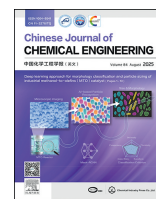




Contents lists available at ScienceDirect

Chinese Journal of Chemical Engineering

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

Full Length Article

Prediction of mass transfer performance in gas-liquid stirred bioreactor using machine learning

Feifei Chen¹, Zhenyuan Xiao¹, Zhongfan Luo¹, Peng Jiang¹, Jingjing Chen^{1,*}, Yuanhui Ji², Jiahua Zhu¹, Xiaohua Lu¹, Liwen Mu^{1,*}¹ State Key Laboratory of Material-Oriented Chemical Engineering, College of Chemical Engineering, Nanjing Tech University, Nanjing 211816, China² Jiangsu Province Hi-Tech Key Laboratory for Biomedical Research, School of Chemistry and Chemical Engineering, Southeast University, Nanjing 211189, China

ARTICLE INFO

Article history:

Received 8 December 2024

Received in revised form

19 June 2025

Accepted 26 June 2025

Available online 2 July 2025

Keywords:

Machine learning

Volumetric mass transfer coefficient

Gas-liquid stirred bioreactor

Multi-parameter optimization

ABSTRACT

The structural and operational optimization of gas-liquid stirred bioreactors presents both complexity and critical importance for enhancing mass transfer performance. This study proposes a machine learning (ML)-driven approach to identify key features and predict the volumetric mass transfer coefficient (k_La). Four ML models were adopted and compared for k_La prediction in Newtonian and non-Newtonian fluids by evaluative indices, with CatBoost and XGBoost emerging as the optimal models, respectively. Specifically, it is demonstrated that Catboost has higher prediction accuracy (AARD = 18.84%) than empirical equations by effectively incorporating multidimensional features (structural, impeller, and operational), while simultaneously extending applicability to diverse Newtonian fluids. For non-Newtonian fluids, XGBoost outperforms empirical equations by effectively incorporating fluid rheological parameters (consistency coefficient, power-law index), thereby better capturing shear-thinning behavior. Feature importance analysis further identified rotational speed (for Newtonian fluids) and liquid height (for non-Newtonian fluids) as the key features, while 2D partial dependence analysis establishes quantitative optimization ranges. This ML approach provides an efficient predictive tool for gas-liquid stirred bioreactor design and optimization.

© 2025 The Chemical Industry and Engineering Society of China, and Chemical Industry Press Co., Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

1. Introduction

In biological fermentation process, mass transfer performance plays a pivotal role in determining both product yield and quality. Particularly, oxygen mass transfer serves as a key parameter for most microbial metabolic activities, with its efficiency directly influencing the success of fermentation reactions [1–5]. The volumetric mass transfer coefficient (k_La) represents a fundamental indicator for assessing gas-liquid mass transfer performance in the stirred bioreactors [6,7]. A high k_La value indicates faster mass transfer rates, which helps to sustain microbial growth and metabolism.

However, the variation of k_La is influenced by multiple factors. Firstly, the physical properties of fluids (e.g., viscosity and density) have a significant impact on k_La . Specially, for non-Newtonian fluids, viscosity and flow behavior exhibit shear-rate dependence, often resulting in dead zones or localized flow regions that compromise mass transfer uniformity and efficiency [8,9]. Secondly, the geometric structures of bioreactors and impeller design, coupled with operational parameters significantly influences mass transfer efficiency. Variations in reactor scale induce modifications in stirring power, consequently transforming key fermentation parameters including shear rate, flow velocity, and mixing intensity, which may ultimately compromise system performance [10]. Operational parameters such as rotational speed and gas flow rate directly govern bubble formation, rise, and rupture behavior, consequently modifying mass transfer characteristics [6,7,11–15].

However, identifying and maintaining optimal operational parameters to achieve desired k_La remains particularly challenging. In biological fermentation systems, increased rotational speed enhances liquid-phase mixing, thereby improving nutrient

This article is part of a special issue entitled: AI for Chemical Engineering published in Chinese Journal of Chemical Engineering.

* Corresponding authors.

E-mail addresses: jingjingchen@njtech.edu.cn (J. Chen), lwmu@njtech.edu.cn (L. Mu).

distribution and oxygen mass transfer to stimulate microbial growth and metabolism. However, beyond critical speed thresholds, the resultant shear forces can compromise cellular integrity and ultimately reduce fermentation productivity [16]. Therefore, precise rotational speed calibration is imperative to simultaneously maximize oxygen transfer efficiency and maintain microbial viability, ultimately governing fermentation performance and product yield. Moreover, the inherent complexity of multiphase systems introduces additional control challenges. Gas–liquid interfacial phenomena, particularly surface tension effects, govern bubble dynamics including size distribution, spatial dispersion, and trajectory patterns [17]. These interdependent features create a highly nonlinear system where k_La prediction and optimization demand multivariate analysis accounting for their synergistic interactions.

To systematically characterize and optimize k_La , researchers have employed comprehensive experimental and computational approaches to evaluate mass transfer coefficients across various bioreactor structures and operational parameters, including impeller types, rotational speed, gas flow rate, and fluid properties [18–22]. For example, Liu *et al.* [23] systematically evaluated k_La variations in xanthan gum solutions across different impellers, demonstrating that Fullzone impeller provide superior oxygen mass transfer in stirred reactor. Aline *et al.* [24] optimized operational parameters to achieve a fivefold k_La enhancement (from 0.001 s^{-1} to 0.005 s^{-1}) through coordinated adjustment of rotational speed and gas flow rate. Rahimzadeh *et al.* [25] explored the effects of the fluid rheology on local mass transfer coefficient in aerated coaxial mixer reactor. Morchain *et al.* [26] systematically evaluated scale-up effects, revealing that reactor diameter influences k_La up to 55% in shear-thinning fluids, particularly at higher viscosities. However, existing empirical equations for k_La are typically derived from specific bioreactor structures, impeller designs and fluid properties. Consequently, their predictive accuracy diminishes when applied to substantially different bioreactor structures or scaled-up impeller dimensions.

The rapid advancement of artificial intelligence and computer science has positioned machine learning (ML) as a transformative approach for analyzing complex, high-dimensional bioprocess data. ML techniques excel at autonomously identifying nonlinear relationships and extracting critical features from multivariate systems [27]. Unlike traditional empirical equations, ML models offer superior adaptability, enabling rapid adjustment for diverse operational parameters and equipment configurations. This capability substantially reduces the experimental iterations required for bioreactor optimization, offering practical advantages for process development and scale-up. Zhao *et al.* [28] employed computational fluid dynamics (CFD) to systematically optimize four critical impeller parameters (e.g., impeller diameter, blade width, blade length, and blade numbers). To simultaneously optimize gas–liquid specific interfacial area while minimizing energy input, Ali *et al.* [29] pioneered a stacking-ensemble artificial neural network approach for precise k_La estimation. These studies have validated the effectiveness of multiple ML approaches, including Artificial Neural Networks (ANN), Random Forest (RF), eXtreme Gradient Boosting (XGBoost), Support Vector Machines (SVM), Categorical Boosting (CatBoost), and hybrid models. Among these, ANN and RF have been extensively applied in fermentation simulations, while XGBoost and CatBoost are particularly valued for their high computational efficiency and robust mechanisms to prevent overfitting [30,31]. These have been effectively employed to predict power consumption and product yields (e.g., ethanol) while optimizing process conditions. However, the predictive capabilities of these models remain constrained by the limited dimensionality of training datasets and exhibit pronounced

condition-specific dependencies that restrict their generalizability.

In this study, the database containing 303 datasets was systematically categorized into two distinct classes based on rheological properties, i.e., Newtonian and non-Newtonian. Four ML models (e.g., ANN, XGBoost, CatBoost and RF) were adopted to predict k_La in gas–liquid stirred bioreactors, accounting for the influence of multiple interacting features (rotational speed, gas flow rate, and viscosity). 2D partial dependence analysis was used to further elucidate the relationship between key features (rotational speed, gas flow rate, and blade number) and k_La , providing mechanistic insights into their combined effects. This ML-based framework serves as a robust and versatile tool for optimizing k_La in gas–liquid stirred bioreactors, streamlining the exploration of critical process parameters.

2. Data and Methods

Fig. 1 presents the ML modelling workflow, which employs a two-stage optimization approach (pre-optimization and re-optimization) to enhance prediction accuracy. During pre-optimization, permutation importance and SHAPley Additive exPlanation (SHAP) analysis were conducted to identify and eliminate features demonstrating minimal influence on k_La prediction. This feature selection process generated refined datasets that were subsequently used for model re-optimization, thereby enhancing the predictive performance.

ML can be categorized into three types: supervised, semi-supervised, and unsupervised learning [32]. Specifically, supervised learning constructs a mapping relationship between input features and output targets based on fully labeled training data to enable prediction or classification of new data [33]. Semi-supervised learning optimizes models by combining a small amount of labeled data with a large amount of unlabeled data [34]. Unsupervised learning aims to mining intrinsic structural features from unlabeled data [35]. The present study adopts supervised learning for its regression task of predicting k_La from experimental data, employing models including RF, XGBoost, CatBoost, and ANN.

2.1. Dataset construction

Initially, 303 datasets (experiment and simulation values) were collected directly from published literatures. Based on the rheological properties of fluids, the whole datasets were divided into two categories: Newtonian fluids dataset, comprising 175 datasets with 15 features, and non-Newtonian fluids dataset, consisting of 128 datasets with 16 features, as shown in Table S1 and S2 in the Supplementary Material.

The features affecting k_La can be categorized into three groups: (1) structural parameters including bioreactor dimensions (volume (V), diameter (D), and liquid height (H)), impeller parameters (diameter (d_0), blade number (b), number (m), disk diameter (d), clearance ratio (z/H), and position (h_0/H)), and blade specifications (length (L), width (W), and inclination angle (θ)). (2) Fluid rheological parameters (viscosity (μ), consistency coefficient (K), and power-law index (n)); and (3) operational parameters (gas flow rate (Q_g) and rotational speed (N)).

Moreover, this study systematically evaluates bioreactor systems across multiple scales, specifically laboratory-scale and pilot-scale configurations. For Newtonian fluids, the bioreactor volumes span from 1 L laboratory-scale to 100 L pilot-scale. In the case of non-Newtonian fluids, the relevant size is represented by 50 L pilot-scale bioreactors. The precise range of these bioreactor sizes are elaborately outlined in Table 1.

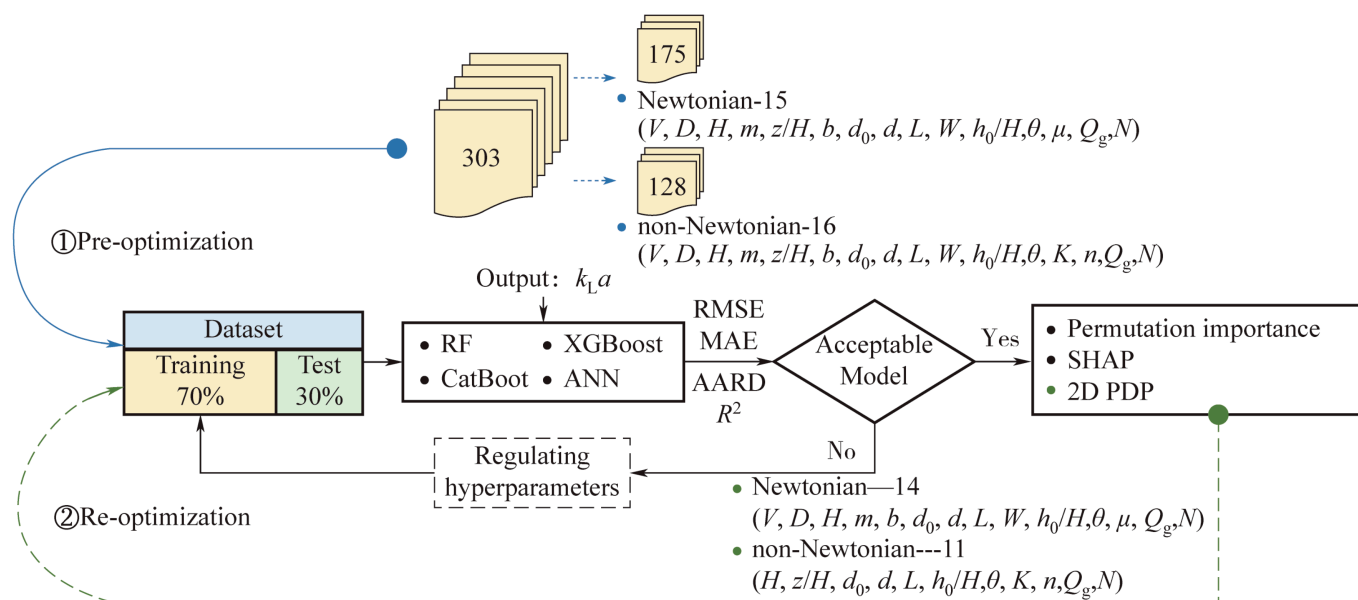


Fig. 1. Workflow of collection curation, ML model training and testing, and feature importance analysis.

Table 1
Structural parameters of the stirred bioreactors.

	Newtonian fluid	non-Newtonian fluid
V/L	1–100	1–50
H/mm	125–880	83–880
D/mm	100–480	120–400

2.2. Data preprocessing and PCC analysis

Data preprocessing serves as a critical step in the data analysis pipeline, ensuring high-quality input data. For fluid analysis, the initial datasets were rigorously processed. Newtonian fluid data were refined from 179 to 175 datasets, while non-Newtonian fluid data were reduced from 151 to 128 datasets through systematic removal of missing values and outliers. Moreover, all features were standardized to eliminate dimensional differences, enabling uniform comparison across parameters. This standardization was consistently applied to both training and test sets, as shown in Eq. (1).

$$z = (x - u) / \sigma \quad (1)$$

where z is the normalized value of initial x , x represents the data points, u is the mean value of x , and σ is the standard deviation of x .

The feature selection process was conducted meticulously. Previous studies indicated that many researchers utilized configurations with four baffles in their experiments, while only a small subset of samples employed no baffles. To maintain model training accuracy and avoid bias from more samples with baffles, a screening was done. Specifically, all datasets without baffles were excluded, resulting in a dataset exclusively comprising datasets with four-baffle configurations. Consequently, due to this uniformity in baffle count across all remaining datasets, the feature was ultimately removed from the dataset as it lacked meaningful diversity.

The features and output were statistically described by using a semi-box plot. The Pearson Correlation Coefficient (PCC) analysis represents a cornerstone methodology in statistical analysis, with its mathematical formulation provided in Eq. (2). This powerful analytical tool establishes a rigorous framework for evaluating linear interdependencies between paired features [36].

$$r = \frac{\sum_{i=1}^n (x - \bar{x}) \sum_{i=1}^n (y - \bar{y})}{\sqrt{\sum_{i=1}^n (x - \bar{x})^2} \sqrt{\sum_{i=1}^n (y - \bar{y})^2}} \quad (2)$$

where r is the value of PCC, $\bar{x}$ or $\bar{y}$ denotes the mean values of variable x or y , n is the number of samples, respectively. Additionally, the p -value is derived from a two-tailed t -distribution with $N-2$ degrees of freedom, representing the reliability of the linear correlation assessment.

The significance level of the correlation coefficient between any two features were calculated using Eq. (3).

$$t = r\sqrt{N-2} / \sqrt{1-r^2} \quad (3)$$

where t represents the t -test, N is the number of samples.

2.3. Machine learning models

Random Forest (RF) [37] is an ensemble learning model that constructs multiple decision trees and aggregates their predictions to produce accurate and robust results. The foundational concept of RF was first introduced in 1995 by Ho [38], who employed the random subspace method. Later, in 2001, Breiman [39] significantly advanced the model by integrating his “bagging” concept. This approach involves bootstrap sampling which generates multiple subsets of the training data by sampling with replacement, combined with random feature selection where a random subset of features is chosen at each node when splitting decision trees. These mechanisms work together to reduce correlation among individual trees while enhancing the model’s predictive performance and generalization capability. As shown in Fig. 2, RF begins with a training set of m samples from which k (for $i = 1, \dots, k$) new training subsets are created via bootstrap sampling, each containing m samples with replacement. A separate decision tree is then trained on each subset, maintaining model independence, before the final prediction is obtained by aggregating the outputs of all k models.

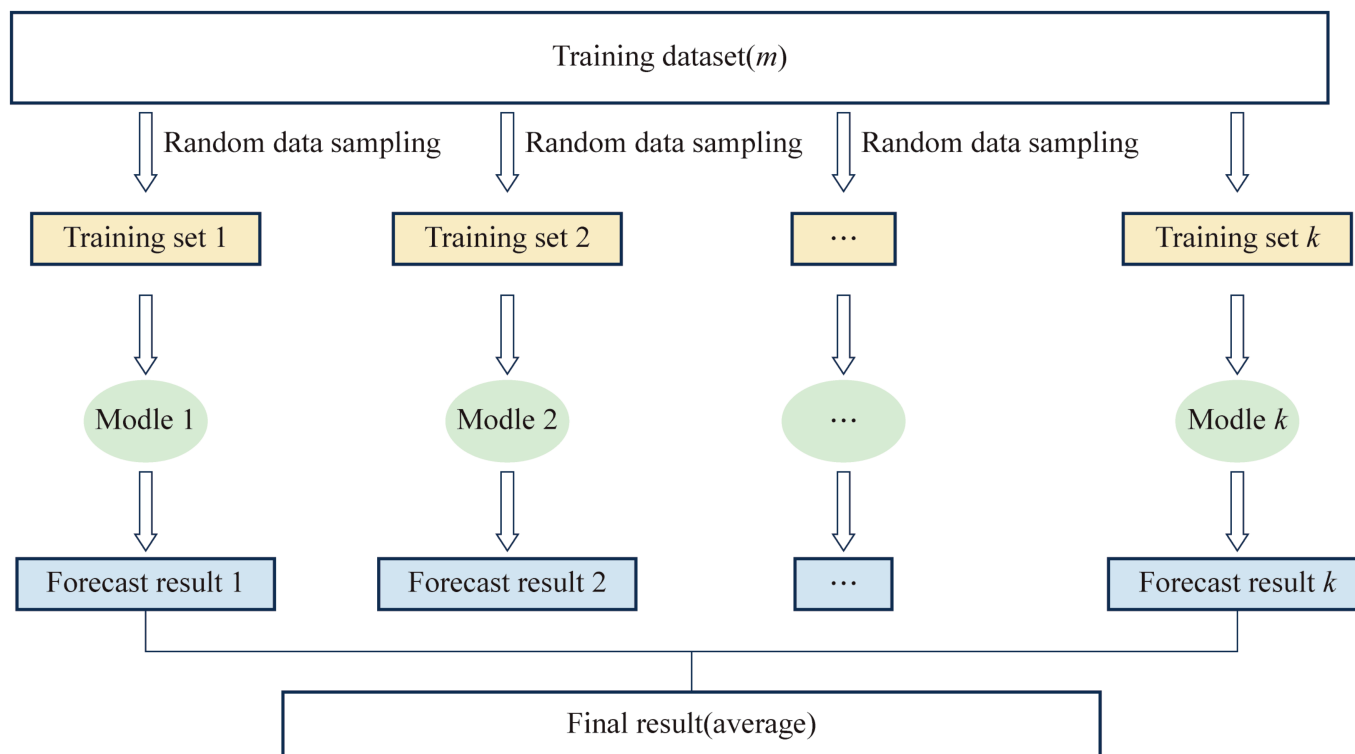


Fig. 2. Model description of random forest.

The prediction of RF, denoted as $\hat{y}$, is obtained by averaging the prediction results from all individual decision trees. This ensemble output can be mathematically expressed as Eq. (4):

$$\hat{y} = \frac{1}{k} \sum_{i=1}^k y_i \quad (4)$$

eXtreme Gradient Boosting (XGBoost) [40,41] is an advanced boosting tree model that integrates multiple decision tree models to form a powerful classifier through an additive modelling approach. Each tree is generated based on the results of the previous tree, aiming to gradually reduce prediction errors and improve overall model performance. XGBoost employs efficient algorithms and regularization techniques, which not only accelerate training speed but also significantly enhance the model's generalization ability.

The optimization objective of XGBoost is the loss function, and the fitting objective is to minimize the “residual” (residual = predicted value – true value), so that the model's prediction results gradually approach the true value (the smaller the residual, the smaller the loss, and the better the model). The objective function formula for setting the t -th decision tree are given by Eqs. (5) and (6):

$$\text{Obj}^{(t)} = \sum_{k=1}^n l(y_k, \hat{y}_k^{(t)}) + \sum_{i=1}^t \Omega(f_i) \quad (5)$$

$$\hat{y}_k^{(t)} = \hat{y}_k^{(t-1)} + f_t(x_k) \quad (6)$$

where “Obj” refers to the objective function, $l(y_k, \hat{y}_k^{(t)})$ is the loss function used to measure the difference between the predicted value $\hat{y}_k^{(t)}$ and the true value y_k , n indicates the number of observations, $\hat{y}_k^{(t)}$ and $\hat{y}_k^{(t-1)}$ are the prediction at phase t and $t-1$,

$f_t(x_k)$ is the learner at phase t , $\Omega(f_i)$ represents the regularization term, which is calculated according to Eq. (7):

$$\Omega(f_i) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T \omega_j^2 \quad (7)$$

where T is the depth of the current subtree, ω denotes vector scores in leaves, γ is the minimum loss necessary to further divide the leaf node, and λ represents the regularization parameter.

Categorical Boosting (CatBoost) [42,43] offers a gradient boosting framework that utilizes symmetric trees, making it especially adept at handling categorical features. In addition, CatBoost introduces the “ordered boosting” strategy, which effectively reduces the risk of overfitting caused by data leakage by sorting the data. This model also fully utilizes multithreading and GPU acceleration, significantly improving training efficiency and making large-scale data processing more efficient.

CatBoost gradually optimizes the model through iterative methods, with the following specific steps. The initial model is first denoted as $\hat{y}_0$. At the t -th iteration, both the gradient and second derivative are computed to guide the construction of a new decision tree $\hat{y}_t$, after which the prediction is updated accordingly.

Artificial Neural Network (ANN) [44,45] represent a class of biologically-inspired computational models that emulate the information processing mechanisms of biological neural systems. As fundamental architectures in machine learning and deep learning, ANNs are composed of interconnected processing units (neurons) organized in layered structures. The input data is processed through the weighted sum activation function of each layer of neurons to generate the final output result. The activation function used is usually represented as Eq. (8):

$$y = \sigma \left(\sum_{i=1}^n (w_i x_i) + b \right) \quad (8)$$

where x_i represents a single input variable or feature (there is n such inputs), w_i represents a learnable weight for that input, b represents a learnable bias term, and σ represents a non-linear activation function that takes a single input and returns a single output.

In this study, four ML models (RF, XGBoost, CatBoost, and ANN) were implemented using the scikit-learn library in Python 3.12. The models were adopted in the PyCharm integrated development environment. These ML models were employed to predict $k_L a$ for both the Newtonian and non-Newtonian fluids datasets. The collected experimental datasets were randomly shuffled and subsequently partitioned into training and test sets with a 70:30 ratio. Hyperparameter optimization was required for four ML models to avoid oversimplification and excessive complexity, which could lead to underfitting or overfitting [46]. Hyperparameters were optimized using a grid search strategy combined with 5-fold cross-validation to generate a robust model with high regression accuracy. Meanwhile, an early stopping strategy was introduced to select the optimal number of iterations and adjusted the L2 leaf regularization to effectively prevent overfitting.

2.4. Model evaluation

Following hyperparameter tuning, ML models were retrained using the training dataset and validated with the independent test set. To quantitatively evaluate the prediction performance, four indices were employed to evaluate the predictive ability of four ML models. The coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and average absolute relative deviation (AARD) were calculated according to Eqs. (9)–(12). R^2 , RMSE, MAE, and AARD measure the variability of the dependent variable, standard deviation of prediction errors, mean absolute value of prediction errors, and average relative error, respectively. The mutual validation of the different types of evaluation indices ensures the consistency and reliability of the prediction performance.

$$R^2 = 1 - \left[\frac{\sum_{i=1}^n (Y - Y^{\text{pred}})^2}{\sum_{i=1}^n (Y - \bar{Y})^2} \right] \quad (9)$$

$$\text{MAE} = (1/n) \times \sum_{i=1}^n |Y - Y^{\text{pred}}| \quad (10)$$

$$\text{RMSE} = \sqrt{(1/n) \times \sum_{i=1}^n (Y - Y^{\text{pred}})^2} \quad (11)$$

$$\text{AARD} = (1/n) \times \sum_{i=1}^n |(Y - Y^{\text{pred}})/Y| \times 100\% \quad (12)$$

where n represents the number of test samples, Y^{pred} denotes the predicted value by ML, Y indicates actual values, $\bar{Y}$ is average value of actual values.

2.5. Model interpretation

To explore the relative importance and interactions of input features with the output ($k_L a$) in “black-box” ML models, two methods were employed to explain the optimal ML model: permutation importance [47] and SHAP [48,49]. Permutation

importance provides a highly interpretable feature selection method that effectively captures nonlinear relationships between input features [50], offering an advantage over PCC analysis [51]. Meanwhile, SHAP analysis extends the interpretability framework by quantifying individual feature contributions through Shapley values while accounting for feature interactions [52].

Additionally, to investigate the effect of representative features on $k_L a$ based on the results of SHAP analysis, 2D partial dependence plot (PDP) was employed for univariate analysis [53,54]. This approach allows for the examination of the effect of individual features on $k_L a$ while controlling for the average effects of other features. These plots reveal how the interplay between two features jointly influences $k_L a$, offering deeper insights into the model's underlying mechanisms.

During the Pre-optimization process, the features for the Newtonian and non-Newtonian datasets were optimized to 14 and 11, respectively. The corresponding optimal hyperparameters are detailed in Table S5. The prediction performance of the ML models for the both datasets are illustrated in Fig. S1 and S2. Permutation importance and SHAP values are shown in Fig. S3, respectively. Additionally, the re-optimized hyperparameters are summarized in Table S6.

3. Results and Discussion

3.1. Dataset statistical analysis

The semi-box plots in Figs. 3 and 4 display the distributions of various features, including structural parameters of the stirred bioreactors (V , D , and H), impeller parameters (m , h_0/H , z/H , b , θ , d_0 , d , L , and W), fluids properties (μ , k , and n), and operational parameters (Q_g and N) for Newtonian and non-Newtonian fluids, respectively. The frequently-selected combinations of these features are summarized in Table 2. The frequently-selected H for non-Newtonian fluids is larger than that for Newtonian fluids, while the frequently-selected h_0/H and z/H of non-Newtonian fluids are smaller than those of Newtonian fluids. The values of D for both Newtonian and non-Newtonian fluids are in the similar range. These findings suggest that non-Newtonian fluids benefit from bioreactors with larger aspect ratio, greater distances between the gas distributor and bottom impeller, and specific impeller clearances for enhanced mass transfer performance. The need for larger H with non-Newtonian fluids likely stems from their higher viscosity, which necessitates longer residence times to achieve effective mass transfer compared with Newtonian fluids. This observation accounts for larger V consistently selected for non-Newtonian fluids.

Regarding impeller parameters, non-Newtonian fluids generally require larger blade parameters (d_0 , d , L , and W) than Newtonian fluids to expand the radial mixing region, which is attributed to the significant shear-thinning behavior ($0.31 < n < 0.65$) of non-Newtonian fluids. Specially, straight-blade impellers with $\theta = 90^\circ$ predominate in these applications, as this configuration offers practical advantages in installation and maintenance. Moreover, for operational parameters, non-Newtonian fluids employ both a lower Q_g and higher N compared with Newtonian fluids, indicating that N may has a more obvious promoting impact on $k_L a$ than Q_g for non-Newtonian fluids.

As shown in Fig. 5, $k_L a$ ranges from 0.00125 to 0.09 s^{-1} for Newtonian fluids and 0.001 to 0.12 s^{-1} for non-Newtonian fluids. However, the most frequent $k_L a$ range from 0.00125 to 0.032 s^{-1} for Newtonian fluids (median = 0.01775 s^{-1}) and 0.00074 to 0.029 s^{-1} for non-Newtonian fluids (median = 0.00735 s^{-1}). These results demonstrate significantly poorer mass transfer performance for non-Newtonian fluids compared with Newtonian fluids. Given these substantial differences in mass transfer characteristics

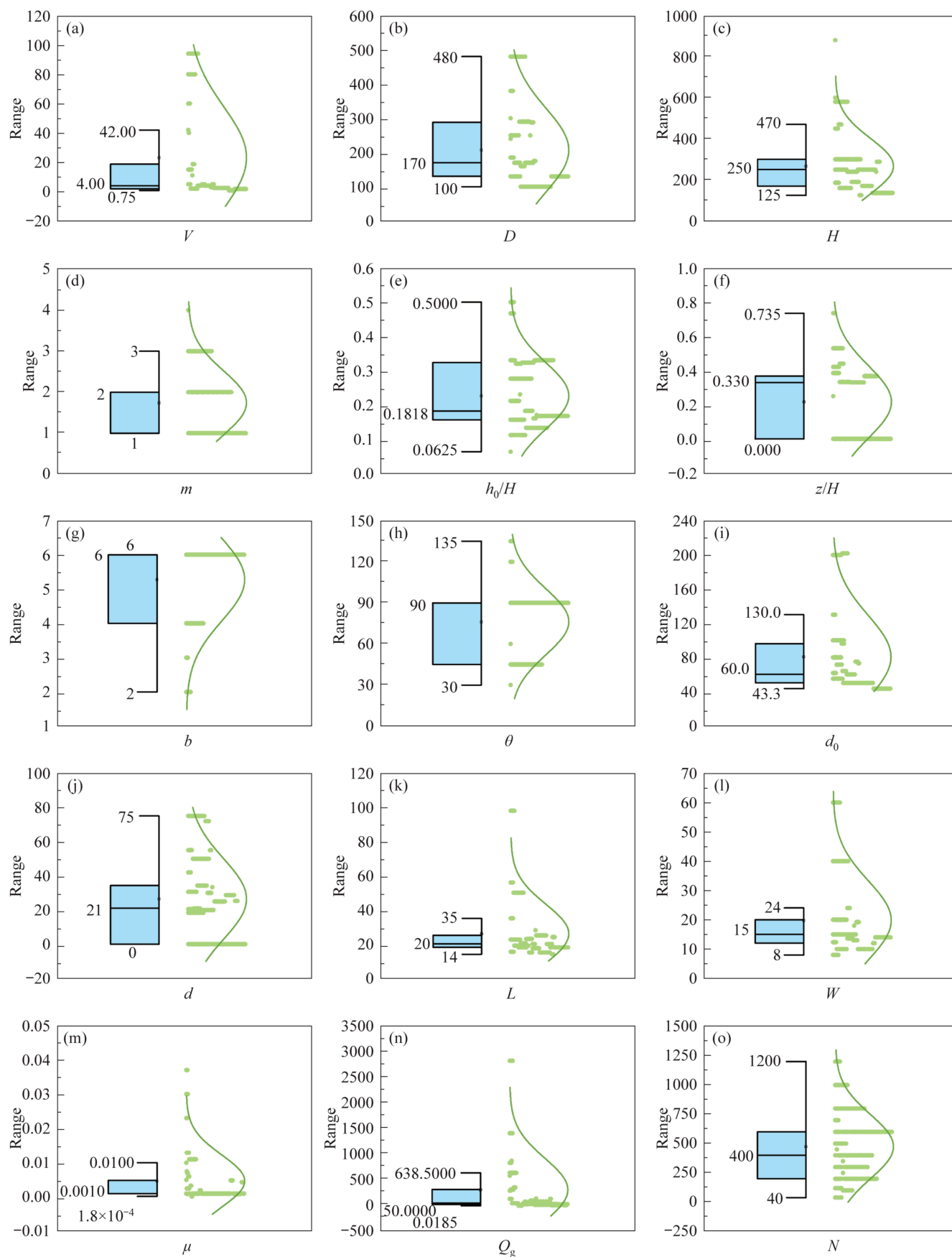


Fig. 3. Semi-box plot of statistical features related to Newtonian fluids: (a) V ; (b) D ; (c) H ; (d) m ; (e) h_0/H ; (f) z/H ; (g) b ; (h) θ ; (i) d_0 ; (j) d ; (k) L ; (l) W ; (m) μ ; (n) Q_g ; and (o) N . The data was depicted on the left side of the half-box plot, and its upper and lower edges of the half-box are the interquartile ranges, which is the range of values in the middle 50% of the dataset, and the lines in the box represent the values in the middle of the dataset. The distribution information of the data is on the right.

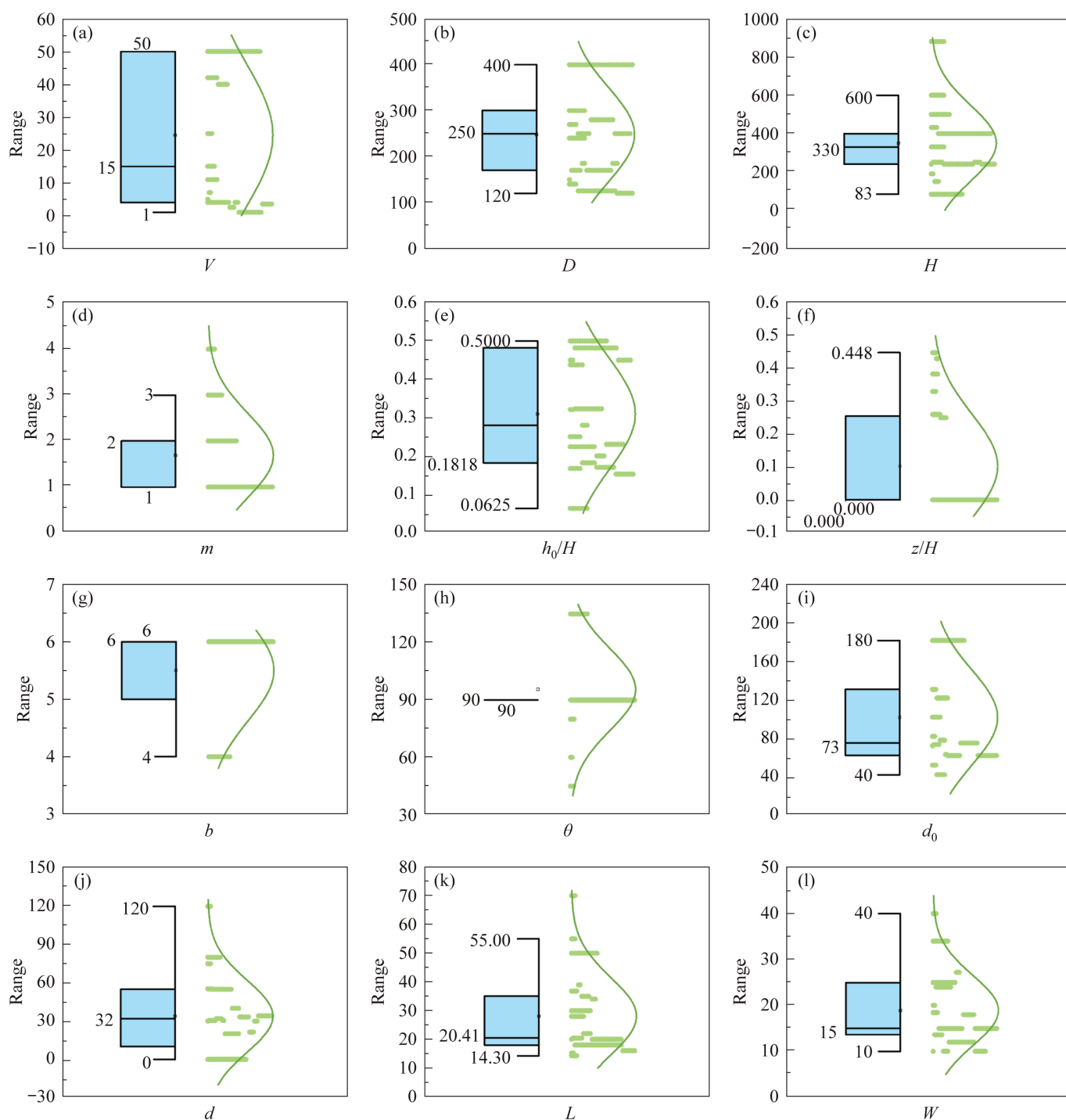


Fig. 4. Semi-box plot of statistical features related to non-Newtonian fluid: (a) V ; (b) D ; (c) H ; (d) m ; (e) h_0/H ; (f) z/H ; (g) b ; (h) θ ; (i) d_0 ; (j) d ; (k) L ; (l) W ; (m) K ; (n) n ; (o) Q_g ; and (p) N .

and the diverse operational parameters involved, a systematic multivariate analysis is warranted.

3.2. Pearson correlation coefficient analysis

Pearson Correlation Coefficients (PCCs) between two features are presented in Fig. 6. The corresponding significance levels (p -values) of the correlation coefficients are listed in Table S3 and S4

in Supplementary Materials ($p < 0.01$ indicates significant correlation). For Newtonian fluids, it is found that $V, D, H, b, d_0, d, \theta, \mu, Q_g$, and N all exhibit weak positive linear correlations with $k_L a$ ($0.2 < r < 0.4$), as shown in Fig. 6(a). It is worth noting that V, D, H, d_0, θ have relatively weak correlations with $k_L a$, which indicates that simple linear correlation analysis may be insufficient to elucidate their underlying mechanistic relationships with mass transfer.

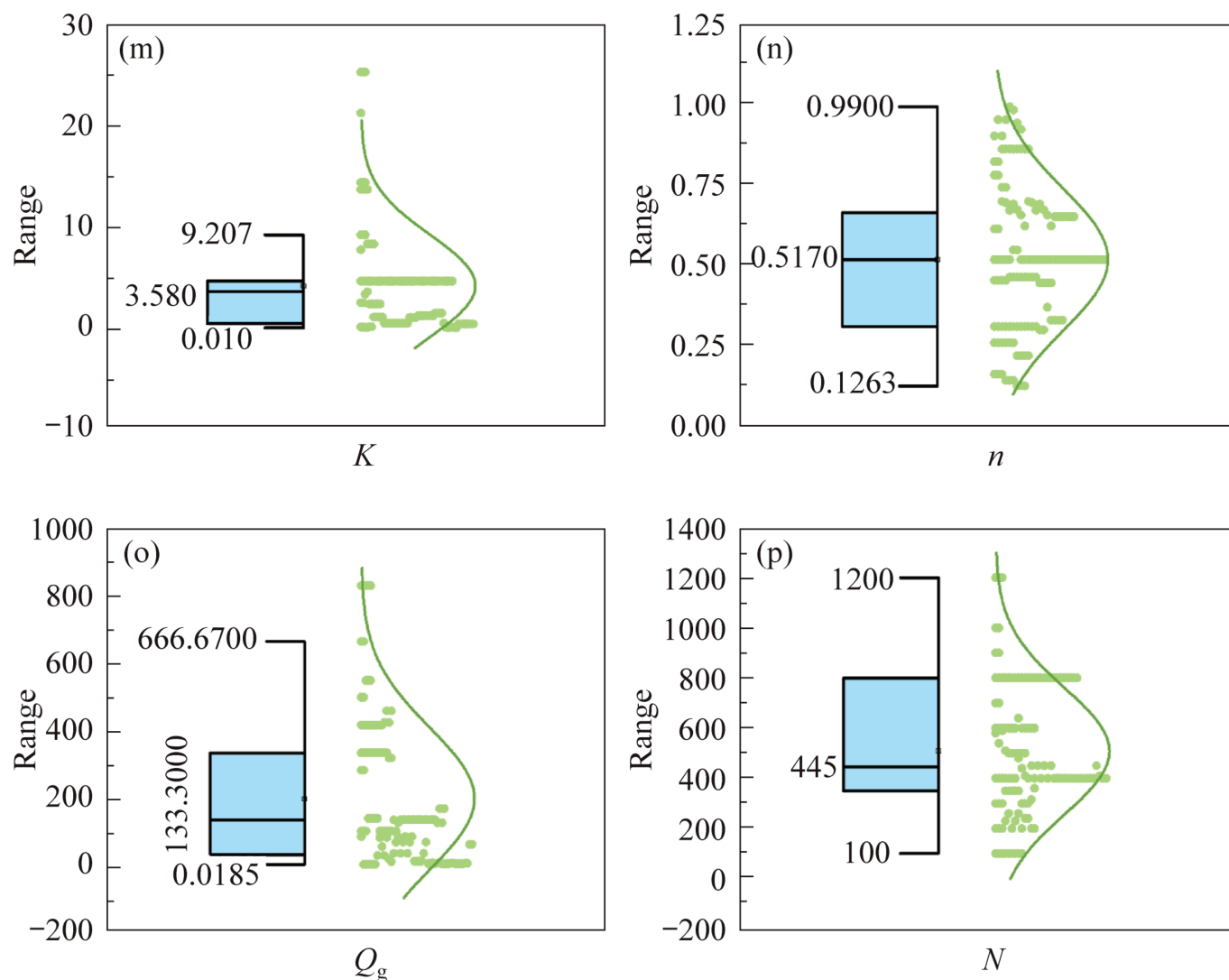


Fig. 4. (continued).

For non-Newtonian fluids, the top six features ranked by their negative correlation with $k_L a$ are V (-0.44) > D (-0.41) > d_0 (-0.39) > θ (-0.36) > L (-0.35) > W (-0.31), as presented in Fig. 6 (b). Interestingly, the positive and negative correlations between V , D , H , d_0 , θ and $k_L a$ are opposite to those of Newtonian fluids. These opposing trends likely reflect shear-rate-induced viscosity variations in larger reactors [55]. While this analysis reveals preliminary linear relationships between features and $k_L a$, it should be noted that PCC only captures linear correlation, whereas actual mass transfer involves complex nonlinearities. Therefore, to address these limitations and better inform bioreactor optimization, subsequent ML approaches will be employed to explore nonlinear relationships and multivariate interactions among features.

3.3. Evaluations of different ML models for $k_L a$

The predictive performance of four ML models (RF, XGBoost, CatBoost, and ANN) on the Newtonian fluids dataset is presented in Fig. 7, with corresponding evaluation indices (R^2 , AARD, RMSE, and MAE) detailed in Table 3. As shown in Fig. 7(a), RF exhibits poor performance on test sets, with R^2 and AARD of 0.8403% and

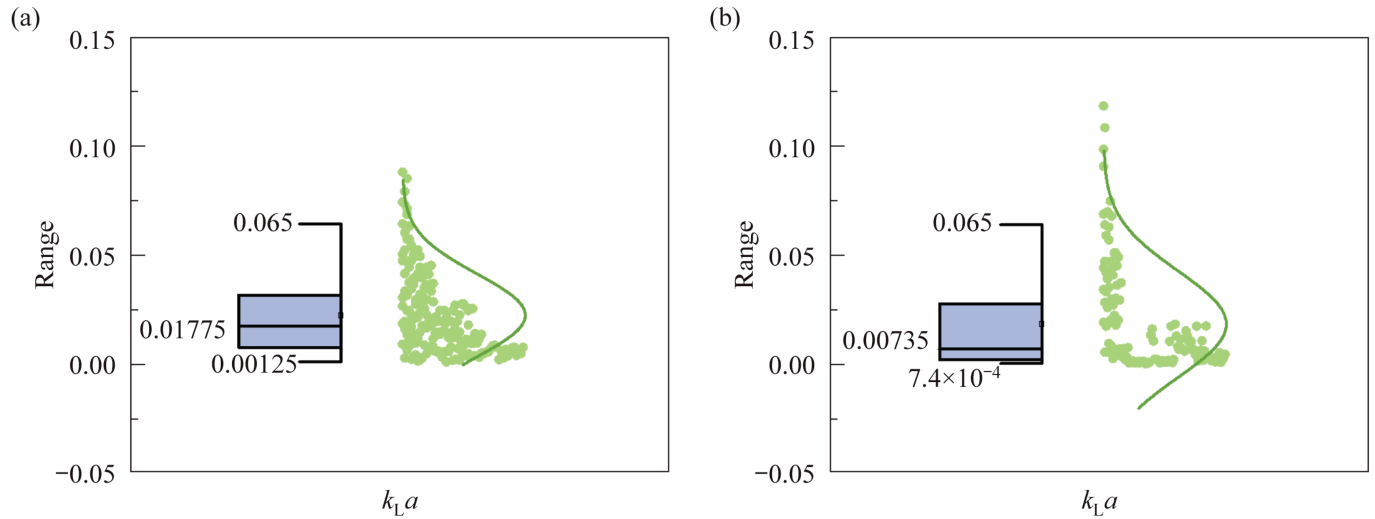
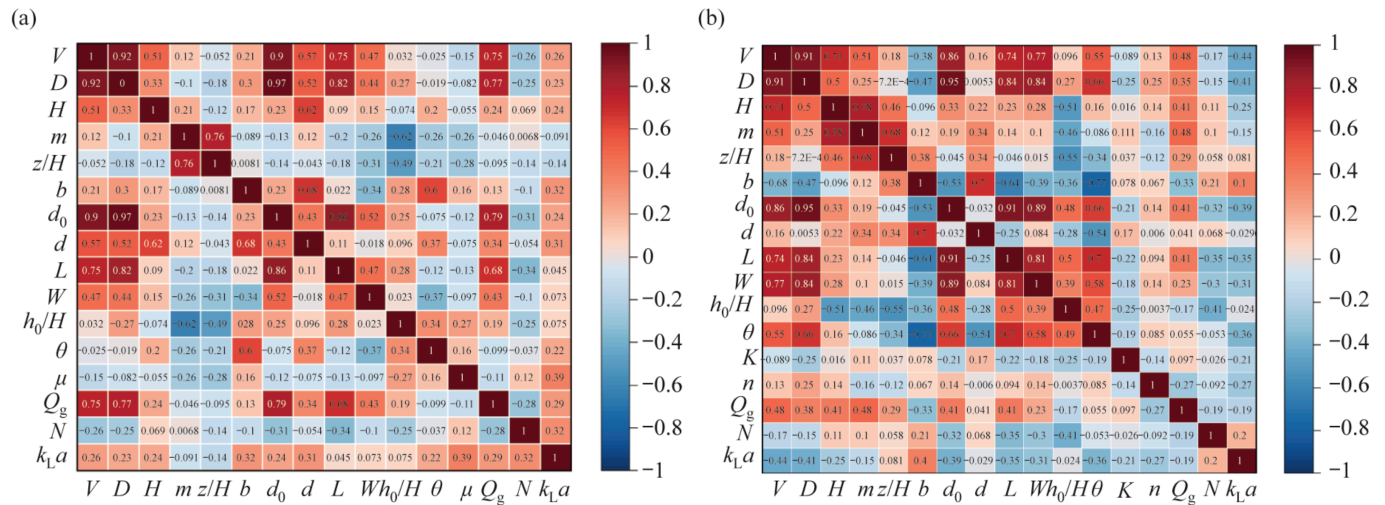
23.59%, respectively. Both XGBoost and ANN underperform relative to RF, as evidenced by their lower test R^2 (<0.8) and higher prediction errors (AARD = 24.28% and 35.84%), shown in Fig. 7(b) and (d). However, CatBoost delivers outstanding performance with training and test R^2 of 0.9985 and 0.8765, respectively (Fig. 7 (c)). This superior performance is further evidenced by its substantially lower AARD (2.6% for training and 18.84% for test). Moreover, CatBoost achieves the lowest RMSE and MAE values across both training and test sets (Table 3), confirming its superior predictive capability among all evaluated models.

Furthermore, Table 4 presents a comprehensive comparison between empirical equations and CatBoost for predicting $k_L a$ in Newtonian fluids. It is demonstrated that CatBoost adopted in this study has superior performance to empirical equations proposed by Vasconcelos *et al.* [56], Petříček *et al.* [57] and Haseidl *et al.* [58], while showing slightly lower accuracy compared with the models by Gabelle *et al.* [26] and Ali *et al.* [59] However, CatBoost incorporates an expanded set of features (e.g., gas flow rate, viscosity, and rotational speed) compared with empirical equations. This enhanced dimensionality enables the model to achieve greater prediction precision for $k_L a$ in Newtonian fluids compared with

Table 2

The frequently-selected combinations of features.

Features		Frequently-selected combinations		Unit
		Newtonian	non-Newtonian	
Structural parameters	V	1–20	4–50	L
	D	150–300	170–300	mm
	H	180–300	250–400	mm
	m	1, 2	1, 2	/
Impeller parameters	h_0/H	0.15–0.32	0.18–0.48	/
	z/H	0–0.35	0.2–0.48	/
	b	4, 5, 6	5, 6	/
	θ	45–90	90	(°)
	d_0	50–95	60–130	mm
	d	0–32	10–50	mm
	L	17–25	17–35	mm
	W	12–20	13–25	mm
	μ	0.001–0.005	/	Pa·s
Fluids properties	K	/	0.086–4.639	Pa·s ^{n}
	n	/	0.31–0.65	/
Operational parameters	Q_g	30–330	50–200	cm ³ ·s ^{−1}
	N	180–625	350–800	r·min ^{−1}

**Fig. 5.** Semi-box plot of $k_L a$: (a) Newtonian fluids; (b) non-Newtonian fluids.**Fig. 6.** Pearson correlation matrix among any two features (a) Newtonian fluids dataset; (b) non-Newtonian fluids dataset.

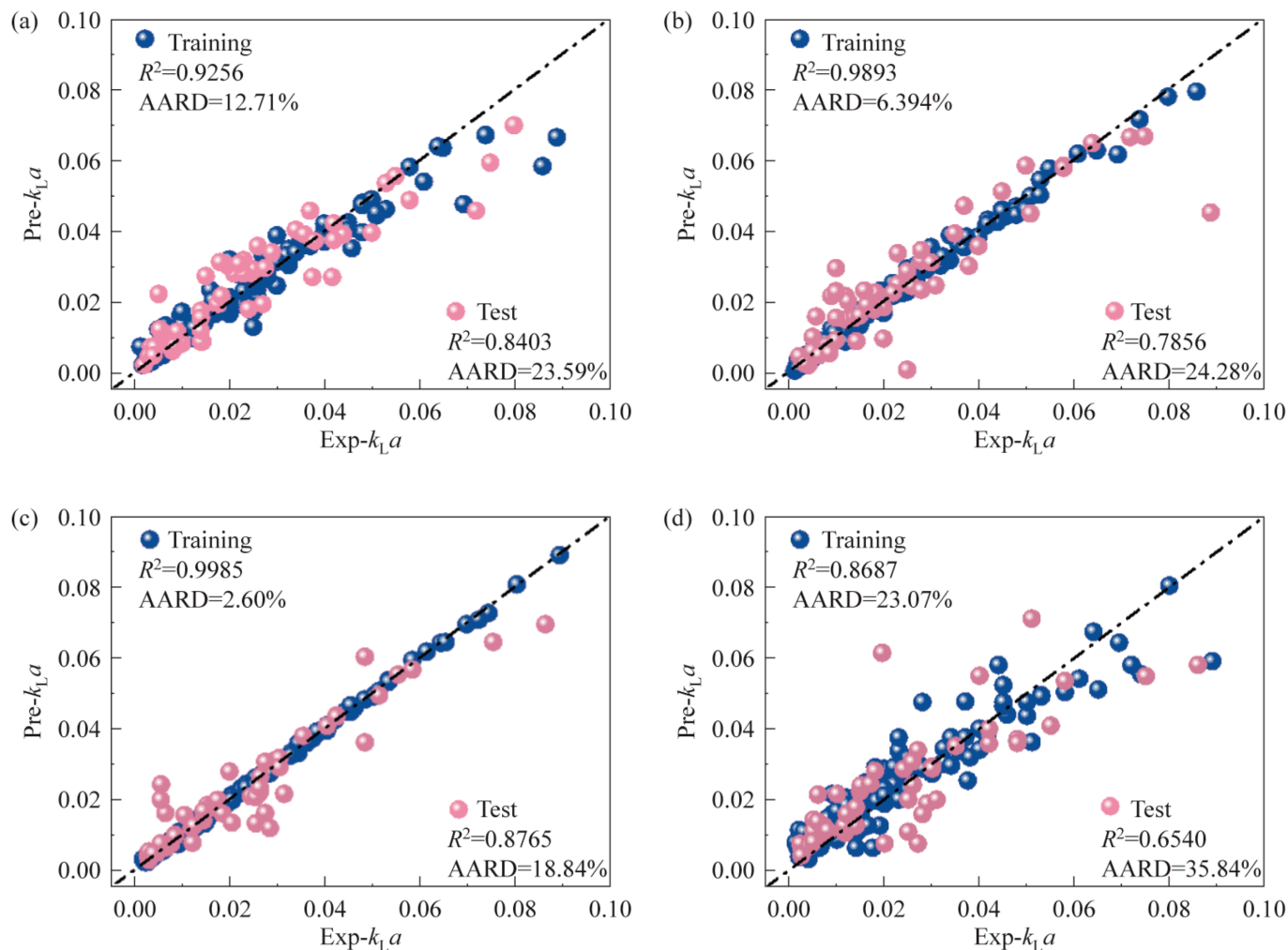


Fig. 7. Prediction performance of $k_L a$ between the true value and (a) RF, (b) XGBoost, (c) CatBoost, and (d) ANN model outputs for Newtonian fluids dataset.

Table 3

Evaluation indices of four ML models for Newtonian fluids dataset.

Model	Training				Test			
	RMSE	MAE	AARD/%	R^2	RMSE	MAE	AARD/%	R^2
RF	0.0052	0.0028	12.71	0.93	0.0078	0.0059	23.6	0.84
XGBoost	0.0019	0.0014	6.39	0.99	0.0093	0.0060	24.28	0.79
CatBoost	0.0007	0.0006	2.60	0.99	0.0066	0.0043	18.84	0.88
ANN	0.0070	0.0053	23.07	0.87	0.0110	0.0081	35.84	0.65

Table 4

Comparison of empirical equations and ML model for predicting $k_L a$ of Newtonian fluids dataset.

Empirical equations and ML model	Gas flow rate/ $\text{m}^3 \cdot \text{h}^{-1}$	Viscosity/ $\text{Pa} \cdot \text{s}$	Rotational speed/ $\text{r} \cdot \text{min}^{-1}$	Error/%	Ref
$k_L a = 0.02 \left(\frac{P_g}{V} \right)^{0.44} (V_s)^{0.448}$	1–3.1	0.001	250–550	11	[26]
$k_L a = 0.006 \left(\frac{P_g}{V} \right)^{0.66} (V_s)^{0.51}$	0–2.840	0.001	/	25	[56]
$k_L a = 0.007 \left(\frac{P_g}{V} \right)^{0.12} (U_g)^{0.004} (\mu_{ap})^{-0.079}$	0.480–0.600	0.001	800–900	15.2	[59]
$k_L a = 0.0027 (P_{TOT})^{0.78} (v_s)^{0.44}$	2.120–8.480	0.010	250–850	35	[57]
$k_L a = 0.011 \left(\frac{P_g/V}{1} \right)^{0.82} \left(\frac{Q/v}{g^2} \right)^{0.43}$	2.120–8.480	0.010	250–850	38	[57]
$k_L a = 0.02 (\varepsilon)^{0.45} (Q_g)^{0.57}$	1.670–6.680	0.001	100	20	[58]
CatBoost	0.001–10	0.001–0.037	40–1200	18.84	This work

empirical equations, with the additional advantage of broader applicability beyond water-based systems (e.g., glycerol solutions).

The predictive performance of four ML models (RF, XGBoost, CatBoost, and ANN) on the non-Newtonian fluids dataset is presented in Fig. 8, with corresponding evaluation indices (R^2 , AARD, RMSE, and MAE) detailed in Table 5. Compared with RF, CatBoost, and ANN models, XGBoost demonstrates superior predictive performance, achieving higher R^2 (training: 0.9957; test: 0.9746) and

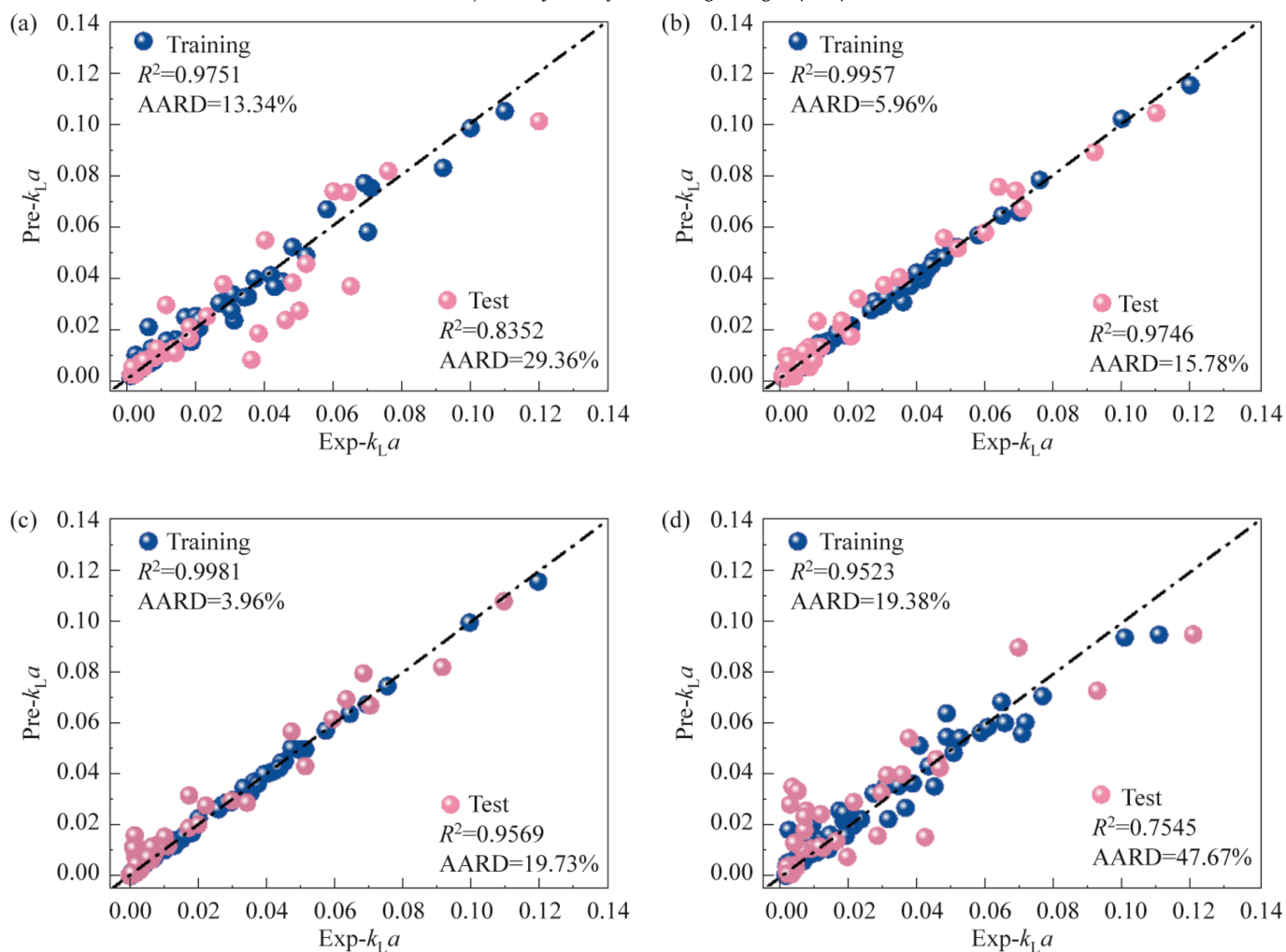


Fig. 8. Prediction performance of k_La between the true value and (a) RF, (b) XGBoost, (c) CatBoost, and (d) ANN model outputs for non-Newtonian fluids dataset.

Table 5

Evaluation indices of four ML models for non-Newtonian fluids dataset.

Model	Training				Test			
	RMSE	MAE	AARD/%	R^2	RMSE	MAE	AARD/%	R^2
RF	0.0036	0.0023	13.34	0.98	0.0109	0.0068	29.36	0.84
XGBoost	0.0015	0.0011	5.96	0.99	0.0044	0.0034	15.78	0.97
CatBoost	0.0010	0.0007	3.96	0.99	0.0058	0.0042	19.73	0.96
ANN	0.0052	0.0037	19.38	0.95	0.0130	0.0090	47.67	0.75

Table 6

Comparison of empirical equations and ML model for predicting k_La of non-Newtonian fluids dataset.

Empirical equations and ML model	Gas flow rate/ $m^3 \cdot h^{-1}$	Viscosity/ $Pa \cdot s$	Rotational speed/ $r \cdot min^{-1}$	Error/%	Ref
$k_La = 0.023 \left(\frac{P_g}{V} \right)^{0.44} v_s^{0.47} \left(1 + \frac{\mu_{ap}}{0.018} \right)^{-1}$	1	0.007–0.250	400	15	[26]
$k_La = 0.115 \left(\frac{P_{eff}}{V} \right)^{0.105} U_g^{0.688} \mu_{ap}^{-0.226}$	0.018–0.036 0.060–0.300	0.003–0.008	200–600	15	[62]
$k_La \left(\frac{v}{g^2} \right)^{\frac{1}{3}}$		0.097–0.269	200–400	20	[9]
$= 0.278 \left(\frac{p/V}{\rho(vg^4)^{\frac{1}{3}}} \right)^{1.05} \left(\frac{Q}{V} \left(\frac{v}{g^2} \right)^{\frac{1}{3}} \right)^{0.895} \left(d_0 \left(\frac{g}{v^2} \right)^{\frac{1}{3}} \right)^{-0.252}$					
$k_La = 4.64 \times 10^{-5} (P_g/V)^{0.48} v_s^{0.26} \mu_{ap}^{-0.92}$	1.200–2.400 1.008	0.028–0.051	350–500	15	[29]
$k_La = 0.42 \left(\frac{\Delta \rho v g}{\rho_L} \right)^{\frac{1}{3}} \left(\frac{D_L}{V} \right)^{\frac{1}{2}}$		0.020–0.088	40–110	24	[63]
XGBoost	0.018–2.400	0.550–41.890	100–1200	15.78	This work

lower AARD (training: 5.96%; test: 15.78%). RF and ANN exhibit poor performance, with R^2 below 0.9 on the test set. Moreover, CatBoost exhibits poor performance compared with XGBoost, with training indices ($R^2 = 0.99$, AARD = 3.96%) and test indices ($R^2 = 0.96$, AARD = 19.73%).

Furthermore, as shown in Table 5, although XGBoost and CatBoost have similar RMSE and MAE values on the training set, XGBoost demonstrates significantly better performance on the

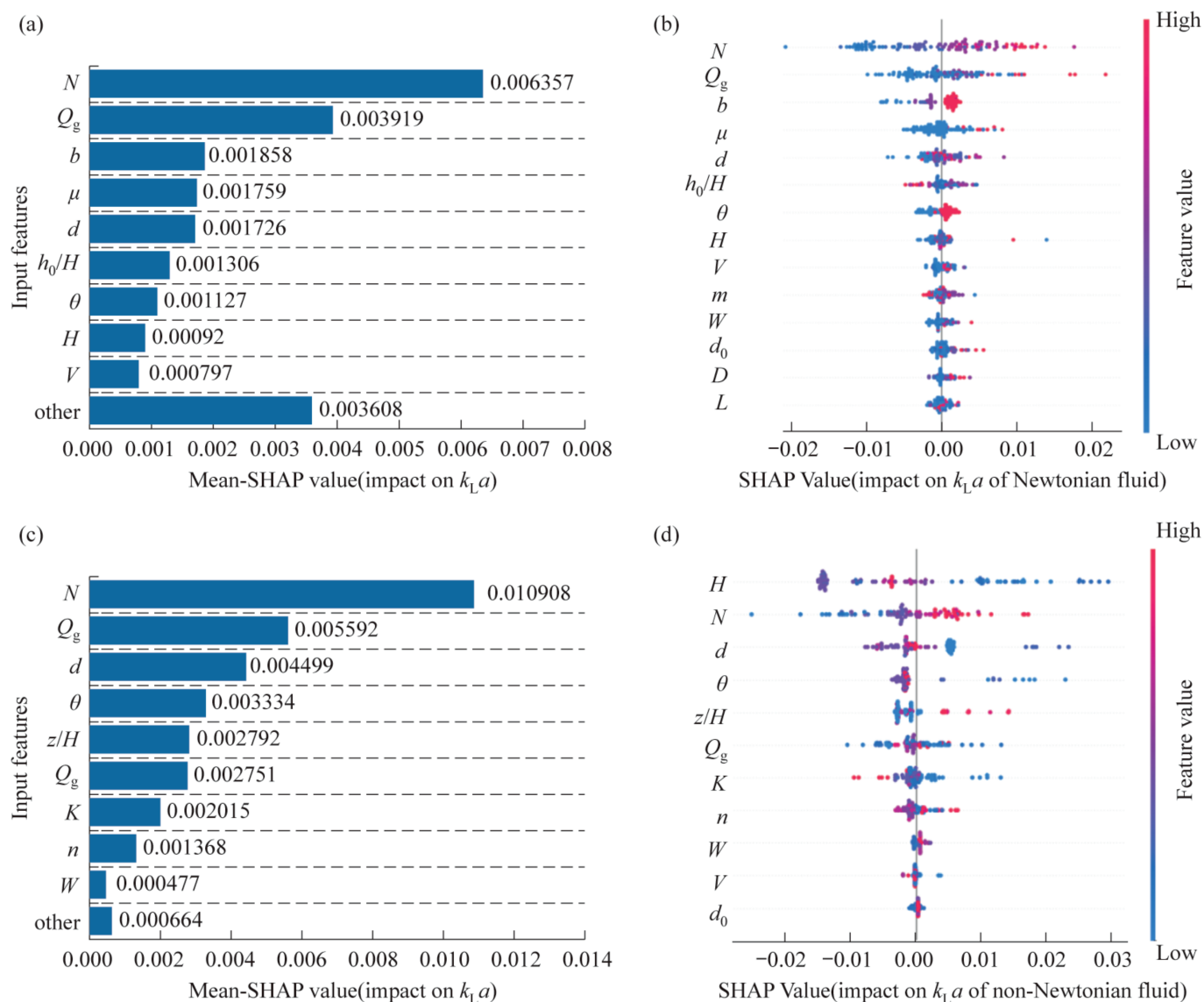


Fig. 9. Model interpretation of CatBoost and XGBoost bases on $k_L a$. (a) Permutation importance of input features for Newtonian fluids dataset, (b) SHAP summary plot showing feature importance for Newtonian fluids dataset, (c) Permutation importance of input features for non-Newtonian fluids dataset, (d) SHAP summary plot showing feature importance for non-Newtonian fluids dataset.

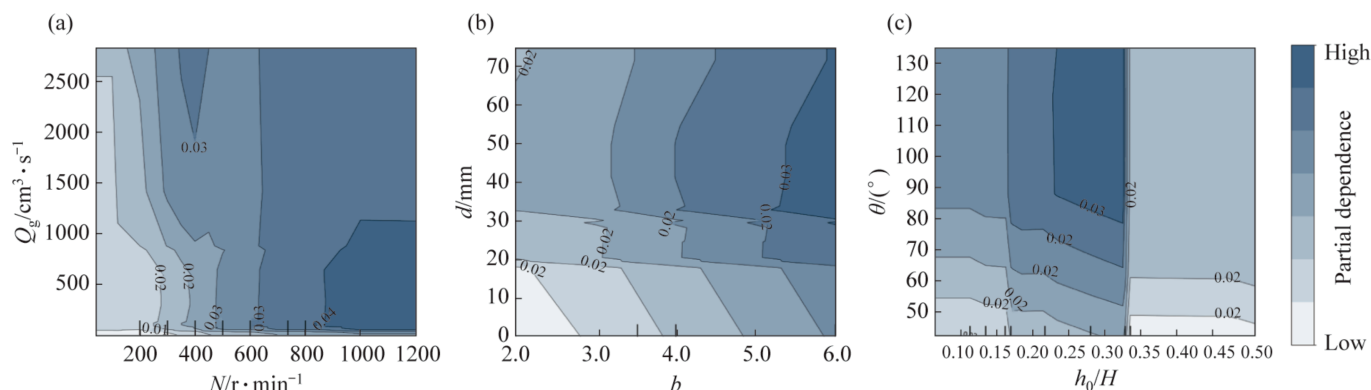


Fig. 10. 2D PDP of the representative features for Newtonian fluids, (a) Operational parameters (N and Q_g represent rotational speed and gas flow rate); (b) impeller parameters (b and d represent number of blades and turbine diameter); (c) impeller parameters (h_0/H and θ represent position and inclination angle).

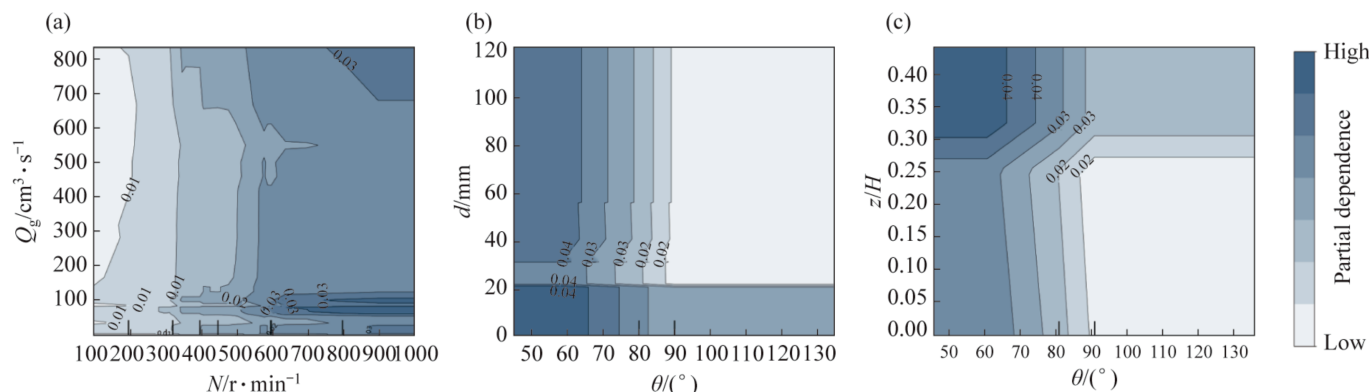


Fig. 11. 2D PDP of the representative features for non-Newtonian fluids, (a) Operational parameters (N and Q_g represent rotational speed and gas flow rate); (b) impeller parameters (d and θ represent turbine diameter and inclination angle); (c) impeller parameters (z/H and θ represent clearance ratio and inclination angle).

testing set. This suggests that XGBoost possesses superior generalization capabilities, making it the most effective model for predicting k_{La} among these ML models in non-Newtonian fluids. XGBoost's outstanding predictive performance stems from its ensemble learning approach, which combines weak learners through additive training to mitigate overfitting while maintaining computational efficiency [60]. Additionally, XGBoost utilizes a second-order Taylor expansion to approximate the loss function, which optimizes the model training process by calculating the gradient and Hessian matrix to determine the best tree split points. This methodological advantage enhances training efficiency and predictive accuracy, particularly when handling complex datasets [61].

For non-Newtonian fluids, Table 6 provides a comprehensive comparison between empirical equations and XGBoost for predicting k_{La} . XGBoost adopted in this study achieves comparable prediction accuracy to empirical equations while offering significantly wider parameters coverage. The complex rheological properties of non-Newtonian fluids result in shear-rate-dependent viscosity behavior, which cannot be adequately characterized by the constant apparent viscosity (μ_{ap}) typically employed in empirical equations. This fundamental limitation significantly restricts the applicability of such empirical equations. In contrast, XGBoost incorporates key rheological features (K , n) of non-Newtonian fluids for describing viscosity variations, thereby overcoming the inability of empirical equations in characterizing shear-thinning behavior. These advantages collectively demonstrate that XGBoost achieves superior accuracy and broader applicability in predicting k_{La} for non-Newtonian fluids compared with empirical equations.

Table 7

The applicable conditions for PBTD impeller.

Applicable conditions	Lower bound	Upper bound
V/L	60	100
H/mm	600	880

Table 8

The applicable conditions of the PBTU impeller.

Applicable conditions	Lower bound	Upper bound
H/mm	83	245
n	0.1263	0.99
$K/Pa \cdot s^n$	0.01	0.04

3.4. Feature importance analysis

To investigate the qualitative relationship between the input features and the model output (k_{La}), permutation importance and SHAP values are employed to analyze the relative impact of each feature. Fig. 9 illustrates the quantitative impact of various features on k_{La} for both Newtonian and non-Newtonian fluids datasets. Fig. 9 (a) illustrates distinct impacts of input features on k_{La} , revealing that N has the dominant effect, followed closely by Q_g , b , μ , and d , while h_0/H , θ , H , and V exhibit relatively weaker effects. In Fig. 9(b), the relative importance of features on k_{La} decreases from N to Q_g to b , with N and b both exhibiting positive relationships with k_{La} . At low Q_g values, the impact on k_{La} remains ambiguous, but a pronounced positive correlation emerges as Q_g increases. The increased rotational speed enhances k_{La} through fundamental mass transfer mechanisms: intensified liquid turbulence thins the boundary layer per the two-film theory, while simultaneously promoting bubble fragmentation to expand interfacial area and accelerating surface renewal dynamics. These synergistic effects compound to dramatically improve gas–liquid mass transfer efficiency. However, the regulation of rotational speed in actual fermentation needs to consider the tolerance threshold of microorganisms [64,65]. From a mechanistic perspective, Q_g influences k_{La} via bubble dynamics. Below critical flow, coalescence reduces interfacial area, while above, turbulent kinetic energy suppresses coalescence, promotes breakup, and accelerates renewal, causing exponential gas–liquid area growth and high k_{La} sensitivity to Q_g [66].

For non-Newtonian fluids, Fig. 9(c) illustrates that H has the most significant effect on k_{La} , followed by N , d , θ , z/H , Q_g , K , and n . Notably, according to Fig. 9(d), H , d , and θ show negative correlations with k_{La} , and N demonstrates a positive correlation. From boundary layer mass transfer theory, decreasing reactor height reduces liquid static pressure, lowering wall viscosity and thus significantly decreasing mass transfer resistance ($1/k_L$) [67–69]. A higher H requires the impeller to overcome the resistance from a longer liquid column for driving gas–liquid mixing, leading to a decrease in axial flow velocity [68]. In addition, an increase in H is accompanied by an increase in V , causing significant spatial heterogeneity in the internal flow field (e.g., flow pattern, shear rate, and mass transfer efficiency), which makes the regulation of mass transfer more complex [70].

3.5. 2D partial dependence of representative features

To elucidate the complex relationships between representative features and k_{La} for both Newtonian and non-Newtonian fluids,

representative features, including operational parameters (Q_g and N) and impeller parameters (b , d , θ , z/H and h_0/H), are systematically visualized on the 2D PDP in Figs. 10 and 11. Fig. 10 presents the 2D PDP for Newtonian fluid. As illustrated in Fig. 10(a), the analysis reveals a significant combined effect of N and Q_g on $k_L a$. A distinct high-dependency region (dark blue) exists when N exceeds $850 \text{ r} \cdot \text{min}^{-1}$ coupled with Q_g ranging from 65 to $1135 \text{ cm}^3 \cdot \text{s}^{-1}$, indicating superior mass transfer performance. This enhanced performance arises from synergistic turbulence by rotational speed and gas velocity. The two-film theory indicates that the thickness of the mass transfer boundary layer is negatively correlated with the Reynolds and Froude number, and the coupling of the two features makes the reduction range of boundary layer exceed the linear superposition of single feature [18,71].

As shown in Fig. 10(b), the 2D PDP analysis of impeller parameters reveal a high-response region when d exceeds 3 mm with six blades ($b = 6$). This configuration maximizes $k_L a$ by generating superior shear forces that enhance bubble fragmentation and gas–liquid interfacial area compared with impellers with fewer blades [72]. The analysis further demonstrates that reduced impeller diameters decrease mixing region effectiveness, consequently impairing bubble dispersion efficiency [28]. Fig. 10(c) reveals a high-dependence region characterized by θ exceeding 90° and h_0/H between 0.23 and 0.33. It is demonstrated that such inclined blade configurations significantly intensify the near-bottom flow field, enhancing local turbulence [73]. The vertical position of the bottom impeller significantly impacts bubble formation and breakup. An impeller positioned too high in the bioreactor reduces the upward acceleration of bubbles, while one placed too low leads to non-uniform axial mixing and incomplete fluid homogenization [74]. This optimal positioning enhances system performance through improved bubble dispersion, expanded gas–liquid interfacial area, and increased mass transfer efficiency.

Based on hydrodynamic analysis, the optimized impeller configuration adopts a pitched blade turbine down (PBTd) design with 120° – 135° blade angles, located at a normalized height h_0/H between 0.23 and 0.33 relative to the bioreactor bottom. This specific geometry achieves an optimal balance between axial flow development and radial mixing intensity while maintaining operational suitability for 60–100 L pilot-scale bioreactors (H : 600–880 mm) (Table 7), as confirmed by 2D PDP analysis (Fig. S5). The determined parameters ensure efficient mass transfer region formation across the specified reactor scale range.

Fig. 11 presents 2D PDP for non-Newtonian fluids dataset. As shown in Fig. 11(a), high mass transfer performance occurs at higher N ($> 600 \text{ r} \cdot \text{min}^{-1}$) and specific Q_g (80 – $120 \text{ cm}^3 \cdot \text{s}^{-1}$). The enhanced mass transfer performance stems primarily from the shear-induced viscosity reduction at higher N , which concurrently decreases both the liquid-film thickness and the associated mass transfer resistance [75]. Generally, high Q_g cause bubble aggregation, shrinking the mass-transfer area, while Q_g may result in inadequate gas supply for efficient mass transfer. Fig. 11(b) presents the 2D PDP analysis of impeller parameters, revealing a high-response region when θ ranges from 45° to 60° and d is below 20 mm. This indicates that impellers with $\theta < 90^\circ$ and $d < 20$ mm significantly enhance mass transfer efficiency, while larger turbine diameters impede bubble rise. Notably, impellers with $\theta < 90^\circ$ show superior performance by generating strong upward jets above the impeller, which not only control fluid flow direction and velocity but also effectively direct bubble movement [76]. Fig. 11(c) identifies $z/H > 0.3$ as the high-dependence region for optimal mass transfer performance, where this configuration minimizes axial mass-transfer resistance and achieves globally optimized mixing [77]. However, both excessive and insufficient z/H can result in

non-uniformity mixing, causing localized inefficiencies near the stirred bioreactor ends or central region [78].

Consequently, a pitched blade turbine up (PBTU) impeller was designed with inclined blades (45° or 60°), a smaller turbine diameter, and $z/H > 0.3$. For non-Newtonian fluids, high N and low Q_g represent optimal conditions, significantly enhancing mass transfer performance. To assess the PBTU impeller's operational range, partial dependence analysis (Fig. S6) was analyzed to establish boundary conditions for H and fluid properties relative to θ (Table 8). The PBTU impeller demonstrates superior mass transfer performance for shear-thinning fluids within specific parameters (H : 83–245 mm; K : 0.01–0.04).

4. Conclusions

This study presents a ML approach to predict $k_L a$ in gas-liquid stirred bioreactors for Newtonian and non-Newtonian fluids. The results show that CatBoost has superior accuracy compared with empirical equations while extending applicability to diverse Newtonian fluids beyond water. For non-Newtonian fluids, XGBoost enhances predictive capability by incorporating rheological features (K , n), overcoming empirical equation's limitations in characterizing shear-thinning behavior. Feature importance analysis demonstrates that N is the dominant feature positively correlated with $k_L a$ in Newtonian fluids, while H emerges as the key feature showing negative correlation in non-Newtonian fluids. 2D PDP analysis reveals optimal operational parameters that include high N , low Q_g , and a PBTd impeller with $0.23 < h_0/H < 0.33$ for maximizing $k_L a$ in pilot-scale Newtonian bioreactors (V : 60–100 L, H : 600–880 mm). For non-Newtonian fluids, a PBTU impeller with $\theta < 90^\circ$ and $d < 20$ mm significantly improves $k_L a$, particularly during processing of shear-thinning fluids ($K < 0.04$) at H ranging 83 to 245 mm. This ML approach efficiently optimizes mass transfer in gas–liquid stirred bioreactors, demonstrating robust predictive capability for both laboratory-scale and pilot-scale systems handling non-Newtonian fluids with shear-thinning behavior. These findings provide quantitative design criteria for scaling up bioreactor structure, demonstrating how impeller and operational parameters must be precisely tailored to fluid rheological properties to achieve optimal mass transfer efficiency.

CRediT Authorship Contribution Statement

Feifei Chen: Writing – original draft, Validation, Software, Methodology, Investigation, Data curation. Zhenyuan Xiao: Software, Data curation. Zhongfan Luo: Software, Methodology. Peng Jiang: Software, Methodology. Jingjing Chen: Writing – review & editing, Supervision, Funding acquisition, Conceptualization. Yuanhui Ji: Conceptualization. Jiahua Zhu: Conceptualization. Xiaohua Lu: Conceptualization. Liwen Mu: Writing – review & editing, Resources, Methodology, Funding acquisition, Conceptualization.

Data Availability

Data will be made available on request.

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.

Acknowledgements

This research is supported by the National Natural Science Foundation of China (22494713, 22178160, 22327809 and 22208141), Natural Science Foundation of Jiangsu Province, China (BK20220349).

Supplementary Material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cjche.2025.06.010>.

References

- [1] Z.M. Kang, L.F. Feng, J.J. Wang, Optimization of a gas–liquid dual-impeller stirred tank based on deep learning with a small data set from CFD simulation, *Ind. Eng. Chem. Res.* 63 (1) (2024) 843–855.
- [2] A.Q. Li, Y. Yao, X.Y. Tang, P.Q. Liu, Q. Zhang, Q. Li, P. Li, F. Zhang, Y.D. Wang, C. Y. Tao, Z.H. Liu, Experimental and computational investigation of chaotic advection mixing in laminar rectangular stirred tanks, *Chem. Eng. J.* 485 (2024) 149956.
- [3] M. Jamshidzadeh, A. Kazemzadeh, F. Ein-Mozaffari, A. Lohi, Analysis of power consumption for gas dispersion in Non-Newtonian fluids with a coaxial mixer: new correlations for Reynolds and power numbers, *Chem. Eng. J.* 401 (2020) 126002.
- [4] A. Kazemzadeh, C. Elias, M. Tamer, A. Lohi, F. Ein-Mozaffari, Mass transfer in a single-use angled-shaft aerated stirred bioreactor applicable for animal cell culture, *Chem. Eng. Sci.* 219 (2020) 115606.
- [5] Y. Kawase, B. Halard, M. Moo-Young, Liquid-phase mass transfer coefficients in bioreactors, *Biotechnol. Bioeng.* 39 (11) (1992) 1133–1140.
- [6] F. Garcia-Ochoa, E. Gomez, Bioreactor scale-up and oxygen transfer rate in microbial processes: an overview, *Biotechnol. Adv.* 27 (2) (2009) 153–176.
- [7] H. Chen, Z. Chen, X.Q. Wang, X.G. Yi, X.B. Zhang, Z.H. Luo, Mass transfer and bubble hydrodynamics in stirred tank with multiple properties fluid via a CFD-PBM method, *Can. J. Chem. Eng.* 102 (11) (2024) 4038–4054.
- [8] N. Kumar, A. Bansal, R. Gupta, Shear rate and mass transfer coefficient in internal loop airlift reactors involving Non-Newtonian fluids, *Chem. Eng. Res. Des.* 136 (2018) 315–323.
- [9] X.M. Liu, J. Wan, J.N. Sun, L. Zhang, F. Zhang, Z.B. Zhang, X.Y. Li, Z. Zhou, Effect of bubble morphology and behavior on power consumption in Non-Newtonian fluids' aeration process, *Chin. J. Chem. Eng.* 65 (2024) 243–254.
- [10] Y.H. Du, L.L. Tong, Y. Wang, M.Z. Liu, L. Yuan, X.Y. Mu, S.J. He, S.X. Wei, Y.D. Zhang, Z.L. Chen, Z.D. Zhang, D.S. Guo, Development of a kinetics-integrated CFD model for the industrial scale-up of DHA fermentation using *Schizochytrium* sp., *AIChE J.* 68 (9) (2022) e17750.
- [11] J.Y. Xia, G. Wang, M. Fan, M. Chen, Z.Y. Wang, Y.P. Zhuang, Understanding the scale-up of fermentation processes from the viewpoint of the flow field in bioreactors and the physiological response of strains, *Chin. J. Chem. Eng.* 30 (2021) 178–184.
- [12] P. Moilanen, M. Laakkonen, O. Visuri, V. Alopaes, J. Aittamaa, Modelling mass transfer in an aerated 0.2m³ vessel agitated by rushton, phasejet and combijet impellers, *Chem. Eng. J.* 142 (1) (2008) 95–108.
- [13] M.H. Xie, J.Y. Xia, Z. Zhou, G.Z. Zhou, J. Chu, Y.P. Zhuang, S.L. Zhang, H. Noorman, Power consumption, local and average volumetric mass transfer coefficient in multiple-impeller stirred bioreactors for xanthan gum solutions, *Chem. Eng. Sci.* 106 (2014) 144–156.
- [14] S. Bun, K. Wongwailikhit, N. Chawaloeshonsiya, J. Lohwacharin, P. Ham, P. Painmanakul, Development of modified airlift reactor (MALR) for improving oxygen transfer: optimize design and operation condition using 'design of experiment' methodology, *Environ. Technol.* 41 (20) (2020) 2670–2682.
- [15] A. Pan, M.H. Xie, J.Y. Xia, J. Chu, Y.P. Zhuang, Gas-liquid mass transfer studies: the influence of single- and double-impeller configurations in stirred tanks, *Korean J. Chem. Eng.* 35 (1) (2018) 61–72.
- [16] T.J. Bondancia, L.J. Corrêa, A.J.G. Cruz, A.C. Badino, L.H.C. Mattoso, J.M. Marconcini, C.S. Farinas, Enzymatic production of cellulose nanofibers and sugars in a stirred-tank reactor: determination of impeller speed, power consumption, and rheological behavior, *Cellulose* 25 (8) (2018) 4499–4511.
- [17] Y.F. Zhou, P.X. Kang, Z.L. Huang, P. Yan, J.Y. Sun, J.D. Wang, Y.R. Yang, Experimental measurement and theoretical analysis on bubble dynamic behaviors in a gas-liquid bubble column, *Chem. Eng. Sci.* 211 (2020) 115295.
- [18] H. Ali, J. Solsvik, Axial distributions of bubble–liquid mass transfer coefficient in laboratory-scale stirred tank with viscous newtonian and non-Newtonian fluids 32 (12) (2020) 123308.
- [19] L. Niño, R. Gelves, J. Solsvik, Viscous effects on gas-liquid hydrodynamics for bubble size determinations in different newtonian and non-Newtonian fluids using a CFD-PBM model, *Chem. Eng. Sci.* 282 (2023) 119324.
- [20] S. Mishra, V. Kumar, J. Sarkar, A.S. Rathore, Mixing and mass transfer in production scale Mammalian cell culture reactor using coupled CFD-Species transport-PBM validation, *Chem. Eng. Sci.* 267 (2023) 118323.
- [21] A. Rahimzadeh, F. Ein-Mozaffari, A. Lohi, New insights into the gas dispersion and mass transfer in shear-thinning fluids inside an aerated coaxial mixer via analysis of flow hydrodynamics and shear environment, *Ind. Eng. Chem. Res.* 61 (10) (2022) 3713–3728.
- [22] X. Xu, L. Wang, H.L. Wang, H.L. Liu, Q. Yang, Circulating jet for enhancing the mass transfer in a gas–liquid stirred tank reactor, *AIChE J.* 68 (1) (2022) e17392.
- [23] B.Q. Liu, Q. Xiao, N. Sun, P.F. Gao, F.Y. Fan, B. Sundén, Effect of gas distributor on gas–liquid dispersion and mass transfer characteristics in stirred tank, *Chem. Eng. Res. Des.* 145 (2019) 314–322.
- [24] S.S. de Jesus, J.M. Neto, A. Santana, R.M. Filho, Influence of impeller type on hydrodynamics and gas-liquid mass-transfer in stirred airlift bioreactor, *AIChE J.* 61 (10) (2015) 3159–3171.
- [25] A. Rahimzadeh, F. Ein-Mozaffari, A. Lohi, A methodical approach to scaling up an aerated coaxial mixer containing a shear-thinning fluid: effect of the fluid rheology, *Ind. Eng. Chem. Res.* 62 (21) (2023) 8454–8476.
- [26] J.C. Gabelle, F. Augier, A. Carvalho, R. Rousset, J. Morchain, Effect of tank size on kLa and mixing time in aerated stirred reactors with Non-Newtonian fluids, *Can. J. Chem. Eng.* 89 (5) (2011) 1139–1153.
- [27] X. Xiao, J.Y. Yin, J. Xu, T. Tat, J. Chen, Advances in machine learning for wearable sensors, *ACS Nano* 18 (34) (2024) 22734–22751.
- [28] X.Z. Zhao, H.A. Fan, G.B. Lin, Z.C. Fang, W.L. Yang, M. Li, J.H. Wang, X.Y. Lu, B.L. Li, K.J. Wu, J. Fu, Multi-objective optimization of radially stirred tank based on CFD and machine learning, *AIChE J.* 70 (3) (2024) e18324.
- [29] P.L. Barros, F. Ein-Mozaffari, A. Lohi, S. Upreti, Exploiting the prediction of mass transfer performance in aerated coaxial mixers containing biopolymer solutions using empirical correlations and neural networks, *Can. J. Chem. Eng.* 103 (3) (2025) 1258–1275.
- [30] I. Andrade Cruz, W. Chuenchart, F. Long, K.C. Surendra, L. Renata Santos Andrade, M. Bilal, H. Liu, R. Tavares Figueiredo, S.K. Khanal, L. Fernando Romanholo Ferreira, Application of machine learning in anaerobic digestion: perspectives and challenges, *Bioresour. Technol.* 345 (2022) 126433.
- [31] A. Shehadeh, O. Alshboul, R.E. Al Mamlook, O. Hamedat, Machine learning models for predicting the residual value of heavy construction equipment: an evaluation of modified decision tree, LightGBM, and XGBoost regression, *Autom. Constr.* 129 (2021) 103827.
- [32] L.T. Zhu, X.Z. Chen, B. Ouyang, W.C. Yan, H. Lei, Z. Chen, Z.H. Luo, Review of machine learning for hydrodynamics, transport, and reactions in multiphase flows and reactors, *Ind. Eng. Chem. Res.* 61 (28) (2022) 9901–9949.
- [33] S.L. Brunton, B.R. Noack, P. Koumoutsakos, Machine learning for fluid mechanics, *Annu. Rev. Fluid Mech.* 52 (2020) 477–508.
- [34] R. Chen, L.L. Luo, X.D. Wang, B.B. Ren, D.K. Guo, S. Zhu, Knowing the unknowns: network traffic detection with open-set semi-supervised learning, *Comput. Netw.* 251 (2024) 110630.
- [35] T.Z. Si, F.Z. He, Z. Zhang, Y.S. Duan, Hybrid contrastive learning for unsupervised person re-identification, *IEEE Trans. Multimed.* 25 (2022) 4323–4334.
- [36] X.S. Feng, H. Zahed, J. Onwuka, M.E.J. Callister, M. Johansson, R. Etzioni, H.A. Robbins, Cancer stage compared with mortality as end points in randomized clinical trials of cancer screening, *JAMA* 331 (22) (2024) 1910–1917.
- [37] S. Ali Shah, S.T. Ai, H.X. Yuan, Predicting water level fluctuations in glacier-fed lakes by ensembling individual models into a quad-meta model, *Eng. Appl. Comput. Fluid. Mech.* 19 (1) (2025), <https://doi.org/10.1080/19942060.2024.2449124>.
- [38] T.K. Ho, Random Decision Forest, In: Proceedings of 3rd International Conference on Document Analysis and Recognition, IEEE, Montreal, QC, Canada, 1995.
- [39] L. Breiman, Random forests, *Mach. Learn.* 45 (2001) 5–32.
- [40] Q.L. Lu, H. Zhang, R. Fan, Y.H. Wan, J.Q. Luo, Machine learning-based Bayesian optimization facilitates ultrafiltration process design for efficient protein purification, *Sep. Purif. Technol.* 363 (2025) 132122.
- [41] M.F. Mahmoud, M. Arabi, S. Pallickara, Harnessing ensemble machine learning models for improved salinity prediction in large river basin scales, *J. Hydrol.* 652 (2025) 132691.
- [42] H. Lamane, L. Mouhir, R. Moussadek, B. Baghdad, O. Kisi, A. El Bilali, Interpreting machine learning models based on SHAP values in predicting suspended sediment concentration, *Int. J. Sediment Res.* 40 (1) (2025) 91–107.
- [43] L. Shen, Y.P. Jin, A.X. Pan, K. Wang, R.Z. Ye, Y.K. Lin, S. Anwar, W.C. Xia, M. Zhou, X.G. Guo, Machine learning-based predictive models for perioperative major adverse cardiovascular events in patients with stable coronary artery disease undergoing noncardiac surgery, *Comput. Methods Progr. Biomed.* 260 (2025) 108561.
- [44] B. Yuan, Z.L. Yang, P. Wu, X.B. Yin, C.J. Liu, F.J. Sun, J. He, W. Jiang, Efficient treatment of chromium-containing wastewater based on auxiliary intelligent model with rapid-response adsorbents, *Sep. Purif. Technol.* 363 (2025) 132037.
- [45] M. Maraschin, T.D. Metzka Lanza, N.P. Gonçalves Salau, Predictions of heat of combustion and formation by interpretable machine learning algorithms, *Fuel* 390 (2025) 134699.
- [46] M.F. Tahir, M.Z. Yousaf, A. Tzes, M.S. El Moursi, T.H.M. El-Fouly, Enhanced solar photovoltaic power prediction using diverse machine learning

- algorithms with hyperparameter optimization, *Renew. Sustain. Energy Rev.* 200 (2024) 114581.
- [47] A. Altmann, L. Tološi, O. Sander, T. Lengauer, Permutation importance: a corrected feature importance measure, *Bioinformatics* 26 (10) (2010) 1340–1347.
- [48] P. Jiang, J. Fan, L. Li, C.H. Wang, S.J. Tao, T. Ji, L.W. Mu, X.H. Lu, J.H. Zhu, A hybrid approach combining mechanism-guided data augmentation and machine learning for biomass pyrolysis, *Chem. Eng. Sci.* 296 (2024) 120227.
- [49] P. Jiang, C.H. Wang, J. Fan, T. Ji, L.W. Mu, X.H. Lu, J.H. Zhu, Hybrid residual modelling of biomass pyrolysis, *Chem. Eng. Sci.* 293 (2024) 120096.
- [50] H. Mesghali, B. Akhlaghi, N. Gozalpour, J. Mohammadpour, F. Salehi, R. Abbassi, Predicting maximum pitting corrosion depth in buried transmission pipelines: insights from tree-based machine learning and identification of influential factors, *process, Saf. Environ. Prot.* 187 (2024) 1269–1285.
- [51] Z.G. Li, H.Z. Shi, X. Yang, H. Tang, Investigating the nonlinear relationship between surface solar radiation and its influencing factors in North China plain using interpretable machine learning, *Atmos. Res.* 280 (2022) 106406.
- [52] J. Li, L.J. Pan, M. Suvarna, Y.W. Tong, X.N. Wang, Fuel properties of hydrochar and pyrochar: prediction and exploration with machine learning, *Appl. Energy* 269 (2020) 115166.
- [53] B.Y. Liang, H.Y. Liu, E.L. Cressey, C.Y. Xu, L. Shi, L. Wang, J.Y. Dai, Z. Wang, J. Wang, Uncertainty of partial dependence relationship between climate and vegetation growth calculated by machine learning models, *Remote. Sens.* 15 (11) (2023) 2920.
- [54] Q.Y. Zhao, T. Hastie, Causal interpretations of black-box models, *J. Bus. Econ. Stat.* 39 (1) (2019) 272–281.
- [55] B. Triveni, B. Vishwanadham, T. Madhavi, S. Venkateshwar, Mixing studies of Non-Newtonian fluids in an anchor agitated vessel, *Chem. Eng. Res. Des.* 88 (7) (2010) 809–818.
- [56] J.M.T. Vasconcelos, S.C.P. Orvalho, A.M.A.F. Rodrigues, S.S. Alves, Effect of blade shape on the performance of six-bladed disk turbine impellers, *Ind. Eng. Chem. Res.* 39 (1) (2000) 203–213.
- [57] R. Petříček, T. Moucha, F.J. Rejl, L. Valenz, J. Haidl, Č. Tereza, Volumetric mass transfer coefficient, power input and gas hold-up in viscous liquid in mechanically agitated fermenters. Measurements and scale-up, *Int. J. Heat Mass Transf.* 124 (2018) 1117–1135.
- [58] F. Haseidl, J. Pottböcker, O. Hinrichsen, Gas–liquid mass transfer in a rotor–stator spinning disc reactor: experimental study and correlation, *Chem. Eng. Process. Process Intensif.* 104 (2016) 181–189.
- [59] H. Ali, S. Zhu, J. Solsvik, Effects of geometric parameters on volumetric mass transfer coefficient of Non-Newtonian fluids in stirred tanks, *Int. J. Chem. React. Eng.* 20 (7) (2022) 697–711.
- [60] S.H. Zhou, Y. Li, X.Y. Zhang, Optimization modeling of anti - breast cancer candidate drugs, *Biotechnol. Genet. Eng. Rev.* 40 (2) (2024) 1334–1352.
- [61] X.Y. Zhang, P.C. Xiao, Y.Z. Yang, Y.J. Cheng, B. Chen, D.Z. Gao, W.R. Liu, Z.W. Huang, Remaining useful life estimation using CNN-XGB with extended time window, *IEEE Access* 7 (2019) 154386–154397.
- [62] D. Gómez-Díaz, J.M. Navaza, Analysis of carbon dioxide gas/liquid mass transfer in aerated stirred vessels using non-newtonian media, *J. Chem. Technol. Biotechnol.* 79 (10) (2004) 1105–1112.
- [63] V. Cappello, C. Plais, C. Vial, F. Augier, Bubble size and liquid-side mass transfer coefficient measurements in aerated stirred tank reactors with Non-Newtonian liquids, *Chem. Eng. Sci.* 211 (2020) 115280.
- [64] S.R. Lone, V. Kumar, J.R. Seay, D.L. Englert, H.T. Hwang, Mass transfer and rheological characteristics in a stirred tank bioreactor for cultivation of *escherichia coli* BL21, *Biotechnol. Bioproc. Eng.* 25 (5) (2020) 766–776.
- [65] A.K. Pandey, J. Park, J. Ko, H.H. Joo, T. Raj, L.K. Singh, N. Singh, S.H. Kim, Machine learning in fermentative biohydrogen production: advantages, challenges, and applications, *Bioresour. Technol.* 370 (2023) 128502.
- [66] É. Dumont, Mass transfer in multiphase gas/liquid/liquid systems. KLa determination using the effectiveness-number of transfer unit method, *Processes* 6 (9) (2018) 156.
- [67] D.L. Li, W. Chen, Effects of impeller types on gas-liquid mixing and oxygen mass transfer in aerated stirred reactors, *process, Saf. Environ. Prot.* 158 (2022) 360–373.
- [68] A. Ordaz, I. Figueroa-González, P. San-Valero, C. Gabaldón, G. Quijano, Effect of the height-to-diameter ratio on the mass transfer and mixing performance of a biotrickling filter, *J. Chem. Technol. Biotechnol.* 93 (1) (2018) 121–126.
- [69] M.R. Valverde, R. Bettega, A.C. Badino, Numerical evaluation of mass transfer coefficient in stirred tank reactors with Non-Newtonian fluid, *Theor. Found. Chem. Eng.* 50 (6) (2016) 945–958.
- [70] A. Rahimzadeh, F. Ein-Mozaffari, A. Lohi, Development of a scale-up strategy for an aerated coaxial mixer containing a Non-Newtonian fluid: a mass transfer approach, *Phys. Fluids* 35 (7) (2023) 073103.
- [71] L. Shu, M.J. Yang, H. Zhao, T.F. Li, L. Yang, X. Zou, Y.W. Li, Process optimization in a stirred tank bioreactor based on CFD-Taguchi method: a case study, *J. Clean. Prod.* 230 (2019) 1074–1084.
- [72] G.M. Mule, R. Lohia, A.A. Kulkarni, Effect of number of branches on the performance of fractal impeller in a stirred tank: mixing and hydrodynamics, *Chem. Eng. Res. Des.* 108 (2016) 164–175.
- [73] H.N. Wang, X.X. Duan, X. Feng, Z.S. Mao, C. Yang, Effect of impeller type and scale-up on spatial distribution of shear rate in a stirred tank, *Chin. J. Chem. Eng.* 42 (2022) 351–363.
- [74] S. Salehi, A. Heydarinasab, F.P. Shariati, A.T. Nakhjiri, K. Abdollahi, Parametric numerical study and optimization of mass transfer and bubble size distribution in a gas-liquid stirred tank bioreactor equipped with rushton turbine using computational fluid dynamics, *Int. J. Chem. React. Eng.* 19 (10) (2021) 1115–1131.
- [75] P. Lins Barros, F. Ein-Mozaffari, A. Lohi, Gas dispersion in non-newtonian fluids with mechanically agitated systems: a review, *Processes* 10 (2) (2022) 275.
- [76] T. Kumaresan, J.B. Joshi, Effect of impeller design on the flow pattern and mixing in stirred tanks, *Chem. Eng. J.* 115 (3) (2006) 173–193.
- [77] A. Gabriele, A.W. Nienow, M.J.H. Simmons, Use of angle resolved PIV to estimate local specific energy dissipation rates for up- and down-pumping pitched blade agitators in a stirred tank, *Chem. Eng. Sci.* 64 (1) (2009) 126–143.
- [78] M. Moayeri Kashani, S.H. Lai, S. Ibrahim, P. Moradi Bargani, Design factors affecting the dynamic performance of soil suspension in an agitated, baffled tank, *Chin. J. Chem. Eng.* 24 (12) (2016) 1664–1673.