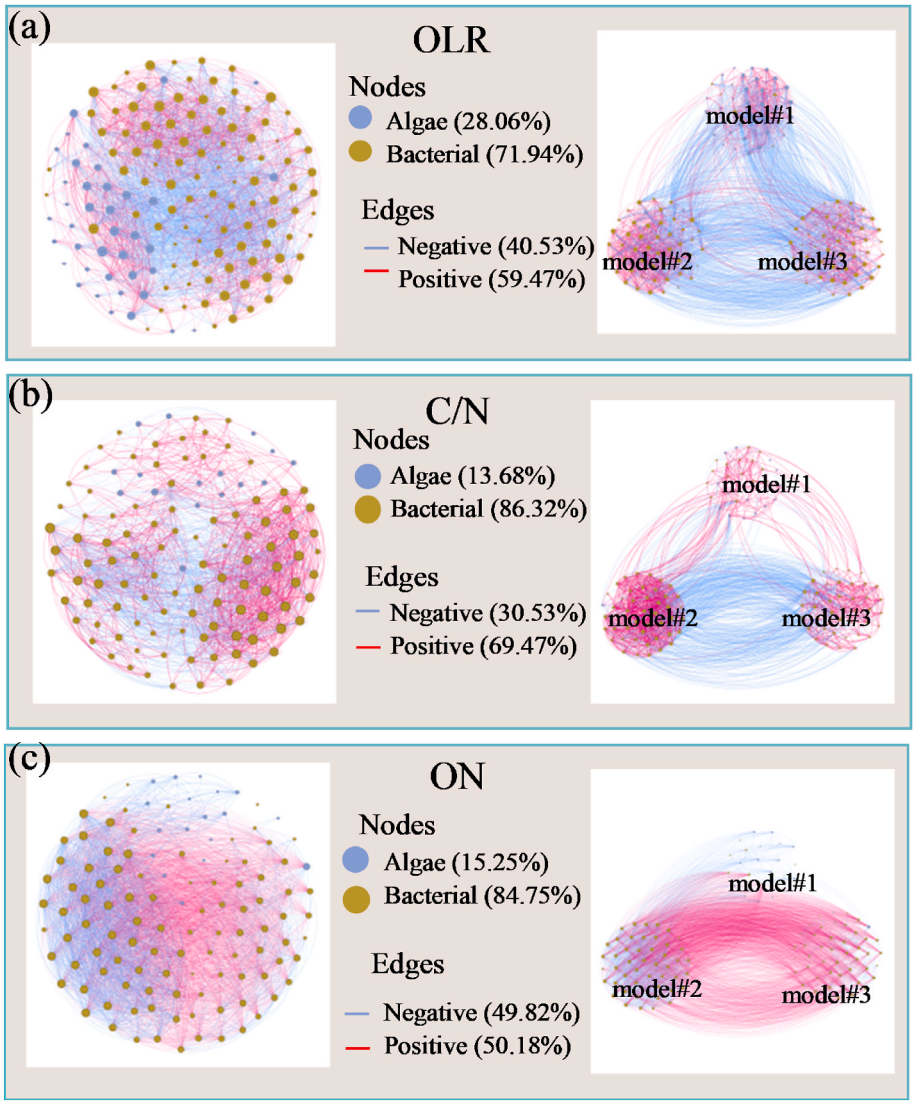

Fig. 3. Molecular ecological networks of microbial communities under the regulation of different indicators.(a) Molecular ecological networks in microbial community succession under OLR regulation, (b)C/N regulation, (c) ON regulation.

performance (Zhu et al., 2023), it was possible to select easily measurable indicators to represent the entire category of indicators when constructing the model, thereby achieving dimensionality reduction of the input indicators (Jiang et al., 2024). The correlation analysis results in Fig. 2a show that among the regulatory indicators and the pollutant removal efficiency of the ABGS system, C/N has the highest correlation with TN-OUT ($r = 0.69$), OLR has the highest correlation with COD-OUT ($r = 0.67$), and ON has the highest correlation with TP-OUT ($r = 0.70$). This indicates that the modeling strategy based on system parameter regulation to enhance pollutant removal efficiency is theoretically feasible. Considering the analysis results and the practical difficulties in regulating each indicator in engineering applications, this study performed dimensionality reduction on the influent indicators in the pollutant removal efficiency prediction model for the second stage, focusing on the regulatory indicators as the main research emphasis (Vincent et al., 2023).

Fig. 2b further reveals the correlation between microbial community structure indicators and the pollutant removal efficiency of the ABGS system. Specifically, the highest correlation with COD-OUT is observed with the microbial species ranked 14th in abundance (Ab14) ($r = 0.48$), while TN-OUT shows the highest correlation with Ab13 ($r = -0.41$), and TP-OUT has the highest correlation with Ba ($r = -0.77$). These findings indicate a significant correlation between microbial community indicators and effluent indicators. Thus, it is feasible to construct a correlation function between microbial community structure and the pollutant removal efficiency of the ABGS system.

Given the complexity of the relationships between microbial community indicators and regulatory indicators, this study employed molecular ecological networks for an in-depth exploration. The results presented in Fig. 3(a–c) reveal that under three different regulatory indicators, bacterial microorganisms dominate the assembly process of the microbial community. Notably, the proportion of bacterial microorganisms is highest under C/N regulation, reaching 86.32 %. Furthermore, the interactions among microbial species under the three regulatory indicators primarily exhibit positive correlations, with the most significant positive correlation observed in the microbial community assembly under C/N regulation, accounting for 69.47 %. Although there is a general consistency in the trends of species proportions and inter-species relationships, the assembly processes of microbial communities under different regulatory indicators still show certain differences. To further analyze the species differences in microbial community assembly under the three regulatory indicators, this study utilized LEfSe (Jia et al., 2024). The results indicate that the number of significantly different bacterial species under ON, OLR, and C/N regulatory indicators are 15, 13, and 16, respectively, while the number of non-bacterial species are 10, 40, and 20, respectively. Specific differential microorganisms are detailed in Fig. S1, highlighting the significant impact of external environmental factors on microbial community succession. Therefore, to construct a predictive model for the microbial community, it is essential to clarify the regulatory indicators influencing the microbial community and incorporate them as input indicators in the predictive model.

3.2. Performance evaluation of a succession prediction model of microbial community structure

Given that both ABGS systems and other water treatment processes are constructed based on microbial community-mediated substrate metabolism principles, belonging to biological water treatment systems, the establishment of predictive models for microbial community structural succession becomes particularly crucial when developing a universally applicable optimization framework for biological water treatment systems. In this study, six ML algorithms were employed to train the predictive model for microbial community structure succession in the first phase, and Bayesian optimization techniques were utilized for hyperparameter tuning of these ML models (Ullah et al., 2024). The Bayesian optimization processes for each model are illustrated in Fig. S5, and the optimized hyperparameter results are summarized in Table S3 (Yin et al., 2023). After a comprehensive evaluation, the best-performing ML algorithms were selected, specifically: the PCT-Ba predictive model (RMSE = 1.387, $R^2 = 0.947$, MAE = 1.146), the Ba predictive model (RMSE = 105.486, $R^2 = 0.947$, MAE = 80.223), and the PCT-Uba predictive model (RMSE = 0.423, $R^2 = 0.977$, MAE = 0.326), all of which utilized the TE algorithm. The Ab predictive model employed the MTNN algorithm (RMSR = 102.004, $R^2 = 0.951$, MAE = 98.445). The performance evaluation results of the microbial community structure succession predictive models trained with other ML algorithms are presented in Table S4 and Fig. 4. The performance of the predictive models trained using ML demonstrated reliability ($R^2 > 0.94$). Except for the Ab predictive model, the optimal ML algorithms for the microbial community structure succession predictive models were all based on the TE algorithm, highlighting its advantages in training on small sample datasets (Chen et al., 2024a). Therefore, during the training process of the predictive models, when dealing with costly data such as high-throughput sequencing (Shi et al., 2020), selecting appropriate ML algorithms can help mitigate the issue of limited data availability to some extent.

3.3. Predictability analysis of microbial community succession in ABGS system

This study utilized a neutral community model to further explore the microbial community assembly process of the predictive model from the first phase (Smith et al., 2024). As shown in Fig. 5 and a-c present the results of the neutral community model analysis, which includes all abundant species. The fitting degrees of the neutral community model for the microbial community assembly of the ABGS system under the regulatory indicators OLR, C/N, and ON were 0.79, 0.86, and 0.23, respectively. When considering all microbial species, the assembly process of microbial community succession was influenced by randomness, making it challenging to establish an accurate model for predicting the abundance of all species in the ABGS system. To address this, the study reduced the microbial community to the top 15 abundant species and conducted a neutral community model analysis, with results shown in Fig. 5d-f. The fitting degrees of the neutral community model for the microbial community assembly under the regulatory indicators OLR, C/N, and ON all dropped to 0, indicating a significant reduction in the influence of randomness on species assembly in the ABGS system. This result can be explained by the notion that "the succession of biological communities in resource-rich environments is often driven by randomness" (Dini-Andreote et al., 2015). In the ecological environment of the ABGS system, environmental resources are typically abundant for low-abundance microbial species, leading to a more random succession process for these species. In contrast, high-abundance microbial species occupy more ecological niches, making their succession more susceptible to environmental fluctuations (Pan et al., 2024b). Therefore, the succession of high-abundance species is more predictable and better reflects the impact of the environment on the microbial community (Dini-Andreote et al., 2015).

Based on the aforementioned conclusions, this study constructed a predictive model for the abundance characteristics of microbial community structure succession in the first phase by focusing on the feedback from high-abundance species. This was achieved by compressing the microbial community to include only the top 15 abundant species. Given that the microbial communities in the reactors under the three different regulatory indicators exhibit variations, this study conducted separate sorting and selection of the initial microbial community abundance characteristics for the ABGS system corresponding to each of the three regulatory indicators. The selected species are detailed in Table S6–S8.

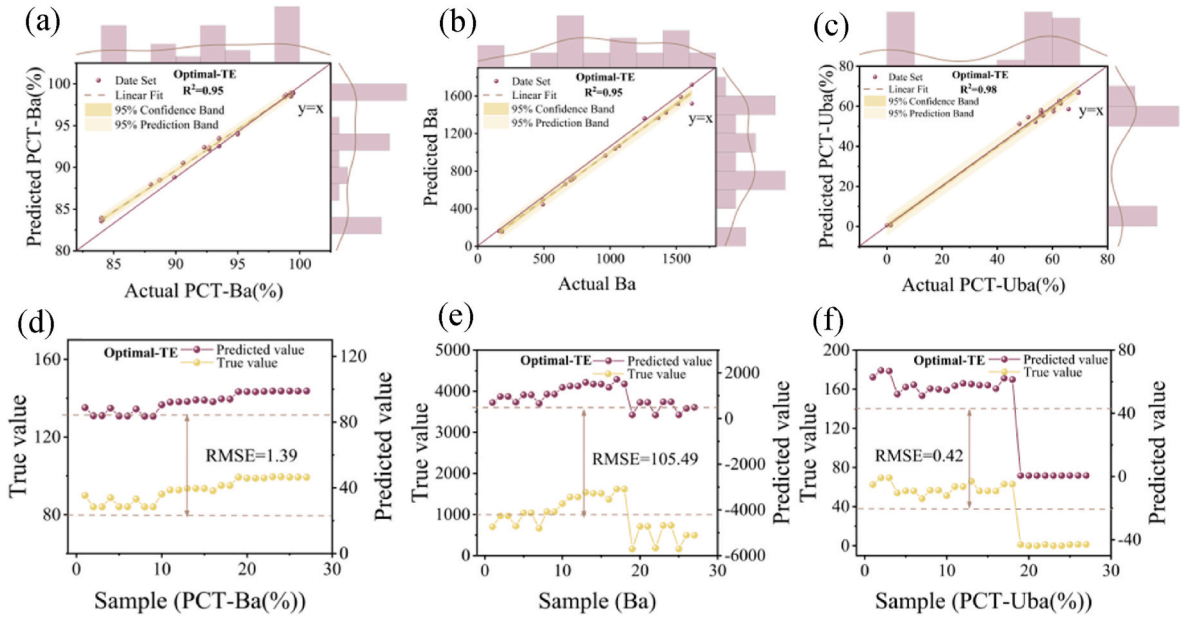

Fig. 4. Performance indicators for the prediction models of microbial community structure, (a) R^2 performance metric for the optimal machine learning model in predicting PCT-Ba, (d) RMSE performance metric for the optimal machine learning model in predicting PCT-Ba,(b, e) Ba, (c, f) PCT-Uba.

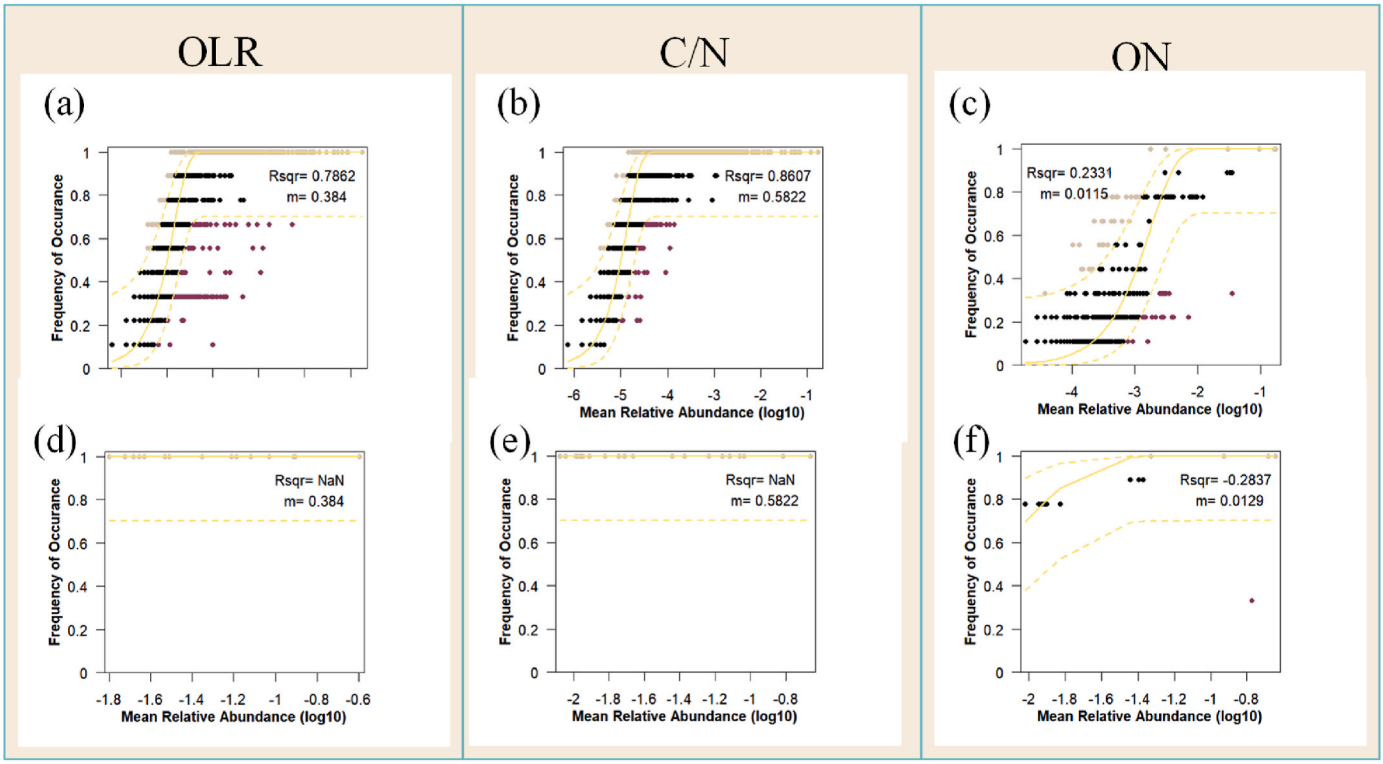

Fig. 5. Analysis of Neutral Community Model for Microbial Communities under Different Regulatory Indices. (a) Analysis of the neutral community model for all microbial community species regulated by OLR, (d) Analysis of the neutral community model for abundant microbial community species regulated by OLR, (b,e) C/N, (c,f) C/N.

3.4. Performance evaluation of a predictive model for pollutant removal efficacy of ABGS system based on microbial community structure

In this study, six ML algorithms were employed to train the predictive model for the second phase, and Bayesian optimization techniques were utilized for hyperparameter optimization of these ML models (Gomez et al., 2024). The Bayesian optimization processes for each model are illustrated in Fig. S5, and the optimized hyperparameter results are summarized in Table S3. After a comprehensive evaluation, the best-performing ML algorithms were selected. Specifically, the COD-OUT predictive model utilized a NN model (RMSE = 0.314, R^2 = 0.980, MAE = 0.285), the TN-OUT predictive model employed a TR model (RMSE = 1.288, R^2 = 0.944, MAE = 1.269), and the TP-OUT predictive model selected a GPR model (RMSE = 0.200, R^2 = 0.983,

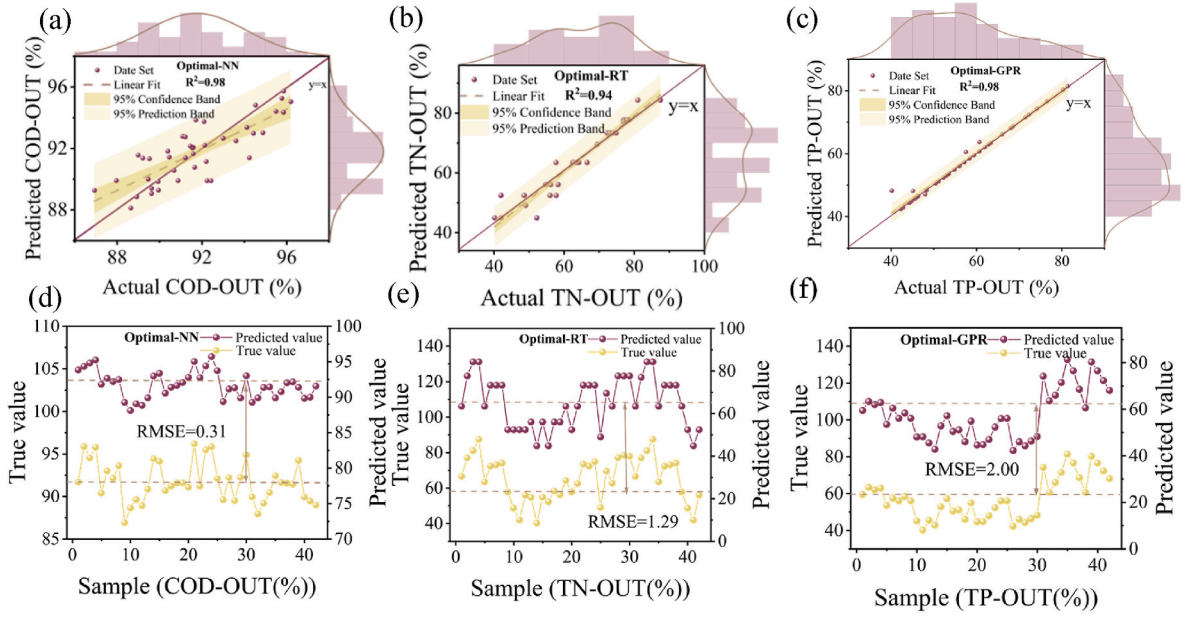

Fig. 6. Performance indicators for the prediction models of pollutant removal efficiency for the ABGS system in Phase Two, (a) R^2 performance metric for the optimal machine learning model in predicting COD-OUT, (d) RMSE performance metric for the optimal machine learning model in predicting COD-OUT, (b, e) TN-OUT, (c, f) TP-OUT.

MAE = 0.199). The performance metrics for the three pollutant removal rate predictive models are detailed in Table S4 and Fig. 6. Based on the output indicators of the microbial community structure from the first phase predictive model as input indicators for the second phase, the ABGS system pollutant removal efficiency predictive model demonstrated reliable performance ($R^2 > 0.94$). Therefore, through exhaustive training of various ML algorithms, the fitting performance of the predictive model can be improved to some extent. To further understand

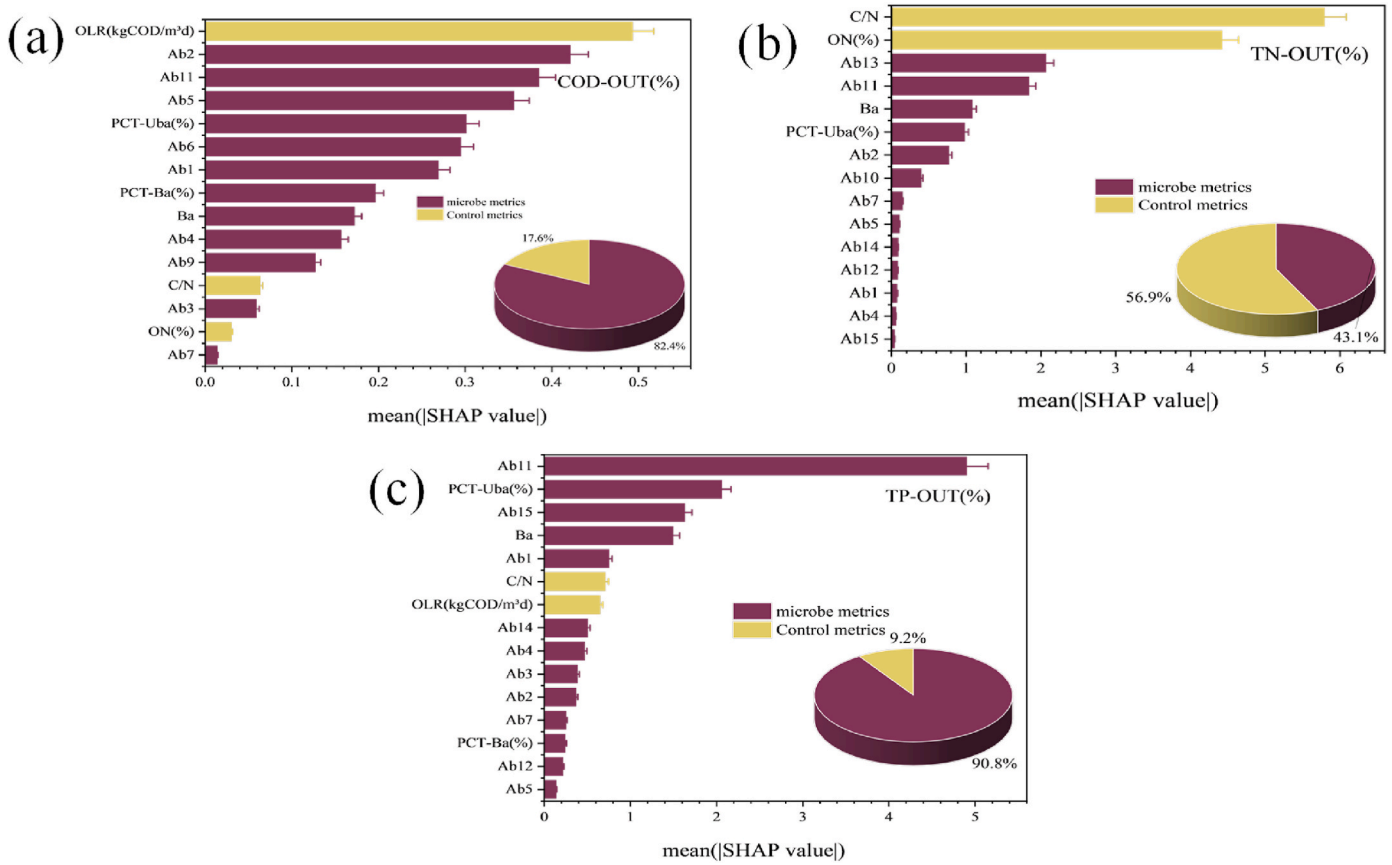

Fig. 7. Analysis of the prediction model of pollutant removal efficiency of ABGS system.(a) Ranking of correlation of input indicators of COD-OUT prediction model based on SHAP (top 15), (b) TN-OUT, (c) TP-OUT.

the regulatory strategies for enhancing the pollutant removal efficiency of the ABGS system, this study first conducted a sensitivity analysis of the pollutant removal efficiency predictive model in the second phase (Liu et al., 2024b).

3.5. Feature analysis of pollutant removal efficiency prediction model

This study conducted an in-depth analysis of the pollutant removal efficiency predictive model for the second phase using SHAP and PDP. Considering the diversity of input indicators, this study comprehensively analyzes the importance of each input indicator by SHAP method (see Fig. S3) (Zhang et al., 2020). As shown in Fig. 7a, microbial indicators accounted for 82.4 % of the influence on COD-OUT, while regulatory indicators contributed 17.6 %. Fig. 7b indicates that microbial indicators accounted for 43.1 % of the influence on TN-OUT. Furthermore, Fig. 7c reveals that microbial indicators had a substantial impact on TP-OUT, with an influence percentage as high as 90.8 %. These results demonstrate that microbial indicators significantly affect the pollutant removal efficiency of the ABGS system. Therefore, optimizing the ABGS system's efficiency based on microbial communities is entirely feasible. Furthermore, as demonstrated in Fig. 7a, the OLR among regulatory indicators exhibited the most pronounced influence on COD effluent concentration. Fig. 7b reveals that the C/N exerted the strongest impact on TN effluent concentration, with ON ranking as the second most influential factor. Fig. 7c illustrates that C/N demonstrated the highest correlation with TP removal efficiency. Considering the enhanced economic viability of adjusting regulatory indicators compared to modifying influent characteristics in practical engineering applications, this study prioritizes regulatory indicators as critical levers for optimizing pollutant removal performance in ABGS systems, establishing their optimization as the central focus of this investigation.

Through PDP, this study investigated the effects of three regulatory indicators on the removal efficiencies of C, N, and P in the ABGS system. The specific experimental results are illustrated in Fig. S4. The findings revealed a positive correlation between the OLR and the removal efficiencies of C and P. Similarly, the C/N showed a positive correlation with the removal efficiencies of both C and N. Additionally, the ON indicator exhibited a positive correlation with the removal efficiencies of C and P; however, it demonstrated a negative correlation with the removal efficiency of N. The influence of C/N on P removal efficiency was not monotonic. Furthermore, substantial interactive effects exist between operational parameters and microbial indicators. However, given that PDPs are incapable of analyzing interactions involving more than two-dimensional metrics, this study refrains from employing PDPs for interaction analysis. A comprehensive analysis indicated that the effects of OLR, C/N, and ON on the removal efficiencies of C, N, and P in the ABGS system are interrelated and sometimes contradictory (Zhang et al., 2020). If regulatory indicators are manipulated directly without fully considering the dynamic structure of the microbial community, it may lead to instability in the final regulatory outcomes of the ABGS system. Relying solely on PDP may not achieve optimal regulation of the ABGS system for C, N, and P removal efficiencies (Wang et al., 2024a). Therefore, further exploration is necessary.

3.6. Simulation of the optimal control model of a two-stage ABGS system

Given the advantages of intelligent algorithms in solving non-convex optimization problems (Chen et al., 2024b), this study employed the NSGA to integrate the predictive model of microbial community structure succession from the first phase with the pollutant removal efficiency predictive model of the ABGS system from the second phase. This integration aimed to optimize the pollutant removal efficiency of the ABGS system through the manipulation of regulatory indicators and to analyze the regulatory mechanisms at the microbial level. The study simulated the influent conditions for the small-scale ABGS cultivation experiments, specifically including COD-IN, TN-IN, TP-IN, and NH_3-IN, with values of 551.66 mg/L, 74.94 mg/L, 8.22 mg/L, and 73.44 mg/L, respectively. The initial structure of the microbial community was set to that of the R4 reactor after one month of operation (see Table S5 and S7). The cultivation time (Δt) for the microbial community was established as the minimum time interval of the experimental data in this study: 10 days. Subsequently, the optimal regulatory strategy for the pollutant removal efficiency of the ABGS system was solved using MATLAB, leading to an analysis of the microbial community structure succession under this regulatory strategy (Chen et al., 2024b), ultimately resulting in 250 PS (Regulation Strategy) (Hossain et al., 2022). The results are presented in Fig. 8. Fig. 8a-c reveal that if the focus is solely on the removal rate of a single pollutant, the NSGA algorithm can identify solutions achieving a 99 % removal rate for that pollutant. However, if only the removal rate of a single indicator is optimized, the removal efficiencies of other pollutants fall below the levels achieved through empirical operations, which does not meet the requirements of most practical engineering applications (Liu et al., 2024b). Therefore, this study introduced the concept of Eq (see Eq. (3)) to evaluate the PS. For visualization purposes, the study described the problems solved by the NSGA algorithm using Eqs. (4) and (5).

$$\text{Main object}\begin{cases}\text{Max}:\text{COD}-\text{OUT}\left(M_{t_0+\Delta t}\left(M_{t_0},D,W,\Delta t\right),D\right)\\ \text{Max}:\text{TN}-\text{OUT}\left(M_{t_0+\Delta t}\left(M_{t_0},D,W,\Delta t\right),D\right)\\ \text{Max}:\text{TP}-\text{OUT}\left(M_{t_0+\Delta t}\left(M_{t_0},D,W,\Delta t\right),D\right)\end{cases} \tag{4}$$

$$\text{St}\begin{cases}\begin{cases}W=\text{Actual water ingress}\\ M_{t_0}=\text{Initial state}\\ \Delta t=10\end{cases}\\ D_{min}<D<D_{max}\end{cases} \tag{5}$$

In the equations, D_{min} represents the lower limit of the regulatory indicators, whileD_{max} denotes the upper limit of the regulatory indicators (see Table S2).

By solving the aforementioned three-objective optimization problem using NSGA, the study evaluated the results based on the Eq values. Ultimately, four different engineering solutions were simulated (see Fig. 8d, e), with specific details as follows.

1. EQ Engineering: The objective was to achieve a balanced removal rate for C, N, and P. The lowest Eq point selected from the 250 PS yielded the following removal rates: COD-OUT, TN-OUT, TP-OUT, and Eq values of 87.16 %, 88.11 %, 89.97 %, and 3.50, respectively. In terms of regulatory strategy, when C/N, OLR, and ON were adjusted to 13.23, 1.13, and 3.09, respectively, the microbial community structure changed. The indicator reflecting microbial community diversity, Ba, decreased from an initial value of 1423 to 1260, while PCT-Ba(%) slightly adjusted from 92.7 to 92.89, indicating that this regulatory strategy would slightly reduce microbial diversity. Among the indicators reflecting microbial community abundance, the abundances of Ab4 (*Rhodobacterales*) and Ab13 (*Burkholderiales*) increased significantly by 23.74 times and 108.86 times, respectively. Previous studies have shown that *Rhodobacterales* possess a high capacity for degrading organic pollutants (Sinha and Mukherji, 2024), while *Burkholderiales* are closely related to the simultaneous removal of nitrogen and phosphorus (Dong et al., 2020). Therefore, this regulatory strategy balances the efficient removal of all three pollutants (with removal rates all exceeding 85 %) by slightly reducing microbial diversity and promoting the enrichment of Ab4 and Ab13 species.
2. NC Engineering: The goal was to maximize the removal rates of N and P, with C removal as a secondary objective. The lowest Eq point selected from the 250 PS yielded removal rates of COD-OUT, TN-OUT, TP-OUT, and Eq values of 81.25 %, 90.76 %, 94.00 %, and 16.23, respectively. This approach enabled N and P removal rates to exceed 90 %.

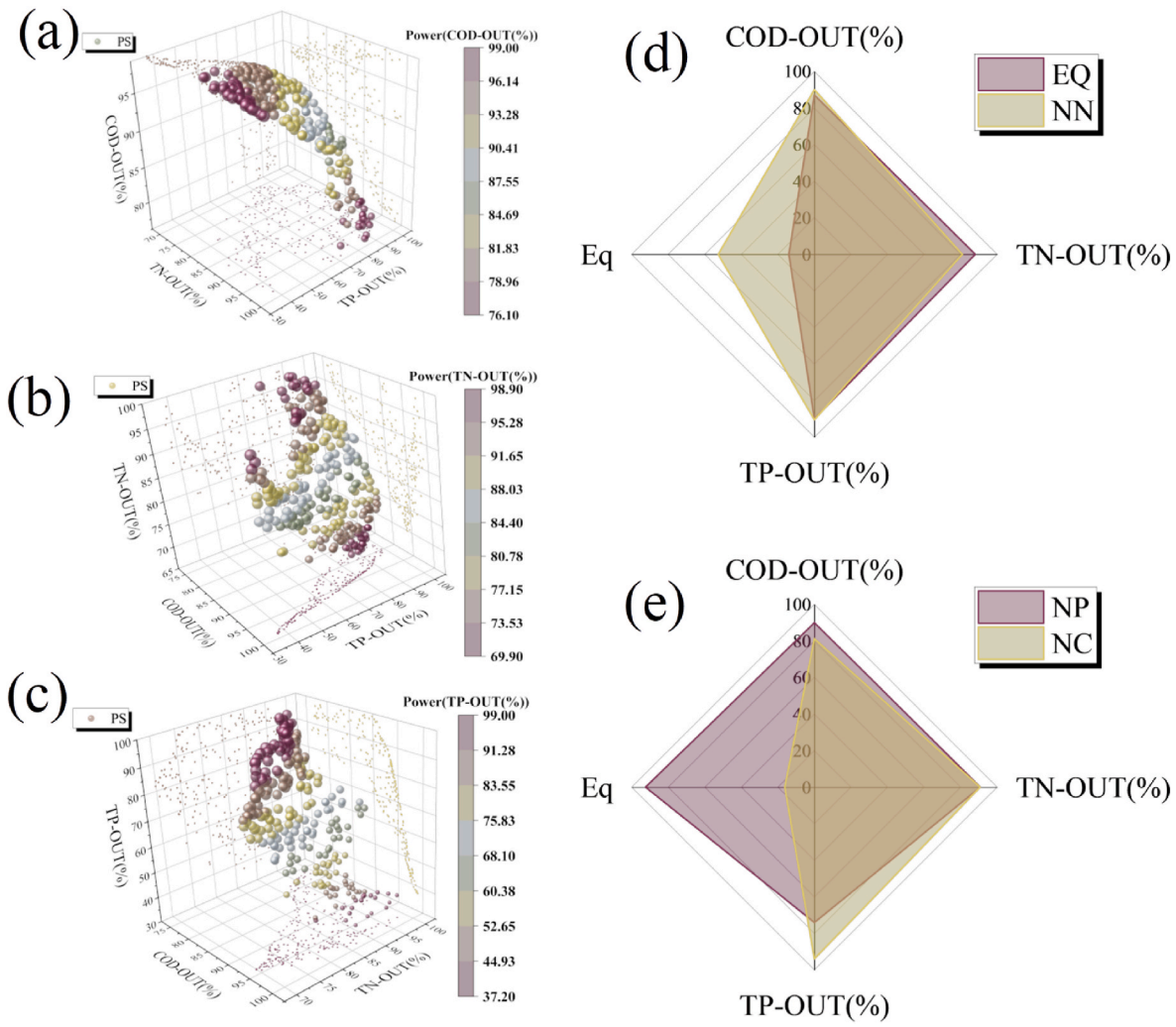

Fig. 8. Solution set of the optimal removal efficiency of three pollutants in the two-stage ABGS system.(a) View of COD-OUT values for 250 optimal solution sets, (b) TN-OUT, (c) TP-OUT, (d, e) Optimal solution of different engineering simulation requirements. Note: EQ is: the strategy of balancing carbon, nitrogen and phosphorus removal, NC is: optimal nitrogen and phosphorus removal rate, NP is: optimal carbon and nitrogen removal rate, NN is: optimal carbon and phosphorus removal rate.

3. NN Engineering: The objective was to maximize the removal rates of C and P, with N removal as a secondary goal. This method achieved C and P removal rates exceeding 90 %.

Table 1
NSGA modulation output results under specific conditions.

Particular solution	C/N	OLR	ON_	COD-OUT (%)	TN-OUT (%)	TP-OUT (%)	Eq
NC	13.44	0.98	13.52	81.25	90.76	94.00	16.23
	13.57	0.97	12.29	81.70	90.44	95.22	16.79
	13.58	0.97	16.16	80.14	92.23	92.79	17.50
NP	12.07	1.45	3.20	90.00	90.84	74.09	23.12
	11.75	1.48	3.73	90.33	90.75	71.45	27.01
	11.94	1.49	3.60	90.28	91.48	71.33	27.69
NN	10.67	1.09	3.47	90.08	81.01	90.54	13.17
	10.61	1.07	2.57	90.50	80.22	92.15	15.84
	9.90	0.99	3.12	91.57	77.02	91.55	20.56
EQ	13.23	1.13	3.09	87.16	88.11	89.97	3.50
	12.56	1.21	4.69	87.34	88.23	85.27	3.72
	14.10	1.12	3.16	86.77	89.87	89.27	4.03

*The full name and interpretation of each acronym variable are given in Table S1.

4. NP Engineering: This strategy aimed to achieve C and N removal rates exceeding 90 %.

The specific regulatory strategies and results for the above solutions are detailed in Table 1, while the microbial community structure after regulation is presented in Table S5–S8. This study provides suitable regulatory strategies for the ABGS system based on the perspective of microbial community structure, tailored to meet different engineering requirements.

4. Potential application and future work

This study established a two-phase optimal control model framework for the ABGS system. The first phase involves a predictive model for microbial community structure succession, which can forecast changes in microbial community structure based on known regulatory indicators and influent parameters. The second phase consists of a predictive model for pollutant removal efficiency in the ABGS system, which utilizes the predicted results of microbial community structure from the first phase, along with regulatory indicators, to estimate the pollutant removal efficiency of the ABGS system. Ultimately, by integrating the NSGA algorithm, the study aimed to explore optimal regulatory strategies for the removal efficiencies of carbon, nitrogen, and phosphorus

pollutants in the ABGS system, while also analyzing the succession patterns of the microbial community under these regulatory strategies. To address the "black box" nature of ML, the study exhaustively evaluated seven different ML algorithms and employed strategies such as five-fold cross-validation, regularization, and Bayesian optimization for hyperparameter tuning. This approach resulted in the development of a high-performance predictive model with excellent generalization capability ($R^2 > 0.94$), providing a comprehensive ML training process guideline for the field of environmental science. Regarding the microbial community abundance characteristics in the first-phase microbial community structure succession prediction model, the strategy of excluding low-abundance microbial species significantly reduced the randomness of species assembly during community succession (the R^2 of the neutral community model fitting decreased from approximately 0.8 to 0). This finding offers a new perspective for research on microbial community succession.

The two-phase optimal control model framework for the ABGS system developed in this study not only enhances the pollutant removal efficiency within the ABGS system but also provides an analysis of the regulatory mechanisms at the microbial level. Additionally, it offers a new research paradigm and model architecture for the integration of AI technologies in the field of environmental science. This model framework is suitable for addressing various complex biological environmental issues related to microorganisms and their engineered regulation. Looking ahead, by integrating fluid dynamics theory, large-scale database information processing, and image recognition technologies, it is anticipated that the model architecture can be further improved and refined. The goal is to construct a universal model framework applicable across various branches of environmental science, facilitating more comprehensive and effective solutions to environmental challenges.

5. Conclusions

This study established a two-phase optimal control model framework for ABGS systems based on microbial community structure prediction. By treating microbial communities as transitional phases, the framework enhances tri-pollutant removal through community regulation. Constructing the first-phase succession model revealed that screening the top 15 species Significantly lowered randomness in community succession (neutral community model fit reduced to $R^2 = 0$). Post-screening community structures effectively predicted pollutant removal efficiency ($R^2 > 0.94$). The NSGA-integrated optimization model achieved: 99 % single-pollutant removal, >90 % dual-pollutant removal, and balanced tri-pollutant removal (>85 % for all), demonstrating precision control of ABGS systems.

CRediT authorship contribution statement

Zhe Liu: Writing – review & editing, Supervision, Project administration, Funding acquisition, Data curation, Conceptualization. **Jie Lei:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Data curation, Conceptualization. **Rushuo Yang:** Writing – review & editing. **Linshan Cheng:** Writing – review & editing. **Ying Du:** Writing – review & editing. **Yuhang Zhang:** Writing – review & editing. **Jiaxuan Wang:** Writing – review & editing. **Yongjun Liu:** 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.

Acknowledgments

The authors gratefully acknowledge financial support from the National Natural Science Foundation of China (52170051, 51808432 and 51808442), the Natural Science Basic Research Plan of Shaanxi Province (2023-YBSF-301 and 2024JC-YBMS-432), the Youth Science and Technology Nova Project of Shaanxi Province (2021KJXX-106 and 2023KJXX-134), the Young Talent fund of University Association for Science and Technology in Shaanxi (20200421 and 20230456), Grant from Youth Innovation Team of Shaanxi Universities in 2021 (21JP061), and the Youth Innovation Team of Shaanxi Universities in 2021(PI: Zhang Haihan).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jenvman.2025.126648.

Data availability

Data will be made available on request.

References

Asif, U., Javed, M.F., Abuhussain, M., Ali, M., Khan, W.A., Mohamed, A., 2024. Predicting the mechanical properties of plastic concrete: an optimization method by using genetic programming and ensemble learners. Case Stud. Constr. Mater. 20. https://doi.org/10.1016/j.cscm.2024.e03135.

Badawi, A.K., Hassan, R., Farouk, M., Bakhoum, E.S., Salama, R.S., 2024. Optimizing the coagulation/flocculation process for the treatment of slaughterhouse and meat processing wastewater: experimental studies and pilot-scale proposal. Int. J. Environ. Sci. Technol. 21, 8431–8446. https://doi.org/10.1007/s13762-024-05591-y.

Bao, R., Zheng, Y., Ma, C., Xue, L., Cheng, W., Ruan, A., Li, X., 2024. A comparative study of algal-bacterial granular sludges and aerobic granular sludge at different C/N ratio: Granule characteristics, SND progress and microbial community. J. Environ. Chem. Eng. 12. https://doi.org/10.1016/j.jece.2024.113245.

Chen, X., Lin, S., Chen, X., Li, W., Li, Y., 2025. Timestamp calibration for time-series single cell RNA-seq expression data. J. Mol. Biol. 437. https://doi.org/10.1016/j.jmb.2025.169021.

Chen, X., Liu, S., Zhao, J., Wu, H., Xian, J., Montewka, J., 2024a. Autonomous port management based AGV path planning and optimization via an ensemble reinforcement learning framework. Ocean Coast Manag. 251. https://doi.org/10.1016/j.ocecoaman.2024.107087.

Chen, X., Wang, J., Wang, Q., Li, Z., Yuan, T., Lei, Z., Zhang, Z., Shimizu, K., Lee, D.-J., 2022. A comparative study on simultaneous recovery of phosphorus and alginate-like exopolymers from bacterial and algal-bacterial aerobic granular sludges: effects of organic loading rate. Bioresour. Technol. 357. https://doi.org/10.1016/j.biortech.2022.127343.

Chen, Z., Cheng, H., Wang, X., Chen, B., Chen, Y., Cai, R., Zhang, G., Song, C., He, Q., 2024b. Development and application of an intelligent nitrogen removal diagnosis and optimization framework for WWTPs: low-carbon and stable operation. Water Res. 266. https://doi.org/10.1016/j.watres.2024.122337.

Dao, F., Zeng, Y., Qian, J., 2024. Fault diagnosis of hydro-turbine via the incorporation of bayesian algorithm optimized CNN-LSTM neural network. Energy 290. https://doi.org/10.1016/j.energy.2024.130326.

Dini-Andreote, F., Stegen, J.C., van Elsas, J.D., Salles, J.F., 2015. Disentangling mechanisms that mediate the balance between stochastic and deterministic processes in microbial succession. Proc. Natl. Acad. Sci. U. S. A 112, E1326–E1332. https://doi.org/10.1073/pnas.1414261112.

Divine, D.C., Hubert, S., Epelle, E.I., Ojo, A.U., Adeleke, A.A., Ogbaga, C.C., Akande, O., Okoye, P.U., Giwa, A., Okolie, J.A., 2024. Enhancing biomass pyrolysis: predictive insights from process simulation integrated with interpretable machine learning models. Fuel 366. https://doi.org/10.1016/j.fuel.2024.131346.

Dong, Y., Lin, H., Zhang, X., 2020. Simultaneous ammonia nitrogen and phosphorus removal from micro-polluted water by biological aerated filters with different media. Water Air Soil Pollut. 231. https://doi.org/10.1007/s11270-020-04616-9.

Edelmann, D., Mori, T.F., Szekely, G.J., 2021. On relationships between the pearson and the distance correlation coefficients. Stat. Probab. Lett. 169. https://doi.org/10.1016/j.spl.2020.108960.

Ghaziasgar, M., Mohammadi, H.M., Adibi, P., 2025. A new geometry-aware non-euclidean distance metric. Mach. Learn. 114. https://doi.org/10.1007/s10994-024-06702-z.

Gomez, W., Wang, F.-K., Chou, J.-H., 2024. Li-ion battery capacity prediction using improved temporal fusion transformer model. Energy 296. https://doi.org/10.1016/j.energy.2024.131114.

Guo, M., Wang, J., Fu, C., You, J., Zong, Y., 2023. Microbial community structure and metabolic pathways in temperature- controlled anaerobic biofilm processes for the

treatment of municipal wastewater in Alpine regions. Desalination Water Treat. 294, 29–39. https://doi.org/10.5004/dwt.2023.29555.
Han, M., Wang, N., Han, W., Liu, X., Sun, T., Xu, J., 2024. Specific vaginal and gut microbiome and the anti-tumor effect of butyrate in cervical cancer women. Transl. Oncol. 44. https://doi.org/10.1016/j.tranon.2024.101902.
Hossain, S.M.Z., Sultana, N., Jassim, M.S., Coskuner, G., Hazin, L.M., Razzak, S.A., Hossain, M.M., 2022. Soft-computing modeling and multiresponse optimization for nutrient removal process from municipal wastewater using microalgae. J. Water Proc. Eng. 45. https://doi.org/10.1016/j.jwpe.2021.102490.
Jia, Z., Ou, C., Sun, S., Sun, M., Zhao, Y., Li, C., Zhao, S., Wang, J., Jia, S., Mao, P., 2024. Optimizing drip irrigation managements to improve alfalfa seed yield in semiarid region. Agric. Water Manag. 297. https://doi.org/10.1016/j.agwat.2024.108830.
Jiang, J., Xiang, X., Zhou, Q., Zhou, L., Bi, X., Khanal, S.K., Wang, Z., Chen, G., Guo, G., 2024. Optimization of a novel engineered ecosystem integrating carbon, nitrogen, phosphorus, and sulfur biotransformation for saline wastewater treatment using an interpretable machine learning approach. Environ. Sci. Technol. 58, 12989–12999. https://doi.org/10.1021/acs.est.4c03160.
Jin, B., Xu, X., 2024. Forecasting Wholesale Prices of Yellow Corn Through the Gaussian Process Regression. Neural Computing & Applications. https://doi.org/10.1007/s00521-024-09531-2.
Keshun, Y., Guangqi, Q., Yingkui, G., 2024. Remaining useful life prediction of lithium-ion batteries using EM-PF-SSA-SVR with gamma stochastic process. Meas. Sci. Technol. 35. https://doi.org/10.1088/1361-6501/acfbef.
Lesnik, K.L., Cai, W., Liu, H., 2020. Microbial community predicts functional stability of microbial fuel cells. Environ. Sci. Technol. 54, 427–436. https://doi.org/10.1021/acs.est.9b03667.
Liu, W., Zhang, Z., Zhang, B., Zhu, Y., Zhu, C., Chen, C., Zhang, F., Liu, F., Ai, J., Wang, W., Kong, W., Xiang, H., Wang, W., Gong, D., Meng, D., Zhu, L., 2024a. Role of bacterial pathogens in microbial ecological networks in hydroponic plants. Front. Plant Sci. 15. https://doi.org/10.3389/fpls.2024.1403226.
Liu, X., Salles, J.F., 2024. Drivers and consequences of microbial community coalescence. ISME J. 18. https://doi.org/10.1093/ismejo/wrae179.
Liu, Z., Lei, J., Cheng, L., Yang, R., Yang, Z., Shi, B., Wang, J., Zhang, A., Liu, Y., 2024b. Intelligent optimal control model of selection pressure for rapid culture of aerobic granular sludge based on machine learning and simulated annealing algorithm. Bioresour. Technol. 413. https://doi.org/10.1016/j.biortech.2024.131509.
Pan, B., Song, T., Yue, M., Chen, S., Zhang, L., Edlmann, K., Neil, C.W., Zhu, W., Iglauer, S., 2024a. Machine learning- based shale wettability prediction: implications for H2, CH4 and CO2 geo-storage. Int. J. Hydrogen Energy 56, 1384–1390. https://doi.org/10.1016/j.ijhydene.2023.12.298.
Pan, Y., Hua, T.-W., Sun, R.-Z., Fu, Y.-Y., Xiao, Z.-C., Wang, J., Yu, H.-Q., 2024b. Machine learning-assisted optimization of mixed carbon source compositions for high-performance denitrification. Environ. Sci. Technol. 58, 12498–12508. https://doi.org/10.1021/acs.est.4c01743.
Pederzoli, F., Riba, M., Venegoni, C., Marandino, L., Bandini, M., Alchera, E., Locatelli, I., Raggi, D., Giannatempo, P., Provero, P., Lazarevic, D., Moschini, M., Luciano, R., Gallina, A., Briganti, A., Montorsi, F., Salonia, A., Necchi, A., Alfano, M., 2024. Stool microbiome signature associated with response to neoadjuvant pembrolizumab in patients with muscle-invasive bladder cancer. Eur. Urol. 85, 417–421. https://doi.org/10.1016/j.eururo.2023.12.014.
Shi, H., Shi, Q., Grodner, B., Lenz, J.S., Zipfel, W.R., Brito, I.L., De Vlaminck, I., 2020. Highly multiplexed spatial mapping of microbial communities. Nature 588, 676. https://doi.org/10.1038/s41586-020-2983-4.
Sinha, P., Mukherji, S., 2024. Efficient treatment of secondary treated refinery wastewater using sand biofiltration: removal of hazardous organic pollutants. Water Res. 259. https://doi.org/10.1016/j.watres.2024.121874.
Smith, S.K., Weaver, J.E., Ducoste, J.J., de los Reyes III, F.L., 2024. Microbial community assembly in engineered bioreactors. Water Res. 255. https://doi.org/10.1016/j.watres.2024.121495.
Sun, W., Chang, L.-C., Chang, F.-J., 2024a. Deep dive into predictive excellence: transformer's impact on groundwater level prediction. J. Hydrol. 636. https://doi.org/10.1016/j.jhydrol.2024.131250.
Sun, Z., Li, Y., Yang, Y., Su, L., Xie, S., 2024b. Splitting tensile strength of basalt fiber reinforced coral aggregate concrete: optimized XGBoost models and experimental validation. Constr. Build. Mater. 416. https://doi.org/10.1016/j.conbuildmat.2024.135133.
Swarnam, T.P., Velmurugan, A., Subramani, T., Ravisankar, N., Subash, N., Pawar, A.S., Perumal, P., Jaisankar, I., Dam Roy, S., 2024. Climate smart crop-livestock integrated farming as a sustainable agricultural strategy for humid tropical islands. Int. J. Agric. Sustain. 22. https://doi.org/10.1080/14735903.2023.2298189.
Ullah, M.S., Khan, M.A., Masood, A., Mzoughi, O., Saidani, O., Alturki, N., 2024. Brain tumor classification from MRI scans: a framework of hybrid deep learning model with Bayesian optimization and quantum theory-based marine predator algorithm. Front. Oncol. 14. https://doi.org/10.3389/fonc.2024.1335740.
Vincent, F., Rao, T.S., Kumar, R., Nancharaiah, Y.V., 2023. Exploring the effects of organic loading rate and domestic wastewater treatment by algal-bacterial granules under natural daylight conditions. Water Environ. Res. 95. https://doi.org/10.1002/wer.10831.
Wang, M., Xie, Y., Gao, Y., Huang, X., Wei, C., 2024a. Machine learning prediction of higher heating value of biochar based on biomass characteristics and pyrolysis conditions. Bioresour. Technol. 395. https://doi.org/10.1016/j.biortech.2024.130364.
Wang, S., Xia, P., Gong, F., Zeng, Q., Chen, K., Zhao, Y., 2024b. Multi objective optimization of recycled aggregate concrete based on explainable machine learning. J. Clean. Prod. 445. https://doi.org/10.1016/j.jclepro.2024.141045.
Wu, L., Wang, X.-W., Tao, Z., Wang, T., Zuo, W., Zeng, Y., Liu, Y.-Y., Dai, L., 2024. Data-driven prediction of colonization outcomes for complex microbial communities. Nat. Commun. 15, 2406. https://doi.org/10.1038/s41467-024-46766-y.
Yang, H., Zhao, X., Wang, L., 2023. Review of data normalization methods. Computer Engineering and Application 59, 13–22.
Yang, R., Liu, Z., Liu, Y., Yang, Z., Zhang, Y., Lei, J., Wang, J., Zhang, A., Li, Z., 2025. High-throughput community and metagenomic elucidate systematic performance variation and functional transition mechanisms during morphological evolution of aerobic sludge. Bioresour. Technol. 429, 132550. https://doi.org/10.1016/j.biortech.2025.132550.
Yin, M., Zhang, X., Li, F., Yan, X., Zhou, X., Ran, Q., Jiang, K., Borch, T., Fang, L., 2023. Multitask deep learning enabling a synergy for cadmium and methane mitigation with biochar amendments in Paddy soils. Environ. Sci. Technol. 58, 1771–1782. https://doi.org/10.1021/acs.est.3c07568.
Zhang, Y., Dong, X., Liu, S., Lei, Z., Shimizu, K., Zhang, Z., Adachi, Y., Lee, D.-J., 2020. Rapid establishment and stable performance of a new algal-bacterial granule system from conventional bacterial aerobic granular sludge and preliminary analysis of mechanisms involved. J. Water Proc. Eng. 34. https://doi.org/10.1016/j.jwpe.2019.101073.
Zhao, J., Wei, H., Gao, H., Jiang, Y., 2017. Cl⁻-interference elimination method in the COD mesurement of waste water. Appl. Chem. Ind. 46, 1630–1634.
Zhu, D., Yu, B., Wang, D., Zhang, Y., 2024a. Fusion of finite element and machine learning methods to predict rock shear strength parameters. J. Geophys. Eng. 21, 1183–1193. https://doi.org/10.1093/jge/gxae064.
Zhu, J.-J., Boehm, A.B., Ren, Z.J., 2024b. Environmental machine learning, baseline reporting, and comprehensive evaluation: the EMBRACE checklist. Environ. Sci. Technol. 58, 19909–19912. https://doi.org/10.1021/acs.est.4c09611.
Zhu, J.-J., Yang, M., Ren, Z.J., 2023. Machine learning in environmental research: common pitfalls and best practices. Environ. Sci. Technol. https://doi.org/10.1021/acs.est.3c00026.