Chinese Journal of Chemical Engineering 84 (2025) 77–85

Contents lists available at ScienceDirect

Chinese Journal of Chemical Engineering

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

Full Length Article

Reaction process optimization based on interpretable machine learning and metaheuristic optimization algorithms

Dian Zhang, Bo Ouyang *, Zheng-Hong Luo *

Department of Chemical Engineering, School of Chemistry and Chemical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

ARTICLE INFO

Article history:
Received 24 October 2024
Received in revised form
10 February 2025
Accepted 19 February 2025
Available online 1 March 2025

Keywords:
Reaction process optimization
Interpretable machine learning
Metaheuristic optimization algorithm
Biodiesel

ABSTRACT

The optimization of reaction processes is crucial for the green, efficient, and sustainable development of the chemical industry. However, how to address the problems posed by multiple variables, nonlinearities, and uncertainties during optimization remains a formidable challenge. In this study, a strategy combining interpretable machine learning with metaheuristic optimization algorithms is employed to optimize the reaction process. First, experimental data from a biodiesel production process are collected to establish a database. These data are then used to construct a predictive model based on artificial neural network (ANN) models. Subsequently, interpretable machine learning techniques are applied for quantitative analysis and verification of the model. Finally, four metaheuristic optimization algorithms are coupled with the ANN model to achieve the desired optimization. The research results show that the methanol: palm fatty acid distillate (PFAD) molar ratio contributes the most to the reaction outcome, accounting for 41%. The ANN-simulated annealing (SA) hybrid method is more suitable for this optimization, and the optimal process parameters are a catalyst concentration of 3.00% (mass), a methanol: PFAD molar ratio of 8.67, and a reaction time of 30 min. This study provides deeper insights into reaction process optimization, which will facilitate future applications in various reaction optimization processes.

1. Introduction

Optimization of chemical reaction processes is a complex, multidimensional challenge involving the interplay of various parameters, such as catalysts, raw materials, solvents, temperature, pressure, and reactor types. These factors collectively determine efficiency, selectivity, and environmental impact of the reaction, necessitating comprehensive evaluation and precise control by experts. Through optimization, it is possible to enhance yield and selectivity, reduce by-products, lower costs, and minimize environmental impact. Therefore, optimization of chemical reaction processes is crucial for promoting industrial sustainability, improving resource efficiency, and reducing environmental footprints.

Intuition-based optimization is quite common in academic research [1,2]. However, this method relies on expert observation and judgment, and intuition is not always reliable. An alternative is the one-factor-at-a-time method. For instance, Zhong *et al.* [3] optimized the selective oxidation of *p*-chlorotoluene to *p*-chlorobenzaldehyde by fixing other parameters and sequentially adjusting reaction time, catalyst loading, oxygen flow rate, and reaction temperature. This method follows a straightforward procedure that requires no modeling. Nevertheless, the parameter space explored by this method is often limited, and by neglecting the synergistic interactions between factors, it frequently misidentifies the optimal reaction conditions [4]. Unlike the one-factor-at-a-time method, design of experiments provides data through structured experiments, constructs statistical models to depict the relationship between inputs and outputs, and is frequently combined with response surface methodology to guide the optimization of reaction processes. This approach allows for simultaneous consideration of multiple factors and offers direct guidance for optimization. Maiti *et al.* [5] utilized Box-Behnken design and response surface methodology to optimize the

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: bouy93@sjtu.edu.cn (B. Ouyang), luozh@sjtu.edu.cn (Z.-H. Luo).

https://doi.org/10.1016/j.cjche.2025.02.001

formulation of active catalysts and the polymerization process of 1,3-butadiene. Similarly, Ma *et al.* [6] applied the same methodology to optimize the process parameters for the reaction of ethylene oxide and hydrazine hydrate to produce β-hydroxyethyl hydrazine. Despite its increasing popularity, design of experiments still faces practical limitations, such as difficulties in handling categorical variables, limited model response ranges due to the lack of support from physical-chemistry principles and an exponential increase in the number of experiments required as the number of variables grows. Kinetic modeling, grounded in physical chemistry principles, is also frequently used in reaction process optimization. Guo *et al.* [7] utilized this method to optimize the synthesis of *N*,*O*-dimethyl-*N*′-nitroisourea from *O*-methyl-*N*-nitroisourea and methylamine. This approach quantifies the effects of parameters on the reaction through mathematical equations, providing direct and actionable guidance for optimization. Additionally, the mechanistic models offer extrapolation capabilities to facilitate process scale-up. Nonetheless, solving complex calculus equations poses significant challenges in terms of numerical stability and computational cost. Guo *et al.* [7] also successfully applied Bayesian optimization to optimize the reaction process. Bayesian optimization demonstrates strong global search capabilities by balancing exploration and exploitation at each iteration. Based on the constructed probabilistic model of the objective function, relatively accurate optimization results can be obtained with a small number of function evaluations. However, it also has drawbacks, such as high computational cost, poor interpretability, reliance on the probabilistic model, and the potential to get trapped in local optima. In this context, the advantages of optimization algorithms capable of both local and global search, as well as handling complex nonlinear relationships, become evident. Among these, metaheuristic optimization algorithms, such as genetic algorithm (GA), particle swarm optimization (PSO), artificial bee colony algorithm (ABC), and simulated annealing (SA) are suitable choices. Silva and Biscaia [8] applied GA to optimize the free radical polymerization process of styrene. Charoenpanich, Anantawaraskul *et al.* [9] compared GA, PSO, ant colony optimization, and ABC for optimizing ethylene/1-olefin copolymerization processes. Chen *et al.* [10] compared GA, PSO, and SA in terms of performance for optimizing cross-coupling yields. However, these optimization algorithms typically rely on mathematical models of the reaction process or experimental data, making the accuracy of the models and data quality critical for reliable optimization. Although successfully applied, traditional approaches still struggle with issues related to multiple variables, nonlinearities, and uncertainties. Additionally, optimization algorithms place higher demands on modeling. Therefore, new methods are urgently needed to tackle these challenges more effectively.

The introduction of artificial intelligence technologies, particularly machine learning (ML) algorithms, offers new insights and tools to tackle these challenges. With the ability to automatically discover complex nonlinear relationships between variables, ML has been extensively applied in chemical engineering, including in multiphase closure modeling [11–13], correlation establishment [14,15], and flow regime identification [16,17]. In reaction process optimization, ML models can capture relationships between reaction conditions and key parameters such as product selectivity, yield, and energy consumption [18]. These models can predict outcomes under new reaction conditions and guide optimization when combined with metaheuristic optimization algorithms. For instance, Kim *et al.* [19] used eXtreme Gradient Boosting (XGBoost) and artificial neural networks (ANN) for modeling, followed by ABC to optimize the non-oxidative conversion of methane. Zhu and Kenig [20] employed various ML techniques for modeling and used GA to optimize the methanol-to-olefins process. Among them, ANN algorithms have been widely applied in modeling and have demonstrated outstanding performance [21,22]. While traditional ML algorithms can deliver high-precision predictions, their black-box nature makes it difficult to further explain the relationships between model inputs and outcomes, which has led to the development of interpretable ML techniques [23–25]. In this study, we introduce interpretable ML to uncover the key factors influencing reaction outcomes and their interactions, and to verify the developed ANN model.

To reduce greenhouse gas emissions and dependence on petroleum diesel fuels, research in the field of renewable energy is receiving significant attention from researchers [26]. Biodiesel, as a renewable, biodegradable, and environmentally friendly energy source, has been regarded as a potential alternative to fossil fuels [27,28]. However, the high production cost hinders the popularization and application of biodiesel, and the optimization of reaction process is an effective strategy to solve this problem.

Therefore, this study adopts a strategy that combines interpretable ML and metaheuristic optimization algorithms for reaction process optimization. First, an ML predictive model is constructed based on experimental data and interpretable ML is introduced for quantitative analysis and model verification. Then the ANN model is coupled with metaheuristic optimization algorithms to solve the objective optimization problem. Finally, the optimal combination of operational parameters corresponding to the best reaction outcome is obtained.

2. Methodology

The research workflow of this study is illustrated in Fig. 1. First, experimental data are collected, followed by thorough preprocessing to ensure high data quality. Next, ANN-based predictive models are constructed, trained, and evaluated using the collected data to obtain the optimal predictive model. Subsequently, the Shapley additive explanations (SHAP) method, a technique from interpretable ML, is introduced to quantitatively analyze the influence of operational parameters on the reaction results and verify the model. Finally, four metaheuristic optimization algorithms, GA, PSO, ABC and SA, are integrated with the ANN model to achieve the desired optimization objectives.

2.1. Data collection and preprocessing

To begin the process, the necessary data are collected from existing literature. Since the raw data come in different ranges and units, the input parameters are normalized during preprocessing. In this study, the min-max normalization technique is applied to scale the actual data into a range of 0 to 1, as per the following equation:

$$x_i' = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{1}$$

Here, x_i' represents the normalized data, x_i is the actual data, $x_{\min}$ is the minimum value of input data, and $x_{\max}$ is the maximum value. Since the output parameters are already within the range from 0 to 1, normalization for them is unnecessary.

2.2. ANN-based predictive model

ANNs are computational models inspired by biological neural systems, capable of learning complex nonlinear relationships from large datasets. In this study, ANNs serve as predictive models that can capture intricate relationships between inputs and outputs

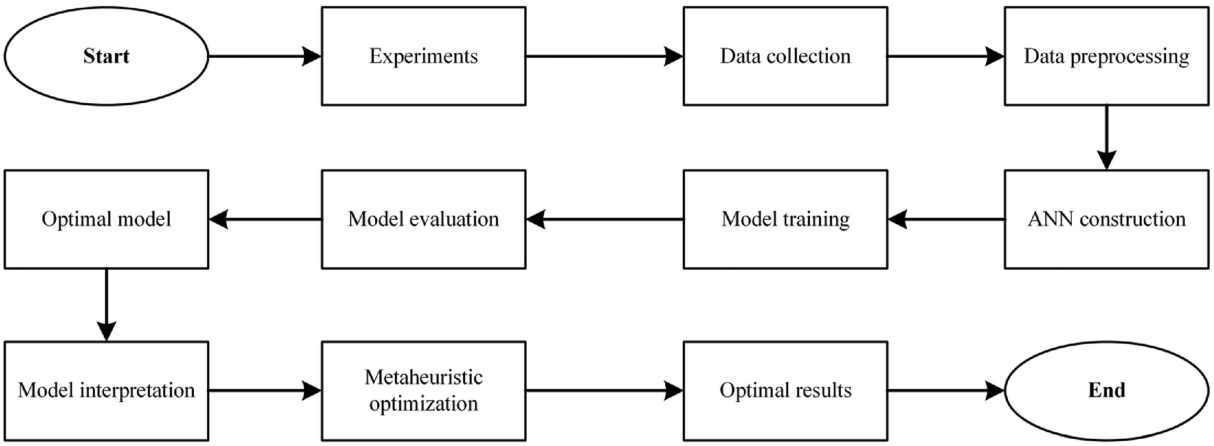

Fig. 1. Illustration of the workflow for prediction, analysis, and optimization of reaction process.

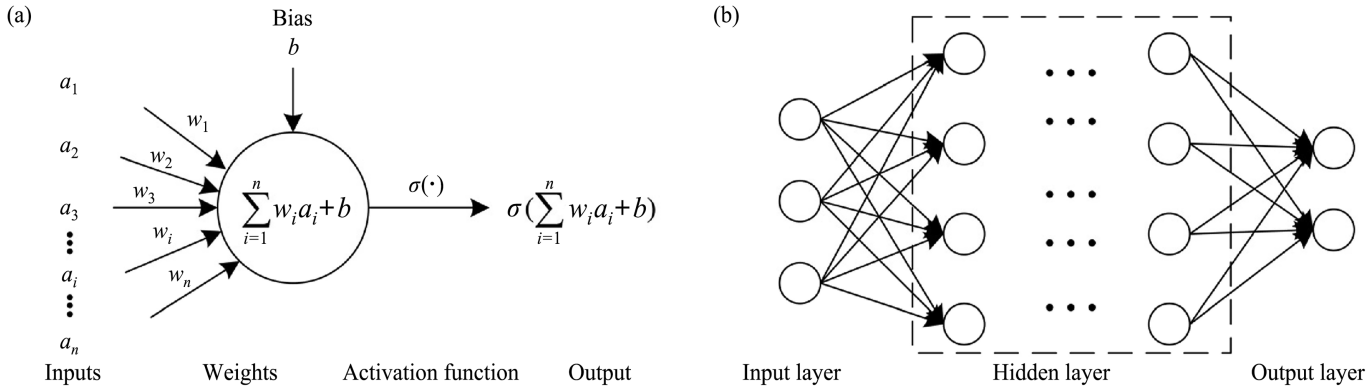

Fig. 2. Architecture of neurons and ANNs: (a) neurons, (b) ANNs.

based on existing data, and predict reaction outcomes under new conditions.

2.2.1. Neuron and ANN architecture

Neurons are the fundamental units of ANNs, simulating the functions of biological neurons. Each neuron receives input signals (a_i) from other neurons and performs a weighted (w_i) sum of these inputs, adding a bias term (b) to obtain the total input value. This value is then processed through a nonlinear activation function ($\sigma(\cdot)$) to generate the output, as illustrated in Fig. 2(a). The structure of ANNs mainly consists of an input layer, hidden layers, and an output layer. The input layer receives external data, where each node represents an input feature; the hidden layers, located between the input and output layers, consist of multiple neurons that extract and process complex features from the data; and the output layer generates the final prediction, as shown in Fig. 2(b).

In this study, the activation functions considered are the rectified linear unit (ReLU) function ($\sigma(z) = \max(0,z)$) and the sigmoid function $\left(\sigma(z) = \frac{1}{1+\exp(-z)}\right)$. This is because the ReLU function is simple and efficient, helps to mitigate the vanishing gradient problem, and enhances network performance and training speed by promoting sparse activation. And the sigmoid function has an output range from 0 to 1, which aligns with the requirements of this study.

2.2.2. Loss function and evaluation criteria

To provide an optimization objective for model training, quantify the difference between predictions and actual values, and guide parameter updates and performance evaluation, it is essential to select an appropriate loss function. The Huber loss function combines the advantages of mean squared error and absolute error, offering robustness, smoothness, and flexibility. It effectively handles outliers, enhancing the stability and convergence speed of the model. Therefore, in this study, the Huber loss function is chosen as the loss function, expressed as follows:

$$\text{Huber loss} = \begin{cases} \frac{1}{2}\left(y_{\text{exact}} - y_{\text{predict}}\right)^2, & \text{if} \left|y_{\text{exact}} - y_{\text{predict}}\right| \le \delta \\ \delta\left|y_{\text{exact}} - y_{\text{predict}}\right| - \frac{1}{2}\delta^2, & \text{if} \left|y_{\text{exact}} - y_{\text{predict}}\right| > \delta \end{cases} \quad (2)$$

Table 1
Selected parameters for GA, PSO, ABC, and SA.

Metaheuristic optimization algorithm	Parameter	Value
GA	Population size	30
	Generations	50
	Mutation rate	0.1
PSO	Swarm size	30
	Max iterations	50
	Inertia weight	0.7
	Cognitive component	1.5
	Social component	1.5
ABC	Employed bees	24
	Scout bees	6
	Max iterations	50
SA	Initial temperature	10
	Cooling rate	0.99
	Step size	0.1
	Max iterations	1500

Table 2
Prediction performance of ANNs.

Training	Learning rate	Hidden layer	Hidden layer activation function	Output layer activation function	Optimizer	Batch	RMSE	R^2	AAD
1	0.0001	256, 128, 128, 64, 32	ReLU	Sigmoid	Adam	1	1.561	0.992	1.158
2	0.001	256, 128, 128, 64, 32	ReLU	Sigmoid	Adam	1	1.032	0.997	0.780
3	0.01	256, 128, 128, 64, 32	ReLU	Sigmoid	Adam	1	2.251	0.983	1.672
4	0.1	256, 128, 128, 64, 32	ReLU	Sigmoid	Adam	1	1.634	0.991	1.314
5	1	256, 128, 128, 64, 32	ReLU	Sigmoid	Adam	1	1.519	0.992	1.164
6	0.001	256, 128, 128, 64, 32	ReLU	Sigmoid	Adadelta	1	2.782	0.974	2.296
7	0.001	256, 128, 128, 64, 32	ReLU	Sigmoid	Adagrad	1	1.970	0.987	1.648
8	0.001	256, 128, 128, 64, 32	ReLU	Sigmoid	RMSprop	1	1.322	0.994	1.035
9	0.001	256, 128, 128, 64, 32	ReLU	Sigmoid	SGD	1	2.426	0.981	2.010
10	0.001	256, 128, 128, 64, 32	ReLU	Sigmoid	Adam	2	1.259	0.995	0.956
11	0.001	256, 128, 128, 64, 32	ReLU	Sigmoid	Adam	5	1.373	0.994	1.084
12	0.001	256, 128, 128, 64, 32	ReLU	Sigmoid	Adam	10	1.379	0.994	1.024
13	0.001	256, 128, 128, 64, 32	ReLU	Sigmoid	Adam	20	1.286	0.995	1.009
14	0.001	8	ReLU	Sigmoid	Adam	1	3.800	0.952	2.970
15	0.001	16	ReLU	Sigmoid	Adam	1	3.583	0.958	2.705
16	0.001	64	ReLU	Sigmoid	Adam	1	1.578	0.992	1.242
17	0.001	256	ReLU	Sigmoid	Adam	1	1.762	0.990	1.451
18	0.001	128, 128	ReLU	Sigmoid	Adam	1	1.150	0.996	0.878
19	0.001	256, 128	ReLU	Sigmoid	Adam	1	1.147	0.996	0.864
20	0.001	256, 128, 128	ReLU	Sigmoid	Adam	1	1.099	0.996	0.830
21	0.001	256, 256,128	ReLU	Sigmoid	Adam	1	1.3326	0.994	1.004
22	0.001	256, 128, 128, 64	ReLU	Sigmoid	Adam	1	1.243	0.995	0.953
23	0.001	256, 128, 128, 32	ReLU	Sigmoid	Adam	1	1.188	0.995	0.919
24	0.001	256, 256, 128, 64, 32	ReLU	Sigmoid	Adam	1	1.253	0.995	0.962
25	0.001	256, 128, 128, 64, 32, 32	ReLU	Sigmoid	Adam	1	1.309	0.994	0.910
26	0.001	256, 128, 128, 64, 32	Sigmoid	Sigmoid	Adam	1	1.445	0.993	1.074
27	0.001	256, 128, 128, 64, 32	Sigmoid	ReLU	Adam	1	1.289	0.995	0.998
28	0.001	256, 128, 128, 64, 32	ReLU	ReLU	Adam	1	Predicted values are 0.		

Here, y_{exact} is the actual value and y_{predict} is the predicted value. δ is a threshold that can be customized, with a default value of 1. To evaluate the model performance, three metrics are used: root mean square error (RMSE), coefficient of determination (R^2), and average absolute deviation (AAD).

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_{\text{exact}} - y_{\text{predict}}\right)^2} \tag{3}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_{\text{exact}} - y_{\text{predict}}\right)^2}{\sum_{i=1}^{n}\left(y_{\text{exact}} - \overline{y}_{\text{exact}}\right)^2} \tag{4}$$

$$\text{AAD} = \frac{1}{n}\sum_{i=1}^{n}\left|y_{\text{exact}} - y_{\text{predict}}\right| \tag{5}$$

2.3. Interpretable ML

In this study, we use the SHAP method to interpret and verify the ANN model. SHAP is based on Shapley values from cooperative game theory, which were originally used to assess the contribution of each participant to the overall payoff. In the field of ML, SHAP explains the behavior of ML models by assigning the contribution of each feature to the prediction output. Shapley values are used to quantify the contribution of each feature to the prediction. The calculation formula for Shapley values is defined as follows [29]:

$$\phi_i = \frac{1}{N!}\sum_{S\subseteq N\setminus\{i\}} |S|!(|N| - |S| - 1)![f(S\cup\{i\}) - f(S)] \tag{6}$$

Here, ϕ_i represents the Shapley value for feature i, N is the set of all features, and S is a subset of features excluding feature i. The term $f(S \cup \{i\})$ the prediction of the model when feature i is included, while $f(S)$ is the prediction when it is not. The formula calculates the weighted average contribution of each feature by considering all possible subsets of features.

2.4. Metaheuristic optimization algorithms

Metaheuristic optimization algorithms are a flexible and powerful class of optimization techniques widely used to address complex optimization problems. These algorithms often simulate phenomena or behaviors found in nature, employing population-based search strategies to explore solution spaces. Among various metaheuristic optimization algorithms, four typical representatives are GA, PSO, ABC, and SA. GA simulates the process of natural selection, utilizing selection, crossover, and mutation operations, making it suitable for high-dimensional discrete and continuous optimization problems [30]. PSO, based on swarm intelligence, quickly approaches optimal solutions through the movement of particles in the solution space, especially excelling in continuous problems [31]. ABC mimics the foraging behavior of bees, combining global and local search capabilities to handle various optimization tasks [19]. SA, grounded in the physical annealing process, allows for a certain probability of accepting worse solutions, which demonstrates excellent local search capabilities and is often applied in scenarios requiring detailed optimization [32]. These four algorithms exhibit strong performance in solving nonlinear and complex problems, providing diverse options to meet different optimization needs.

In this study, the ANN-based predictive model is used as the fitness function and integrated with GA, PSO, ABC, and SA for optimizing operational parameters. These four algorithms are all constructed and implemented using the Python programming language. Table 1 lists the selected parameters for GA, PSO, ABC, and SA during the optimization process.

3. Results and Discussion

3.1. Predictive performance of the ANN model

Data are collected from Ref. [33] for the modeling of ANNs. The selected input parameters are catalyst concentration, methanol: palm fatty acid distillate (PFAD) molar ratio, and reaction time, while the output is fatty acid methyl ester (FAME) content. After preprocessing, the splits for the training, validation, and test sets are aligned with the literature to facilitate model performance evaluation. The entire process is conducted on a personal laptop using open-source tools such as Python, TensorFlow, Keras, and Matplotlib. A GPU (GeForce RTX 1650) is utilized to accelerate the training process. During training, RMSE of the validation set is monitored to evaluate whether the model achieves optimal performance, with 20000 training epochs set. To reduce overfitting and save computational resources, an early stopping mechanism is employed. This technique involves regularly checking the model's RMSE on the validation set and monitoring for improvements over a span of 100 consecutive epochs. If no significant progress is observed during these 100 epochs, the training process is halted. A comprehensive analysis of the impact of hyperparameters on the prediction performance of the ANN is conducted, with the results presented in Table 2. For uniform comparison, RMSE, R^2, and AAD of the FAME percentage content are used.

Table 2 indicates that the batch size (1–20) has only a slight impact on model performance. For the choice of learning rate (0.0001–1), lower values tend to produce better results, while excessively high learning rates may lead to suboptimal outcomes. The selection of different optimizers (Adam, Adadelta, Adagrad, RMSprop, SGD) also affects model performance, with Adam proving to be the most suitable. Furthermore, as the number of hidden layers and neurons per layer increases, model performance gradually improves. However, an excessive number of hidden layers and neurons can lead to a decline in performance. The selections of activation functions also impact model performance, and in some cases, may even cause the model to fail. Based on above results, the optimal configuration is achieved in training 2, where the Adam optimizer is used with a learning rate of 0.001, five hidden layers with 256, 128, 128, 64, and 32 neurons in each respective layer, ReLU activation function for the hidden layers, sigmoid activation function for the output layer, and a batch size of 1. At this configuration, RMSE is 1.032, R^2 is 0.997, and AAD is 0.780.

3.2. Comparison of model performance

A further comparison between the trained model and models (quick propagation (QP), incremental backpropagation (IBP), batch backpropagation (BBP), GA, and Levenberg–Marquardt (LM)) from the literature is presented in Table 3. It can be observed that the trained model achieves the highest R^2 and the lowest AAD, though its RMSE is not the lowest. Overall, performance differences between the models are minimal. However, unlike models in the literature, which use professional software (NeuralPower software version 2.5), this study utilizes open-source tools, demonstrating that comparable performance can be achieved using more accessible tools. The scatter plots of the predicted FAME percentage content versus the actual values for the training, validation, and test sets are displayed in Fig. 3. This visually demonstrates the excellent performance of the trained model.

3.3. Interpretation of the ANN model

After completing the model construction in the previous section, which uses catalyst concentration, methanol: PFAD molar ratio, and reaction time as features, this section aims to provide a deeper understanding of each feature's contribution to model predictions and to further verify the model. The shap-lib is employed to handle the core computation tasks for model interpretation. Fig. 4 illustrates the impact of individual features on model predictions. Each point represents a sample, and its position along the *x*-axis indicates the magnitude and direction of the feature's contribution to the prediction (SHAP value). The color gradient reflects the relative feature value, where red points correspond to high feature values and blue points correspond to low feature values. It can be observed that for catalyst concentration and reaction time, there is a general trend where larger feature

Table 3
Comparison of performance with models in the literature.

Model	RMSE	R^2	AAD
ANN	1.032	0.997	0.780
QP	0.666	0.996	0.901
IBP	1.040	0.664	1.290
BBP	0.959	0.995	1.221
GA	5.902	0.885	7.765
LM	0.705	0.996	0.988

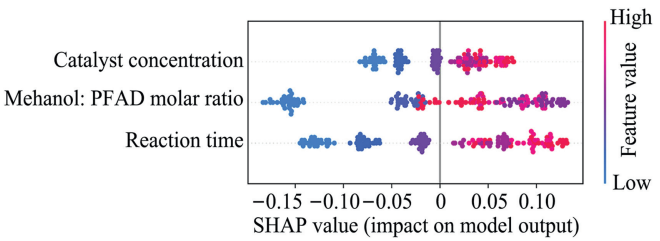

Fig. 4. Summary plot of individual feature contributions (SHAP values).

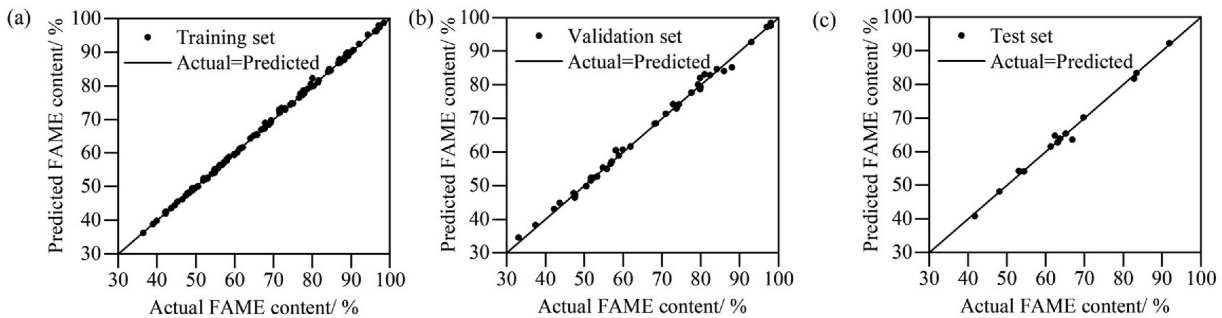

Fig. 3. The scatter plots of the predicted FAME percentage content versus the actual values for the training, validation, and test sets: (a) training set; (b) validation set; (b) test set.

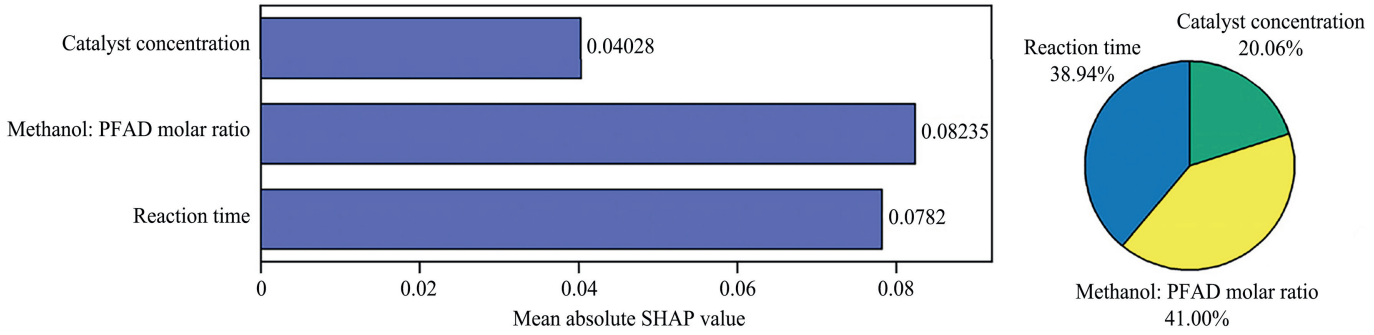

Fig. 5. Global feature importance plot (Mean absolute of SHAP values).

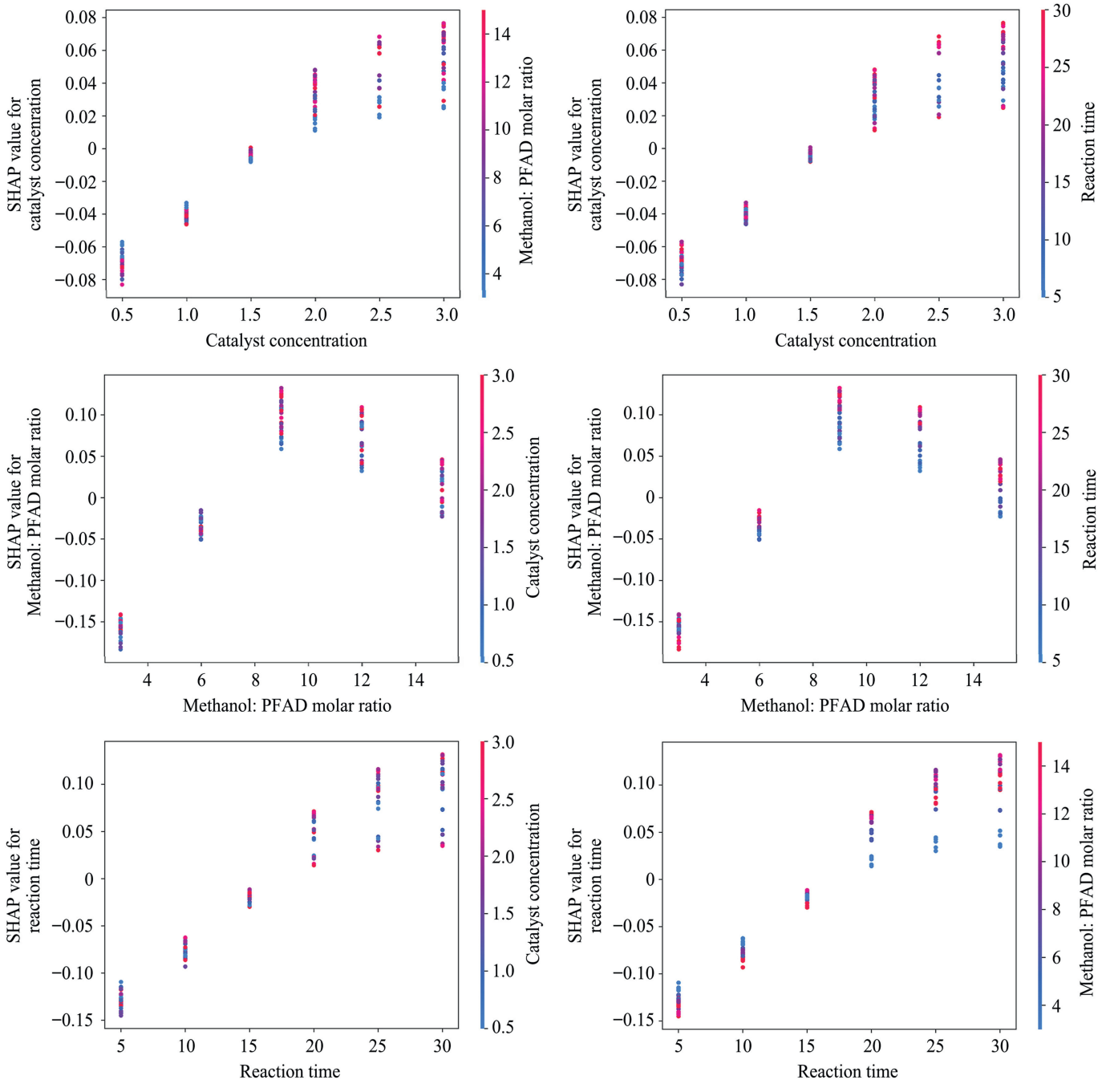

Fig. 6. Feature interaction dependence plot.

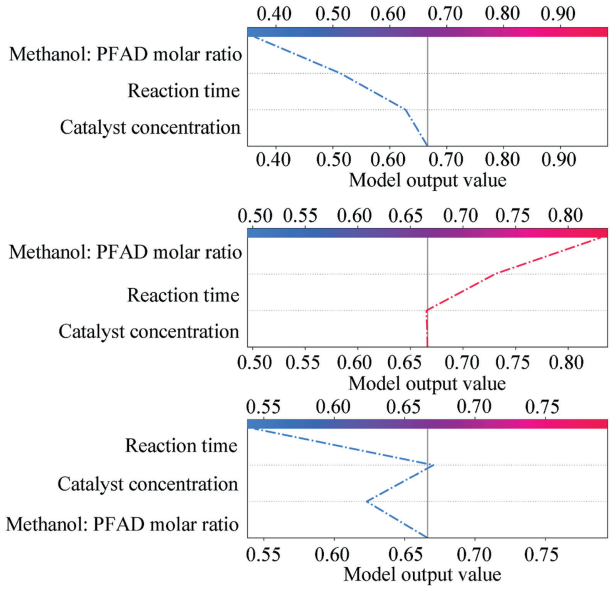

Fig. 7. SHAP decision plot for individual data points.

values have a more significant positive impact on model predictions. However, for certain points, larger feature values result in a smaller positive impact on the model's predictions. This indicates that the relationship is not strictly linear, and the influence of these features on the model is more complex, possibly involving interactions with other variables. For the methanol: PFAD molar ratio, moderate feature values, rather than excessively high or low ones, have a significant positive impact on model predictions. This is an important aspect to consider in the subsequent optimization process.

Fig. 5 displays the mean absolute SHAP values for each feature, which can represent their importance. It is evident that all three features contribute to model predictions to some extent, and it is correct to consider the three factors during the modeling process. Among them, methanol: PFAD molar ratio and reaction time contribute more significantly to the model, with similar levels of importance, while catalyst concentration contributes relatively less.

Fig. 6 illustrates the interaction dependencies of the features, helping to understand the model sensitivity to specific features and the interactions between them. By analyzing the relationship between the feature values (*X*-axis) and corresponding SHAP values (*Y*-axis), the same conclusion as shown in Fig. 4 can be obtained more intuitively. Specifically, for catalyst concentration and reaction time, the trend is that the larger feature values, the more significant positive impact on model prediction. For methanol: PFAD molar ratio, it is moderate feature values that have a pronounced positive effect on model prediction. Further analysis of the features represented on the *X*-axis and the interaction features indicated by the color gradient reveals that the interaction between any two features is quite complex. This also shows that these three

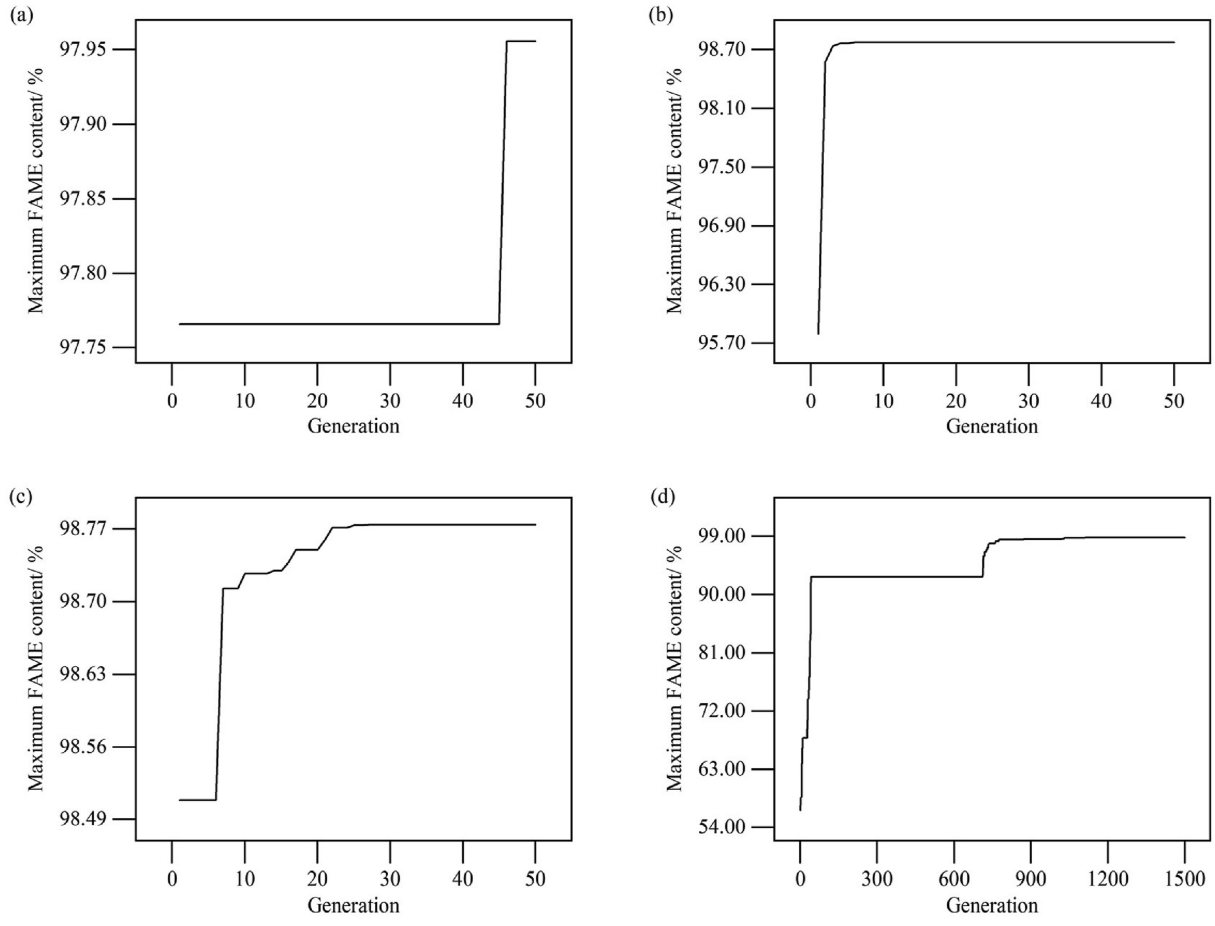

Fig. 8. Four algorithm optimization processes: (a) ANN-GA, (b) ANN-PSO, (c) ANN-ABC, (d) ANN-SA.

Table 4
Final optimization results.

Optimization method	Catalyst concentration/% (mass)	Methanol: PFAD molar ratio	Reaction time/min	FAME content/%	Time/s
ANN-GA	2.27	8.75	28.43	97.96	189.65
ANN-PSO	3.00	8.67	30.00	98.77	85.06
ANN-ABC	3.00	8.67	30.00	98.77	85.53
ANN-SA	3.00	8.67	30.00	98.77	84.54

factors need to be considered comprehensively in the optimization process. Specifically, for catalyst concentration, at lower levels, a decrease in methanol: PFAD molar ratio makes its positive contribution to predictions more evident, while at higher levels, an increase in methanol: PFAD molar ratio enhances this contribution. Additionally, as reaction time increases, the positive contribution of catalyst concentration to predictions becomes clearer. For methanol: PFAD molar ratio, increases in both catalyst concentration and reaction time amplify its positive impact on predictions. Regarding reaction time, when it is low, decreases in both catalyst concentration and methanol: PFAD molar ratio significantly enhance their positive contribution, while at longer reaction time, increases in these factors lead to a more pronounced positive impact on predictions. Combined with professional knowledge, these results are consistent with the expected behavior of the reaction.

Additionally, several data points are randomly selected from the dataset for decision-making process analysis. Fig. 7 illustrates the cumulative contribution of individual features to the final model prediction. Each line represents a prediction trajectory for a specific instance, starting from the base value (the average model output). As each feature is added along the *y*-axis, its contribution (positive or negative) is accumulated, eventually reaching the final prediction value. Features with higher contributions create steeper slopes in the plot, indicating a stronger influence on the model decision. Overall, the results align with the conclusions drawn in the previous sections, indicating that the methanol: PFAD molar ratio and reaction time have a greater influence on model predictions, while catalyst concentration has a relatively smaller impact.

3.4. Reaction process optimization results

Through above research, it is evident that there are complex interactions among the three operational parameters. Analyzing them through single or pairwise factors is insufficient to achieve the optimal combination of operational parameters. Therefore, ANN is coupled with GA, PSO, ABC, and SA to maximize FAME content. Fig. 8 illustrates the optimization processes of the four algorithms, showing the progression of each algorithm toward the optimal solution. Meanwhile, Table 4 presents the final optimization results, providing a comprehensive comparison of each algorithm. Based on the maximum FAME content obtained and the runtime of each algorithm, the efficiency of the optimization techniques can be ranked in descending order as follows: ANN-SA, ANN-PSO, ANN-ABC and ANN-GA. More specially, the top three optimization algorithms all achieve the same optimized results, with only slight differences in runtime. Because all three methods are based on the same ANN predictive model. Although different metaheuristic optimization algorithms are used, all three algorithms follow similar fundamental principles during the optimization process. Furthermore, in this study, the total number of iterations is set to be the same. Only ANN-GA yields relatively poorer results and takes longer to run. Based on previous analysis, for a fixed methanol: PFAD molar ratio, increasing the catalyst concentration and reaction time can enhance its positive contribution to the prediction. This does not imply that ANN-GA is ineffective; rather, it indicates that the current number of iterations is insufficient, and the algorithm requires further iterations for improvement. The longer runtime is due to the relatively complex nature of the ANN-GA code.

ANN-SA identifies the optimal process variable combination at a catalyst concentration of 3% (mass), a methanol: PFAD molar ratio of 8.67, and a reaction time of 30 min, achieving a maximum FAME content of 98.77%. Compared to the best experimental data where a catalyst concentration of 3% (mass), a methanol: PFAD molar ratio of 9, and a reaction time of 30 min resulted in a FAME content of 98.34%, and the results obtained through factor analysis with a catalyst concentration of 3% (mass), a methanol: PFAD molar ratio of 9, and a reaction time of 25 min yielding 97.74% FAME content, the optimization process has improved the FAME content to 98.77%, reflecting a noticeable enhancement in performance.

4. Conclusions and Outlook

In this work, in order to optimize the reaction process efficiently and deepen insights into the process, a coupling approach of interpretable ML and metaheuristic optimization algorithms is adopted. First, an experimental dataset of a biodiesel production process is constructed to train the ANN model. Then, the SHAP values of the operating parameters are used to quantify their contributions and interaction dependencies and to verify the ANN model. Subsequently, four metaheuristic optimization algorithms (GA, PSO, ABC, and SA) are integrated with the ANN-based predictive model to achieve the desired optimization. Finally, the optimal results and corresponding combinations of operating parameters are obtained. The main conclusions of this study can be summarized as follows:

(1) The ANN model achieves an R^2 of 0.997, an RMSE of 1.032, and an AAD of 0.780, demonstrating its capability to reasonably predict reaction outcomes.
(2) The contributions of the methanol: PFAD molar ratio, reaction time, and catalyst concentration to the reaction outcome are 41.0%, 38.94%, and 20.06%, respectively, with the methanol: PFAD molar ratio having the greatest impact.
(3) The optimization performance is ranked from highest to lowest as follows: ANN-SA, ANN-PSO, ANN-ABC, and ANN-GA.
(4) The ANN-SA hybrid method predicts a maximum FAME content of 98.77%, with the optimal combination of operational parameters being a catalyst concentration of 3.00% (mass), a methanol: PFAD molar ratio of 8.67, and a reaction time of 30 min.

This coupling strategy of interpretable ML and metaheuristic optimization algorithms has successfully optimized the biodiesel production process and is expected to be applied to the optimization of various reaction processes. Despite current efforts, we acknowledge that this study has certain limitations. Since the model is built on existing data, it cannot be updated in real-time with newly acquired data. In the future, we hope to improve this technology and integrate it with advanced monitoring and

automation techniques to achieve real-time online optimization of models and reaction processes.

CRediT Authorship Contribution Statement

Dian Zhang: Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. Bo Ouyang: Writing – review & editing, Visualization, Project administration, Data curation, Conceptualization. Zheng-Hong Luo: Writing – review & editing, Supervision, Software, Resources, Project administration, Funding acquisition, Formal analysis, Conceptualization.

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 work was supported by the National Natural Science Foundation of China (22408227, 22238005) and the Postdoctoral Research Foundation of China (GZC20231576).

References

[1] J.B. Roque, Y. Kuroda, L.T. Göttemann, R. Sarpong, Deconstructive diversification of cyclic amines, *Nature* 564 (7735) (2018) 244–248.
[2] J. Derosa, R. Kleinmans, V.T. Tran, M.K. Karunananda, S.R. Wisniewski, M.D. Eastgate, K.M. Engle, Nickel-catalyzed 1, 2-diarylation of simple alkenyl amides, *J. Am. Chem. Soc.* 140 (51) (2018) 17878–17883.
[3] S.D. Zhong, C. Xu, Q.F. Long, H. Zhou, J.H. Yang, J. Guo, X.Z. Wang, L.Y. Dai, Highly catalytic performance for the selective C–H bond oxidation of p-chlorotoluene over Co, Cr Co-doped OMS-2 catalyst, *Chem. Eng. Sci.* 302 (2025) 120789.
[4] C.J. Taylor, A. Pomberger, K.C. Felton, R. Grainger, M. Barecka, T.W. Chamberlain, R.A. Bourne, C.N. Johnson, A.A. Lapkin, A brief introduction to chemical reaction optimization, *Chem. Rev.* 123 (6) (2023) 3089–3126.
[5] M. Maiti, V.K. Srivastava, S. Shewale, R.V. Jasra, A. Chavda, S. Modi, Process parameter optimization through Design of Experiments in synthesis of high cis-polybutadiene rubber, *Chem. Eng. Sci.* 107 (2014) 256–265.
[6] H.Y. Ma, L.X. Yang, C.Q. Yao, S.N. Zhao, F.J. Jiao, G.S. Luo, G.W. Chen, Liquid-liquid addition reaction of ethylene oxide with hydrazine hydrate in microreactors: kinetics and process optimization, *Chem. Eng. J.* 497 (2024) 154511.
[7] J.P. Guo, G.H. Luo, K.J. Chai, W.K. Su, A. Su, Continuous flow synthesis of *N, O*-dimethyl-*N*'-nitroisourea monitored by inline Fourier transform infrared spectroscopy: Bayesian optimization and kinetic modeling, *Ind. Eng. Chem. Res.* 63 (23) (2024) 10162–10171.
[8] C.M. Silva, E.C. Biscaia, Genetic algorithm development for multi-objective optimization of batch free-radical polymerization reactors, *Comput. Chem. Eng.* 27 (8–9) (2003) 1329–1344.
[9] T. Charoenpanich, S. Anantawaraskul, J.B.P. Soares, Using artificial intelligence techniques to design ethylene/1-olefin copolymers, *Macromol. Theory Simul.* 29 (6) (2020) 2000048.
[10] J.L. Chen, W.J. Xu, R.Q. Zhang, Optimization of chemical synthesis with heuristic algorithms, *Phys. Chem. Chem. Phys.* 25 (5) (2023) 4323–4331.
[11] B. Ouyang, L.T. Zhu, Z.H. Luo, Data-driven modeling of mesoscale solids stress closures for filtered two-fluid model in gas–particle flows, *AIChE J.* 67 (7) (2021) e17290.
[12] Z. Yang, B.N. Lu, W. Wang, Coupling Artificial Neural Network with EMMS drag for simulation of dense fluidized beds, *Chem. Eng. Sci.* 246 (2021) 117003.
[13] C.Z. Du, C.X. Han, Z. Yang, H. Wu, H. Luo, L. Niedzwiecki, B.N. Lu, W. Wang, Multiscale CFD simulation of an industrial diameter-transformed fluidized bed reactor with artificial neural network analysis of EMMS drag markers, *Ind. Eng. Chem. Res.* 61 (24) (2022) 8566–8580.
[14] J.B. Zhou, D.P. Liu, M. Ye, Z.M. Liu, Data-driven prediction of minimum fluidization velocity in gas-fluidized beds using data extracted by text mining, *Ind. Eng. Chem. Res.* 60 (37) (2021) 13727–13739.
[15] J.B. Zhou, M. Ye, Z.M. Liu, Towards a general correlation for minimum fluidization velocity in gas-fluidized beds: based on a database mining from the literature, *Chem. Eng. Sci.* 251 (2022) 117455.
[16] C. Wang, S.S. Li, Y. Yang, Z.L. Huang, J.Y. Sun, J.D. Wang, Y.R. Yang, Z.W. Liao, B.B. Jiang, On flow regime transition in trickle bed: development of a novel deep-learning-assisted image analysis method, *AIChE J.* 66 (2) (2020) e16833.
[17] D. Zhang, B. Ouyang, Z.H. Luo, Identification of gas-solid flow regimes using convolutional neural network techniques, *Powder Technol.* 442 (2024) 119848.
[18] T.T. Qi, G.H. Luo, H.T. Xue, F. Su, J.L. Chen, W.K. Su, K.J. Wu, A. Su, Continuous heterogeneous synthesis of hexafluoroacetone and its machine learning-assisted optimization, *J. Flow Chem.* 13 (3) (2023) 337–346.
[19] H.W. Kim, S.W. Lee, G.S. Na, S.J. Han, S.K. Kim, J.H. Shin, H. Chang, Y.T. Kim, Reaction condition optimization for non-oxidative conversion of methane using artificial intelligence, *React. Chem. Eng.* 6 (2) (2021) 235–243.
[20] L.T. Zhu, E.Y. Kenig, A study of methanol-to-olefins packed bed reactor performance using particle-resolved CFD and machine learning, *AIChE J.* 70 (10) (2024) e18520.
[21] O.S. Stamenković, K. Rajković, A.V. Veličković, P.S. Milić, V.B. Veljković, Optimization of base-catalyzed ethanolysis of sunflower oil by regression and artificial neural network models, *Fuel Process. Technol.* 114 (2013) 101–108.
[22] E. Betiku, O.R. Omilakin, S.O. Ajala, A.A. Okeleye, A.E. Taiwo, B.O. Solomon, Mathematical modeling and process parameters optimization studies by artificial neural network and response surface methodology: a case of non-edible neem (*Azadirachta indica*) seed oil biodiesel synthesis, *Energy* 72 (2014) 266–273.
[23] M.N. Du, N.H. Liu, X. Hu, Techniques for interpretable machine learning, *Commun. ACM* 63 (1) (2019) 68–77.
[24] C. Jiang, R. Vinuesa, R.L. Chen, J.Y. Mi, S.J. Laima, H. Li, An interpretable framework of data-driven turbulence modeling using deep neural networks 33 (5) (2021) 055133.
[25] B. Ouyang, L.T. Zhu, Z.H. Luo, Interpretable machine learning analysis and automated modeling to simulate fluid-particle flows, *Particuology* 80 (2023) 42–52.
[26] N.P. Asri, S. Machmudah, S. Wahyudiono, K. Budikarjono, A. Roesyadi, M. Goto, Palm oil transesterification in sub- and supercritical methanol with heterogeneous base catalyst, *Chem. Eng. Process. Process. Intensif.* 72 (2013) 63–67.
[27] F. Kusumo, A.S. Silitonga, H.H. Masjuki, H.C. Ong, J. Siswantoro, T.M.I. Mahlia, Optimization of transesterification process for Ceiba pentandra oil: a comparative study between kernel-based extreme learning machine and artificial neural networks, *Energy* 134 (2017) 24–34.
[28] A.S. Yusuff, N.B. Ishola, A.O. Gbadamosi, T.M. Azeez, M.O. Onibonoje, An artificial intelligence approach to model and optimize biodiesel production from used cooking oil using CaO incorporated zeolite catalyst, *Energy Convers. Manag. X* 20 (2023) 100452.
[29] R. Rodríguez-Pérez, J. Bajorath, Interpretation of compound activity predictions from complex machine learning models using local approximations and shapley values, *J. Med. Chem.* 63 (16) (2020) 8761–8777.
[30] D. Sarkar, S. Rohani, A. Jutan, Multiobjective optimization of semibatch reactive crystallization processes, *AIChE J.* 53 (5) (2007) 1164–1177.
[31] Z.W. Fu, Z.Y. Wang, G. Chen, Enhanced parameter estimation with improved particle swarm optimization algorithm for cell culture process modeling, *AIChE J.* 70 (4) (2024) e18388.
[32] M.F. Cardoso, R.L. Salcedo, S.F. de Azevedo, D. Barbosa, Optimization of reactive distillation processes with simulated annealing, *Chem. Eng. Sci.* 55 (21) (2000) 5059–5078.
[33] S. Soltani, T. Roodbar Shojaei, N. Khanian, T. Shean Yaw Choong, N. Asim, Y. Zhao, Artificial neural network method modeling of microwave-assisted esterification of PFAD over mesoporous TiO_2–ZnO catalyst, *Renew. Energy* 187 (2022) 760–773.