Heliyon 10 (2024) e30821

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Ultra-early prediction of the process parameters of coal chemical production

Zheng Li, Min Yao *, Zhenmin Luo, Qianrui Huang, Tongshuang Liu

College of Safety Science and Engineering, Xi'an University of Science and Technology, Xi'an 710054, China

ARTICLE INFO

Keywords:
Coal chemical industry
Process parameters
Bidirectional long short-term memory network
Loss function
Ultra-early prediction

ABSTRACT

Most accidents in a chemical process are caused by abnormal or deviations of the process parameters, and the existing research is focused on short-term prediction. When the early warning time is advanced, many false and missing alarms will occur in the system, which will cause certain problems for on-site personnel; how to ensure the accuracy of early warning as much as possible while the early warning time is a technical problem requiring an urgent solution. In the present work, a bidirectional long short-term memory network (BiLSTM) model was established according to the temporal variation characteristics of process parameters, and the Whale optimization algorithm (WOA) was used to optimize the model's hyperparameters automatically. The predicted value was further constructed as a Modified Inverted Normal Loss Function (MINLF), and the probability of abnormal fluctuations of process parameters was calculated using the residual time theory. Finally, the WOA-BiLSTM-MINLF process parameter prediction model with inherent risk and trend risk was established, and the fluctuation process of the process parameters was transformed into dynamic risk values. The results show that the prediction model alarms 16 min ahead of distributed control systems (DCS), which can reserve enough time for operators to take safety protection measures in advance and prevent accidents.

1. Introduction

The coal chemical industry involves the conversion of coal into various forms, such as gas, liquid, solid products, and semifinished goods, through chemical processes. These products are subsequently refined into chemical and energy-related products [1]. The production process of the coal chemical industry is characterized high toxicity, high temperature, high pressure, strong production continuity, high risk, difficult safety control, *etc*. It has strict requirements on production process parameters, and mastering the changing trend of key process parameters can greatly eliminate potential fluctuations and maintain the stability of working conditions [2,3]. According to an investigative report by the Center for Chemical Process Safety of the American Institute of Chemical Engineers, nearly all accidents related to chemical processes are caused by abnormal or deviating process parameters. Abnormalities or deviations of key process parameters such as temperature, pressure, flow rate, velocity, level, vibration, displacement, and concentration can increase the risk of accidents, particularly when they exceed design safety limits [4]. Although many automated instruments and control equipment have been introduced into coal chemical plants, the production process has yet to be fully automated, and many operations still require manual completion. Due to the hazards present in materials, equipment, and process flow during production, if operators fail to detect and handle abnormal working conditions on time, it is easy for risks to spread and accidents to occur [5].

* Corresponding author.
E-mail address: ndglym@163.com (M. Yao).

https://doi.org/10.1016/j.heliyon.2024.e30821
Received 29 November 2023; Received in revised form 18 April 2024; Accepted 6 May 2024
Available online 7 May 2024

Prediction of the process parameters is crucial in implementing dynamic risk warnings for coal chemical production processes [6]. Specifically, the prediction of process parameters enables the early detection of potential deviations or anomalies in the production process; this can provide operators with sufficient time to implement preventive measures in advance, thereby reducing the probability of abnormal situations [7–10]. The higher the prediction accuracy of key parameters and the longer the prediction cycle, the more time can be provided for on-site operators to eliminate accident hazards and reduce the possibility of hazard conversion accidents [11].

Various monitoring and prediction technologies for coal chemical process parameters have gradually developed in recent years. The known methods for predicting coal chemical process parameters include mechanism model-based, knowledge-based, and data-driven methods [12]. Zhong et al. comprehensively studied the characteristics of various forecasting methods and expounded suitable methods according to different data characteristics [13]. The mechanism model-based method requires prior physical and chemical knowledge to construct an accurate mathematical model that can fully reflect the changing mechanism of the process. There are mainly Bayesian analysis algorithms and grey prediction models such as GM (1,1) [14,15]. However, this method is challenging to model and is suitable for the early stages of coal chemical production, where only a single product is involved, and the production line is relatively simple. For modern coal chemical production processes that are far more complex and large-scale than in the past, it is unrealistic to use mechanisms to establish precise mathematical models of the production process due to the high coupling between variables and the strong temporal variability among the variables [16,17]. In addition, knowledge-based methods require a deep understanding of complex coal chemical production processes, necessitating substantial expertise and years of accumulated experience to study and interpret anomalies [18]. Data-driven methods involve analyzing historical data reflecting actual working conditions and using a machine learning model to establish the mapping relationships between target parameters and associated variables. Using a machine learning model in such methods obviates the need to consider the complex mechanism of coal chemical systems; instead, the methods should focus on only the information contained within the data, thus allowing for flexible prediction [19]. Accordingly, data-driven approaches have been extensively applied for early fault detection in crucial components (*e.g.* rolling bearings and gear-boxes) [20,21], equipment (*e.g.* fans, pumps, and wind turbines) [22,23], and entire industrial systems (*e.g.* power plant and chemical process) [24].

With the popularization of intelligent instruments, sensor networks, and distributed control systems (DCS), massive amounts of data are generated and stored in coal chemical processes [25]. These stored large quantities of historical data represent the entire production process or production process, providing rich and useable digital resources for analyzing and organizing industrial processes [26]. Therefore, using data-driven methods for predicting coal chemical process parameters is practical [27]. In recent years, scholars have mainly conducted related research into process parameter prediction using data-driven research methods. Among them, machine learning models have been extensively used to predict anomalies in industrial production processes due to their ease of implementation, comprehensibility, and well-established theoretical foundations. Common machine learning models include support vector machine (SVM) [28], support vector regression (SVR) [29], random forest (RF) [30], multi-layer perceptron (MLP) [31–33], extreme gradient boosting (XGBoost) [34], and extreme learning machine models (ELM) [35]. However, most machine learning models rely on superficial learning mechanisms and are limited by single-layer hidden-layer network structures, making it difficult to predict complex sequence data with significant non-linear and dynamic characteristics, which limits their ability to capture underlying information in long-term sequence data [36]. Consequently, such machine learning models may not detect patterns in long-time series data, rendering them unsuitable for process parameter prediction [37]. Researchers have increasingly focused on data-driven deep learning models to address these limitations, demonstrating strong learning ability and adaptability and excellent performance in time series prediction *inter alia* [38,39].

Deep learning models are revolutionary technologies in artificial intelligence, which endow them with infinite feature mining and function fitting capabilities by cascading multiple non-linear processing layers. Firstly, deep learning models can directly process diverse types of raw monitoring signals and transform and extract essential features layer-by-layer without relying on manual experience [40]. Secondly, the mechanism of the coal chemical process is complex, and the modes are diverse. In addition, external disturbances such as equipment failures and raw material changes result in high correlation, coupling, and uncertainty of parameters. Deep learning models' strong non-linear mapping ability can better characterize this complex function distribution [41]. Finally, coal chemical production has dynamic time-varying characteristics, and there is a potential correlation between data at different time points. Deep learning models can better capture, interpret, and store the temporal attributes of variables [42]. Common deep learning models include convolutional neural networks [43], recurrent neural networks [44], long and short-term memory networks (LSTM) [45], bidirectional long short-term memory networks (BiLSTM) [46,47], gated recurrent units [48], and temporal convolutional networks [49]. Among these models, LSTM networks have been extensively adopted, particularly for tasks involving time series data, because they can learn long-term historical information from time series data while mitigating gradient explosion and the vanishing problem. For example, Ma et al. [50] proposed a novel deep neural network based on a convolutional LSTM network to predict the remaining useful life of bearings. This network uses time-frequency signals as input data; it thus preserves the advantages of LSTM while leveraging the characteristics of time-frequency data. Wang et al. [51] proposed a data-driven multivariate regression method based on an LSTM network with residual filtering (LSTM-RF). This method employs a filter to smooth the residuals between actual flight data and predicted values and then compares them with statistical thresholds to detect faults in unmanned aerial vehicles. Kumari et al. [52] integrated uniform manifold approximation and projection methods with an LSTM model for predicting fire conditions in enclosed areas of coal mines. This approach provides early warnings of potential mine accidents, thereby enhancing the safety of miners.

In addition, time series prediction models contain numerous parameters. Researchers generally select these parameters according to their experience, which may only sometimes obtain the optimal parameters. However, advancements in intelligent algorithms have resulted in new approaches to address this challenge. Researchers increasingly integrate neural network models with intelligent

algorithms to determine optimal parameters [53]. This integration has led to substantial improvements in the accuracy of integrated predictive models compared with single predictive models [54].

Numerous time series forecasting models and methods have been developed, however, most studies have focused on short-term forecasting. In coal chemical production processes, on-site personnel often need help to address some complex operations within a limited time. Therefore, extending the warning time beyond only a few minutes is essential to ensure that operators have sufficient time to take appropriate actions. However, extending the warning time may result in numerous false alarms and missed alarms, which may cause inconvenience for on-site personnel. Therefore, striking a balance between the extension of early warning times and the provision of accurate warnings is a key technical problem that must be addressed [55]. In addition, existing early warning research mainly focuses on prediction of production process parameters and has yet to achieve further risk quantification. The present study proposes an ultra-early prediction model for coal chemical production process parameters. This model predicts trends of variation in risks associated with abnormal key parameters over a period exceeding 5 min [56]. The contributions of this study are outlined as follows:

(1) A BiLSTM model was established using time series data of process parameters in coal chemical production. This model considers historical and future data to identify variation patterns of process parameters and to predict their variation trend;
(2) The study used the whale optimization algorithm (WOA) to optimize hyperparameters in the BiLSTM network model, including the learning rate, dropout rate, number of iterations, and number of hidden layer neuron nodes. This optimization eliminated subjectivity associated with empirical parameter selection and addressed the problem of traditional BiLSTM algorithms falling into local optima, thus improving prediction accuracy and convergence speed;
(3) The WOA-BiLSTM model was used to predict process parameters, which were adopted to construct a modified inverted normal loss function (MINLF). On the basis of residual time theory, this function was used to calculate the risk, including inherent risk and trend risk, of abnormal fluctuations of process parameters. Thus, the study derived an ultra-early prediction model, namely, WOA-BiLSTM-MINLF, which can transform parameter fluctuations into dynamic risk values, thereby enabling the long-term prediction of parameters of coal chemical production processes. By incorporating both inherent risk and trend risk, this model addresses the problem of false alarms and missed alarms associated with long-term forecasts;
(4) To assess the predictive performance of the WOA-BiLSTM model, RNN, LSTM, BiLSTM, and WOA-LSTM models were trained using the same sample data, and the predictive performance of these models was comprehensively analyzed and compared.

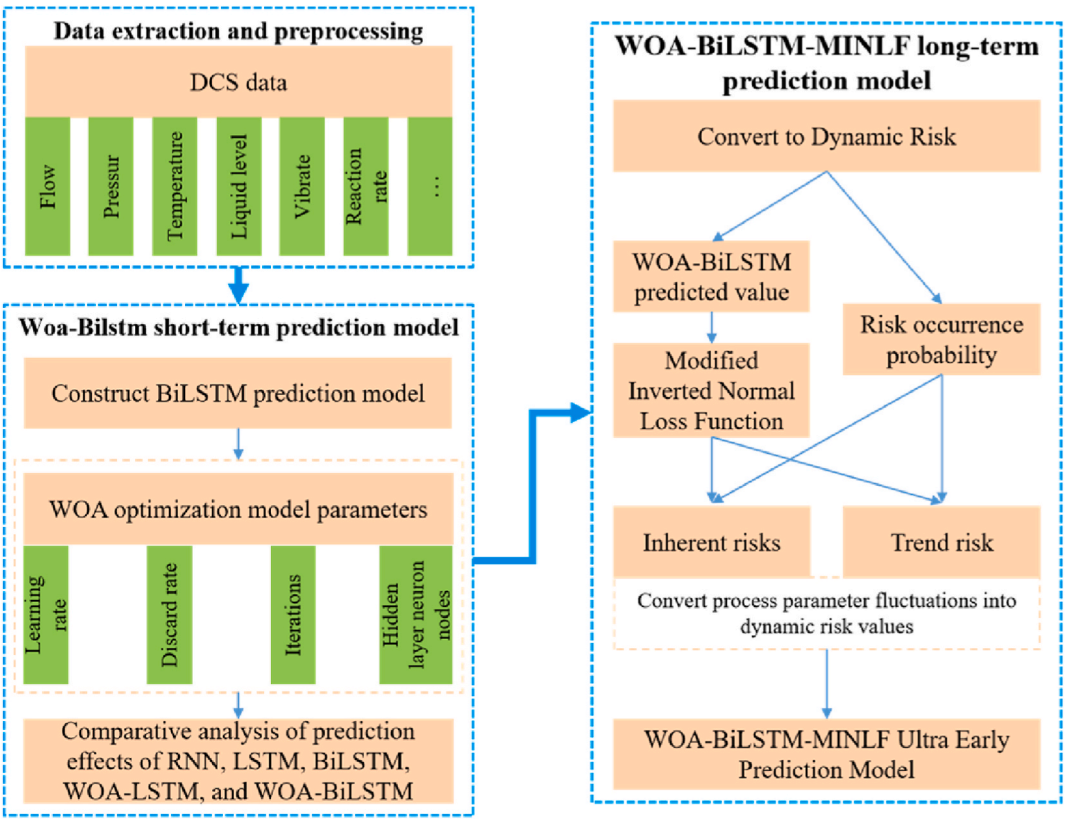


Fig. 1. Flowchart through the proposed model.

2. Models and methods

A BiLSTM predictive model based on the temporal variation of process parameters in coal chemical production was established. The Whale Optimization Algorithm (WOA) was used to optimize hyperparameters such as learning rate, dropout rate, training batch, and number of neurons in the BiLSTM network, forming a WOA-BiLSTM process parameter prediction model. Furthermore, the predicted values of process parameters were constructed in the form of a MINLF, and residual time theory was used to calculate the probability of abnormal fluctuations in process parameters, and ultimately construct a WOA-BiLSTM-MINLF process parameter ultra early prediction model that includes inherent and trend risks, converting the process parameter fluctuation process into dynamic risk values. Monitoring real-time changes in process parameters can predict dynamic risks during production. In practical applications, historical data of a key parameter is extracted from DCS as a dataset, which has time attributes. The first 80 % of the dataset is used as the training set, and the last 20 % is used as the testing set. By training and testing, the established WOA-BiLSTM model can obtain the optimal hyperparameter prediction. Real-time monitoring values in DCS are then used as inputs to the WOA-BiLSTM-MINLF model to obtain the real-time predicted values of the WOA-BiLSTM model. The constructed MINLF is further used to convert the above-predicted values into dynamic risk values, which are the outputs of the WOA-BiLSTM-MINLF model, thereby quantifying the risk caused by parameter fluctuations. The overall flow of this model is presented in Fig. 1.

2.1. Long short-term memory (LSTM)

Neural networks are primarily divided into feedforward and feedback neural networks [57]. In a feedforward neural network, each network level operates relatively independently. Although this architecture has its merits, its shortcomings can be magnified when assessing the logical relationships within data, often resulting in poor fitting [58]. Conversely, feedback neural networks, after enhancements, are more effective in analyzing the logical relationships within complex data samples [59]. RNNs constitute a type of feedback neural network. Unlike feedforward networks, the levels of RNNs are not independent, and information flows bidirectionally, enabling the output of one layer to serve as the input for subsequent hidden layer nodes [60]. The network structure is presented in Fig. 2.

The introduction of RNNs marked a breakthrough in neural network technology, particularly for tasks involving time series data. However, with the progress of research on RNNs, some notable limitations of such networks have been identified, such as gradient disappearance and gradient explosion. Hochreiter [61] proposed a novel recursive network architecture called LSTM, which replaces hidden layer neurons in the RNN with memory functions to address these limitations. In this architecture, introducing "gates" to control the loss or increase of information mitigates the problem of vanishing and exploding gradients in RNNs [62]. The LSTM architecture is presented in Fig. 3. Numerous studies have demonstrated that LSTM networks offer substantial improvements in time series prediction compared with conventional methods such as BPNN, SVR, and RNN models [63–65].

The LSTM unit comprises three gates: forget, input, and output. These gates play a central role in controlling the transmission of internal information within the unit, allowing it to forget or retain historical details in a selective manner.

2.2. Bi-directional long short-term memory (BiLSTM)

Although LSTM networks have proven more effective in processing long time series data when compared with conventional RNNs, both types of networks can process data in only a unidirectional manner; that is, they rely solely on past data to predict future data, ignoring the information contained in the upcoming data points. A BiLSTM network is an extension of the one-way LSTM network, composed of a forward LSTM model and backward LSTM models. This model architecture not only retains the advantages of LSTM in addressing sequences with long-term dependencies but also overcomes the LSTM network's shortcomings concerning its inability to consider future data [66]. The BiLSTM network structure is presented in Fig. 4. In the context of predicting process parameters in the coal chemical industry, the output of a BiLSTM model contains information on both the past and future changes in process parameters, thereby enhancing the accuracy of the model [67].

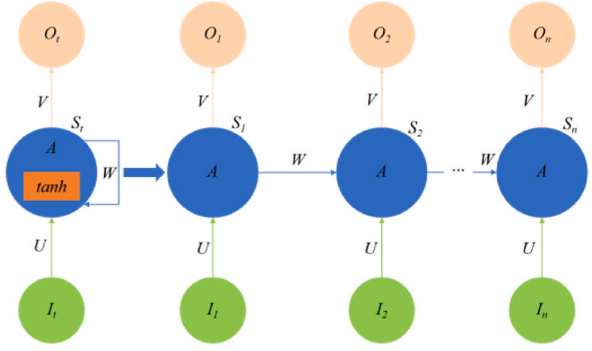


Fig. 2. The structure of recurrent neural network.

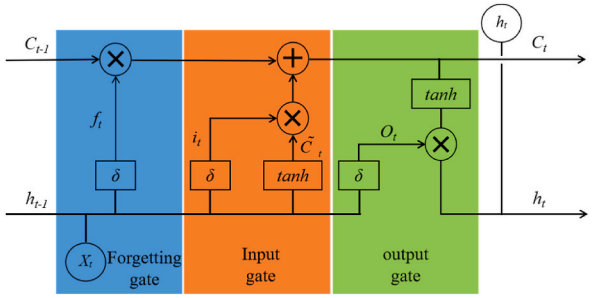


Fig. 3. LSTM cell structure.

The BiLSTM model analyses forward and backward sequences by employing two hidden layers. The final predicted output value Y_t is jointly determined by the forward and reverse hidden layers [68].

2.3. WOA

In BiLSTM neural networks, network parameters are usually continuously adjusted using control variable methods or based on the experience of previous researchers. However, these approaches may lack a sound theoretical foundation. The swarm intelligence optimization algorithm is a rapidly developing heuristic algorithm whose basic theory is to simulate the collective behavior of various biological populations in nature, utilize information exchange and cooperation between groups, and achieve the goal of optimizing model parameters through interaction between limited individuals. The swarm intelligence optimization algorithms mainly include particle swarm optimization algorithm, whale optimization algorithm, ant colony algorithm, grey wolf optimization algorithm, sparrow search algorithm, *etc*. [69]. The unique position update mechanism of the Whale Optimization Algorithm enables the entire optimization process to smoothly transition from the global search stage to the local search stage, balancing the exploration and development capabilities of the algorithm, resulting in faster convergence [70]. In the present study, the WOA was adopted to address the problems above and optimize BiLSTM network parameters, including the learning rate, dropout rate, number of iterations, and number of hidden layer neuron nodes, thereby improving the speed and accuracy of parameter selection.

The WOA is characterized by a rapid optimization speed, robust global convergence, straightforward principles, ease of implementation, and the requirement of a minimal number of parameters [71]. During their hunt, humpback whales randomly collaborate to encircle a shoal of fish or disperse their prey by blowing bubbles. Additionally, these whales might randomly choose to approach other whales in more advantageous geographical locations. In the WOA, the position of each whale represents a viable solution to an optimization problem.

In the WOA search space, the search particle is first initialized. When $|A| < 1$, the WOA initiates a local search, however, when $|A| > 1$, the WOA enters a global search.

$$a(t) = 2 - \frac{2t}{T} \tag{1}$$

$$A(t) = 2a(t)r - a(t) \tag{2}$$

$$C(t) = 2r \tag{3}$$

where t represents the number of iterations that have been completed and T represents the maximum number of iterations allowed, $r \in [0,1]$, $A \in [-a, a]$. From the second iteration, a continues to decrease until it finally reaches 0.

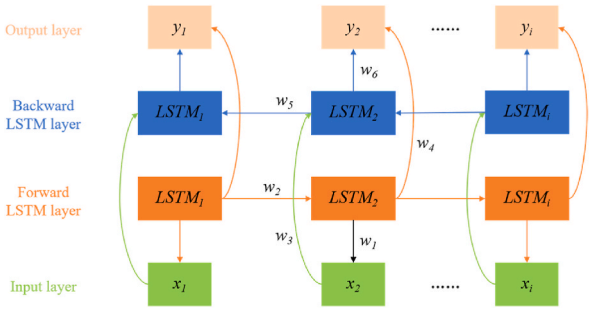


Fig. 4. The structure of BiLSTM.

(1) Local search phase

In this stage, other particles move closer to the optimal particle and update their position information concurrently. This phase can be modelled using contraction enveloping and spiral updating. X^* represents the optimal solution so far.

① Contraction enveloping phase

In this phase, a gradually decreases, which leads to a decrease in the range of A. The search particle X randomly selects the direction to move, and its updated position is determined using the following formula:

$$\vec{X}(t+1) = \vec{X}^*(t) - A(t) \cdot \vec{D}(t) \tag{4}$$

$$\vec{D}(t) = \left| C(t) \cdot \vec{X}^*(t) - \vec{X}(t) \right| \tag{5}$$

where $\vec{D}$ is the random distance (*i.e.* the distance between the target and the search particle).

② Spiral renewal phase

After the search particle is updated, the position can be determined using the spiral equation as follows:

$$\vec{X}(t+1) = \vec{D'}(t) \cdot e^{bl} \cdot \cos(2\pi l) + \vec{X}^*(t) \tag{6}$$

$$\vec{D'}(t) = \left| \vec{X}^*(t) - \vec{X}(t) \right| \tag{7}$$

where $\vec{D'}$ is the distance between the optimal solution and the searched particle, b is a constant, and l is any random number within $[-1, 1]$.

The probability $p \in [0, 1]$ is introduced to indicate how whales hunt. The value and position update of p are expressed in Formula (8) and Table 1.

$$\vec{X}(t+1) = \begin{cases} \vec{X}^*(t) - A(t) \cdot \vec{D}(t), & p < 0.5 \\ \vec{D'}(t) \cdot e^{bl} \cdot \cos(2\pi l) + \vec{X}^*(t), & p \geq 0.5 \end{cases} \tag{8}$$

(2) Global search phase

The population randomly selects a search particle to update in the global search phase. Other particles within the population then move away from this selected particle, enabling a global search. The formulae governing this process are as follows:

$$\vec{X}(t+1) = \overrightarrow{X_{rand}} - A(t) \cdot \vec{D}(t) \tag{9}$$

$$\vec{D}(t) = \left| C(t) \cdot \overrightarrow{X_{rand}} - \vec{X}(t) \right| \tag{10}$$

where $\overrightarrow{X_{rand}}$ denotes the randomly selected search particle in the population. X^* is updated by comparing the randomly selected with the current optimal particle until the termination condition is met. Additionally, A is influenced by a. The global and local search processes occur alternately, which ensures that the WOA can avoid local optimal solutions and accurately identify global optimal solutions.

The operating flow of the WOA is presented in Fig. 5.

2.4. WOA-BiLSTM model

The WOA addresses the problem of the traditional BiLSTM network's susceptibility to local optima and enhances the accuracy of parameter optimization. The operating flow of the WOA-BiLSTM model for analyzing the time series characteristics of coal chemical production processes is illustrated in Fig. 6.

The BiLSTM network hyperparameters in this model, such as the learning rate, dropout rate, number of iterations, and number of

Table 1
Correspondence between p and the position update mode.

p value	$p < 0.5$	$p \geq 0.5$
Location update mode	Contraction enveloping	Spiral renewal

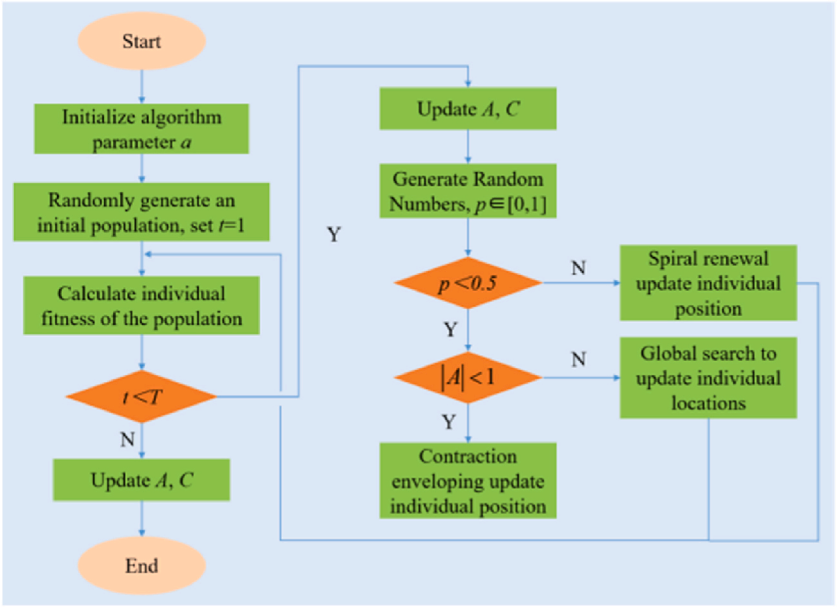


Fig. 5. The WOA optimization algorithm flow.

hidden layer neuron nodes, are treated as an optimization problem for the WOA. The objective function used for optimization is the RNN training error function. The global optimal solution is the whale parameter corresponding to the minimum value of the objective function. The optimal BiLSTM network structure can be determined by automatically optimizing the BiLSTM network hyper-parameters using the WOA, after which the model is trained and predicted. Finally, the trained model is employed to perform predictions of the variation trend of the parameters of coal chemical production processes. These steps are detailed as follows:

Step 1. Preprocessing of the data. Specifically, to obtain the time series data from the DCS, abnormal data are processed and normalized first. Subsequently, the normalized data are divided into training and test sets.

Step 2. The learning rate, number of iterations, number of hidden layer neuron nodes, and other relevant BiLSTM network parameters are initialized in addition to the BiLSTM architecture.

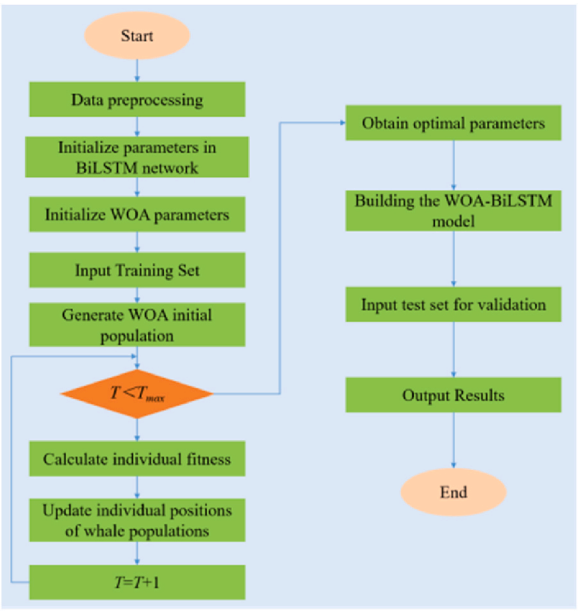


Fig. 6. The prediction model flow of WOA-BiLSTM process parameters.

Step 3. Initialization of WOA parameters. Specifically, WOA parameters such as population size and the number of iterations can be defined. The dimensions of individuals in the optimization search space are determined based on the hyperparameters to be optimized. The error function of the BiLSTM network model is used as the fitness function for the WOA;

Step 4. The position of each whale is encoded based on the initial parameters of the BiLSTM network. The fitness of each whale is calculated. The best search location is identified by comparing the fitness of individuals;

Step 5. The position of each whale individual is updated using an update formula. The fitness of the individuals is recalculated after the update. Whether an individual's fitness is lower than the historical optimal fitness is determined. If lower, the position and fitness of the individual should be updated to the current value. Otherwise, leave them unchanged;

Step 6. Whether the iteration termination condition is met is determined. If the condition is met, the iteration is terminated, yielding the optimal solution. If not, we return to the previous step and continue the iteration until the termination condition is met;

Step 7. The optimized hyperparameters are assigned to the BiLSTM model again, and then the test data set is used to test and analyze the accuracy of the prediction results.

2.5. WOA-BiLSTM-MINLF model

2.5.1. Loss function

Process deviations combined with failure of protection layers and control systems induce failures that increase the likelihood of an accident and subsequently increase the operational risk associated with process systems. Traditional risk assessment techniques cannot account for risk variations due to process deviations because the conventional methods are static [71]. The loss function is commonly used to quantify losses related to deviation from the target value and has been applied by scholars in many cases, but more research into process safety risk assessment is required [72].

Taguchi proposed a quadratic loss function (QLF) to illustrate losses to society associated with deviations of quality characteristics from their operational targets in industrial applications [73]. However, it is unrealistic to use this function in many manufacturing processes due to the unbounded and symmetric characteristic of Taguchi's loss function [74]. Scholars have improved QLF by truncating it at the point where the function intersects with the maximum loss to provide a quantifiable maximum loss. Spiring proposed the inverted normal loss function (INLF), which has a bounded value from above, and its supremum can be specified by the user. It is more flexible and provides a more reasonable assessment of the loss associated with deviations from the target [75]. Sun et al. refined the INLF further and developed the modified INLF (MINLF), which has a shape parameter specified by the user, and its value determines the slope of the function in the neighborhood of the target value [76]. MINLF provides greater flexibility in handling symmetric loss cases by specifying its shape parameters [77]. Spiring and Leung further extended the concept of INLF to other inverse probability density functions, including the inverted beta loss function (IBLF) and the inverted gamma loss function (IGLF) [78]. IBLF provides both the traditional features of loss functions and asymmetrical loss situations; IGLF can be adopted to represent processes with continuous asymmetric loss.

Although loss functions have been applied in many cases, the key to using them is to choose a suitable loss function that fits the characteristics of the evaluation object. Using inappropriate loss functions may lead to inaccurate evaluation results [79]. Hashemi et al. extensively investigated the applicability of various loss functions in process operation safety assessment and derived that MINLF and IBLF are more adaptable in representing losses related to process deviations [80]. However, construction of the MINLF is easier when compared to IBLF due to the simplicity of the formulation, and also, there is no need for transformation of the scales. The most widely used univariate loss functions are listed in Table 2.

In Table 2, y denotes the status parameter value, $L(y)$ is the actual loss at y, T is the target value, EML denotes the estimated maximum loss, Δ is the distance from the target to the point where the maximum loss EML first occurs, and α and γ are shape parameters and need to be determined from additional information for MINLF, IBLF, and IGLF.

Considering the characteristic of random deviation between process parameters and actual values in coal chemical production processes, the MINLF was adopted to represent dynamic risks under abnormal working conditions. This function enables adjusting the shape of the loss function through the adjustment of a shape factor, thus providing a better representation of the actual loss. The formula for this function can also be simplified as follows:

Table 2
Listing of univariate loss function formulations.

Type of loss function	Formulation of loss function	Formula
QLF	$L(y) = B(y-T)$ *where* $B = ELM/\Delta^2$	(11)
INLF	$L(y) = ELM\{1-\exp(-(y-T)^2/2\gamma^2)\}$ *where* $\gamma = \Delta/4$	(12)
MINLF	$L(y) = \frac{EML_\Delta}{1-\exp\{-0.5(\Delta/\gamma)^2\}}\{1-\exp(-(y-T)^2/2\gamma^2)\}$	(13)
IBLF	$L(y) = ELM\{1-C[y(1-y)^{(1-T)/T}]^{(\alpha-1)}\}$ *where* $C = \{T(1-T)^{(1-T)/T}\}^{1-\alpha}$	(14)
IGLF	$L(y) = ELM\left\{1-\left[\frac{y\exp(1-(y/T))}{T}\right]^{\alpha-1}\right\}$	(15)

$$L = \frac{1 - e^{-(y-T)^2/2\gamma^2}}{1 - e^{-\Delta^2/2\gamma^2}} \tag{16}$$

To derive a loss function that is tailored to the characteristics of measured parameters, various shape factors should be determined based on random fluctuations of these parameters. This approach enables the precise description of the expected loss resulting from parameter deviations. As indicated by the aforementioned formula, when the parameter is equal to the target value, the loss value is 0; the further the parameter deviates from the target value, the larger the expected loss value.

2.5.2. Probability of abnormal fluctuation of process parameters

Residual time theory was employed to calculate the probability of abnormal fluctuations of process parameters [81]. This theory can quantify the risk associated with deviations from normal operating conditions by calculating the time required for process parameters to reach the alarm threshold. If more time remains, then more adequate measures can be taken against abnormal working conditions; thus, the safety of the production process is enhanced. Conversely, if less time is available, the probability of abnormal conditions and subsequent hazards is higher. The remaining time t can be calculated as follows:

$$t = \frac{y_{alarm} - y}{\Delta V} \tag{17}$$

where y_{alarm} is the alarm threshold of the process parameters and ΔV is the instantaneous change rate of the process parameters, which is obtained by fitting the data curve by using the least square method and calculating the first-order differential of the fitting polynomial.

The probability density function $f(t)$ for abnormal conditions is expressed as follows:

$$f(t) = \lambda e^{-\lambda t}, \quad t \geq 0 \tag{18}$$

where t is the remaining time and λ is the reciprocal of the remaining time allowed when process parameters fluctuate. This reciprocal is calculated as follows:

$$\lambda = \frac{\Delta V_{\max}}{y_{alarm} - y} \tag{19}$$

where $\Delta V_{\max}$ denotes the maximum rate of change of process parameters under normal working conditions. For a specific process parameter, $\Delta V_{\max}$ is a constant.

Combining Formulae (17) to (19) can yield the probability P of the occurrence of abnormal conditions:

$$P = \int_t^{\infty} \lambda e^{-\lambda t} dt = e^{-\frac{\Delta V_{\max}}{\Delta V}} \tag{20}$$

2.5.3. Risk associated with abnormal fluctuations of process parameters

The risk of abnormal fluctuations of process parameters can be categorized into inherent risk and trend risk. Inherent risk refers to the risk that is already present because of process parameter deviations. The severity of consequences, denoted by L, is measured using the loss function. The probability of the occurrence of abnormal working conditions is $P = 1$. Therefore, according to the risk calculation formula, the inherent risk R_1 can be expressed as follows:

$$R_1 = L \times P, \quad (P = 1) \tag{21}$$

The maximum value of the loss function is 1; accordingly, the severity of the maximum consequence of process parameter deviation is calculated as $L^{'} = 1 - L$. The probability P of occurrence of abnormal conditions can be computed using Formula (20). Therefore, on the basis of the risk calculation formula, the trend risk R_2 of the process parameters can be expressed as follows:

$$R_2 = (1 - L) \times P, \quad \left(P = e^{-\frac{\Delta V_{\max}}{\Delta V}}\right) \tag{22}$$

In summary, the risk of abnormal fluctuations of process parameters can be expressed as follows:

$$R = R_1 + R_2 \tag{23}$$

On the basis of the aforementioned methodology, the process parameter values predicted by the WOA-BiLSTM model, as detailed in Section 2.4, can be transformed into risk values. These values can thus be used to execute dynamic risk prediction.

2.5.4. WOA-BiLSTM-MINLF ultra-early prediction model

The WOA-BiLSTM model can predict the state values of the parameters of coal chemical production processes. The alarm threshold for the process parameters is denoted as y_{alarm}, the interlock alarm threshold is denoted as I, and the average value is denoted as $\overline{Y}$. On the basis of Formula (16), let I be the value of the process parameters causing maximum loss and $\overline{Y}$ be the target value of the process parameters. Hence, $T = \overline{Y}$, $\Delta = I - \overline{Y}$. The expected loss caused by abnormal fluctuations of process parameters in coal chemical production is expressed as follows:

$$L=\frac{1-e^{-(y-\overline{Y})^2/2\gamma^2}}{1-e^{-(I-\overline{Y})^2/2\gamma^2}} \tag{24}$$

According to the ALARP principle and a grading strategy based on the Pareto distribution, the region between the unacceptable risk value and the maximum risk value is designated as the unacceptable high-risk region, accounting for 20 % of the total risk range [56]; because $L \in [0,1]$, the unacceptable risk value is set as the risk alarm threshold; the risk alarm threshold can be set to $L_{alarm} = 0.8$.

When the predicted process parameters y reach the interlock alarm threshold I, the loss function reaches its maximum value (*i.e.* 1). When $y = \overline{Y}$, it indicates that the equipment and hazardous materials are in the most stable state; therefore, the loss function is 0. As mentioned, the risk of abnormal fluctuations of process parameters includes inherent risk R_1 and trend risk R_2. When $y \geq y_{alarm}$ and $R_2 = 0$, the risk value should be greater than or equal to the risk threshold (*i.e.* $R_1 \geq 0.8$) to prevent false negatives. When $y \leq y_{alarm}$ and $R_2 = 0$, the risk value should be less than or equal to the risk threshold (*i.e.* $R_1 \leq 0.8$) to prevent false positives. On the basis of these conditions, when $y = y_{alarm}$ and $R_1 = 0.8$, then the loss function $L = 0.8$, according to Formula (21).

The ideal loss value L_i can be calculated according to the specific state of the process parameters, represented as y_i. The expected loss L_f can be calculated using Formula (24). To establish the minimum loss equations, as indicated in Formula (25), the form factor γ that minimizes the sum of squared deviations δ can be used, and the loss function closest to the ideal state can be determined:

$$\min_{\gamma>0} = \sum_{i=1}^{n} \left(L_i - L_{fi}\right)^2 \tag{25}$$

In summary, the risk of abnormal fluctuations of process parameters in coal chemical production can be derived as follows:

$$R=\begin{cases} \dfrac{1-e^{-(y-\overline{Y})^2/2\gamma^2}}{1-e^{-(I-\overline{Y})^2/2\gamma^2}} + \left(1-\dfrac{1-e^{-(y-\overline{Y})^2/2\gamma^2}}{1-e^{-(I-\overline{Y})^2/2\gamma^2}}\right) \times e^{-\frac{\Delta V_{\max}}{\Delta V}}, & \Delta V_i \times (y-\overline{Y}) \geq 0 \\ \dfrac{1-e^{-(y-\overline{Y})^2/2\gamma^2}}{1-e^{-(I-\overline{Y})^2/2\gamma^2}}, & \text{Others} \end{cases} \tag{26}$$

The trend in the variation of process parameters is bidirectional, so the trend risk R_2 is included in the calculation of the overall risk only when the direction of the variation of the process parameters is the same as the direction of the deviation of the parameters from the target value (*i.e.* ΔV_i has the same sign as $y - \overline{Y}$). Accordingly, the variation trend of coal chemical production process parameters can be predicted by assessing the risk of the predicted parameters of coal chemical production processes and comparing it with the established risk threshold.

2.6. Model performance indicators

To assess the performance of the proposed model, the following metrics were used: mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), and coefficient of determination R^2.

The MAE is used to estimate accuracy by deriving the absolute values of errors and then calculating their average to determine the difference between predicted and actual values. The MAE is in the range of $[0, +\infty)$. When the predicted value is exactly equal to the actual value, the MAE is 0; conversely, when the predicted value is significantly different from the actual value, the MAE is larger. The MAE is expressed as follows:

$$MAE=\frac{\sum_{i=1}^{n}|y_i-\widehat{y_i}|}{n} \tag{27}$$

The RMSE is calculated by taking the square root of the mean square error to prevent excessively large errors from affecting assessments of model performance. A larger RMSE value means a larger error, signifying poorer model performance. The RMSE can be derived as follows:

$$RMSE=\sqrt{\frac{\sum_{i=1}^{n}(y_i-\widehat{y_i})^2}{n}} \tag{28}$$

The MAPE is a metric used to calculate relative errors to prevent the magnitude of the data from influencing performance assessment processes. The MAPE value is in the range of $[0, +\infty)$. A lower MAPE refers to a more accurate model. The MAPE can be derived as follows:

$$MAPE=\frac{\sum_{i=1}^{n}|y_i-\widehat{y_i}|}{y_i}\times\frac{100\%}{n} \tag{29}$$

R^2 is the proportion of the variance explained by a model and is a relative measure, $R^2 \in [0,1]$. A lower R^2 value indicates a poorer

model fit. When R^2 approaches 1, the model is at risk of overfitting. R^2 can be derived as follows:

$$R^2 = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y}_i)^2} \tag{30}$$

3. Empirical analysis

3.1. Data sources

A DCS is a digitized and intelligent system that is connected to other production systems; it has been widely applied in the industrial sector, particularly in complex process control systems such as those in the coal chemical industry. The reliability of a DCS is pivotal in ensuring the safe, stable, and efficient operation of coal chemical production processes. This study developed a model for ultra-early prediction of process parameters in coal chemical production processes. The model achieves early detection of abnormal conditions in coal chemical production processes by real-time monitoring of the process parameters through a DCS, identifying abnormal phenomena in the production process, and sending alarm signals only when the parameter values exceed the designated thresholds; the model does not send warnings before the occurrence of abnormal phenomena [27]. To demonstrate the performance of the proposed model, the study used data obtained from a DCS on crude methanol feed flow monitored at flow point FI101 within the methanol pre-rectification tower of a coal chemical enterprise. The DCS interface for the methanol pre-rectification tower is depicted in Fig. 7.

The crude methanol raw material from outside the storage tank undergoes heat exchange with the top gas of the pre-distillation tower before entering the pre-distillation tower to cut light and heavy components, separating the non-condensable gas and light components in the crude methanol raw material. The pre-distillation tower is operated by the total reflux, and the gas phase at the top of the tower is condensed through the feed preheater and two-stage condenser of the pre-distillation tower. The non-condensable gas in the raw materials is sent to the exhaust gas washing tower for washing and discharge. The condensed liquid phase is pressurized by a reflux pump and sent to the top of the tower. The bottom pump pressurizes the materials at the bottom of the tower and then sent to the atmospheric tower. A partition is installed on the reflux tank at the top of the tower to separate methanol oil from the raw materials and expel them from the device. Simultaneously, addition of alkaline solution is required to neutralize acidic substances in the material to reduce corrosion of the equipment.

The FI101 point in the methanol rectification tower records flow data at 1-s intervals. This study used a total of 41,760 flow data sets recorded over a specific period. In the DCS, the alarm threshold for the flow rate of crude methanol was set to 130 m^3/h, meaning that a flow rate above or below this threshold would trigger an alarm. To facilitate the visual examination of the variation trend of this parameter, this study selected a subset of 1400 data points, including the initial alarm occurrence, for visualization, as depicted in Fig. 8.

The collected time series data need to be pre-processed. The preprocessing procedures involved the processing of abnormal data and normalization of the data. A general overview of these procedures is provided as follows:

(1) Abnormal data processing. In general, collected data can be abnormal due to sensor failure or communication network problems, leading to data values far outside the supercritical interval range. These values do not accurately represent the actual state of key parameters at the detection point and fail to reflect the dynamic risk level in the production process. So, it is necessary to replace these abnormal data. The replacement method depends on the specific scenario. If a single data point exhibits abnormality in the historical time series, the average value of the two data points before and after the anomaly is calculated, and

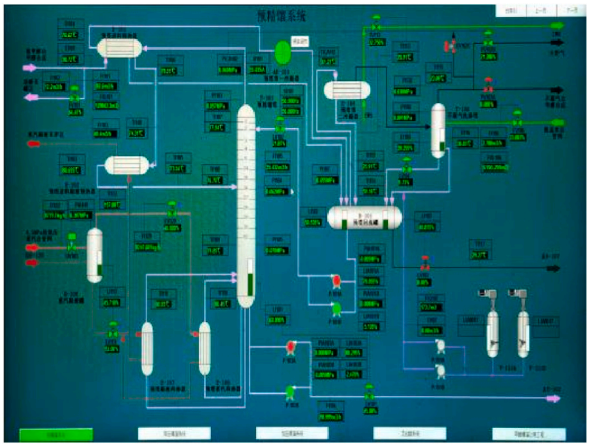


Fig. 7. DCS interface of methanol pre-rectification tower.

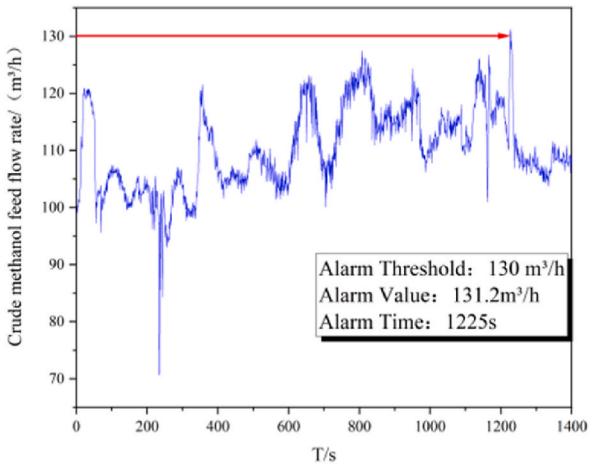


Fig. 8. Variations of the crude methanol feed flow in a methanol pre-rectification column.

then the abnormal data point is replaced with this mean average value. In the case of a single data point exhibiting abnormality in real-time data, anomalous data are replaced with values from the previous time;

(2) Normalization. Normalization is generally performed to adapt the value range of different dimensions within multi-dimensional data by means of translation and scaling. This adjustment helps prevent small numerical signals from being overshadowed by larger ones. Although normalization is not mandatory for artificial neural networks to learn, it is beneficial for transforming input variables into a range of data where the activation function resides. The BiLSTM model selected in this study employs a sigmoid activation function with an S-shaped curve; whose range is determined to be (0,1). Therefore, the collected data require normalization to ensure their effective use in the RNN model.

3.2. Results and analysis

3.2.1. Prediction results of WOA-BiLSTM model

Optimizing BiLSTM hyperparameters, including initial learning rate, number of neural units in the hidden layers, number of iterations, and training batch size, during model training can enhance the capability of the model to learn complex data features. This optimization can ensure that the output distribution of the hidden layer closely matches the actual input features, ultimately improving prediction accuracy. This study used data on crude methanol feed flow rates collected in a pre-rectification tower. Unlike classification prediction, time series prediction requires sample data to have continuity during model training. In this paper, the first 80 % of the dataset is used as the training set, the last 20 % as the testing set, and MAPE is chosen as the loss function. The Adam optimizer was selected to update model parameters. The number of whale populations was initialized to 10. Five BiLSTM hyperparameters were selected as optimization parameters: learning rate, dropout rate, training batch size, number of first-layer neurons, and number of second-layer neurons. Each whale in the population had an individual dimension of five, and the search range for the BiLSTM hyperparameters is as listed in Table 3.

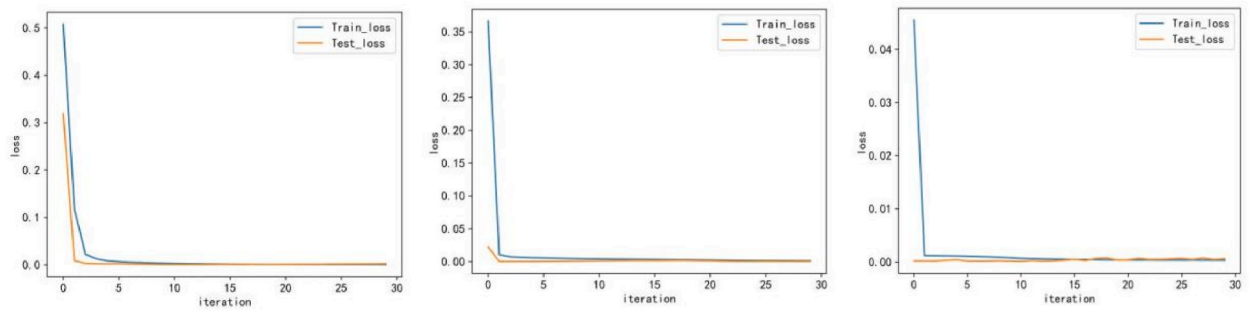
Using the WOA optimization algorithm to optimize the hyperparameters in the BiLSTM network, MAPE is selected as the loss function to determine the prediction accuracy of the training model. As the number of iterations increases, the model parameters are continuously optimized, and the loss value of the model on the training set decreases. The loss curve of the model training is shown in Fig. 9(e). When the model iterates to the 20th iteration, the loss curve tends to stabilize and approaches 0, indicating that the trained model has reached its optimal state. The optimal hyperparameters obtained are 0.00899, 0.5, 328, 64, and 128, respectively.

After training to obtain the optimal prediction model, input the test dataset to verify the model. To comprehensively estimate the predictive performance of the WOA-BiLSTM model, compare the predictive performance with RNN, LSTM, BiLSTM, and WOA-LSTM models (Fig. 10).

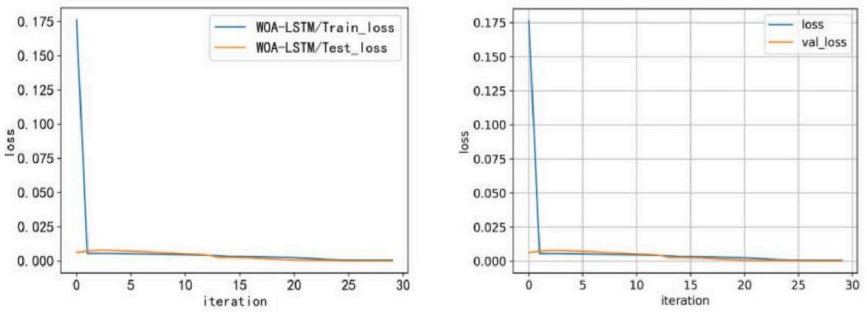
Due to the large sample size of the test dataset, only 1000 predicted values were plotted. To clearly observe the differences in the predicted results of each model, some areas in the graph were enlarged (Figs. 11 and 12). As illustrated in Fig. 11, when WOA is not

Table 3
BiLSTM hyperparameter search range.

Hyperparameter	Optimization range
Learning rate	0.001–0.01
Discard rate	0.1–0.5
Training batch	32–512
Number of neurons in the first layer	16–64
Number of neurons in the second layer	32–128

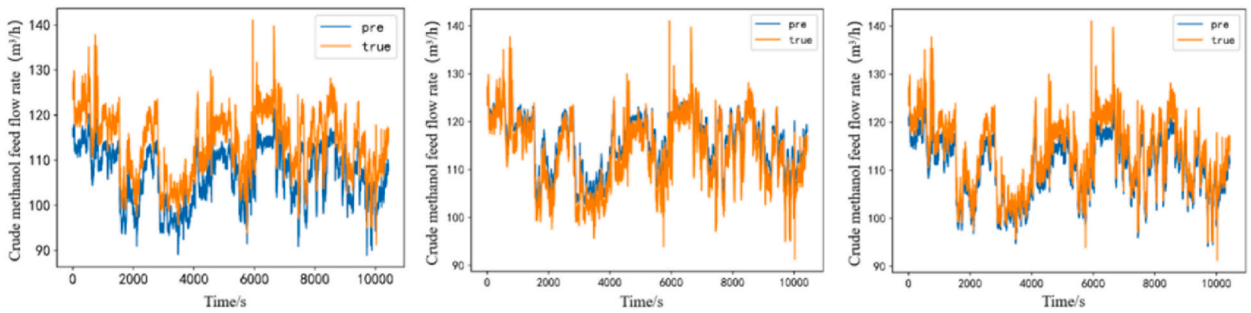


a) RNN model loss curve b) LSTM model loss curve c) BiLSTM model loss curve

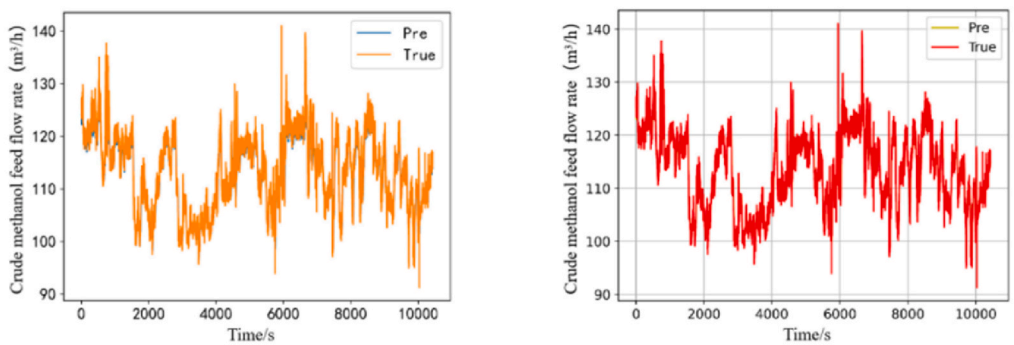


d) WOA-LSTM model loss curve e) WOA-BiLSTM model loss curve

Fig. 9. Loss function curves of each model.



a) RNN model prediction b) LSTM model prediction c) BiLSTM model prediction



d) WOA-LSTM model prediction e) WOA-BiLSTM model prediction

Fig. 10. Predictions from each model.

used to optimize model parameters, the trends of the prediction curves of each model were broadly consistent with the actual values. However, in the local-amplification region, the prediction curve of the RNN model deviated the most from the actual values. In contrast, the curve of the BiLSTM model was closest to the actual values. These observations prove that the BiLSTM model outperforms the other two models in terms of time series prediction. As illustrated in Fig. 12, the predicted values of the WOA-LSTM and WOA-BiLSTM models are highly consistent with the actual values. The prediction performance of the model using WOA is significantly improved, and the prediction performance of the WOA-BiLSTM model is better than that of the WOA-LSTM model.

It is not easy to quantify the actual predictive performance of the model solely based on the curve shape. MSE, MAE, RMSE, and R^2 are selected as evaluation indicators to evaluate the predictive performance of each model, and the metrics derived for each model are presented in Table 4.

According to the MSE, MAE, RMSE, and R^2 values, the RNN model shows the poorest predictive performance and fit among the models. By contrast, the WOA-BiLSTM model demonstrates the best predictive performance among the models. The MAE value of the BiLSTM model is 3.1880, slightly higher than that of the LSTM model (2.9856). Moreover, the BiLSTM model outperformed the LSTM model regarding the MSE, RMSE, and R^2. The WOA-LSTM model exhibits considerable improvements over the LSTM model, with a 67.5 % decrease in MSE, a 50.3 % reduction in MAE, and a 42.3 % reduction in RMSE. Additionally, the R^2 value of the WOA-LSTM model is increased by 21.4 % compared with that of the LSTM model. The WOA-BiLSTM model exhibits even more substantial improvements than the BiLSTM model, with a 77.0 % reduction in MSE, 64.2 % in MAE, and 52.0 % in RMSE. Furthermore, the R^2 value of the WOA-BiLSTM model is increased by 21.5 % compared with that of the BiLSTM model. These results demonstrate that integrating the WOA can significantly improve the predictive performance of the model. In summary, the WOA-BiLSTM model developed in this study shows remarkable advantages in analyzing time series data for predicting the parameters of coal chemical production processes; the model can accurately capture the underlying patterns of the time series and predict the variation trends of the process parameters.

3.2.2. Prediction results of the WOA-BiLSTM-MINLF model

To compare and analyze the ultra-early prediction performance of the WOA-BiLSTM-MINLF model more intuitively, 1400 sets of data, including the first alarm in the process parameters of crude methanol feed flow in the methanol pre-distillation tower are still used as examples for analysis. The prediction model of WOA BiLSTM process parameters was used to predict 1400 data sets, and the results are shown in Fig. 13. The model accurately predicted the trends in this data set.

The alarm threshold for the crude methanol feed flow rate is $y_{alarm} = 130m^3/h$, the alarm interlock threshold is $I = 150m^3/h$, and the historical data mean is $\overline{Y} = 111.7m^3/h$. According to Formula (24), the loss function can be expressed as follows:

$$L = \frac{1 - e^{-(y-111.7)^2/2\gamma^2}}{1 - e^{-38.3^2/2\gamma^2}} \tag{31}$$

The shape factor $\gamma = 10.21$ is obtained using the least squares method, and the maximum instantaneous rate of change in flow is determined to be $\Delta V_{\max} = 13.267$. By substituting the predicted process parameters y_i into the aforementioned formula, this study obtains the corresponding loss function value. The inherent and trend risk of abnormal fluctuations of process parameters can be expressed as follows: $R_1 = L \times P, P = 1$ and $R_2 = (1 - L) \times P$, respectively, where $P = e^{-13.267/\Delta V_i}$. The risk R of abnormal fluctuations of predicted process parameters can be calculated using Formula (26). On the basis of these formulas, this study derived the risk of abnormal variations of crude methanol feed flow, and the results are illustrated in Fig. 14.

According to Fig. 8, the crude methanol feed flow rate reaches the alarm threshold at 1225 s and triggers a DCS alarm. Through the ultra-early prediction model of process parameters constructed in this article, the predicted value of process parameters is converted into a dynamic risk value, causing the risk value to start triggering the risk threshold at 233 s and then exceed the alarm limit multiple times. The results show that the WOA-BiLSTM-MINLF prediction model constructed in the present research sounds alarm some 16 min (960 s) earlier than DCS, achieving ultra-early prediction of process parameters in coal chemical processes. It can reserve sufficient time for operators to take safety protection measures in advance and prevent the occurrence of safety production accidents.

4. Conclusion

Coal chemical production has strong continuity and dynamism, and there is a certain degree of danger in equipment and hazardous materials involved in the production process. This study combines data collected during coal chemical production to construct a WOA-BiLSTM-MINLF-based ultra-early prediction model for process parameters, and the effectiveness of the model is assessed. The key conclusions are as follows:

(1) The BiLSTM prediction model was established according to the temporal variation characteristics of process parameters in the coal chemical production and the WOA-BiLSTM process parameter prediction model was formed by using WOA to optimize the learning rate, discard rate, training batch, number of neurons, and other hyperparameters in the BiLSTM network. The prediction performance of the RNN, LSTM, BiLSTM, WOA-LSTM, and WOA-BiLSTM models were compared and analyzed. According to the error evaluation indices of MSE, MAE, RMSE, and R^2, the prediction performance of the BiLSTM model is better than that of the RNN and LSTM models. The prediction performance of the WOA-LSTM and WOA-BILSTM models using WOA to optimize network parameters is better than that of the LSTM and BiLSTM models. The WOA-BiLSTM model is superior to other models as evinced by the four error-evaluation indices, which indicates that the model can mine the trends of variations in the process parameters more accurately;

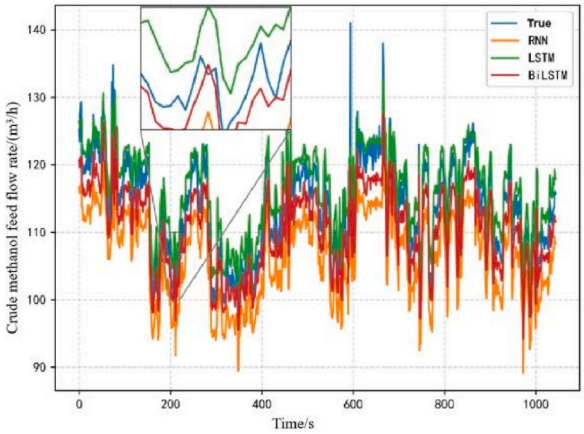


Fig. 11. Model prediction performance without WOA.

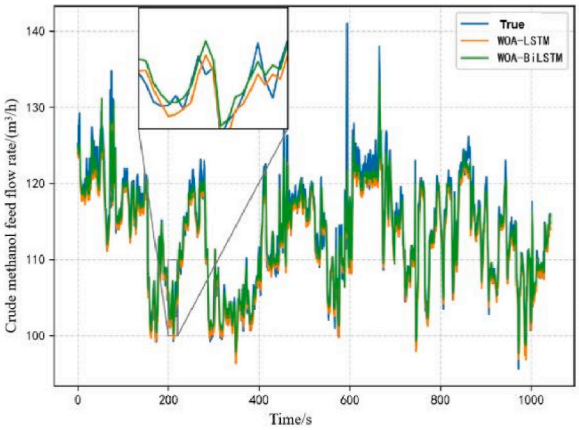


Fig. 12. Model prediction performance with WOA.

Table 4
Prediction errors of different models.

Evaluation index	RNN	LSTM	BiLSTM	WOA-LSTM	WOA-BiLSTM
MSE	50.7041	12.6173	11.4301	4.1025	2.6289
MAE	6.9271	2.9856	3.1880	1.4825	1.1399
RMSE	7.1207	3.5521	3.3801	2.0255	1.6214
R^2	0.0336	0.7595	0.7821	0.9218	0.9499

(2) To derive the risk of abnormal fluctuations of process parameters based on residual time theory, including inherent risk and trend risk, process parameter values predicted by the WOA-BiLSTM model were used to construct a MINLF. This comprehensive model, referred to as WOA-BiLSTM-MINLF, was established to convert abnormal fluctuations of process parameters into dynamic risk values, enabling the real-time monitoring of equipment and hazardous materials and facilitating early risk prediction in the production process. Through case analysis, the study demonstrated the effectiveness of the model. Specifically, the WOA-BiLSTM-MINLF model could detect abnormal conditions approximately 16 min earlier than the DCS, thereby achieving ultra-early prediction of process parameters;

(3) This study selects the process parameters related to equipment and materials when developing the model for predicting process parameters. These parameters comprise continuous data; hence, future research should explore the use of discrete data for risk prediction. To more comprehensively and accurately predict the evolving risks in coal chemical production processes, can ultimately guide enterprises in managing and controlling risks.

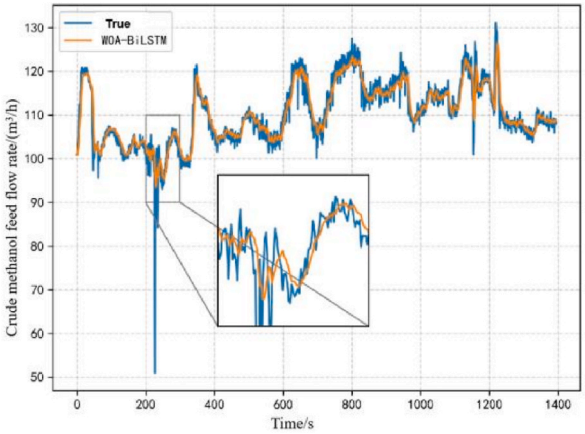


Fig. 13. WOA-BiLSTM model predicts the results of selected process state parameters.

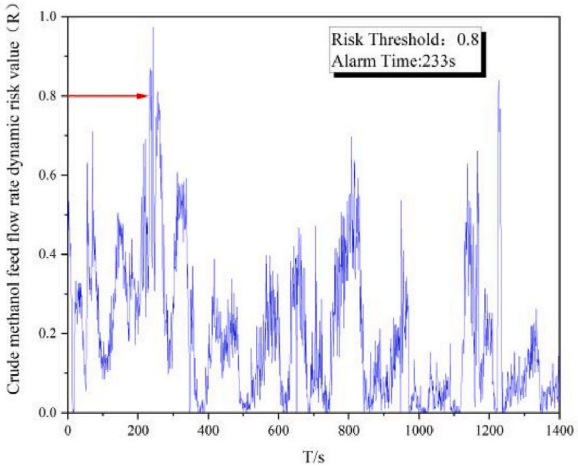


Fig. 14. Risk to crude methanol feed flow.

Funding statement

The work is funded by the Natural Science Foundation of China (Project Code: 51674193) and the Key R&D Project in Ningxia Hui Autonomous Region (Project Code: 2022BEE02001).

Data availability statement

The labelled dataset used to support the findings of this study is available from the corresponding author upon request.

CRediT authorship contribution statement

Zheng Li: Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation. **Min Yao:** Project administration, Funding acquisition, Conceptualization. **Zhenmin Luo:** Supervision, Resources, Funding acquisition, Data curation. **Qianrui Huang:** Validation, Investigation. **Tongshuang Liu:** Visualization, Investigation.

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.

Acknowledgment

The authors thank the reviewers for their valuable suggestions on improving the quality of the paper. At the same time, we thank the relevant leaders of the safety department of the research enterprise for their support during the authors' on-site research.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e30821.

References

[1] K.C. Xie, Breakthrough and innovative clean and efficient coal conversion technology from a chemical engineering perspective, Chem. Eng. Sci. X 10 (2021) 100092.
[2] L. Ma, J. Fan, R.Z. Guo, et al., Characteristics of fires in coal mine roadways and comparative analysis of control effectiveness between longitudinal ventilation and cross-section sealing, Case Stud. Therm. Eng. 53 (2024) 103878.
[3] J. Deng, C. Su, Z.M. Zhang, et al., Evolutionary game analysis of chemical enterprises' emergency management investment decision under dynamic reward and punishment mechanism, J. Loss Prev. Process. Ind. 87 (2024) 105230.
[4] S. Kenan, S. Kadri, Process safety leading indicators survey-february 2013: center for chemical process safety-white paper, Process Saf. Prog. 33 (2014) 247–258.
[5] T. Wende, H.U. Minggang, L.I. Chuankun, Fault prediction based on dynamic model and grey time series model in chemical processes, Chin. J. Chem. Eng. 22 (6) (2014) 643–650.
[6] M. Khashei, M. Bijari, A new class of hybrid models for time series forecasting, Expert Syst. Appl. 3 (4) (2012) 4344–4357.
[7] M.T. Amin, F. Khan, S. Ahmed, et al., A data-driven Bayesian network learning method for process fault diagnosis, Process Saf. Environ. Protect. 150 (2021) 110–122.
[8] C. Benson, C.D. Argyropoulos, C. Dimopoulos, et al., Safety and risk analysis in digitalized process operations warning of possible deviating conditions in the process environment, Process Saf. Environ. Protect. 149 (2021) 750–757.
[9] D.Y. Wu, J.S. Zhao, Process topology convolutional network model for chemical process fault diagnosis, Process Saf. Environ. Protect. 150 (2021) 93–109.
[10] Q.S. Yin, J. Yang, M. Tyagi, et al., Field data analysis and risk assessment of gas kick during industrial deepwater drilling process based on supervised learning algorithm, Process Saf. Environ. Protect. 146 (2021) 312–328.
[11] X.F. Yuan, C. Ou, Y.L. Wang, et al., A novel semi-supervised pre-training strategy for deep networks and its application for quality variable prediction in industrial processes, Chem. Eng. Sci. 217 (2020) 115509.
[12] F. Yang, C. Dai, J. Tang, et al., A hybrid deep learning and mechanistic kinetics model for the prediction of fluid catalytic cracking performance, Chem. Eng. Res. Des. 155 (2020) 202–210.
[13] K. Zhong, M. Han, B. Han, Data-driven based fault prognosis for industrial systems: a concise overview, IEEE-CAA Journal of Automatica Sinica 7 (2020) 330–345.
[14] K. Yao, Bayesian inference with uncertain data of imprecise observations, Commun. Stat. Theor. Methods 51 (15) (2022) 5330–5341.
[15] F. Qi, L. Zhang, K. Zhou, X. Ma, Early waning for manufacturing supply chain resilience based on improved grey prediction model, Sustainability 14 (20) (2022) 215–226.
[16] K. Wang, J. Ding, J. Deng, et al., Hydrogen generation mechanism of oil-rich coal oxidation in low temperature, Energy (2024) 130739.
[17] A. Abid, M.T. Khan, J. Iqbal, A review on fault detection and diagnosis techniques: basics and beyond, Artif. Intell. Rev. 54 (5) (2021) 3639–3664.
[18] Z. Yin, J. Hou, Recent advances on SVM based fault diagnosis and process monitoring in complicated industrial processes, Neurocomputing 174 (2016) 643–650.
[19] Y. Liu, C. Yang, Z.L. Gao, et al., Ensemble deep kernel learning with application to quality prediction in industrial polymerization processes, Chemometr. Intell. Lab. Syst. 174 (2018) 15–21.
[20] W.G. Wang, J.B. Shen, J.W. Xie, et al., Revisiting video saliency prediction in the deep learning era, IEEE Trans. Pattern Anal. Mach. Intell. 43 (1) (2021) 220–237.
[21] L.S. Liu, L.Q. Chen, Z.L. Wang, et al., Early fault detection of planetary gearbox based on acoustic emission and improved variational mode decomposition, IEEE Sensor. J. 21 (2) (2021) 1735–1745.
[22] Y. Lv, F. Fang, T.T. Yang, et al., An early fault detection method for induced draft fans based on MSET with informative memory matrix selection, ISA (Instrum. Soc. Am.) Trans. 102 (2020) 325–334.
[23] L. Wei, Z. Qian, H. Zareipour, Wind turbine pitch system condition monitoring and fault detection based on optimized relevance vector machine regression, IEEE Trans. Sustain. Energy 11 (4) (2020) 2326–2336.
[24] W. Zhang, J.Z. Liu, M.M. Gao, et al., A fault early warning method for auxiliary equipment based on multivariate state estimation technique and sliding window similarity, Comput. Ind. 107 (2019) 67–80.
[25] L. Yao, Z.Q. Ge, Deep learning of semisupervised process data with hierarchical extreme learning machine and soft sensor application, IEEE Trans. Ind. Electron. 65 (2) (2018) 1490–1498.
[26] P. Bangalore, M. Patriksson, Analysis of SCADA data for early fault detection, with application to the maintenance management of wind turbines, Renew. Energy 115 (2018) 521–532.
[27] G. Dorgo, A. Palazoglu, J. Abonyi, Decision trees for informative process alarm definition and alarm-based fault classification, Process Saf. Environ. Protect. 149 (2021) 312–324.
[28] M.J. Mayer, G. Gróf, Extensive comparison of physical models for photovoltaic power forecasting, Appl. Energy 283 (2021) 116239.
[29] S. Mozaffari, S. Javadi, H.K. Moghaddam, et al., Forecasting groundwater levels using a hybrid of support vector regression and particle swarm optimization, Water Resour. Manag. 36 (6) (2022) 1955–1972.
[30] S.A. Jasmin, P. Ramesh, M. Tanveer, An intelligent framework for prediction and forecasting of dissolved oxygen level and biofloc amount in a shrimp culture system using machine learning techniques, Expert Syst. Appl. 199 (2022) 117160.
[31] M.A. Ghorbani, R.C. Deo, V. Karimi, et al., Implementation of a hybrid MLP-FFA model for water level prediction of Lake Egirdir, Turkey, Stoch. Environ. Res. Risk Assess. 32 (6) (2018) 1683–1697.
[32] M.N.A. Zakaria, M.A. Malek, M. Zolkepli, et al., Application of artificial intelligence algorithms for hourly river level forecast: a case study of Muda River, Alex. Eng. J. 60 (4) (2021) 4015–4028.
[33] S. Zhu, S. Heddam, Prediction of dissolved oxygen in urban rivers at the Three Gorges Reservoir, China: extreme learning machines (ELM) versus artificial neural network (ANN), Water Quality Research Journal 55 (1) (2020) 106–118.
[34] Y. Wu, L. Sun, X. Sun, B. Wang, A hybrid XGBoost-ISSA-LSTM model for accurate short-term and long-term dissolved oxygen prediction in ponds, Environ. Sci. Pollut. Control Ser. 29 (12) (2021) 18142–18159.

[35] M. Yaqub, W. Lee, Modeling nutrient removal by membrane bioreactor at a sewage treatment plant using machine learning models, J. Water Proc. Eng. 46 (2022) 102521.
[36] S. Liu, X.J. Liu, Q. Lyu, et al., Comprehensive system based on a DNN and LSTM for predicting sinter composition, Appl. Soft Comput. 95 (2020) 106574.
[37] C. Xu, X. Chen, L. Zhang, Predicting River dissolved oxygen time series based on stand-alone models and hybrid wavelet-based models, J. Environ. Manag. 295 (2021) 113085.
[38] N. Guo, Z. Wang, A combined model based on sparrow search optimized BP neural network and Markov chain for precipitation prediction in Zhengzhou City, China, AQUA-Water Infrastructure Ecosystems and Society 71 (6) (2022) 78–800.
[39] J. Xu, Y. Zhao, H. Chen, et al., ABC-GSPBFT: PBFT with grouping score mechanism and optimized consensus process for flight operation data-sharing, Inf. Sci. 624 (2023) 110–127.
[40] Q.Q. Sun, Z.Q. Ge, Probabilistic sequential network for deep learning of complex process data and soft sensor application, IEEE Trans. Ind. Inf. 15 (5) (2019) 2700–2709.
[41] W.M. Shao, L. Yao, Z.Q. Ge, et al., Parallel computing and SGD-based DPMM for soft sensor development with large-scale semisupervised data, IEEE Trans. Ind. Electron. 66 (8) (2019) 6362–6373.
[42] X.F. Yuan, L. Li, Y.A.W. Shardt, et al., Deep learning with spatiotemporal attention-based LSTM for industrial soft sensor model development, IEEE Trans. Ind. Electron. 68 (5) (2021) 4404–4414.
[43] D. Li, X. Zhang, Utilizing a two-dimensional data-driven convolutional neural network for long-term prediction of dissolved oxygen content, Front. Environ. Sci. 10 (2022) 904939.
[44] Y. Huang, C.H. Chen, C.J. Huang, Motor Fault detection and feature extraction using RNN-based variational autoencoder, IEEE Access 7 (2019) 139086–139096.
[45] Z. Chen, H. Xu, P. Jiang, et al., A transfer learning-based LSTM strategy for imputing largescale consecutive missing data and its application in a water quality prediction system, J. Hydrol. 602 (2021) 126573.
[46] N. Zhu, X. Ji, J. Tan, et al., Prediction of dissolved oxygen concentration in aquatic systems based on transfer learning, Comput. Electron. Agric. 180 (2021) 105888.
[47] W. Yang, W. Liu, Q. Gao, Prediction of dissolved oxygen concentration in aquaculture based on attention mechanism and combined neural network, Math. Biosci. Eng. 20 (1) (2023) 998–1017.
[48] J. Guo, J. Dong, B. Zhou, et al., A hybrid model for the prediction of dissolved oxygen in seabass farming, Comput. Electron. Agric. 198 (2022) 106971.
[49] Y. Fu, Z. Hu, Y. Zhao, et al., A long-term water quality prediction method based on the temporal convolutional network in smart mariculture, Water 13 (20) (2021) 2907.
[50] M. Ma, Z. Mao, Deep-convolution-based LSTM network for remaining useful life prediction, IEEE Trans. Ind. Inf. 17 (3) (2021) 1658–1667.
[51] B.K. Wang, D.T. Liu, Y. Peng, et al., Multivariate regression-based fault detection and recovery of UAV flight data, IEEE Trans. Instrum. Meas. 69 (6) (2020) 3527–3537.
[52] K. Kumari, P. Dey, C. Kumar, et al., UMAP and LSTM based fire status and explosibility prediction for sealed-off area in underground coal mine, Process Saf. Environ. Protect. 146 (2021) 837–852.
[53] X. Zhou, J. Wang, H. Zhang, et al., Application of a hybrid improved sparrow search algorithm for the prediction and control of dissolved oxygen in the aquaculture industry, Appl. Intell. 53 (7) (2023) 8482–8502.
[54] W. Cao, J. Zhou, Q. Xu, et al., Short-term forecasting and uncertainty analysis of photovoltaic power based on the FCM-WOA-BILSTM model, Front. Energy Res. 10 (2022) 926774.
[55] K. Yu, Q.G. Cao, C.Z. Xie, et al., Analysis of intervention strategies for coal miners' unsafe behaviors based on analytic network process and system dynamics, Saf. Sci. 118 (2019) 145–157.
[56] J.Q. Hu, F. Guo, L.B. Zhang, Research on ultra early monitoring and early warning of chemical abnormal conditions by combining improved PSO algorithm and LSSVM, J. Electron. Meas. Instrum. 32 (2) (2018) 36–41.
[57] R. Tan, Y. Hu, Z.C. Wang, A multi-source data-driven model of lake water level based on variational modal decomposition and external factors with optimized bi-directional long short-term memory neural network, Environ. Model. Software 167 (2023) 105766.
[58] H. Fan, M. Jiang, L. Xu, et al., Comparison of long short term memory networks and the hydrological model in runoff simulation, Water 12 (1) (2020) 175.
[59] G. Li, H. Wang, S. Zhang, et al., Recurrent neural networks based photovoltaic power forecasting approach, Energies 12 (13) (2019) 2538.
[60] A.K.L. Rufus, A Goa-RNN controller for a stand-alone photovoltaic/wind energy hybrid-fed pumping system, Soft Comput. 23 (2019) 12255–12276.
[61] S. Hochreiter, J. Schmidhuber, LSTM can solve hard long time lag problems, Adv. Neural Inf. Process. Syst. 9 (1996) 1–8.
[62] F.A. Gers, D. Eck, J. Schmidhuber, Applying LSTM to time series predictable through time-window approaches, in: International Conference on Artificial Neural Networks, Springer Berlin Heidelberg, Berlin, Heidelberg, 2002.
[63] C.H. Ho, I. Park, J. Kim, et al., PM2.5 forecast in korea using the long short-term memory (LSTM) model, Asia-Pacific Journal of Atmospheric Sciences 59 (5) (2023) 563–576.
[64] S.Y. Yang, B.C. Jhong, Y.D. Jhong, et al., Long short-term memory integrating moving average method for flood inundation depth forecasting based on observed data in urban area, Nat. Hazards 116 (2) (2022) 2339–2361.
[65] G. Ozdogan-Sarikoc, M. Sarikoc, M. Celik, et al., Reservoir volume forecasting using artificial intelligence-based models: artificial neural networks, support vector regression, and long short-term memory, J. Hydrol. 616 (2023) 128766.
[66] S. Lee, K.K. Lee, H. Yoon, Using artificial neural network models for groundwater level forecasting and assessment of the relative impacts of influencing factors, Hydrogeol. J. 27 (2) (2019) 567–579.
[67] J. Wu, Z. Wang, Y. Hu, et al., Runoff forecasting using convolutional neural networks and optimized bi-directional long short-term memory, Water Resour. Manag. 37 (2) (2023) 937–953.
[68] M. Schuster, K.K. Paliwal, Bidirectional recurrent neural networks, IEEE Trans. Signal Process. 45 (11) (1997) 2673–2681.
[69] M. Xu, L. Cao, D. Lu, Z. Hu, Y. Yue, Application of swarm intelligence optimization algorithms in image processing: a comprehensive review of analysis, synthesis, and optimization, Biomimetics 8 (2023) 235.
[70] J. Du, J. Hou, H.Y. Wang, Z. Chen, Application of an improved whale optimization algorithm in time-optimal trajectory planning for manipulators, Math. Biosci. Eng. 20 (9) (2023) 16304–16329.
[71] S. Dey, Bayesian estimation of the parameter and reliability function of an inverse Rayleigh distribution, Malaysian Journal of Mathematical Sciences 6 (2012) 113–124.
[72] M. Abimbola, F. Khan, Dynamic blowout risk analysis using loss functions, Risk Anal. 38 (2) (2018) 255–271.
[73] G. Taguchi, Quality Engineering in Production Systems, McGraw-Hill, New York, 1989.
[74] O. Zadakbar, F. Khan, S. Imtiaz, Development of economic consequence methodology for process risk analysis, Risk Anal. 35 (4) (2015) 713–731.
[75] F.A. Spiring, The reflected normal loss function, Can. J. Stat. 21 (3) (1993) 321–330.
[76] F.B. Sun, J.Y. Laramee, J.S. Ramber, On Spiring's normal loss function, Can. J. Stat. 24 (1996) 241–249.
[77] F. Khan, S.J. Hashemi, et al., Dynamic risk management: a contemporary approach to process safety management, Current Opinion in Chemical Engineering 14 (2016) 9–17.
[78] B.P.K. Leung, F.A. Spiring, Some properties of the family of inverted probability loss functions, Quality Technology & Quantitative Management 1 (1) (2004) 125–147.
[79] W.M. Chan, R.N. Ibrahim, P.B. Lochert, Quality evaluation model using loss function for multiple S-type quality characteristics, Int. J. Adv. Manuf. Technol. 26 (2005) 98–101.
[80] S.J. Hashemi, S. Ahmed, F. Khan, Loss functions and their applications in process safety assessment, Process Saf. Prog. 33 (3) (2014) 285–291.
[81] H. Wang, F. Khan, S. Ahmed, et al., Dynamic quantitative operational risk assessment of chemical processes, Chem. Eng. Sci. 142 (2016) 62–78.