Chinese Journal of Chemical Engineering 90 (2026) 348–360

Contents lists available at ScienceDirect

Chinese Journal of Chemical Engineering

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

Full Length Article

Improved neural network automation design method for energy saving and carbon emission reduction of petrochemical production processes

Guocong Lin [1, 2, #], Qingxu Ni [1, 3, #], Xuehai Liu [1, 2], Zhiqiang Geng [1, 2, *], Feng Pan [3], Jun Tian [4, *], Yongming Han [1, 2, *]

[1] College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China
[2] Engineering Research Center of Intelligent PSE, Ministry of Education of China, Beijing, 100029, China
[3] Sinopec Sales Co., Ltd., North China Branch, Tianjin 300384, China
[4] School of Economics and Management, Beijing University of Chemical Technology, Beijing 100029, China

ARTICLE INFO

Article history:
Received 4 June 2025
Received in revised form
26 August 2025
Accepted 10 September 2025
Available online 26 November 2025

Keywords:
Prediction
Process control
Model
Production guidance
Automation design
Energy efficiency optimization

ABSTRACT

Ethylene yield serves as a key metric in petrochemical production, where optimizing its energy efficiency remains a critical challenge for sustainable production. Meanwhile, manual hyperparameter tuning of deep learning based yield prediction models often results in suboptimal configurations, which reduces prediction accuracy and reliability due to extreme operating conditions generating outliers in ethylene production processes. Therefore, a novel neural network automatic design method (NNADM) is proposed, which incorporates the neural network parameters automatic optimization and loss function adaptive construction. An innovative adaptive loss formulation is proposed to strategically integrate the complementary strengths of the mean squared error (MSE) and the Log-Cosh functions, featuring dynamic outlier resistance through self-adjusting weight coefficients. Then, the Bayesian optimization search algorithm is utilized to discover optimal hyperparameters of the neural network, including hidden layer unit, epoch, batch size, and the loss function. Finally, the NNADM is integrated with several classical neural networks for ethylene yield prediction. Experimental results show that several classical neural networks realize an average of 16.86% decrease in MSE index after integrating the NNADM. In addition, the proposed model offers direction and development blueprints for ethylene production facilities that have low energy efficiency. By implementing this model, it is possible to cut approximately 8376.4 tons of carbon emissions and simultaneously secure an extra 499 tons of ethylene output.

1. Introduction

The ethylene output in petrochemical processes is very large and the ethylene production is energy intensive [1]. Enhancing the energy utilization efficiency of ethylene production units is the key to achieve energy conservation, emission reduction and sustainable economic development. At present, relevant research focuses on establishing a yield prediction model to enhance energy-saving potential [2]. Yield prediction models can be categorized into machine learning-based and deep learning-based models.

Traditional machine learning-based models establish the input-output relationship through multiple fully connected nonlinear maps, such as the back propagation (BP) network, the extreme learning machines (ELM) and the radial basis function (RBF). Samuel and Muggeridge [3] introduced the edge orthogonal decomposition into the RBF neural network to provide data processing for reservoir prediction with constant rate injection through a non-invasive dimensionality reduction framework with strong robustness and stability possessed. Liu *et al.* [4] optimized the principal component analysis (PCA) by replacing the covariance matrix with the Spearman rank correlation coefficient matrix, and introduced a Gaussian distribution transformation *via* the

* Corresponding authors.
E-mail addresses: gengzhiqiang@mail.buct.edu.cn (Z. Geng), 2010500027@buct.edu.cn (J. Tian), hanym@mail.buct.edu.cn (Y. Han).
# These authors contributed equally to this work.

https://doi.org/10.1016/j.cjche.2025.09.030

Johnson transform for chemical process detection and variable state detection. Jahirul *et al.* [5] used the PCA to understand the relationship between the important properties of biodiesel and its chemical composition to predict the specific properties. Chen *et al.* [6] combined the kernel extreme learning machine (KELM) with the robust weight mean of vectors (INFO) algorithm for building the water quality modeling. Afzal *et al.* [7] used the response surface method (RSM) combining the ELM, the RBF, and the multilayer perceptron (MLP) to construct a comprehensive framework for accurate estimation of heat and cold loads. Li *et al.* [8] employed perceptron neural networks and Gaussian process regression modeling to predict biochar yield and carbon and nitrogen content. Godwin *et al.* [9] achieved the efficient prediction of the characteristics of gasoline engines combining pine oil and mixtures using an integrated least squares boosting machine learning framework. Jani *et al.* [10] trained and validated various machine learning algorithms using a polyethylene dataset, demonstrating that optimized Gaussian process regression models with appropriate kernel functions and ensemble bagged tree models provided the highest accuracy in predicting polyethylene melt flow index, melt flow index, flow rate ratio, and density. However, traditional machine learning methods tend to fall into local optimization in practice and often poor at handling high dimensional, multi-noise and nonlinear data [11].

The deep learning-based models have better feature representation, nonlinear data convergence, and the ability to learn from large amounts of data, which is widely used for yield forecasting and energy saving potential analysis. Işık *et al.* [12] used multiple seasonal-trend decomposition using locally estimated scatterplot smoothing (LOESS) technique for disposing the electricity demand forecasting problem. Compared to the long short-term memory (LSTM) with three gates, gated recurrent unit (GRU) neural networks have fewer parameters, and are widely used for data modeling. Ma *et al.* [13] used the whale optimization algorithm and extreme gradient boosting to extract multi-category probabilistic features, which were then combined with the BiLSTM and attentional Q-networks for energy consumption and carbon emission prediction. Metcalfe *et al.* [14] built an LSTM based soft sensor model of penicillin concentrations in real time and predict penicillin concentrations over an 8-h forecast horizon. Sun *et al.* [15] improved the accuracy of carbon emission predictions and enhanced the interpretability of carbon emission patterns through a combination of multiscale decomposition and deep learning models. Zhou *et al.* [16] developed a new dynamic process temperature control strategy using neural networks with memory functions (*e.g.*, RNN, LSTM, GRU) as soft sensors in combination with fuzzy control, thereby improving the accuracy of batch distillation tower products. Gao *et al.* [17] used generative adversarial interpolation networks to interpolate missing data, combining temporal attention and spatial attention with Bidirectional Long-Short Term Memory (Bi-LSTM) networks to improve the prediction accuracy of butane content and steam volume. Li *et al.* [18] proposed a correlation analysis method for variable-length sequences for the feature selection in chemical process sequence modeling and constructed an encoder-decoder based multi-step prediction model on different process variables such as flow, pressure and temperature. Li *et al.* [19] significantly improved the accuracy of soft measurement modeling in Debutanizer columns and continuous catalytic reforming processes by integrating stacked autoencoders, mutual information, and multi-scale feature fusion strategies. Homod *et al.* [20] adopted a new boiler sequencing control (BSC) strategy using deep clustering of reinforcement learning based on a bang-bang manner. Xu *et al.* [21] proposed a deep reinforcement learning (DRL) approach to address the electricity arbitrage problem associated with optimal energy storage system (ESS) management. Deng *et al.* [22] adopted a model-free DRL algorithm soft-actor-critic (SAC) to adjust demand to matching renewable supply with maintaining user satisfaction. Li *et al.* [23] combined a multilinear convolutional neural network (CNN) with a hybrid attention mechanism to achieve individual extraction and unified fusion of multiple features, and designed an integrated multi-task learning shared layer for power price prediction. Song *et al.* [24] used a temporal attention layer, a spatial attention layer and a gated temporal convolutional network (TCN) to achieve error reduction in energy cooling, heating, and electrical load prediction tasks. Brozos *et al.* [25] used the graph neural network (GNN) framework to predict the critical micelle concentration of binary mixtures of surfactants and successfully extended it to ternary mixtures. Han *et al.* [26] used the Pearson-GNN to predict polypropylene and fully exploited the relationship between variables to improve polypropylene production and save energy. Guo *et al.* [27] proposed the global dynamic adjacent layer information enhancement auto encoder model, which effectively improved the spatio-temporal feature extraction and prediction accuracy of industrial soft measurement models by introducing the GRU and uniform manifold approximation and projection and enhancing adjacent hidden layer information. Ban *et al.* [28] used a wavelet threshold noise reduction algorithm to effectively mitigate the effects of noise in wind speed prediction, and introduced the Autoformer model for wind speed prediction to achieve efficient long-term wind speed prediction. Han *et al.* [29] proposed a time series prediction model based on the iTransformer method, which was used to accurately predict the methane production and carbon dioxide emissions during the anaerobic digestion (AD) process of food waste. Guo *et al.* [30] applied frequency-enhanced decomposition transformers to predict total phosphorus concentrations, effectively capturing long-term correlations between water quality parameters and thereby improving the accuracy of total phosphorus predictions.

Data-driven neural network models are widely used in industries, but there are many parameters that need to be adjusted in the application. It is difficult to find a set of parameters through manual adjustment to make neural network models optimal in the modeling process [31]. Therefore, it is necessary to select an appropriate set of parameters for the neural network model to design the network structure. Nakamura *et al.* [32] used a stochastic batch optimization algorithm to achieve adaptive regularization by determining the batch size of each model parameter in each optimization iteration through a stochastic process. Liang *et al.* [33] introduced the Hunger Games search algorithm based on the artificial neural networks (ANN) to optimize their network model for energy consumption prediction of building air conditioning systems. Nawaz *et al.* [34] used the gradient descent optimization to achieve intelligent diagnosis of cardiovascular diseases with high sensitivity and accuracy. Huang *et al.* [35] used an adaptive learning rate approach in the LSTM to achieve optimization of electric vehicle sound packages. Zulfiqar *et al.* [36] introduced an adaptive difference algorithm based on support vector machines for power load forecasting with higher accuracy, stability and convergence rate. Chang *et al.* [37] used a grid search algorithm to optimize the parameters of the LSTM to ensure accurate prediction of non-smooth degradation for predicting the remaining life prediction process of the device. Narmadha and Balaji [38] used the particle swarm optimization (PSO) to optimize the performance of an LSTM model for achieving high accuracy in detecting network intrusions. Tunçel and Duran [39] demonstrated the effectiveness of stochastic search and grid search algorithms for parameter optimization. Babu *et al.* [40] used a stochastic search algorithm for optimizing the beamline parameters of the dominant multi-species beam in space. The birth of

heuristic algorithms led to a new phase of parameter optimization. Ansari *et al.* [41] used a data-driven approach based on a particle swarm optimization algorithm to determine optimal parameters of a neural network for lithium battery lifetime prediction. ElSaid *et al.* [42] used an ant colony optimization algorithm to optimize hyperparameters of the LSTM for predicting the vibration of the turbine generator, which led to a significant reduction in the error of the experimental result. Kedia *et al.* [43] used a genetic algorithm to optimize the surface roughness process parameters of chromium nickel-iron alloy workpieces. Wang *et al.* [44] used the sparrow search algorithm to improve the semi-supervised learning method of Stacked Autoencoders (SAEs), and realized the effective utilization of unlabeled auxiliary variable samples in time series prediction. Wang and Li [45] improved sparrow search algorithm and optimized parameters of the echo state networks (ESN) with a variant of recurrent neural networks (RNN) to achieve effective sulfide emission reduction. Liu *et al.* [46] used the annealing particle swarm optimization algorithm to search for hyperparameters of the BP neural network for predicting the corrosion life of fiberglass, verifying that this method has a stronger global search ability. Ghasemi *et al.* [47] used the firefly algorithm to achieve robustness and high adaptability in engineering optimization. Mahmoodzadeh *et al.* [48] used the Gray Wolf optimization algorithm to optimize hyperparameters of the LSTM for predicting the excavation of tunnel boring machines to obtain better experimental results. Gkonis and Tsakalos [49] proposed an optimized hybrid deep learning model for accurately predicting stock prices *via* the golden jackal optimizer (GJO) algorithm. Long *et al.* [50] used the improved particle swarm optimization with niche technology for the soft measurement in the gasoline blending process to reduce blending costs. Qian *et al.* [51] proposed a time difference PSO localization method to improve the diversity and convergence ability of the particle population through asynchronous dynamic learning factors. However, heuristic algorithms are difficult to determine the time complexity based on guesses and are more sensitive to strategy selection.

In recent years, Bayesian optimization algorithms have been widely used in industries due to the design of appropriate probabilistic models and collection functions from a probabilistic point of view, which require only a small number of evaluations to obtain a better and ideal solution [52]. Li *et al.* [53] used a Bayesian optimization algorithm to optimize a convolutional neural network model for automated model construction to gear point fault diagnosis. Chen *et al.* [54] used a Bayesian algorithm to optimize hyperparameters of a Bi-LSTM neural network model to accurately predict the performance degradation of a fuel cell. Kim *et al.* [55] used five machine learning methods for predicting surface subsidence in urban tunnels and demonstrated the excellent search capability of the Bayesian optimization for parameter tuning. Xu *et al.* [56] combined soft threshold functions with residual networks to address noise issues in industrial processes and used Bayesian optimization to optimize hyperparameters, achieving high-precision soft measurements on the steam dataset and Debutante tower dataset. In complex industrial modeling processes, the design of optimizers and loss functions directly affects the effectiveness of neural network models [57], but the Bayesian efficient parameter optimization capability is difficult to be directly applied in the design process of the loss function.

Therefore, a new weighted loss function is proposed to construct a novel neural network automatic design method, which builds a yield prediction model in a real ethylene production process. First, a loss function adaptive construction method is used to combine the advantages of MSE loss function and Log_Cosh loss function, which allows the Bayesian optimization algorithm to be directly applied in the loss function design process. Then, the Bayesian optimization search algorithms are utilized to discover the optimal hyperparameters of the neural network, including hidden layer unit, epoch, batch size, and the loss function. The advantage of the Bayesian optimization search algorithms is that only a small number of evaluations is required to obtain a better and ideal solution. Finally, the NNADM is applied to several classical neural networks for ethylene yield prediction. The primary contributions of this paper are outlined below.

(1) The NNADM is introduced to tackle the challenge of optimizing model hyperparameters in the process of yield prediction modeling.
(2) A new weighted loss function is designed to construct the neural network in the NNADM, allowing Bayesian optimization algorithms to be directly applied in the loss function design process.
(3) The NNADM is verified on the real industrial production process ethylene dataset. The experimental results show the superiority of NNADM over other methods. The input-output analysis is optimized based on the prediction results to guide production, reduce carbon emissions and increase profits, which can increase ethylene production by 499 tons by adjusting inputs and outputs, and reduce about 8376.4 tons of carbon emissions according to the carbon emission factor.

The subsequent research trajectory is structured as follows. Section 2 formally introduces the groundbreaking NNADM framework. Section 3 delineates the implementation of the developed methodology within an industrial ethylene manufacturing context, while empirically validating its superior performance against state-of-the-art benchmark techniques. The culminating section provides systematic analysis of experimental discoveries and synthesizes consequential research insights.

2. The Proposed Methodology

In this section, a new NNADM method is proposed, which consists of automatic construction of neural network loss functions, and Bayesian optimization-based adaptive learning of neural network hyperparameters. Meanwhile, the NNADM can determine the optimal loss function during the optimization search process and significantly improve the performance of the production prediction model.

2.1. New weighted loss function

In the current process of industrial data modeling, the improvement of the loss function optimization has gradually become hot. Common loss functions of neural networks for complex chemical processes mainly include MSE loss function, the mean absolute error (MAE) loss function and Log_Cosh loss function. The nonlinear relationship between the square of the real value and the predicted value described by the MSE loss function can be expressed by Eq. (1).

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(\mathrm{RealOut}_i - \mathrm{PreOut}_i)^2 \tag{1}$$

which in, N is the batch size, RealOut is the true value and PreOut is the model predicted value. The image of the MSE loss function is described in Fig. 1. The x-axis of Fig. 1 is $\mathrm{RealOut}_i - \mathrm{PreOut}_i$. The curve in Fig. 1 is smooth and differentiable everywhere. However,

when the error value is greater than 1, it makes the anomaly more obvious. Amplified anomalies result in poor model performance.

When contrasted with the MSE loss function, the MAE loss function demonstrates a notable insensitivity to outliers. This characteristic enables it to precisely delineate the linear relationship between the predicted and the true values, as indicated in Eq. (2).

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N} |\mathrm{RealOut}_i - \mathrm{PreOut}_i| \tag{2}$$

As shown in Fig. 2, the MAE loss function is always the same because of its update gradient, and the function at 0 is not derivable, which is not conducive to function convergence and learning of the neural network model.

Log_Cosh loss function uses hyperbolic cosine to estimate the loss, is not easily affected by the outlier, and is globally derivable, which can be expressed by Eq. (3).

$$\mathrm{Log\_Cosh} = \frac{1}{N}\sum_{i=1}^{N} \log\left(\cosh\left(\mathrm{RealOut}_i - \mathrm{PreOut}_i\right)\right) \tag{3}$$

where, $\cosh\left(\mathrm{RealOut}_i - \mathrm{PreOut}_i\right)$ can be implemented by Eq. (4).

$$\cosh(\mathrm{RealOut}_i - \mathrm{PreOut}_i) = \frac{e^{\mathrm{RealOut}_i - \mathrm{PreOut}_i} + e^{\mathrm{PreOut}_i - \mathrm{RealOut}_i}}{2} \tag{4}$$

When $\mathrm{RealOut}_i - \mathrm{PreOut}_i \to +\infty$, Eq. (3) tends to be as shown in Eq. (5).

$$\mathrm{Log\_Cosh} \approx (\mathrm{RealOut}_i - \mathrm{PreOut}_i) - \lg 2 \tag{5}$$

When $\mathrm{RealOut}_i - \mathrm{PreOut}_i \to -\infty$,Eq. (3) tends to be as shown in Eq. (6).

$$\mathrm{Log\_Cosh} \approx -(\mathrm{RealOut}_i - \mathrm{PreOut}_i) - \lg 2 \tag{6}$$

Therefore, when the value of $\mathrm{RealOut}_i - \mathrm{PreOut}_i$ is significantly greater than 0, Log_Cosh Loss function converges to the form in Eq. (7).

$$\mathrm{Log\_Cosh} \approx |\mathrm{RealOut}_i - \mathrm{PreOut}_i| - \lg 2 \tag{7}$$

When $\mathrm{RealOut}_i - \mathrm{PreOut}_i = 0$, Log_Cosh tends to be as shown in Eq. (8).

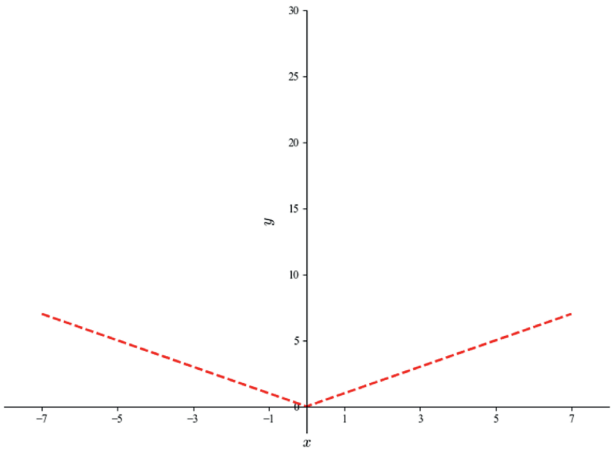

Fig. 2. The loss function of the MAE.

$$\mathrm{Log\_Cosh} \approx \frac{(\mathrm{RealOut}_i - \mathrm{PreOut}_i)^2}{2} \tag{8}$$

The loss function of the Log_Cosh is shown in Fig. 3.

From Eq (7) and Fig. 3, It is evident that when the absolute difference between the $\mathrm{RealOut}_i$ and the PreOut_i is excessively large, the Log_Cosh function experiences a certain level of performance degradation. This degradation decelerates the gradient update rate and exerts an adverse impact on the convergence efficacy of the model.

The value of the MSE loss function decreases with $\mathrm{RealOut}_i - \mathrm{PreOut}_i$, the gradient is also decreasing, which is beneficial for the convergence of the yield prediction model. However, as the model moves away from 0, the error increases, and the update gradient rises abruptly. This, in turn, affects the reliability of the yield prediction model. As illustrated in Fig. 3, The Log_Cosh makes the gradient update slower when it moves away from 0, which degrades into the MAE loss function and affects the yield prediction model effect. Therefore, this subsection improves the MSE loss function and Log_Cosh loss function by introducing a hyperparameter α to form a novel weighted loss function (αMixLoss), as shown in Eq. (9).

$$\alpha\mathrm{MixLoss} = \alpha \times \mathrm{MSE} + (1-\alpha) \times \mathrm{Log\_Cosh} \tag{9}$$

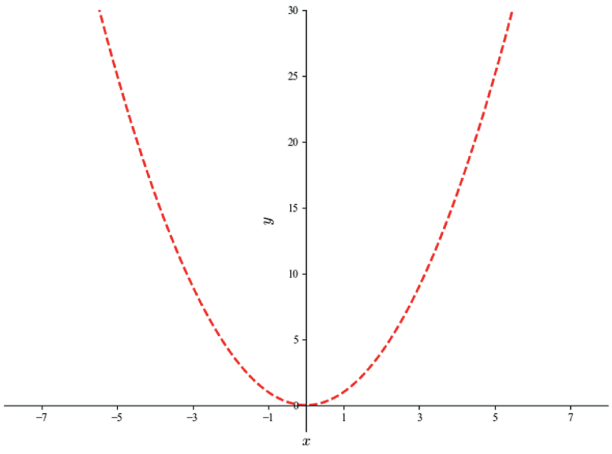

Fig. 1. The loss function of the MSE.

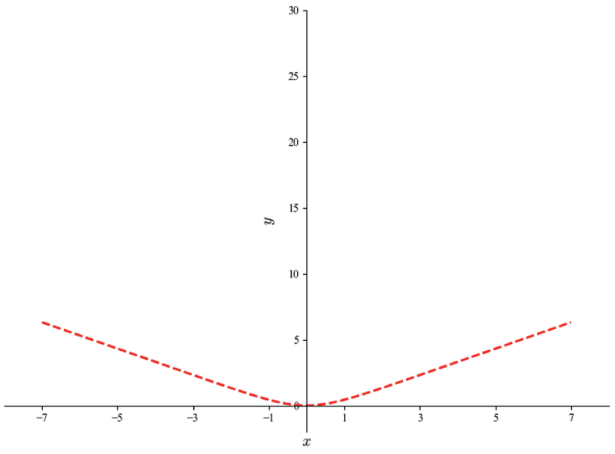

Fig. 3. The loss function of the Log_Cosh.

It is evident that the incorporation of the hyperparameter α enables the weighted loss function to adaptively adjust, thereby mitigating the impact of anomalies on the yield prediction model. This, in turn, leads to a notable enhancement in the performance of the neural network model. Fig. 4 displays the graphs of the αMixLoss function when the values of α are, in sequence, 0, 0.2, 0.4, 0.5, 0.6, 0.8, and 1.

From Fig. 4, it can be seen that αMixLoss degenerates into a Log_Cosh loss function when $\alpha = 0$ and into an MSE loss function when $\alpha = 1$. The shape of the αMixLoss loss function is determined by the hyperparameter α, and the value of the hyperparameter α can be determined by the Bayesian optimization algorithm. By learning the hyperparameter α, the strength of the penalty for outliers is balanced against the performance degradation caused by outliers being amplified (especially outliers larger than one).

2.2. The Bayesian optimization method

In actual industrial production environments, traditional manual parameter adjustment methods have a long-time cycle and are difficult to find the optimal solution from the hyperparameter set, resulting in low reliability of modeling and uncertainty in prediction results. The hyperparameter optimization of neural networks in complex petrochemical process aims to find the optimal set of neural network parameters in the data model established for the petrochemical process, which can be solved using Eq. (10).

$$\theta^* = \mathrm{argmin} f(\theta), \theta \in X \tag{10}$$

where $f(\theta)$ is the objective function being solved, $X = \theta_1, \theta_2, \theta_3, \cdots$ represents a set of hyperparameter combinations, θ representations decision parameter vector enables the objective function to find the optimal solution set.

The Bayesian parameter optimization algorithm is a global optimization method that uses the probabilistic surrogate model and selects the evaluation after fitting according to the results, and reasonably uses historical information to improve the search efficiency. In the application process, it applies the Bayes theorem, as shown in Eq. (11).

$$p(f|X) = \frac{p(X|f)p(f)}{p(X)} \tag{11}$$

where f represents the objective function, $p(f)$ represents the Prior probability of f, X represents the observed set, $p(X)$ represents the marginal likelihood distribution, and $p(X|f)$ represents the likelihood distribution.

The Bayesian optimization primarily revolves around two fundamental components: a probabilistic surrogate model and an acquisition function. Derived from posterior probability distributions, the acquisition function guides the selection of evaluation points through an optimization framework designed to systematically reduce cumulative losses. This mechanism operates by strategically balancing exploration and exploitation during the search process. The operational framework and theoretical foundations of Bayesian optimization are visually summarized in Fig. 5.

As shown in Fig. 5, assuming the model follows a Gaussian distribution in advance, the specific model function representation can be calculated, and the shaded part is the calculated Gaussian distribution model. The yellow dashed line is the predicted objective function, the blue solid line is the true objective function, and the black asterisks indicate the evaluated observations. Bayesian optimization improves the fitting of the objective function by balancing exploration and exploitation step by step to finally find the optimal solution.

The Gaussian process is a typical probability representative model, which is often used for parameter adjustment of neural networks and can be expressed in Eq. (12).

$$f(x) \sim GP(m(x), \mathbb{C}) \tag{12}$$

where $\mathbb{C}$ represents the covariance matrix formed by $k(x,x)$, $m(x) = E(f(x))$ is the mathematical expectation about $f(x)$, which is usually set to 0, and $f(x)$ represents the set of unknown functions. Assuming that historical exploration information has been obtained from the probabilistic surrogate model, from x_t to the next value x_{t+1} can be expressed as the covariance matrix shown in Eq. (13).

$$K = \begin{bmatrix} k(x_1,x_1) & k(x_1,x_2) & \cdots & k(x_1,x_t) \\ k(x_2,x_1) & k(x_2,x_2) & \cdots & k(x_2,x_t) \\ \vdots & \vdots & \cdots & \vdots \\ k(x_t,x_1) & k(x_t,x_2) & \cdots & k(x_t,x_t) \end{bmatrix} \tag{13}$$

According to the properties of Gaussian processes, the joint distribution function can be expressed as Eq. (14).

$$\begin{bmatrix} f_{1:t} \\ f_{t+1} \end{bmatrix} \sim N\left(0, \begin{bmatrix} k & k \\ k^{\mathrm{T}} & k(x_{t+1}, x_{t+1}) \end{bmatrix}\right) \tag{14}$$

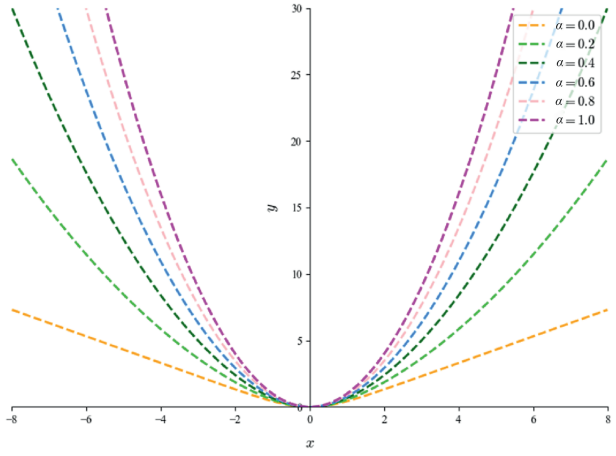

Fig. 4. Graph of the αMixLoss function, where α is adjustable.

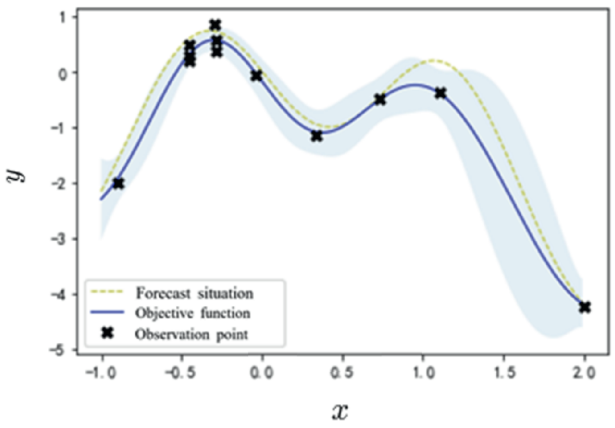

Fig. 5. Bayesian optimization process with Gaussian process regression.

Then, by calculating its edge density function, the posterior probability of f_{t+1} is expressed by Eq. (15).

$$p(f_{t+1}|D_{1:t}) = N\Big(m_t(x_{t+1}), \sigma_t^2(x_{t+1})\Big) \tag{15}$$

where $D_{1:t} = \{x_{1:t}, f_{1:t}\}$, $m_t(x_{t+1})$ and $\sigma_t^2(x_{t+1})$ can be used in the following Eq. (16).

$$\begin{cases} m_t(x_{t+1}) = k^{\mathrm{T}}K^{-1}f_{1:t} \\ \sigma_t^2(x_{t+1}) = k(x_{t+1}, x_{t+1}) - k^{\mathrm{T}}K^{-1}k \end{cases} \tag{16}$$

After calculating x_{t+1} normal distribution at any place, the most likely sample point for the best evaluation of the model at the next position is found. The current popular collection function is mainly the expected improvement (EI) collection function, which can be expressed by the following Eq. (17).

$$\mathrm{EI}(x) = \begin{cases} (o^* - m(x))\varnothing\left(\dfrac{o^* - m(x)}{\sigma_t(x)}\right) + \sigma_t(x)\varnothing\left(\dfrac{o^* - m(x)}{\sigma_t(x)}\right), \sigma_t(x) > 0 \\ 0, \sigma_t(x) = 0 \end{cases} \tag{17}$$

where o^* represents the optimal function value and $\varnothing$ represents the standard normal distribution probability density function.

2.3. The NNADM

There are many hyperparameters of neural network models built in traditional complex petrochemical processes, and the hyperparameters vary among different kinds of neural networks, mainly including initial learning rate, number of hidden layer neuron nodes, number of iterations, batch size, regularization coefficients, and other hyperparameters. Among them, the hyperparameters of most neural networks can be determined by the characteristics of the model itself. In contrast, the hyperparameters of the number of neuron nodes in the hidden layer, the number of iterations, and the batch size have a great influence on the yield prediction model. This subsection focuses on the hyperparameters of these three parameters and the hyperparameter α. As mentioned above a total of 4 hyperparameters are given to the NNADM to optimize. Meanwhile, the key phases involved in the introduced NNADM methodology include:

Step 1: Acquire industrial production data through systematic collection procedures, followed by partitioning the obtained dataset into training and testing subsets for subsequent neural network processing.
Step 2: Construct the loss function αMixLoss as shown in Eq. (9), which effectively balances the penalty for outliers and mitigates the issue of performance degradation in the NNADM. Concurrently, hyperparameter tuning is conducted on the hidden layer node quantity, epoch count, mini-batch dimension, and α coefficient within the αMixLoss function to attain the globally optimal configuration for the NNADM framework.
Step 3: Implement the NNADM framework to enhance diverse neural architectures, specifically benchmarking against conventional optimization techniques. Performance assessment incorporates six quantitative indicators: MSE, MAE, mean absolute percentage deviation (MAPE), root mean squared error (RMSE), R-squared metric (R^2), along with computational efficiency metrics, collectively demonstrating the efficacy of this automated modeling paradigm. These evaluation metrics including the MAPE, the RMSE and R^2 are mathematically defined in Eq. (18)–(20).

$$\mathrm{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\frac{|\mathrm{RealOut}_i - \mathrm{PreOut}_i|}{\mathrm{RealOut}_i} \times 100\% \tag{18}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\mathrm{RealOut}_i - \mathrm{PreOut}_i)^2} \tag{19}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(\mathrm{RealOut}_i - \mathrm{PreOut}_i)^2}{\sum_{i=1}^{N}\left(\mathrm{RealOut}_i - \frac{1}{N}\sum_{i=1}^{N}\mathrm{RealOut}_i\right)^2} \tag{20}$$

In general, smaller values of the MSE, the MAE, the MAPE, and the RMSE are indicative of a more accurate prediction model. Furthermore, an R^2 value that approaches 1 signifies an effective prediction model.

Step 4: The experimental results are systematically evaluated and analytically interpreted using multidimensional validation methods. This systematic process ultimately yields data-driven recommendations to enhance the operational efficiency of industrial manufacturing processes.

The proposed improved automated design method for neural networks in this paper is shown in Fig. 6.

3. Experimentation

3.1. Ethylene synthesis process schematic and predictive dataset profiling

To validate the practical applicability of the NNADM-optimized neural architecture in industrial settings, this study employs a benchmark dataset originating from real-time operational records of ethylene manufacturing systems. The dataset captures multi-source sensor networks spanning thermal cracking to fractionation stages. Key process dynamics and energy transfer mechanisms characterizing ethylene synthesis are systematically analyzed through this data-driven framework. The schematic representation of ethylene synthesis stages with material flows is illustrated in Fig. 7.

The composition of raw materials directly affects the production of ethylene, and crude oil plays a decisive role in the production of ethylene. The main components of crude oil are Lig, Nap, Raf, Hyd, Lhy, C345, Other, *etc*. Ethylene is used as the dependent variable for the output. Table 1 shows the representative meanings of each variable and their carbon emission factors [58].

A subset of the experimental data is presented in Fig. 8. To avoid the randomness of production forecasts in the ethylene prediction industry. All experiments are repeated three times and the results are averaged, implemented in PyTorch and in a single NVIDIA TITAN RTX 24 GB GPU.

The original ethylene production data set contained 3680 records col-lected from continuous industrial operations, which lasted approximately two and a half years with a sampling interval of 6 h. After data cleaning with removing missing values, obvious sensor failures, and extreme outliers, 3521 valid records are retained. These cleaned data are randomly divided into three non-overlapping subsets: 70% is used for model training, 15% is used for validation during hyperparameter tuning, and 15% is used for final testing. This clear division ensures that hyperparameter

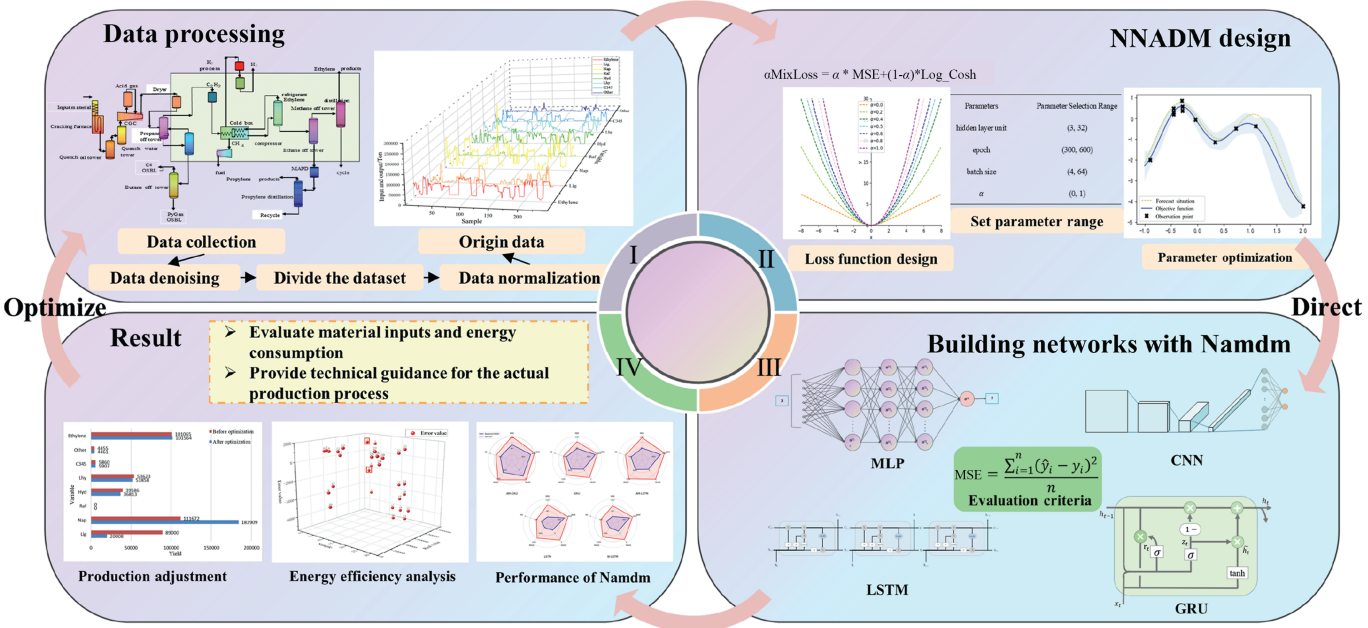

Fig. 6. The process of the NNADM.

optimization and performance evaluation are based on non-overlapping data, thereby enhancing the statistical reliability and practical relevance of the reported results.

3.2. Automation design method

The automatic network-design method (NNADM) mainly optimizes four hyperparameters including hidden layer unit, epoch, batch size, and hyperparameter α that determines the state of the loss function, whose parameters are selected in the range shown in Table 2.

The hyper-parameter search range is set based on the pre-experiment results and the computational resource constraints: the number of nodes in the hidden layer is limited to 3–32, where above 32 and the accuracy improvement is limited and the risk of overfitting increases, the epoch is 300–600 where above 600 and the validation error saturates or rebounds, the batch size is 4–64, where below 4 and the parallel efficiency decreases significantly and the upper limit is constrained by 24 GB GPU memory. The number of hidden layer nodes is limited to 3–32, where above 32 and the accuracy improvement is limited and the risk of overfitting increases, the epoch is 300–600, where above 600 and the validation error tends to be saturated or rebound, the batch size is 4–64, where below 4 and the parallelism efficiency decreases significantly, and the upper limit is constrained by the 24 GB GPU memory, and the range of α is 0–1, where determined by the form of αMixLoss. The remaining hyperparameters are control variables: the activation function is fixed to ReLU, the optimizer is Adam with a learning rate of 0.001, the weight decay is 1×10^{-5}, the Dropout rate is 0.2, and the random seed is fixed to ensure reproducibility.

The value range of hyperparameter α is set to (0,1), and the joint search optimization is performed by the Bayesian optimization algorithm in conjunction with other key hyperparameters including hidden layer unit, epoch and batch size. The process dynamically selects the optimal hyperparameter combination by constructing an agent model based on a Gaussian process and using expected improvement (EI) as a collection function.

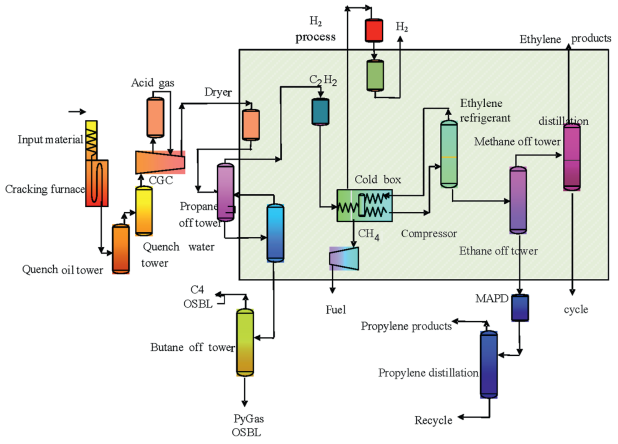

Fig. 7. The production process of ethylene.

3.3. Validation of methodological sophistication

In order to verify the superiority of the αMixLoss function in the proposed NNADM network design method, five loss functions such as Huber, Log_Cosh, MSE and fixed αMixLoss are used as comparisons in Table 3. Each loss function is based on Bayesian optimization, and the tuned parameters are consistent with the NNADM. In this paper, neural networks in complex petrochemical processes such as the GRU with an attention mechanism (AM-GRU), the GRU, the LSTM with an attention mechanism (AM-LSTM), the LSTM and the Bi-LSTM are used as baseline models.

As shown in Table 3, under the ethylene industrial production prediction environment, the αMixLoss function and the Huber loss function perform significantly better compared to other loss functions and are more suitable for predicting ethylene production in the industrial environment. Comparative analysis shows that the proposed αMixLoss function in this paper achieves a MSE

Table 1
The explanation of ethylene production process variables.

No.	Variable (Ton)	Abbreviations	Carbon emission factor /t·t^{-1}
1	Light oil	Lig	0.074
2	Naphtha	Nap	0.082
3	Extractive residue	Raf	–
4	Hydraulic oil	Hyd	0.010
5	Light hydrotreated tail oil	Lhy	0.074
6	Carbon3-5	C345	0.080
7	Other crude oil	Other	0.070
8	Ethylene	Ethylene	–

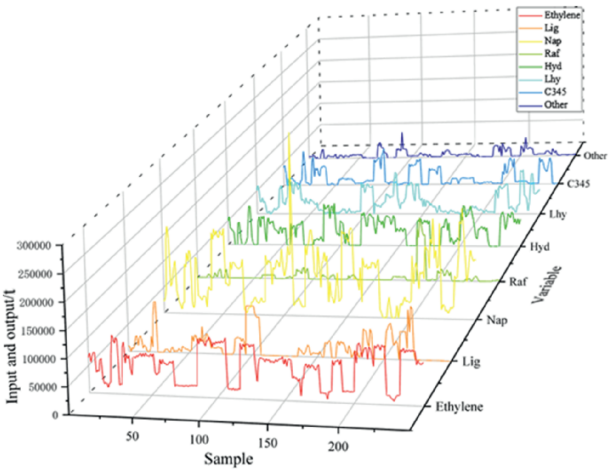

Fig. 8. A portion of the data display in ethylene production process.

Table 2
The selection range of hyperparameters of the NNADM design method.

Parameters	Parameter Selection Range
Hidden layer unit	(3, 32)
Epoch	(300, 600)
Batch size	(4, 64)
α	(0, 1)

reduction of 6.64% on average across the five baseline models compared to the huberLoss function. The experimental results show that by optimizing the α of the loss function, the NNADM can find a better loss function. Thus, the NNADM based prediction model is more accurate.

Fig. 9 systematically shows the MSE variation curves of the AM-GRU and GRU models on the validation set for different fixed α values (from 0 to 1, with an interval of 0.1). The experimental results show that the prediction model performance is optimal when α is in the 0.5–0.7 and 0.7–0.9 intervals, respectively, which verifies that the tunability of αMixLoss is more adaptive to outliers. This ablation experiment also shows that αMixLoss can achieve better performance at a specific α than using only fixed Log_Cosh ($\alpha = 0$) or MSE ($\alpha = 1$).

To further validate the efficacy of the proposed NNADM framework's architectural strategy, comparative analyses of distinct automated design approaches were conducted on AM-GRU neural architectures, with detailed experimental outcomes documented in Table 4.

From Tables 4 and it can be seen that the proposed NNADM network automated design algorithm in this paper outperforms other design methods in terms of function evaluation. MSE is reduced by 27% for the NNADM compared to Manual trial-and-error tuning. The NNADM with suboptimal Bayesian optimization still achieved 15% MSE reduction. This suggests that NNADM-based optimization neural networks can predict outcomes more accurately. In terms of time efficiency, it is clear that Manual trial-and-error tuning is done by experience and observation, which requires constant parameter tuning of the prediction model, and the results obtained are highly uncertain in terms of time and performance compared to other algorithms. Since the Grid Search algorithm uses an exhaustive search to enumerate parameter combinations, it takes the longest time to find the best parameters among all hyperparametric automatic design methods. Meanwhile, the NNADM and Bayesian optimization algorithms make full use of historical search information, the search time is reduced. Compared to the Bayesian optimization algorithm, the NNADM optimizes one more hyperparameter alpha and thus the time comes to be suboptimal, but its search speed is only 155s slower than the Bayesian optimization algorithm.

3.4. The predictive performance of ethylene production

This section builds upon the six baseline models introduced in Section 3.2 and the two most effective optimization design methods including the Bayesian optimization (Only use the MSE as the loss function) and the NNADM for a more comprehensive comparison of prediction performance. The optimized parameters of aforementioned design methods are enumerated in Table 5.

The prediction experiments were performed using the optimized parameters in Table 5, and five evaluation metrics (MSE, MAE, MAPE, RMSE, R^2) were used for a complete comparison. The values obtained from the experiment are shown in Table 6 and the model performance comparison is shown in Fig. 10.

As demonstrated in the experimental results presented in Table 6 and Fig. 10, the proposed NNADM method exhibits a substantial enhancement in prediction performance. Among six benchmark models, the MSE of the proposed NNADM method is reduced by an

Table 3
The MSE of different loss functions in the NNADM network design method, where $\alpha = 0.5$ refers to an α of 0.5 in a fixed αMixLoss.

Baseline models	αMixLoss	huberLoss	Log_Cosh	$\alpha = 0.25$	$\alpha = 0.5$	$\alpha = 0.75$	MSE
AM-GRU	$\mathbf{3.194 \times 10^7}$	3.475×10^7	4.513×10^7	3.995×10^7	3.523×10^7	3.829×10^7	3.795×10^7
GRU	$\mathbf{3.624 \times 10^7}$	3.844×10^7	4.796×10^7	4.35×10^7	3.966×10^7	3.865×10^7	4.124×10^7
AM-LSTM	$\mathbf{3.314 \times 10^7}$	3.421×10^7	3.612×10^7	3.412×10^7	3.786×10^7	3.538×10^7	4.089×10^7
LSTM	$\mathbf{3.615 \times 10^7}$	3.945×10^7	4.987×10^7	4.468×10^7	4.331×10^7	4.215×10^7	4.516×10^7
Bi-LSTM	$\mathbf{3.707 \times 10^7}$	4.024×10^7	4.561×10^7	4.336×10^7	4.689×10^7	4.874×10^7	4.49×10^7

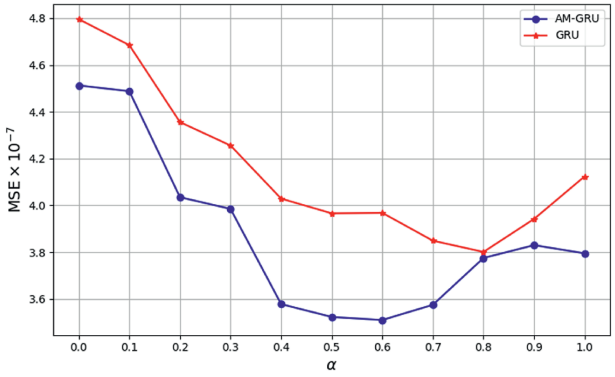

Fig. 9. The MSE of different α in the NNADM network design method.

average of 15.80%, the MAE by 5.08%, the MAPE by 27.21%, the RMSE by 8.07%, and the R^2 is improved by an average of 0.012. A comparison of the model performance with other modeling methods reveals that the AM-GRU exhibits superior performance on the test set, with a R^2 of 0.9821. Notably, only 1.79% of the variance is not explained by the prediction model, suggesting that it effectively captured the regular variations in ethylene production forecasts. Fig. 11 explicitly shows the loss curves of the training set and the validation set under the optimal configuration, and the experimental results show that the training set loss and the validation set loss maintain a consistent downward trend, with no significant signs of overfitting. It further proves that αMixLoss has strong robustness.

3.5. Optimization and energy savings for ethylene production

The NNADM-optimized neural architecture demonstrates superior predictive capabilities across multivariate operational scenarios, enabling precise yield estimation for dynamic production planning. Through multi-parametric simulations of feedstock formulation parameters and energy allocation schemes, optimization algorithms identify Pareto-optimal solutions that balance ethylene output with environmental constraints. This predictive framework provides actionable insights for (a) evaluating production capacity under novel operating conditions through virtual commissioning, and (b) deriving data-driven optimization strategies via historical pattern mining.

A substantial disparity between the simulated yield and the target yield implies that the present proportion of raw material inputs and energy consumption is sub-optimal, preventing the plant from attaining optimal operation. Conversely, a smaller gap indicates that there is little room for improvement. For

Table 4
The effect of different optimization methods on the AM-GRU model.

Optimization methods	Hidden layer unit	Epoch	Batchsize	MSE	Time/s
Manual trial and error	15	350	32	4.395×10^7	–
Grid search	32	400	16	4.243×10^7	15568
Random search	23	550	4	4.134×10^7	2857
Bayesian optimization	22	315	31	3.795×10^7	**1072**
NNADM	31	334	33	$\mathbf{3.194 \times 10^7}$	1227

Table 5
The comparison of hyperparameters of 5 baseline models with the NNADM optimization and the Bayesian optimization.

Baseline Models	Optimization methods	hidden layer unit	epoch	batch size	α
AM-GRU	NNADM	31	334	33	0.62
	Bayesian optimization	22	315	31	–
iTransformer	NNADM	28	305	32	0.39
	Bayesian optimization	28	450	26	–
GRU	NNADM	30	467	36	0.53
	Bayesian optimization	4	550	25	–
AM-LSTM	NNADM	30	320	30	0.69
	Bayesian optimization	18	310	37	–
LSTM	NNADM	4	325	27	0.89
	Bayesian optimization	4	339	21	–
Bi-LSTM	NNADM	15	507	31	0.53
	Bayesian optimization	8	571	39	–

Table 6
The performance of baseline models and different optimization methods in the ethylene production process.

Baseline models	Optimization methods	MSE	MAE	MAPE	RMSE	R^2
AM-GRU	Bayesian (MSE)	3.795×10^7	3770	3.87	6167	0.9705
	NNADM	$\mathbf{3.194 \times 10^7}$	**3504**	**3.07**	**5658**	**0.9821**
iTransformer	Bayesian (MSE)	3.825×10^7	3764	3.68	6173	0.9693
	NNADM	3.424×10^7	3667	3.38	5937	0.9812
GRU	Bayesian (MSE)	4.124×10^7	3824	5.27	6429	0.9681
	NNADM	3.624×10^7	3691	3.62	6019	0.9751
AM-LSTM	Bayesian (MSE)	4.089×10^7	3806	5.25	6396	0.9683
	NNADM	3.314×10^7	3558	3.29	5758	0.9825
LSTM	Bayesian (MSE)	4.516×10^7	3921	5.67	6721	0.9559
	NNADM	3.615×10^7	3692	3.69	6015	0.9689
Bi-LSTM	Bayesian (MSE)	4.49×10^7	3919	5.47	6726	0.9559
	NNADM	3.707×10^7	3723	3.78	6089	0.9704

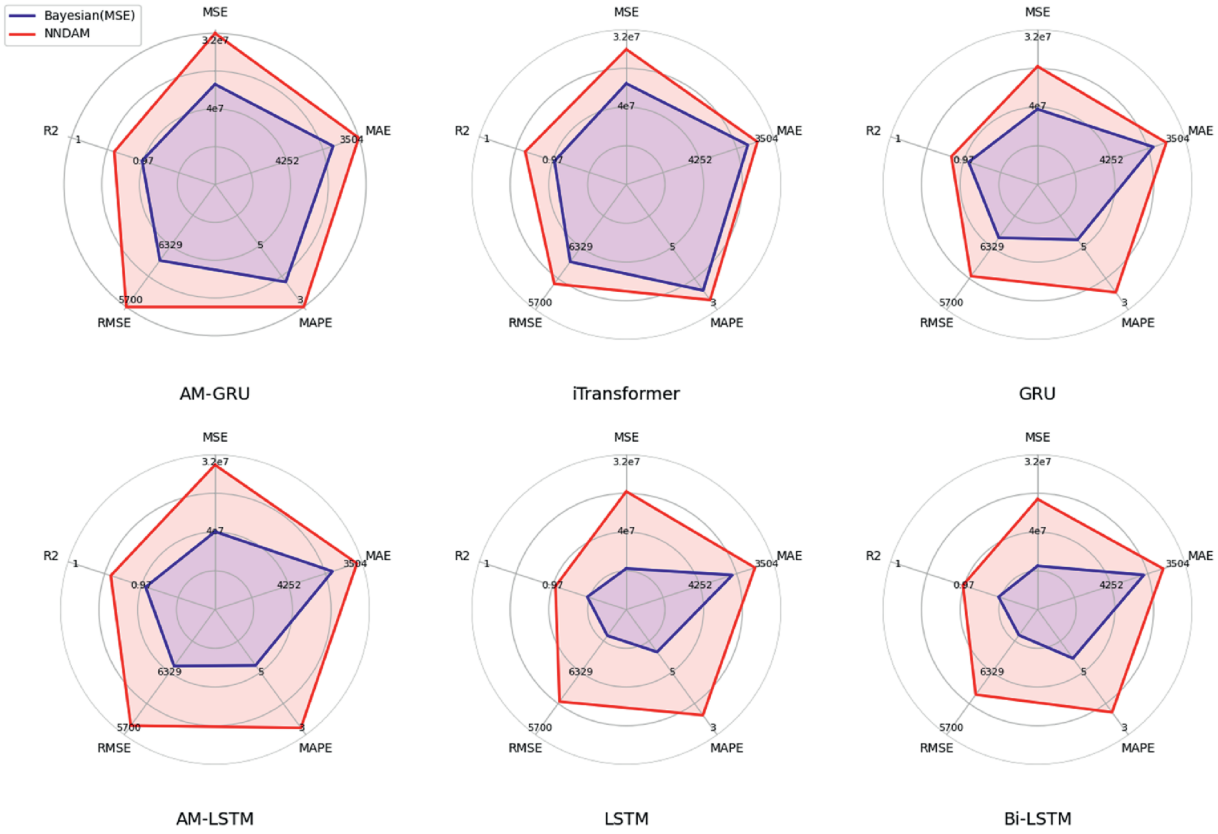

Fig. 10. Model performance comparison.

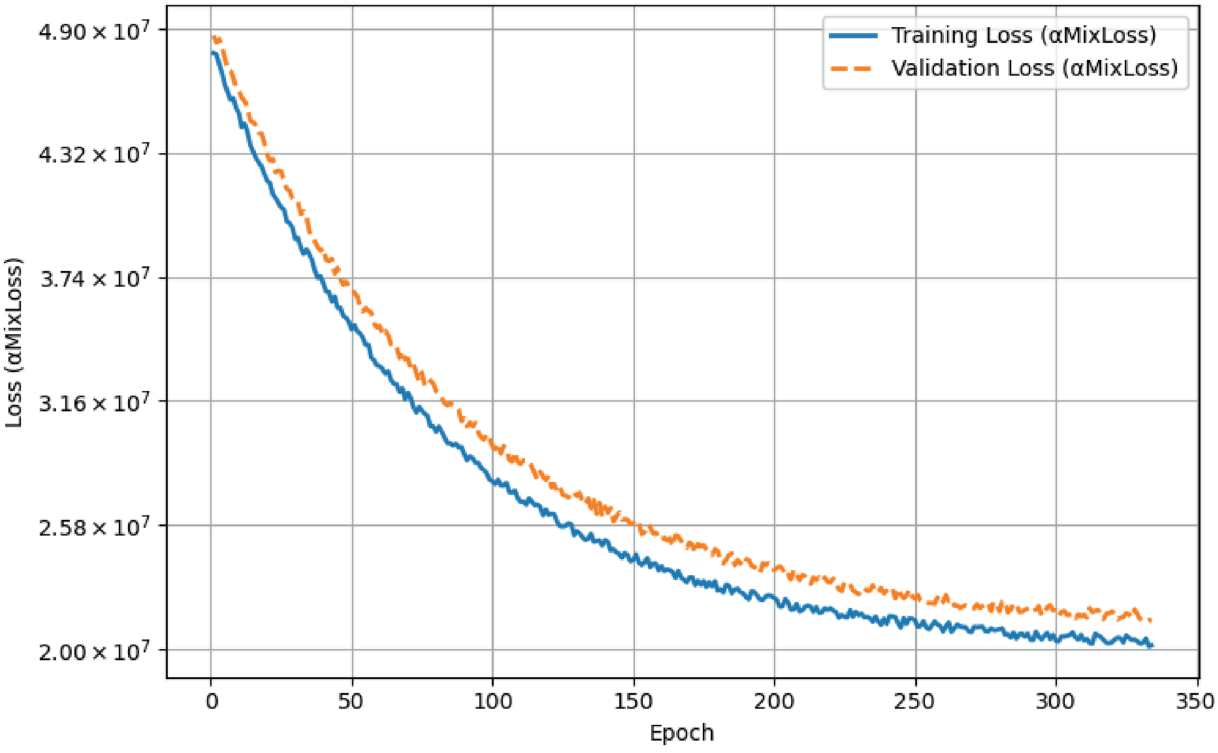

Fig. 11. The loss curves of the training set and validation set of the AM-GRU model integrated by NNADM.

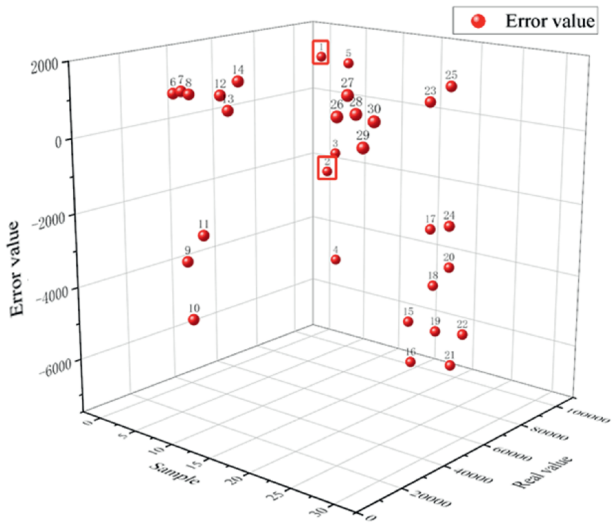

Fig. 12. The comparison between the real value and the predicted value of AM-GRU in the ethylene production process.

instance, the current production situation is in the optimization zone. By constructing predictive models through analytical processing of historical ethylene production records, an accurate forecasting framework is established to optimize raw material allocation cycles, enabling simultaneous enhancement of ethylene output and minimization of carbon footprint in chemical manufacturing processes. From the historical data, the scenarios that are similar to the current production condition and in the optimized operating zone are selected to optimize the current feedstock input and energy consumption. The ratio of energy input and feedstock input in the inefficient production configuration is adjusted, as shown in Fig. 12 for the optimization effect.

Among the randomly selected samples in the test set shown in Fig. 12, the test sample No. 2 have a large gap between the simulated yield and the plant current yield, which indicates that the test sample No.2 overall energy efficiency is low. The operating condition of the test sample No. 2 is not at full capacity, and the raw material ratio of No. 2 is unreasonably wasteful. Meanwhile, the true yield of sample No. 1 is closer to the true yield of sample No. 2, but sample No.2 have a smaller gap between the simulated yield and the plant current yield, indicating that the operating condition of sample No. 1 is closer to full capacity. Therefore, No. 2 inputs can be adjusted according to No. 1 raw material ratios, as shown in Fig. 13 for the input-output adjustment of No. 2 sample.

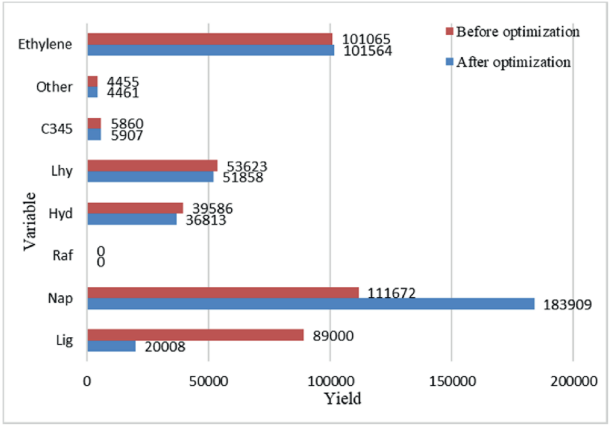

Fig. 13. The production adjustment of sample No. 2 of the ethylene.

Adjusting for No. 2 in Fig. 13 requires reducing the input Lig by 68992 tons, Hyd by 2773 tons, and Lhy by 1765 tons, and increasing Nap by 72237 tons, C345 by 47 tons, and Other by 6 tons. After adjustment, the ethylene production can be increased by 499 tons. Based on the carbon emission factors in Table 1, carbon emissions could be reduced by approximately 5108.1 tons, as shown in Eq. (21).

$$\Delta CO_2 = \sum_i (\text{Change in Input}_i \times \text{Carbon Emission Factor}_i) \tag{21}$$

where i represents each of the different raw materials including Lig, Hyd, Lhy, Nap, C345, and Other.

4. Conclusions

In order to solve the problem that traditional neural network modeling methods require a long parameter tuning process and the results of manual tuning are uncertain, the NNADM is proposed by incorporating the neural network parameters automatic optimization and loss function adaptive construction. Five commonly used yield forecasting models are chosen as baseline models including the AM-GRU, the GRU, the AM-LSTM, the LSTM, and the Bi-LSTM. And Hyperparameter optimization with the NNADM is performed for each baseline model. The experimental results show that the average MSE decreases more after the baseline models incorporating the NNADM. Therefore, the NNADM can effectively discover the optimal neural network hyperparameters. To conclusively demonstrate the operational superiority of the NNADM framework, a methodological comparison was executed encompassing empirical trial-based approaches (manual parameter tuning), systematic grid exploration, random search strategies, and probabilistic Bayesian optimization techniques. the average MSE of the NNADM for these five baseline models decreased by 15.8%, 12.1%, 19.0%, 20.0% and 17.4% respectively. The experimental results demonstrate the efficacy of the proposed loss function adaptive construction method in synthesizing the advantages of the MSE and the Log_Cosh loss function. Moreover, the energy-saving potential and input-output of the plant are analyzed based on the most accurate ethylene yield forecasting model, which can make the ethylene reduce about 8376.4 tons of carbon emission while ensuring 499 tons of additional production. In the future, we will continue to optimize the proposed model and continue to validate it in other production data from different complex petrochemical processes.

CRediT Authorship Contribution Statement

Guocong Lin: Writing – original draft, Visualization, Resources, Methodology, Data curation. Qingxu Ni: Writing – original draft, Resources, Methodology. Xuehai Liu: Software, Resources, Methodology. Zhiqiang Geng: Validation, Supervision, Resources, Project administration, Data curation. Feng Pan: Software, Resources. Jun Tian: Writing – review & editing, Visualization, Resources, Methodology. Yongming Han: Visualization, Resources, Project administration, Methodology.

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 is supported by the Xinjiang Uygur Autonomous Region Key R&D Project, China(2023B01031-3) and the National Natural Science Foundation of China (62422303 and 62373035).

Nomenclature

RealOut	the true value of data
PreOut	the model predicted value
R^2	Coefficient of Determination

References

[1] S.X. Gong, Multi-scale energy efficiency recognition and diagnosis scheme for ethylene production based on a hierarchical multi-indicator system, *Energy* 267 (2023) 126478.
[2] Z.Q. Geng, W.K. Kong, X.T. Wang, L. Wang, Y.M. Han, Adaptive search based Grey Wolf optimization algorithm for multi-objective optimization of ethylene cracking furnace, *Swarm Evol. Comput.* 92 (2025) 101810.
[3] J.S. Samuel, A.H. Muggeridge, Fast modelling of gas reservoir performance with proper orthogonal decomposition based autoencoder and radial basis function non-intrusive reduced order models, *J. Petrol. Sci. Eng.* 211 (2022) 110011.
[4] N. Liu, J. Wang, S.L. Sun, C.K. Li, W.D. Tian, Optimized principal component analysis and multi-state Bayesian network integrated method for chemical process monitoring and variable state prediction, *Chem. Eng. J.* 430 (2022) 132617.
[5] M.I. Jahirul, M.G. Rasul, R.J. Brown, W. Senadeera, M.A. Hosen, R. Haque, S.C. Saha, T.M.I. Mahlia, Investigation of correlation between chemical composition and properties of biodiesel using principal component analysis (PCA) and artificial neural network (ANN), *Renew. Energy* 168 (2021) 632–646.
[6] H.L. Chen, I. Ahmadianfar, G.X. Liang, A.A. Heidari, Robust kernel extreme learning machines with weighted mean of vectors and variational mode decomposition for forecasting total dissolved solids, *Eng. Appl. Artif. Intell.* 133 (2024) 108587.
[7] S. Afzal, A. Shokri, B.M. Ziapour, H. Shakibi, B. Sobhani, Building energy consumption prediction and optimization using different neural network-assisted models; comparison of different networks and optimization algorithms, *Eng. Appl. Artif. Intell.* 127 (2024) 107356.
[8] Y.Z. Li, R. Gupta, W.L. Li, Y. Fang, J. Toney, S.M. You, Machine learning-assisted life cycle assessment of biochar soil application, *J. Clean. Prod.* 498 (2025) 145109.
[9] D.J. Godwin, E.G. Varuvel, M.L. Jesu Martin, A. Jasmine R, F. Josephin Js, A comparative analysis of advanced machine learning models for the prediction of combustion, emission and performance characteristics using endoscopic combustion flame image of a pine oil–gasoline fuelled spark ignition engine, *J. Clean. Prod.* 484 (2024) 144284.
[10] F. Jani, S. Hosseini, A. Sepahi, S.K. Afzali, F. Torabi, R. Ghorbani, S. Houshmandmoayed, Enhancing quality control of polyethylene in industrial polymerization plants through predictive multivariate data-driven soft sensors, *Can. J. Chem. Eng.* 103 (5) (2025) 2234–2250.
[11] J. Taghinezhad, S. Sheidaei, Prediction of operating parameters and output power of ducted wind turbine using artificial neural networks, *Energy Rep.* 8 (2022) 3085–3095.
[12] G. Işık, H. Öğüt, M. Mutlu, Deep learning based electricity demand forecasting to minimize the cost of energy imbalance: a real case application with some fortune 500 companies in Türkiye, *Eng. Appl. Artif. Intell.* 118 (2023) 105664.
[13] X. Ma, J. Wang, J. Huang, Y. Ke, Energy consumption and carbon emission modeling and forecasting study with novel deep learning methods, *Expert Syst. Appl.* 290 (2025) 128314.
[14] B. Metcalfe, J.C. Acosta-Pavas, C.E. Robles-Rodriguez, G.K. Georgakilas, T. Dalamagas, C.A. Aceves-Lara, F. Daboussi, J.J. Koehorst, D.C. Corrales, Towards a machine learning operations (MLOps) soft sensor for real-time predictions in industrial-scale fed-batch fermentation, *Comput. Chem. Eng.* 194 (2025) 108991.
[15] Y.N. Sun, Z.E. Qu, Z.D. Liu, X.Y. Li, Hierarchical multi-scale decomposition and deep learning ensemble framework for enhanced carbon emission prediction, *Mathematics* 13 (12) (2025) 1924.
[16] X.Y. Zhou, E.W. Song, M.M. Wang, E.Q. Wang, Dynamic temperature control of dividing wall batch distillation with middle vessel based on neural network soft-sensor and fuzzy control, *Chin. J. Chem. Eng.* 79 (2025) 200–211.
[17] S.W. Gao, P.X. Yun, W.B. Yang, J. Yan, Dynamic soft sensor modelling based on data imputation and spatiotemporal attention, *Can. J. Chem. Eng.* 103 (10) (2025) 4751–4773.
[18] Y. Li, H.T. Cao, X. Wang, Z.N. Yang, N. Li, W.F. Shen, A new correlation-similarity conjoint algorithm for developing encoder-decoder based deep learning multi-step prediction model of chemical process, *Chem. Eng. Sci.* 288 (2024) 119748.
[19] Z. Li, Y.C. Xia, J. Long, C.S. Liu, L.F. Zhang, Multi-scale feature fused stacked autoencoder and its application for soft sensor modeling, *Chin. J. Chem. Eng.* 81 (2025) 241–254.
[20] R.Z. Homod, B.S. Munahi, H.I. Mohammed, M.A.A. Albadr, A. Abderrahmane, J. M. Mahdi, M.B. Ben Hamida, B.N. Alhasnawi, A.S. Albahri, H. Togun, U.F. Alqsair, Z.M. Yaseen, Deep clustering of reinforcement learning based on the bang-bang principle to optimize the energy in multi-boiler for intelligent buildings, *Appl. Energy* 356 (2024) 122357.
[21] G.Y. Xu, J. Shi, J.M. Wu, C.B. Lu, C.Y. Wu, D. Wang, Z. Han, An optimal solutions-guided deep reinforcement learning approach for online energy storage control, *Appl. Energy* 361 (2024) 122915.
[22] X.T. Deng, Y. Zhang, Y. Jiang, Y. Zhang, H. Qi, A novel operation method for renewable building by combining distributed DC energy system and deep reinforcement learning, *Appl. Energy* 353 (2024) 122188.
[23] K. Li, Y.C. Mu, F. Yang, H.Y. Wang, Y. Yan, C.H. Zhang, Joint forecasting of source-load-price for integrated energy system based on multi-task learning and hybrid attention mechanism, *Appl. Energy* 360 (2024) 122821.
[24] C.R. Song, H.D. Yang, J.Y. Cai, P. Yang, H. Bao, K.K. Xu, X.B. Meng, Multi-energy load forecasting *via* hierarchical multi-task learning and spatiotemporal attention, *Appl. Energy* 373 (2024) 123788.
[25] C. Brozos, J.G. Rittig, E. Akanny, S. Bhattacharya, C. Kohlmann, A. Mitsos, Predicting the temperature-dependent CMC of surfactant mixtures with graph neural networks, *Comput. Chem. Eng.* 198 (2025) 109085.
[26] Y.M. Han, X.H. Liu, C. Guo, H. Wu, M. Liu, Z.Q. Geng, Improved pearson correlation coefficient-based graph neural network for dynamic soft sensor of polypropylene industries, *Ind. Eng. Chem. Res.* 64 (1) (2025) 551–563.
[27] W.L. Guo, Y.Z. Wen, M.H. Liu, H.W. Chen, L.Q. Wen, Real-time, reagent-free total phosphorus soft sensor based on frequency-enhanced decomposed transformer model, *Measurement* 253 (2025) 117509.
[28] G.H. Ban, Y. Chen, Z.H. Xiong, Y.X. Zhuo, K. Huang, The univariate model for long-term wind speed forecasting based on wavelet soft threshold denoising and improved autoformer, *Energy* 290 (2024) 130225.
[29] Y.M. Han, C.K. Zeng, Q.X. Ni, J.Q. Wang, Z.Y. Chu, X.X. Zhang, Z.Q. Geng, L. Tan, Y.D. Liu, Time series prediction of anaerobic digestion yield and carbon emissions from food waste based on iTransformer model, *Chem. Eng. J.* 513 (2025) 163064.
[30] X.P. Guo, S.L. Pan, Y. Li, Global dynamic features and information of adjacent hidden layer enhancement based on autoencoder for industrial process soft sensor application, *Can. J. Chem. Eng.* 103 (5) (2025) 2218–2233.
[31] L.J. Ge, Y.L. Li, J. Yan, Y.Q. Wang, N. Zhang, Short-term load prediction of integrated energy system with wavelet neural network model based on improved particle swarm optimization and chaos optimization algorithm, *J. Mod. Power Syst. Clean Energy* 9 (6) (2021) 1490–1499.
[32] K. Nakamura, S. Soatto, B.W. Hong, Stochastic batch size for adaptive regularization in deep network optimization, *Pattern Recogn.* 129 (2022) 108776.
[33] R. Liang, T. Le-Hung, T. Nguyen-Thoi, Energy consumption prediction of air-conditioning systems in eco-buildings using hunger games search optimization-based artificial neural network model, *J. Build. Eng.* 59 (2022) 105087.
[34] M.S. Nawaz, B. Shoaib, M.A. Ashraf, Intelligent cardiovascular disease prediction empowered with gradient descent optimization, *Heliyon* 7 (5) (2021) e06948.
[35] H.B. Huang, X.R. Huang, W.P. Ding, S.W. Zhang, J. Pang, Optimization of electric vehicle sound package based on LSTM with an adaptive learning rate forest and multiple-level multiple-object method, *Mech. Syst. Signal Process.* 187 (2023) 109932.
[36] M. Zulfiqar, M. Kamran, M.B. Rasheed, T. Alquthami, A.H. Milyani, Hyperparameter optimization of support vector machine using adaptive differential evolution for electricity load forecasting, *Energy Rep.* 8 (2022) 13333–13352.
[37] Z.H. Chang, W. Yuan, K. Huang, Remaining useful life prediction for rolling bearings using multi-layer grid search and LSTM, *Comput. Electr. Eng.* 101 (2022) 108083.
[38] S. Narmadha, N.V. Balaji, Improved network anomaly detection system using optimized autoencoder – LSTM, *Expert Syst. Appl.* 273 (2025) 126854.
[39] M. Tunçel, A. Duran, Effectiveness of grid and random approaches for a model parameter vector optimization, *J. Comput. Sci.* 67 (2023) 101960.
[40] P. Sing Babu, A. Goswami, V.S. Pandit, Optimization of beam line parameters for space charge dominated multi-species beam using random search method, *Phys. Lett.* 376 (45) (2012) 3192–3198.
[41] S. Ansari, A. Ayob, M.S. Hossain Lipu, A. Hussain, M.H.M. Saad, Particle swarm optimized data-driven model for remaining useful life prediction of lithium-ion batteries by systematic sampling, *J. Energy Storage* 56 (2022) 106050.
[42] F. El Jamiy ElSaid, J. Higgins, B. Wild, T. Desell, Optimizing long short-term memory recurrent neural networks using ant colony optimization to predict turbine engine vibration, *Appl. Soft Comput.* 73 (2018) 969–991.
[43] H. Kedia, A. Pandey, A. Kumar, A. Majumdar, J. De, N. Ghosh, Optimization of process parameters in machining of inconel 718 super alloy on HMT (NH22) capstan lathe using genetic algorithm subject to minimization of surface roughness, *Mater. Today Proc.* 66 (2022) 3910–3915.
[44] S. Wang, L.H. Li, H.S. Zhang, X.F. Liu, N.Q. Li, Q.H. Wang, A local semi-supervised soft sensor modeling method based on SAE neural networks for spatiotemporal dynamic chemical process, *Ind. Eng. Chem. Res.* 63 (50) (2024) 21645–21661.
[45] Q.H. Wang, L.H. Li, Research on semi-supervised soft sensor modeling method for sulfur recovery unit based on ISSA-VMD-ESN, *Chem. Eng. Sci.* 298 (2024) 120397.
[46] B. Liu, Y.W. Liu, Corrosion life prediction of glass fiber reinforced plastics by optimized BP neural network, *Inorg. Chem. Commun.* 144 (2022) 109846.

[47] M. Ghasemi, S.K. Mohammadi, M. Zare, S. Mirjalili, M. Gil, R. Hemmati, A new firefly algorithm with improved global exploration and convergence with application to engineering optimization, *Decis. Anal. J.* 5 (2022) 100125.
[48] H.R. Nejati Mahmoodzadeh, M. Mohammadi, H. Hashim Ibrahim, S. Rashidi, T. Ahmed Rashid, Forecasting tunnel boring machine penetration rate using LSTM deep neural network optimized by grey wolf optimization algorithm, *Expert Syst. Appl.* 209 (2022) 118303.
[49] V. Gkonis, I. Tsakalos, A hybrid optimized deep learning model *via* the golden jackal optimizer for accurate stock price forecasting, *Expert Syst. Appl.* 278 (2025) 127287.
[50] J. Long, K. Deng, R.C. He, Closed-loop scheduling optimization strategy based on particle swarm optimization with niche technology and soft sensor method of attributes-applied to gasoline blending process, *Chin. J. Chem. Eng.* 61 (2023) 43−57.
[51] Y.N. Qian, Q.G. Li, W.X. Li, Q.T. Hu, B.Z. Yan, C.J. Yu, S.Y. Peng, Evaluation of hydraulic fracturing in coal seam based on time difference-particle swarm optimization algorithm, *Energy Fuel.* 38 (12) (2024) 10955−10971.
[52] Z. Sun, E. Caetano, S. Pereira, C. Moutinho, Employing histogram of oriented gradient to enhance concrete crack detection performance with classification algorithm and Bayesian optimization, *Eng. Fail. Anal.* 150 (2023) 107351.
[53] J.L. Li, R.X. Chen, X.Z. Huang, Y.Z. Qu, Development of deep residual neural networks for gear pitting fault diagnosis using Bayesian optimization, *IEEE Trans. Instrum. Meas.* 71 (2022) 2521715.
[54] D.F. Chen, W.L. Wu, K.Y. Chang, Y.H. Li, P.C. Pei, X.M. Xu, Performance degradation prediction method of PEM fuel cells using bidirectional long short-term memory neural network based on Bayesian optimization, *Energy* 285 (2023) 129469.
[55] D. Kim, K. Kwon, K. Pham, J.Y. Oh, H. Choi, Surface settlement prediction for urban tunneling using machine learning algorithms with Bayesian optimization, *Autom. Constr.* 140 (2022) 104331.
[56] J.P. Xu, S.W. Gao, X.C. Dang, W.F. Zhao, Q.S. Zhang, S.L. Qiu, BO-MADRSN: bayesian optimized multi-attention residual shrinkage networks for industrial soft sensor modeling, *Measurement* 224 (2024) 113477.
[57] R.M. Zou, J.X. Yang, Y. Wang, F. Liu, M. Essaaidi, D. Srinivasan, Wind turbine power curve modeling using an asymmetric error characteristic-based loss function and a hybrid intelligent optimizer, *Appl. Energy* 304 (2021) 117707.
[58] Z.T. Zhao, K.T. Chong, J.Y. Jiang, K. Wilson, X.C. Zhang, F. Wang, Low-carbon roadmap of chemical production: a case study of ethylene in China, *Renew. Sustain. Energy Rev.* 97 (2018) 580−591.