Reciprocating compressor start-up fault monitoring based on sensor and limit learning machine

Cheng Pan, Lei Guo, Weiwei Zeng, Shihao Hu, Xin Li

Reciprocating compressor start-up fault monitoring based on sensor and limit learning machine

Cheng Pan*, Lei Guo, Weiwei Zeng and Shihao Hu

Petro China Tarim Oilfield Company,
Korla 841000, China
Email: 15276077902@163.com
Email: 18690666278@163.com
Email: 13999617003@163.com
Email: hu546611467@163.com
*Corresponding author

Xin Li

Xinjiang Compressor Branch,
CNPC Jichai Power Company Limited,
Korla 841000, China
Email: 17380439643@163.com

Abstract: Condition monitoring of reciprocating compressors (RC) can improve reliability of equipment operation. To address the problem of unsatisfactory monitoring accuracy of the existing RC start-up monitoring methods, the common types of failures were first analysed to obtain the external influencing factors. Second, the overall architecture of RC starter fault monitoring was designed based on the shock pulse sensor to obtain the intrinsic signal data of the RC starter. The internal and external influence data were then pre-processed, the main variables were extracted using principal element analysis, and the variables were decomposed into eigenmode components using an improved empirical modal decomposition method. Finally, the extreme learning machine (ELM) algorithm (OLEM) is optimised by the regularisation term, and the RC start-up fault state is predicted using OLEM. The experimental outcome indicates that the proposed method has a monitoring accuracy of 92.5% and has a strong monitoring capability.

Keywords: reciprocating compressor; RC; fault monitoring; principal element analysis; empirical modal decomposition; EMD; extreme learning machine; ELM.

Biographical notes: Cheng Pan received his Bachelor's in Engineering from Chongqing University of Science and Technology in 2009. His research interests include the installation of oil and gas field equipment, compressor and pump technology management.

Lei Guo received his Bachelor of Engineering degree from China University of Petroleum, Beijing in 2013. His research interests include natural gas processing and oil and gas gathering and transportation.

Weiwei Zeng received his Bachelor of Engineering degree from China University of Petroleum (East China) in 2015. His interests include information and communication technology for oil and gas fields.

Shihao Hu received his Master's degree from Hebei University of Technology in 2020. His research interests include water injection technology in oil and gas fields and management of technological information.

Xin Li received his Bachelor's in Engineering from Henan University in 2011. His research interests include reciprocating compressor design and manufacturing.

1   Introduction

Reciprocating compressor (RC) is an indispensable unit in chemical production, due to its complicated structure, numerous moving parts, and many sources of vibration shocks, resulting in a high frequency of failures, various types of failures and complex vibration response (Almasi, 2016). There is no heating equipment in all the machine rooms of Tower One Union, and the ambient temperature is only 10°C in winter when there is a unit running in the machine room, and the oil temperature of the standby unit is around 15°C, due to the low ambient temperature in winter and the fact that the external circulating heater cannot be heated up for a long period of time to run. In winter, it takes about 2 hours to warm up the unit from the point of running button to loading. If the unit fails to shut down and needs to be urgently backed up to another unit to ensure the volume of gas lift, the pressure of gas lift will drop drastically due to the long loading time, which not only directly affects the production of Taizhong oil and gas management area, but also has the risk of causing the shutdown of wells. RC start-up failure not only affects the normal process of oil production (Qi et al., 2018), but also leads to the leakage of hazardous gases in serious cases, resulting in a series of malignant failures such as fires, explosions, and other injuries and deaths (Wang et al., 2022). Therefore, it is of great significance to carry out fault diagnosis of compressor start-up.

The traditional RC monitoring method installs sensors on the compressor to collect real-time data on operating conditions; the collected data is transmitted to the monitoring system for storage and analysis, and the results are used to assess the compressor's operating condition and performance, as well as whether there is a risk of failure. However, the monitoring is not timely and accurate, and is highly influenced by environmental factors. Deep learning-based fault monitoring methods have better generalisation ability because they do not require pre-extraction of features and their parameters are automatically tuned. Sharma and Parey (2019) extracted the fault characteristics of valve wear by using empirical modal decomposition (EMD), wavelet packet, and other methods, and established a comprehensive evaluation of valve wear failure index. Pichler et al. (2016) used wavelet transform to pre-process the sensor signals, converting the original signal into multiple detailed signals, then extracting fault features from them, and finally inputting the extracted feature values into a support vector

machine, but the classification accuracy is not high. Nwakpang et al. (2019) performed numerical simulations of the secondary RC and achieved fault diagnosis by monitoring the instantaneous angular velocity and energy. Patil et al. (2021) used a generalised data mining model to monitor all faults in the RC by analysing the signals from the collected sensors, but the monitoring was not satisfactory. Loukopoulos et al. (2019a) established a test bed for analysing the wear mechanism of RCs and used oil analysis to evaluate their wear state, which does not allow for timely diagnosis of faults.

The deep learning-based fault monitoring method does not need to extract features in advance and its parameters have the characteristics of automatic tuning, so its generalisation ability is better. Lu and Wang (2021) collected the common faults of RCs and identified their fault characteristics and corresponding fault types, and utilised a combined BP neural network for fault diagnosis of RC starters. Ahmadianfar et al. (2022) established an improved adaptive neuro-fuzzy inference fault diagnostic system for RCs and applied it to mining air compressors. Cervantes-Bobadilla et al. (2023) used radial-based neural networks to make initial diagnoses of multiple bodies of evidence and then synthesised the results of each initial RC diagnosis using weighted evidence fusion theory, but the accuracy of the diagnoses was not high. Loukopoulos et al. (2019b) fused SOM neural network with BP neural network for nine types of faults in RC, and optimised the fusion algorithm using differential evolutionary algorithm, and the monitoring accuracy of this method reached 85% after experiments.

All of the above methods have the defects of slow training speed and more human intervention in the training process. In order to address the above issues, extreme learning machine (ELM) has emerged, which requires almost no human intervention and has higher learning efficiency and generalisation performance (Ding et al., 2014), and therefore has attracted a lot of attention from researchers recently. Medina et al. (2022) used principal component analysis (PCA) and ELM to analyse the dissolved gas in oil from various sensor data sets to diagnose early RC faults, and compared the results with those of fuzzy logic and BPNN, which showed better diagnostic effect of ELM. Wu et al. (2023) used variational modal decomposition (VMD) to extract features from the vibration signals of the sensors of the unit, and then the acquired multi-domain RC features were transmitted to ELM for fault classification.

By analysing the above research status, it can be seen that the existing research leads to inefficient monitoring due to the high complexity of input variables. Focusing on the above issues, this paper innovatively uses the data from shock pulse sensors (SPSs) as input variables and optimise the ELM algorithm using regularisation terms to achieve efficient monitoring of RCs. Firstly, this paper analyses and elaborates the five common fault types and gives the corresponding coping strategies, so as to obtain the external influencing factors of the RC starter. On this basis, the overall architecture of RC start-up fault monitoring is designed by optimising the acquisition frequency and distribution method of SPSs in fault monitoring to obtain the intrinsic signal data of RC start-up obtained from sensor monitoring. The external influences and intrinsic data are then pre-processed, and the main variables are extracted using principal meta-analysis, and the improved EMD method is used to reduce the complexity of the input variables by decomposing the main variables into a series of intrinsic modal (IMF) components. Finally, the ELM algorithm (OLEM) is optimised by the regularisation term, and the network parameters are reduced to enhance the stability of the model, and the OLEM is

used to predict the failure state of the RC starter. The experimental outcome indicates that the monitoring accuracy of the proposed method is 92.5%, which is improved by 5%–20% compared with the other four models, and the monitoring results of the proposed method are in line with the actual situation, which can better realise the fault monitoring of the RC starter.

2 Relevant theoretical foundations

2.1 Principle of operation of RCs

RC, also known as piston compressors, are volumetric compressors (Winandy et al., 2002). RCs have a complex mechanical structure and a variety of designs, but their core components are more or less the same, consisting of seven main parts: the body, the drive mechanism, the cylinder and piston, the valve, the sliding sealing parts, the lubrication system and the cooling system. When the crankshaft rotates, the piston is driven downward by the connecting rod and the cross head. In the compression process, with the upward movement of the piston, the volume in the cylinder gradually decreases, the gas is compressed, and the pressure rises. At this time, the suction valve is closed to prevent gas backflow, while the exhaust valve is still closed, so that the gas continues to be compressed in the cylinder. When the gas pressure in the cylinder reaches or exceeds the exhaust pressure, the exhaust valve opens, and the gas is discharged from the cylinder and enters the exhaust pipe. With the continuous rotation of the crankshaft, the piston moves back and forth in the cylinder, constantly repeating the process of suction, compression and exhaust. The task workflow can be divided into four processes: gas expansion, suction, compression and discharge (Marchante-Avellaneda et al., 2023), and the cylinder oiling process is shown in Figure 1.

1   Gas expansion: the crankshaft rotates, driving the connecting rod to make the piston reciprocating motion, which causes the effective volume of the cylinder to increase, the working pressure decreases, so that the gas continues to expand.

2   Intake gas: When the operating pressure is reduced to a pressure slightly lower than the actual pressure of the gas in the intake tube, the gas in the intake tube pushes off the intake gas control valve and enters the cylinder. It enters the cylinder with the piston and continues until the piston reaches the end of the left side.

3   Compression: As the piston moves freely to the right, the effective volume of the cylinder is gradually reduced and the rest of the compression workflow is carried out. The non-return action of the gas intake control valve will prevent the gas in the cylinder from backing up into the inlet tube.

4   Discharge of gas: With the piston moving to the right side, the working pressure of the compressed gas is higher than the actual pressure of the gas in the outlet pipe, and the gas in the cylinder starts to be continuously discharged. Each reciprocating cycle of the piston gradually develops into a cyclic working condition, and the effective distance passed by the piston each time it comes or goes back is the stroke.

Figure 1   The cylinder oiling process (see online version for colours)
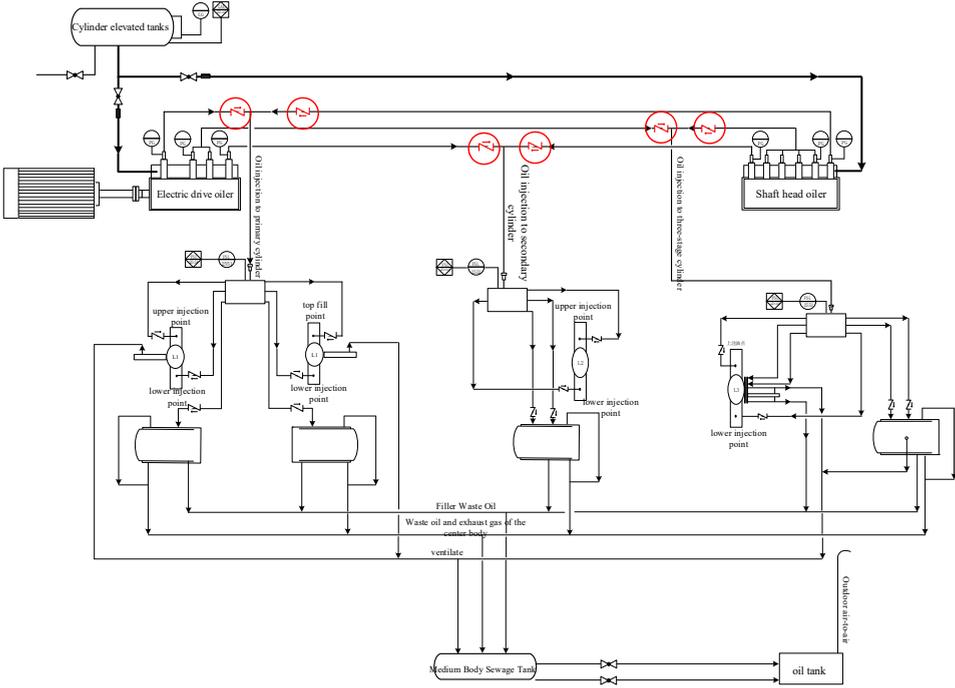
2.2   *Extreme learning machine*

Aiming at the existing deep learning algorithms which have the problems of long training time and simple to fall into native minima, Huang et al. (2006) designed a single obscured-level feed-forward neural network ELM with faster training speed and less human intervention, which contains an input level, a obscured level and an output level. Compared to deep neural networks such as BPNN and CNN, ELM uses fewer parameters, has excellent generalisation ability, and can show better prediction performance on sample data.

First, given an ELM network with l obscured level neurons and activation function $g(x)$, a training set $S = \{(x_i, t_i)|_{i=1}^{N}\}$ containing N arbitrary and non-repeating samples is input to the ELM network for training. In this case, the ELM model is denoted as $H\beta = T$, and the obscured level output matrix H is calculated as follows:

$$H = \begin{bmatrix} h_1^T & h_2^T & \cdots & h_N^T \end{bmatrix}^T = \begin{bmatrix} g(w_1, b_1, x_1) & \cdots & g(w_l, b_l, x_1) \\ \vdots & \ddots & \vdots \\ g(w_1, b_1, x_N) & \cdots & g(w_l, b_l, x_N) \end{bmatrix}_{N \times l} \quad (1)$$

where w is the weight and b is the bias, and the commonly used activation functions are Tanh and ReLU. The implicit layer output of each node of the ELM is as follows, where $i = 1, 2, \ldots, L$.

$$\sum_{i=1}^{L} \beta_i G(a_i \cdot x_N + b_i) = t_N \tag{2}$$

In order for the trained ELM network to fit all the training samples, the error among the network output and the actual is required to be minimised. In addition, to reduce the computational complexity, the minimum norm least squares solution (LS-Norm) $\hat{\beta} = H^+ T$ is solved by the generalised inverse theory, where $H^+ = (H^+ H)^{-1} H^+$ is the generalised inverse of H.

For LS-Norm, it satisfies the minimum training error as a least squares solution, as shown in equation (3).

$$\left\| H(w_1, ..., w_l, b_1, ..., b_l) \hat{\beta} - T \right\| = \min_{\beta} \left\| H(w_1, ..., w_l, b_1, ..., b_l) \beta - T \right\| \tag{3}$$

3 Typical faults of RC start-up and improvement measures

The mechanical structure of RC is complex, and there are many parts, resulting in the complexity of the fault. Therefore, the analysis of common RC faults is the basis of fault diagnosis. Through literature review, field data analysis and discussion with field technicians, the causes of several major fault types and their countermeasures are summarised.

1 Air lifter malfunction. The air lift opens and closes automatically in response to changes in air pressure in the cylinder, and an intake or discharge valve needs to be opened and closed once during a compressor operating cycle. If the unit fails to shut down and needs to be transported to another one to ensure the amount of gas lift in an emergency, the pressure of gas lift will drop sharply due to the long loading time, generally from 10.95 Mpa to 9.2 Mpa. Failure of gas lift may lead to the cylinder temperature, vibration abnormality, insufficient gas pressure and other problems. To shorten the warm-up time of winter start-up, wrap 60 W electric heat tracing around the lower part of the crankcase and add heat preservation at the same time to increase the insulation effect of the crankcase lubricating oil. Can also be outside the circulation pump inlet line loaded with heat and heat preservation, to ensure that in the cycle of heating does not cause heat loss.

2 No oil flow failure. The gas lift compressor of Tower One Union is located in the desert hinterland, and the temperature is low in winter. Together with the lack of heating facilities in the plant, it has a greater impact on the fluidity of the lubricant. The unit has been running for more than 16,400 hours, with a total of 42 shutdowns due to no oil flow failure, resulting in an annual venting of about 28,000 cubic metres. This unit operates as a gas lift to increase the gathering rate and requires one hour for troubleshooting in the event of a failure, affecting oil production by approximately 18 tons. In response to the failure, all the way to the oil supply to two-way oil supply, oil supply line before and after the filter element to increase the ball valve, increase the ball valve can be realised without stopping the clogging of the filter element to clean the situation. Through this optimisation and modification of no oil flow failure downtime from 10 times in 2022 to 1 time in 2023. At the same

time, it provides a guarantee for the completion of the user's production tasks, as shown in Figure 2.

Figure 2   No oil flow failure shutdown response strategies (see online version for colours)

3   Failure of packing. The packing is tightened radially by the axial force of the cylinder head to achieve the goal of sealing to prevent the compressed gas from leaking in the direction of the piston rod. The packing box is subjected to huge pressure difference and thermal load for a long time, and it is very easy to fail due to serious wear and tear over a long period of time. A cut-off valve is added at the 2-stage washing tank discharge line to connect to the vent line, and the cut-off valve is slightly opened to bring out the system liquid when the unit is running, and the valve is closed when the unit is on long-term standby.

4   Slideway failure. The diameter of the oil path from the elevated oil tank to the inlet of the single pump is too small, and at the same time, the lubricant viscosity is high and the circulation is small, resulting in the oil supply not being able to keep up. The elevated mailbox to the oil injection pump oil collection body at the pipeline diameter increased from 10 mm to 22 mm; oil circuit design is unreasonable, resulting in pre-lubrication oil injection pump and oil injection single pump two oil circuit tampering pressure. Increase the check valve at the convergence of the two oil circuits to avoid pressure.

5   Piston ring failure. The temperature of the piston rod has been abnormally high for many times, and the highest temperature detected has reached 185°C. There have been problems such as packing ring flanging, packing seal failure, localised carbonisation of lubricant at high temperatures, and bluing of piston rod. Increase the bore clearance by 0.16–0.20 mm, and also increase the clearance of the first straight opening ring of the sealing ring by 0.20 mm. In addition, the materials of the main sealing ring are all PEEK, which has poor thermal conductivity, and change the support ring in the main sealing ring to brass, which has good thermal conductivity.

By comprehensively analysing and deducing the working principle, structural characteristics and performance in practical applications of the above RCs, and combining with the existing research (Lu and Wang, 2021), the influencing factors of RC

starter, including temperature, pressure, noise, etc., can be obtained and denoted as x_1, x_2, …, x_n.

4 General architecture of SPS-based RC start-up fault monitoring

After obtaining the influencing factors of RC start-up failures, it is essential to accumulate the operational information of the RC in order to understand the condition and damage of the RC for subsequent monitoring models. The SPS is a type of sensor used for condition monitoring (Iqbal and Israr, 2021), which collects signals from the device in order to identify possible safety risks in the RC. The SPS has a high frequency response compared to other sensors and is highly sensitive to small vibration and shock signals. This allows it to capture weak signals that conventional sensors may miss, thus providing more comprehensive monitoring data. The overall architecture is shown in Figure 3.

Figure 3 SPS-based fault monitoring framework for RC starters (see online version for colours)
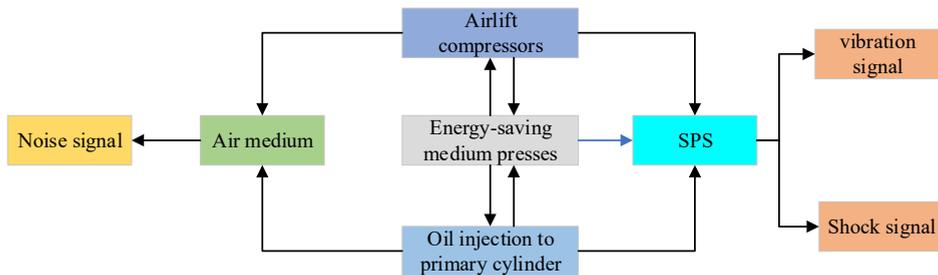
Due to the wide variety of signals collected by conventional SPS, the overlap of different signals can lead to diagnostic errors. For this reason, this paper optimises the SPS for fault monitoring.

1 Optimise the acquisition frequency of the SPS. The SPS uses a high-speed sampling rate, and if the acquisition frequency is too low, it will not be able to acquire enough data, making it tough to perform defect monitoring. Thus, the acquisition frequency of the SPS needs to be optimised to ensure that enough data can be acquired.

$$F_s = \omega 2D = (2\pi f)2Da \tag{4}$$

where F_s is the acquisition frequency of the SPS; w is the angular frequency; D is the displacement; f is the frequency; and a is the acceleration.

2 Optimise the distribution method of SPS. In the monitoring process, the distribution point mode of SPS can be reasonably arranged according to the structure and operation of RC. This can minimise the interference between data collection points and improve the data collection efficiency. Since the standard rod and the SPS contact surface have equal force at both ends, the force distribution points of the sensor are as follows.

$$F(t) = \left[\sigma_I(t) + \sigma_R(t)\right]A \tag{5}$$

where $\sigma_I(t)$ is the incident value, $\sigma_R(t)$ is the reflected stress value, and A is the calibration value. $\sigma_I(t)$ and $\sigma_R(t)$ are calculated as shown in equation (6) and equation (7), respectively.

$$\sigma_1(t) = E \cdot \varepsilon_1(t) = E \cdot \frac{U_1(t)}{S_g K_g} \tag{6}$$

$$\sigma_R(t) = E \cdot \varepsilon_R(t) = E \cdot \frac{U_R(t)}{S_g K_g} \tag{7}$$

where E is the correction variable; $\varepsilon_1(t)$ is the incident signal; $\varepsilon_R(t)$ is the reflected signal; $U_1(t)$ is the incident voltage signal generated by the strain.

5   RC start-up fault monitoring based on limit learning machine

5.1   RC start-up fault input variable pre-processing based on principal element analysis approach

Focusing on the issue that the high complexity of input variables in the current RC start-up fault monitoring method leads to low monitoring accuracy, a sensor- and ELM-based RC start-up fault monitoring method is designed. Firstly, PCA is used to eliminate irrelevant variables, then Gaussian white noise is introduced into the traditional EMD algorithm to improve it (EEMD), EEMD is used to decompose the main input variables to reduce the complexity of input variables, and finally ELM is used to predict the failure type of RC starter by using the improved ELM with regularisation term. The overall flow is shown in Figure 4.

Figure 4   The overall flow of the designed RC starter fault monitoring
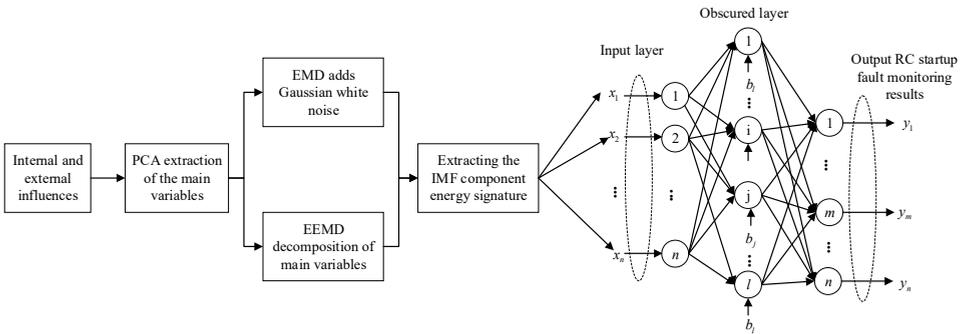
RC starters involve external influences as well as internal operating data (temperature, pressure, etc.), and these variables are large enough to form an unrecognisable database. Therefore, the use of PCA (Beattie and Esmonde-White, 2021) eliminates irrelevant variables and reduces the complexity of the input variables In the RC start-up process there are hundreds of different parameters, how can the main variables be accurately extracted need to use the PCA method. The purpose of calculating the principal elements is to determine the primary and secondary status of the principal elements based on the influencing factors and the magnitude of the variance of the kinetic parameters.

Assuming that the RC is in a normal state during operation, the data samples collected at this time are n groups, denoted as $x = [x_1 \; x_2 \; ... \; x_n]$, and the sample data set consisting of m-dimensional variables is $X \in R^{m \times n}$. In order to eliminate the influence of the variable's magnitude, the standardisation of X is performed. Let $\bar{X}$ be the normalised data matrix as follows.

$$\bar{X} = \left(X - I_m m^T \right) D_s^{-1/2} \tag{8}$$

where I_m is an m-dimensional column vector with all elements being 1; $u = [u_1, u_2, …, u_n]^T$ is a vector of sample means; and $D_s = diag(s_1^2, s_2^2, L, s_n^2)$ is the sample variance matrix.

The process of eigenvalue decomposition of the covariance matrix S of X is actually the process of principal element analysis modelling, and S can be expressed as follows:

$$S = \frac{X^T X}{m - l} \tag{9}$$

The eigenvalue decomposition of S has $Sp_i = \lambda_i p_i$, $i = 1, 2, …, n$, where λ_i is an eigenvalue, p_i is an eigenvector, and there is $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, and the corresponding eigenvector is $p_1, p_2, …, p_n$. The data matrix can be decomposed in terms of principal elements as follows:

$$X = TP^T + \overline{TP}^T = TP^T + E \tag{10}$$

where T is the principal score matrix, $\bar{T}$ is the residual score matrix, E is the residual term, P is the principal loading matrix, $\bar{P}$ is the residual loading matrix, and k is the number of principal elements.

If the residual term of equation (10) is removed, only the k principal elements are retained, resulting in a PCA model.

$$X \approx TP^T = x_1 p_1^T + x_2 p_2^T + \cdots + x_k p_k^T = X_P \tag{11}$$

The main input variable for RC starter failure can be obtained as $\{x_1, x_2, …, x_k\}$ by the above equation.

5.2 Variable feature extraction based on improved EMD

There is a large amount of noise interference in the input variables of RC starter, resulting in the intrinsic mode function IMF after EMD decomposition (Li et al., 2021) containing interference components, which affects the accuracy of fault diagnosis. Therefore, the EMD is improved and the reconstructed variables are formed into an input matrix to realise signal-noise separation and thus realise the noise reduction of variables. The IMF components after EEMD decomposition have modal attributes, and the energy characteristics of the IMF components of the main influencing variables of the RC initiator are used as the extracted features. The Gaussian white noise is added to the original variables to obtain the signal to be decomposed.

$$x_i(t) = x(t) + h_i(t) \tag{12}$$

where $h_i(t)$ is the Gaussian parameter added to the ith decomposition; t is the random variable; and $x(t)$ is the source variable. The decomposition of the variables to be decomposed using EEMD is as follows:

$$x_i(t) = \sum_{j=1}^{K} c_{ij}(t) + r_i(t) \tag{13}$$

where $c_{ij}(t)$ is the jth component of the ith decomposition; K is the mode number; $r_i(t)$ is the residual component of the ith decomposition. The new components are obtained by summing the obtained components as follows, where N is the number of variables.

$$c_j(t) = \frac{1}{N} \sum_{i=1}^{N} c_{ij}(t) \tag{14}$$

The EEMD decomposition values were subsequently obtained as follows:

$$x(t) = \sum_{j=1}^{K} c_j(t) + r(t) \tag{15}$$

where $c_j(t)$ is the jth IMF component and $r(t)$ is the residual component.

Find the energy $E_i = \sum_{k=1}^{n} |x_{ik}|^2$ of each IMF component, where x_{ik} is the amplitude of the discrete point of the ith IMF component, sum E_i to calculate the total energy $E = \sqrt{\left(\sum_{i=1}^{N} |E_i|^2 \right)}$, and replace the previous energy feature E_i with the normalised energy $e_i = E_i / E$, and finally get the extracted feature as $F = [e_1, e_2, \ldots, e_N]$.

5.3 RC start-up fault monitoring based on improved ELM

The traditional ELM algorithm is proposed based on the principle of empirical risk minimisation, which is highly influenced by the amount of features, and the algorithm becomes unusually complex to deal with these features, resulting in overfitting. Therefore, the stability of the model is improved by limiting the complexity of the algorithm and reducing the parameters of the network by adding a regularisation term that represents the structural complexity to the objective function, which is as follows.

$$\min_{\beta} \left\{ L_{regular} = \frac{C}{2} \| \xi_i \|^2 + \frac{1}{2} \| \beta \|^2 \right\} \tag{16}$$

where $\beta = H^T H$, H are the hidden layer output matrices, ξ_i is the training error of the ith sample, and C is the regularisation parameter, which represents the weight of the training error and algorithm complexity in the loss function. The constrained optimisation problem of equation (16) is transformed into a pairwise optimisation problem to obtain the following objective function.

$$\min_{\beta, \alpha, \xi} \left\{ L_{dual} = \frac{1}{2} \| \beta \|^2 + \frac{C}{2} \| \xi_i \|^2 - \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_{ij} \left(\left(h(x_i)^T \beta_j - t_{ij} + \xi_{ij} \right) \right) \right\} \tag{17}$$

where α is the Lagrange multiplier, β is the connection weight between the implicit level and the jth output node, and the corresponding optimisation constraints are as follows:

$$\frac{\partial L_{dual}}{\partial \beta_j} = 0 \rightarrow \beta_j = \sum_{i=1}^{N} \alpha_i h(x_i)^T \rightarrow \beta = H^T \alpha \tag{18}$$

$$\frac{\partial L_{dual}}{\partial \xi_j} = 0 \rightarrow \alpha_i = C\xi_i \tag{19}$$

$$\frac{\partial L_{dual}}{\partial \alpha_j} = 0 \rightarrow h(x_i)\beta - t_i^T + \xi_i^T = 0 \tag{20}$$

Bringing equation (18) into equation (19) yields the following equation.

$$\begin{cases} h(x_1)H^T C\xi_1 - t_1^T + \xi_1^T = 0 \\ \dots \\ h(x_N)H^T C\xi_N - t_N^T + \xi_N^T = 0 \end{cases} \tag{21}$$

Let $T = [t_1^T \ t_2^T \ \dots \ t_N^T]$, $H = [h(x_1) \ h(x_2) \ \dots \ h(x_N)]$ and write $[(I \ / \ C) + HH^T]\alpha = T$ by combining T and H. Equation (22) can be derived.

$$\beta = H^T \left(\frac{I}{C} + HH^T \right)^{-1} T \tag{22}$$

The final classification function for RC start-up fault monitoring is derived as follows.

$$f(x) = h(x)H^T \left(\frac{I}{C} + HH^T \right)^{-1} T \tag{23}$$

where x is the feature extracted from the main input variables and is the final RC start-up fault monitoring result.

6 Experimental results and analyses

In this paper, the Tarim Oilfield Tarzhong oil recovery and gas management area Tarim Yilian DTY1400H 240 × 240 × 145 × 145 gas lift compressor as the RC, impact pulse sensors to monitor the RC at all levels of cylinder inlet pipe and exhaust pipe temperature packing box leakage and so on, and real-time monitoring of the temperature changes, as shown in Figure 5. Heating for 40 minutes at 18:50 on the same day, the temperature went to 37°C and dropped to about 23°C at 4:00 AM. Heating for 40 minutes at 4:00 AM, the temperature dropped to 21°C at 11:00 AM, and returned after 11:00 AM as the ambient temperature increased. A total of 590 sets of laboratory information is adopted as training data of A_1, A_2, A_3, and A_4. Among them, 400 sets of data are selected for the normal state A_1, 50 sets of data are chosen as each of defect condition of A_2, A_3, and A_4 as the training data, and 10 sets of data are chosen for each state as the examination data. This resulted in 550 sets of training data and 40 sets of examination

data. 40 sets of instances were adopted to test the SOELM-trained defect diagnosis method. A description of SOELM monitoring results is shown in Table 1.

Figure 5   Real-time monitoring of temperature changes (see online version for colours)
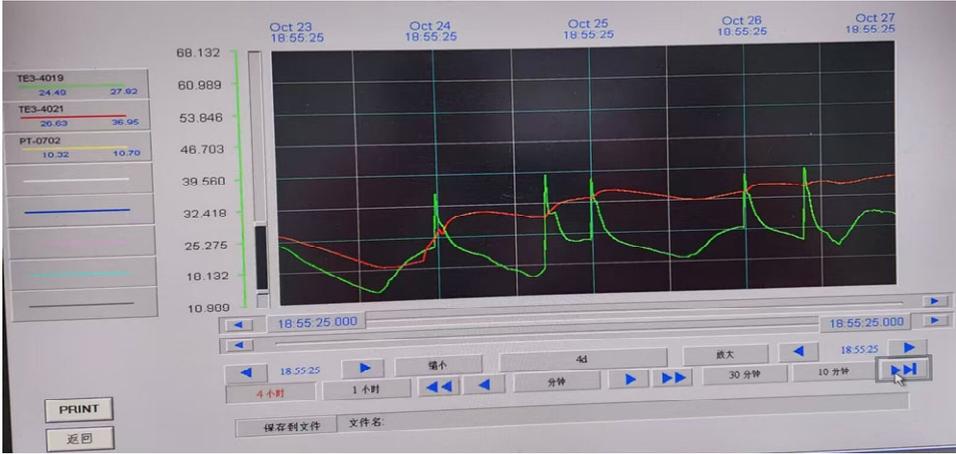
MATLAB was chosen for the test platform, the amount of output neurons was set as 4, the Sigmoid operation was chosen as the stimulation operation, and the number of neurons in the obscured level was 75. The monitoring results of the OELM are implied in Figure 6. By adopting the trained OELM method to diagnose every defect type with ten sets of test data, the test outputs were compared and organised in terms of the fundamental that the test outputs are turned to binary codes, which has already been set up in the previous section. The accuracy of fault monitoring for types A_1, A_2, A_3 and A_4 was 100%, 90%, 100% and 80% respectively. In summary, the ultimate outcome is that 37 groups of data were accurately diagnosed and three groups of data were incorrectly diagnosed, thus it could be deduced that the ultimate defect monitoring accuracy rate obtained from the test diagnosis of 40 groups of test data is 92.5%.

Table 1   Description of SOELM monitoring output results

SOELM monitoring outputs				State of health
A_1	A_2	A_3	A_4	
0	0	0	1	Well-being
0	0	1	0	A_2 fault
0	1	0	0	A_3 fault
1	0	0	0	A_4 fault

To further validate the effectiveness of the proposed monitoring method SOELM, SOELM was compared with the IU-ANN method (Cervantes-Bobadilla et al., 2023), the MRCVF method (Loukopoulos et al., 2019b), the RCCP method (Medina et al., 2022), and the GFDSD method (Wu et al., 2023). GFDSD method (Wu et al., 2023) were used for comparison experiments, and the metrics chosen were monitoring accuracy, MAE, R^2, and AUC of curved surface area under the ROC line. The results of the monitoring accuracy of different methods are shown in Table 2.The monitoring accuracy of SOELM

was 92.5%, which was improved by 20%, 12.5%, 7.5%, and 5% compared with IU-ANN, MRCVF, RCCP, and GFDSD, respectively. SOELM was developed by fusing the external factors and the sensor data as the input variables, utilising the PCA to remove the redundant variables, and by EEMD for the decomposition of input variables reduces the complexity of input variables. And it also improves the ELM, which greatly improves the accuracy of RC starter fault monitoring.

Table 2   Monitoring accuracy of different methods

Method	Total number of test samples	Correctly identify the sample size	Number of misidentified samples	Accuracy
IU-ANN	40	29	11	72.5%
MRCVF	40	32	8	80%
RCCP	40	34	6	85%
GFDSD	40	35	5	87.5%
SOELM	40	37	3	92.5%

Figure 6   SOELM fault monitoring results (see online version for colours)
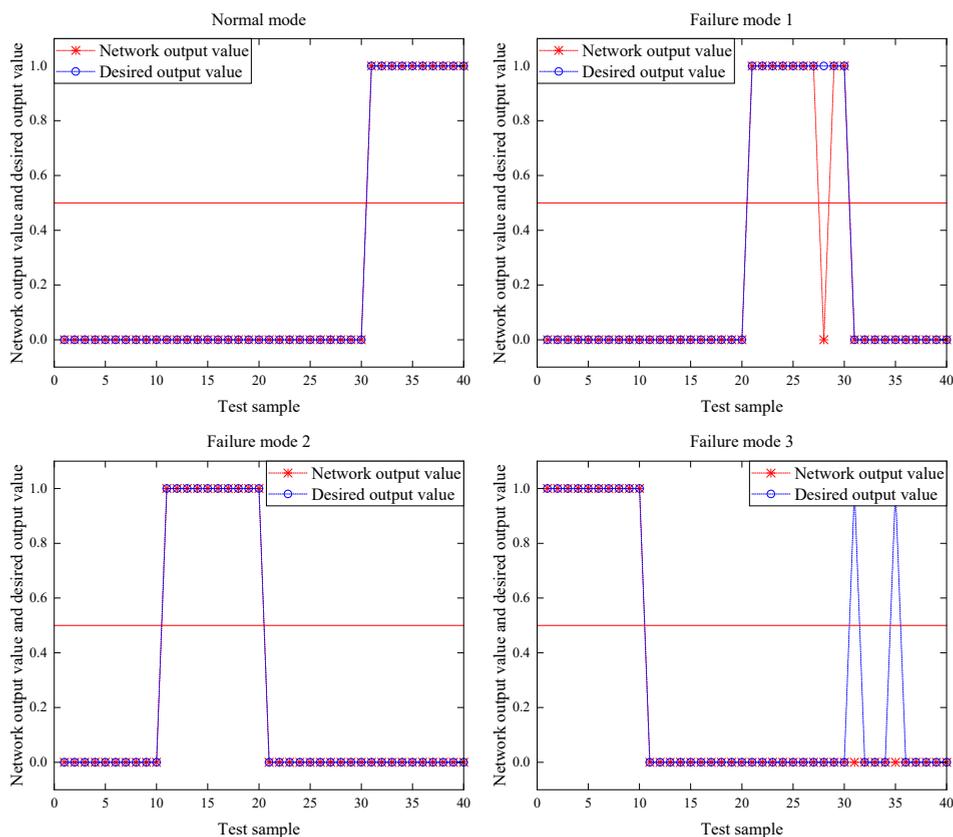
Comparisons of F1, MAE, R^2, and AUC for the different methods are shown in Table 3. The MAE for SOELM is 0.1235, which is at least 29.9% lower compared to the other

four models. R^2 is the coefficient of determination, and the closer it is to 1, the better the fit. The R^2 of SOELM is 0.9712, which is closest to 1, and the fitting degree of monitoring results is the highest. Compared with the AUC of the other four methods, the AUC of SOELM was increased by 17.59%, 9.16%, 4.6% and 1.2% respectively. IU-ANN and MRCVF only considered the external factors affecting RC failure, but did not consider the sensor data, resulting in poor monitoring performance. RCCP and GFDSD do not improve ELM and EMD, resulting in lower monitoring performance than SOELM. According to the above analysis, the monitoring performance of SOELM is the best.

Table 3   Comparison of monitoring performance indicators of different methods

Method	IU-ANN	MRCVF	RCCP	GFDSD	SOELM
MAE	0.3514	0.2965	0.2108	0.1762	0.1235
R^2	0.8104	0.8745	0.9120	0.9527	0.9712
AUC	0.8369	0.9015	0.9408	0.9724	0.9841

7   Conclusions

RC is a key equipment in the production process of petrochemical enterprises, and it is crucial to ensure its safe operation. In order to solve the problems of high complexity of input variables and low monitoring accuracy of existing RC start-up monitoring methods, a sensor and ELM based RC start-up fault monitoring method is designed. Through the analysis of common failure types and their response strategies, the external influencing factors of RC starters are summarised. Secondly, the acquisition frequency and distribution method of the SPS are optimised, based on which the overall architecture of RC starter fault monitoring is designed to obtain the intrinsic signal data of the RC starter and improve the data acquisition efficiency. The external influences and intrinsic data were then pre-processed to reduce the complexity of the input variables by eliminating irrelevant variables using PCA and decomposing the main variables into a series of IMF components using the EEMD method. Finally, the ELM algorithm (OLEM) is optimised by the regularisation term, and the network parameters are reduced to improve the stability of the model, and the OLEM is used to predict the defect state of the RC starter. The experimental outcome implies that the proposed method has a high monitoring accuracy and can greatly improve the monitoring efficiency of EC starters.

Declarations

All authors declare that they have no conflicts of interest.

References

Ahmadianfar, I., Shirvani-Hosseini, S., He, J. et al. (2022) 'An improved adaptive neuro fuzzy inference system model using conjoined metaheuristic algorithms for electrical conductivity prediction', *Scientific Reports*, Vol. 12, No. 1, p.4934.

Almasi, A. (2016) 'Latest practical notes and recent lessons learned on reciprocating compressors', *Australian Journal of Mechanical Engineering*, Vol. 14, No. 2, pp.138–150.

Beattie, J.R. and Esmonde-White, F.W. (2021) 'Exploration of principal component analysis: deriving principal component analysis visually using spectra', *Applied Spectroscopy*, Vol. 75, No. 4, pp.361–375.

Cervantes-Bobadilla, M., García-Morales, J., Saavedra-Benítez, Y. et al. (2023) 'Multiple fault detection and isolation using artificial neural networks in sensors of an internal combustion engine', *Engineering Applications of Artificial Intelligence*, Vol. 117, p.105524.

Ding, S., Xu, X. and Nie, R. (2014) 'Extreme learning machine and its applications', *Neural Computing and Applications*, Vol. 25, pp.549–556.

Huang, G-B., Zhu, Q-Y. and Siew, C-K. (2006) 'Extreme learning machine: theory and applications', *Neurocomputing*, Vol. 70, Nos. 1–3, pp.489–501.

Iqbal, M.Z. and Israr, A. (2021) 'To predict a shock pulse using nonlinear dynamic model of rubber waveform generator', *International Journal of Impact Engineering*, Vol. 147, p.103731.

Li, L., Zhou, H., Liu, H. et al. (2021) 'A hybrid method coupling empirical mode decomposition and a long short-term memory network to predict missing measured signal data of SHM systems', *Structural Health Monitoring*, Vol. 20, No. 4, pp.1778–1793.

Loukopoulos, P., Zolkiewski, G., Bennett, I. et al. (2019a) 'Abrupt fault remaining useful life estimation using measurements from a reciprocating compressor valve failure', *Mechanical Systems and Signal Processing*, Vol. 121, pp.359–372.

Loukopoulos, P., Zolkiewski, G., Bennett, I. et al. (2019b) 'Reciprocating compressor prognostics of an instantaneous failure mode utilising temperature only measurements', *Applied Acoustics*, Vol. 147, pp.77–86.

Lu, Y.J. and Wang, C.H. (2021) 'Integration of wavelet decomposition and artificial neural network for failure prognosis of reciprocating compressors', *Process Safety Progress*, Vol. 40, No. 3, pp.105–115.

Marchante-Avellaneda, J., Navarro-Peris, E., Corberan, J. et al. (2023) 'Analysis of map-based models for reciprocating compressors and optimum selection of rating points', *International Journal of Refrigeration*, Vol. 153, pp.168–183.

Medina, R., Cerrada, M., Yang, S. et al. (2022) 'Fault classification in a reciprocating compressor and a centrifugal pump using non-linear entropy features', *Mathematics*, Vol. 10, No. 17, p.3033.

Nwakpang, I.A., Lebele-Alawa, B.T. and Nkoi, B. (2019) 'Performance assessment of a two-stage reciprocating air compressor', *European Journal of Engineering and Technology Research*, Vol. 4, No. 4, pp.74–82.

Patil, A., Soni, G., Prakash, A. et al. (2021) 'Intelligent valve fault diagnosis approach for reciprocating compressor based on acoustic signals', *Reliability: Theory & Applications*, Vol. 16, No. SI2(64), pp.35–47.

Pichler, K., Lughofer, E., Pichler, M. et al. (2016) 'Fault detection in reciprocating compressor valves under varying load conditions', *Mechanical Systems and Signal Processing*, Vol. 70, pp.104–119.

Qi, G., Zhu, Z., Erqinhu, K. et al. (2018) 'Fault-diagnosis for reciprocating compressors using big data and machine learning', *Simulation Modelling Practice and Theory*, Vol. 80, pp.104–127.

Sharma, V. and Parey, A. (2019) 'Performance evaluation of decomposition methods to diagnose leakage in a reciprocating compressor under limited speed variation', *Mechanical Systems and Signal Processing*, Vol. 125, pp.275–287.

Wang, D., Sun, J., He, Q. et al. (2022) 'Failure analysis and improvement measures for crankshaft connecting rod of refrigerator compressor', *Engineering Failure Analysis*, Vol. 141, p.106585.

Winandy, E., Saavedra, C. and Lebrun, J. (2002) 'Simplified modelling of an open-type reciprocating compressor', *International Journal of Thermal Sciences*, Vol. 41, No. 2, pp.183–192.

Wu, Z., Yan, H., Zhan, X. et al. (2023) 'Gearbox fault diagnosis based on optimized stacked denoising auto encoder and kernel extreme learning machine', *Processes*, Vol. 11, No. 7, p.1936.