Article

Deep Learning-Based Approach for Heat Transfer Efficiency Prediction with Deep Feature Extraction

Yuanhao Shi,* Mengwei Li, Jie Wen, Yanru Yang, and Jianchao Zeng

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Failure to blow ash on the heated surface of the boiler will cause a drop in heat transfer rate and even industrial safety accidents. Nowadays, the shortcomings of the fixed soot blowing operation every hour and every shift are significant, which can be improved by high-precision ash accumulation prediction. Therefore, this paper proposes a deep learning model fused with deep feature extraction. First, a dynamic fouling model and a health index-clearness factor (CF) of the heated surface are established. The data preprocessing method reduces unnecessary forecasting difficulty and makes the degradation trend of the CF time series more obvious. In addition, deep feature extraction is composed of complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) and kernel principal component analysis (KPCA), which completes the multiscale analysis of time series and reduces the training time of deep learning models, and has significant contributions to improving prediction accuracy and reducing time consumption. The adaptive sliding window and the encoder−decoder based on the attention mechanism (EDA) can better mine the internal information of the time series. Compared with long short-term memory (LSTM), taking the 300 MW boiler's various heated surface data sets as an example, multistep forward prediction and different starting point prediction experiments have verified the superiority and effectiveness of the model. Finally, under the variable working condition economizer datasets, the proposed method better completes the predictive maintenance task of the heated surface. The research results provide operational guidance for improving heat transfer rate, energy saving, and reducing consumption.
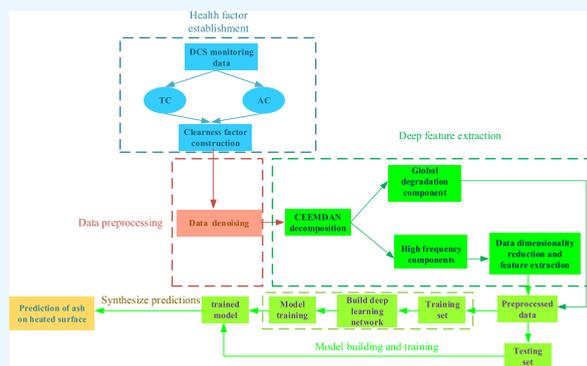
1. INTRODUCTION

With the continuous improvement of living standards around the world, the issues of environmental protection, energy conservation, and emission reduction have become the focus of attention all over the world.[1−3] Although active energy transformation is being carried out, fossil fuels are still the main world energy sources, and their proportion and status are still irreplaceable by new energy sources. The main existing problems of fossil energy are utilization and pollution emissions. Coal is an important part of fossil energy. Energy consumption mainly comes from the consumption of it, and more than half of the coal is supplied to coal-fired power stations every year.[4]

As the basis for the operation of coal-fired power stations, boilers have basically reached a satisfactory level of power generation efficiency with the development of instrumentation and intelligence. However, if the parameters and power of the boiler are large, this situation will occur: after the pulverized coal is burned at a high temperature of thousands of degrees, high-temperature flue gas will be generated to the working fluid side inside the heated surface by means of heat transfer.[5] Ash present in the high-temperature flue gas is in a molten state at this time because it exceeds the melting point. The melted ash will cause

ash accumulation as the high-temperature flue gas flows through each heating surface since the thermal resistance of ash fouling and slagging is much greater than the thermal resistance of the metal heating surface, the working fluid on the working fluid side will need to provide more raw coal in order to meet the required critical requirements. In addition, ash deposits on the heating surface will cause a series of problems, such as the reduction of the operating efficiency of the heat exchanger, the corrosion of the heating surface and metal pipes, the overall shutdown of the unit, and a significant reduction in the service life of the equipment.[6] Due to the poor heat absorption of the heating surface, the flue gas temperature at the outlet of the boiler is relatively high, which reduces the flue gas desulfurization efficiency. The core of tapping the energy-saving potential of the

boiler is to improve the heat transfer efficiency of the various heat exchange equipment and the overall heat transfer of the boiler and to convert the calorific value of the coal into the heat of the working fluid to the greatest extent.

With the popularization and application of distributed control system systems (DCS), power plants began to establish management information systems and plant-level supervisory information systems, which conveniently and quickly recorded the production process of the power plant real-time information of each location, and save complete historical data.[7] It has laid a good foundation for improving the online monitoring of ash pollution and predictive maintenance of the heating surface.

As an effective method to keep the heated surface healthy, soot blowing is used to clean the surface of the heat exchanger through a medium such as high temperature steam. Nowadays, many thermal power stations all over the world adopt the soot blowing method at a fixed time and a fixed operation process.[8,9] This soot blowing method has such a hidden problem: If soot blowing is not timely (under soot blowing), it will lead to aggravation of the ash situation in the heated area, reduction of heat transfer efficiency, and major safety accidents. If the soot blowing frequency is too high (over soot blowing), it will not only cause waste of high-temperature and pressure steam used for soot blowing but also cause corrosion of the heated surface and pipeline. Long-term over soot blowing will greatly shorten the power station equipment life span, and it also brings potential problems with energy utilization and safe operation.

Accidental contamination on the surface of heat transfer boilers has always been one of the main operational problems of coal-fired utility boilers. A large number of studies have shown that in order to develop intelligent soot blowing technology on the heating surface of coal-fired power stations to avoid the heat transfer loss of the heat exchanger and the occurrence of safety accidents caused by the traditional way of empirical soot blowing, research work mainly focuses on the monitoring and prediction of ash deposits.[10] In recent years, the research work has mainly been carried out from two aspects: ash accumulation monitoring, prediction, and soot blowing optimization. In detail, there are usually monitoring devices, actual physical models[5], and data-driven methods for fouling monitoring. Perez et al.,[11] considering the global response time of the system in the polluted state and comparing it with the cleaning state, designed a new transient thermal fouling probe for crossflow tubular heat exchangers, which accurately estimated the convection exchange coefficient and the degree of fouling of the heat exchanger. Shi et al.[12] based on dynamic mass and energy balance to detect contamination on the surface of the heat exchanger's heating surface, in addition to steam flow soft measurement, completed the online evaluation of boiler performance. Zhang et al.[13] proposed an acoustic system that is used to monitor the temperature change near the boiler water wall and a new cleanness factor. Based on this method, the ash fouling and slagging are monitored, which makes a certain contribution to the development of smarter smoke blowers. Ma et al.[14] integrated boiler computational fluid dynamics (CFD) simulation and ash behavior model-developed ash behavior prediction tool AshProSM, which can provide a qualitative and quantitative description of the formation and deposition process of the fireside slag. AshProSM has been applied to the industrial boilers of the Columbia Energy Center of Wisconsin Electric and Lighting Company. These methods monitor from the perspective of mechanism and the results can play a certain role in qualitative analysis.

As there are many factors affecting fouling, such as strong coupling between various factors, complicated and cumbersome calculations in the internal operation of the boiler, etc., the model-driven method has the problems of large prediction errors and time lag. In addition, due to the complexity and uncertainty of coal-fired power plant boiler production, the abovementioned method may not be able to comprehensively reflect the impact of various uncertainties and is limited in accuracy and difficult to apply to actual soot blowing optimization control. Therefore, data-driven methods are becoming more and more mainstream. With the continuous development of big data and artificial intelligence, data-driven methods have gradually become the mainstream method of monitoring the health of the heating surface. Unlike mathematical models, machine learning treats the actual system as a black box and fits the mathematical and physical principles inside the black box through input and output. Although such a pure data-driven algorithm lacks the exploration of the actual internal mechanism, with the continuous intake of intelligent optimization algorithms, further optimization of required parameters can also obtain satisfactory results. Sun et al.[15] selected fouling resistance as an indicator to monitor the pollution status of the heating surface. In addition, they analyzed fouling-related variables (such as working fluid input temperature, working fluid flow rate, etc.) and passed the Support Vector Machine (SVM) algorithm that has completed the monitoring of fouling on the heating surface. Similarly, Tong et al.[16] used Support vector regression (SVR) to complete the nonlinear mapping relationship between 20 related variables of ash formation and actual fouling conditions (characterized by the thermal resistance of the ash layer calculated by the thermal balance mechanism model), which reached the test set 98.5% accuracy rate. Shi and Wang[17] on the basis of characterizing the health status of the heating surface also proposed an artificial neural network-based key variable analysis to study the internal behavior of ash pollution and thermal efficiency. Sivathanu and Subramanian[18] designed a dual extended Kalman filter (DEKF) to estimate the model parameters that affect the pollution of the heating surface of the reheater. According to the estimated parameters, health indicators reflecting the pollution of the heated surface are obtained. DEKF is better than traditional joint EKF (JEKF) in terms of estimating model parameters. At present, many methods are based on artificial neural network technology,[19] which regards the fouling deposition system as a 'black box model', and completes the prediction of fouling and integrated optimization and automatic soot blowing control.

Predicting the future status of ash pollution is another important task. A large number of studies have shown that ash prediction of the heated area is essentially a time-series predicting task, and it can be predicted to a certain extent by using certain reasonable methods. Shi et al.[20] used the measurement data of the distributed control system (DCS) of thermal power plants and basic thermodynamic calculation data to monitor the pollution rate of the heated surface in real time. By analyzing the pollution rate of multiple groups, the incremental distribution of the same measurement point at different times is obtained, and the future state is predicted by the known initial ash pollution. Li et al.[21] decomposed the historical pollution rate data into two parts, the fitted curve data and the difference between the original data and the fitted curve, and then combined the real-time pollution rate data to establish the prediction model. This method does not require additional special instruments or complex computing systems but can use

existing monitoring data to realize economizer fouling monitoring. Compared with the traditional Elman neural network, the traditional Neural network algorithms find it difficult to achieve long-term predictions in multifactor coupled fouling prediction projects. At the time of the explosion of deep learning, due to the inherent deep feature extraction effect of the model, it has begun to show its strength in the application of multifactor coupling such as the time series of ash accumulation degradation.[22] In fact, most current research studies are using sensors, soft sensing,[23] and machine learning methods for online ash accumulation monitoring and short-term prediction. The improved ash cleaning method is generally divided based on predictive maintenance[21] and soot blowing optimization models. However, if any of these two methods are only based on online monitoring and short-term prediction, it is very limited in actual engineering applications.

The high-pressure steam required by the soot blower and the staffing of the soot blowing operation take a certain amount of time. This requires the establishment of health factors that can reflect the health of the heated surface of the heat exchanger to complete the prediction of the future situation. Based on the health factor-clearness factor (CF), this article predicts and analyzes the fouling conditions of the heated surfaces of different devices under the same operating conditions and the same devices under different operating conditions. In this regard, based on the safe operation of the heat exchanger and the need to avoid over-blowing and under-blowing, a method for predicting the health of the heating surface that combines deep feature extraction and deep learning is proposed. First, the wavelet threshold denoising method is used to reduce the burrs and noises in the CF curve, so that the overall trend of the ash accumulation curve is more obvious. The depth feature extraction method is mainly divided into complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) decomposition and kernel principal components analysis (KPCA) dimensionality reduction. CEEMDAN decomposition completes the multiscale analysis of the ash accumulation curve of various devices in order to obtain higher prediction accuracy. In addition, we generally increase the number of forwarding prediction steps in order to obtain a longer soot blowing operation preparation time, although a longer forward prediction time can reserve enough time for soot blowing operation preparation work and complete the 'early warning'. In general, in shallow prediction models, such as SVR, random forest, etc., the model training time is often neglected, so the forward prediction time will be completely used for preparation. Because the deep learning prediction model has the characteristics of a huge overall structure, numerous parameters, and many samples, the training time cannot be ignored, which will indirectly occupy the forward prediction time.

More importantly, in many cases, there may be correlations between various imfs, which increases the complexity of problem analysis. The KPCA not only eliminates redundant information but also reduces the training time of the model by performing dimensionality reduction operations and input reconstruction on the high-frequency components obtained by CEEMDAN decomposition. Therefore, this is a reasonable dimensionality reduction method to ensure the integrity and effectiveness of the original information to the greatest extent on the basis of reducing the number of inputs that need to be analyzed. The adaptive sliding window and the encoder—decoder based attention (EDA) complete the sudden change capture of the fouling time series and the long-term memory

establishes a prediction model for the newly reconstructed input sequence after feature extraction. In the end, this new hybrid model achieves a high-precision prediction of the health of the heating surface of the heat exchanger.

Contributions of this work:

(1) In order to obtain better fouling prediction accuracy, this paper proposes deep feature extraction, which includes multiscale analysis of fouling time series and dimensionality reduction algorithms

(2) Considering the relevance of the fouling time series and in order to mine its potential information, a fusion of the adaptive sliding window and encoder—decoder prediction framework is proposed.

(3) Taking a variety of boiler heating surface datasets of coal-fired power plants as an example, from the perspective of multistep forward prediction, the validity and adaptability of the proposed model in multistep-ahead prediction under different types of data sets are verified.

(4) Starting from multiple sets of variable-condition economizer datasets, the superiority and practicability of the proposed model in predictive maintenance tasks on the heating surface are verified.

The remainder of this paper is organized as follows. Health factor-clearness factor and data preprocessing, deep feature extraction, and deep learning algorithms are introduced in Section 2. In Section 3, we took the datasets of various heated surfaces and economizer variable conditions of coal-fired power stations as an example and conducted detailed verification and discussion on the research results of multistep-ahead prediction and predictive maintenance of the heated surfaces. Finally, the conclusions and prospects for the future are given in Section 4.

2. METHODOLOGY

This paper aims at the monitoring and prediction of ash accumulation in the heated area of coal-fired power station boilers and builds a deep learning model based on actual production data. In order to monitor and predict the ash accumulation on the heating surface, it is first necessary to extract characteristic variables that can reflect the ash accumulation status from a large number of relevant monitoring data in the boiler DCS system. Considering the influence of dynamic factors, a dynamic model is established so that it can better reflect the health status of the heated surface under the influence of ash pollution, that is, the clearness factor.

With the rapid development of Prognostics Health Management (PHM),[24] predictive maintenance of the heated surface of coal-fired power plants has become one of the focuses of power plants because it involves boiler safety issues and economic benefits. However, the traditional shallow model has poor multistep prediction performance, so it is difficult to perform the task of predicting the health of the heated surface. This paper constructs a prediction method based on the fusion of improved feature extraction and deep learning models and completes the feature decoupling and deep feature extraction of the fouling signal. In addition, the deep learning model based on the attention mechanism and the recurrent neural network increases the long-term dependence mining on the time series compared with the shallow model and obtains high-precision prediction results. The framework of the hybrid model we proposed is mainly composed of four parts as follows: First, the theoretical heat transfer coefficient and the actual heat transfer coefficient are calculated according to the DCS system, and then the clean
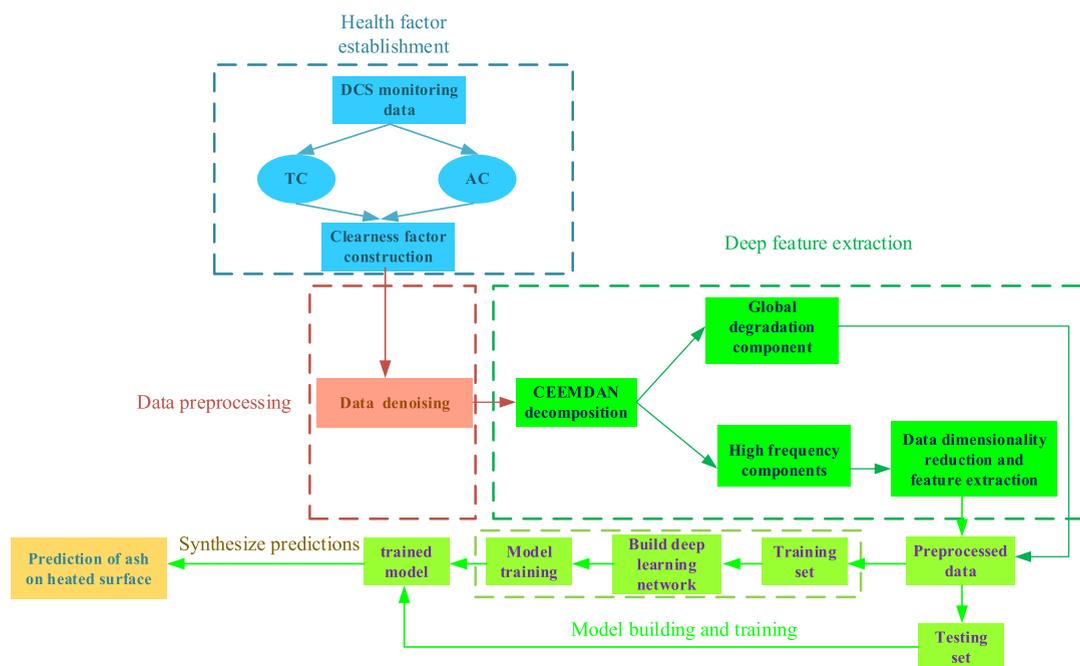
Figure 1. Online prediction of ash accumulation.

factor that characterizes the health of the heated surface is obtained. Then denoise the original cleaning factor degradation curve. By specifying the wavelet basis function, the number of decomposition layers, and the threshold function to complete the denoising and smoothing operation of the original data, the changing trend of the fouling signal is more obvious. Then, we use CEEMDAN decomposition to complete the multiscale analysis of the denoising signal and decompose it into multiple imfs and a trend component. In addition, KPCA is used for dimensionality reduction and deep mining of the decomposed features to complete input reconstruction with high-level abstract features. This dimensionality reduction algorithm reduces the computational cost and further improves the overall performance of the model. Finally, based on the adaptive sliding window and the encoder−decoder model of the attention mechanism, the information mining and accurate multistep-ahead prediction of the ash accumulation time series are completed. Figure 1 shows a complete prediction flow figure.

2.1. Dynamic Monitoring Model and Health Indicator. In this paper, in order to calculate the health status of each heating surface in real time and fully reflect the dynamic status of ash deposits under variable working conditions of the boiler,[12] we combine the basic thermodynamic formula and real-time measured data from the boiler DCS system to obtain the health indicator of the heated surface-clearness factor.

The clearness factor is mathematically composed of the ratio of the actual heat transfer coefficient to the theoretical heat transfer coefficient of the convective heating surface. The data required in the entire calculation process can be collected in real time by the boiler DCS system.

$$CF = \frac{k_r}{k_o} \tag{1}$$

The theoretical heat transfer coefficient is the original state without ash deposits on the heated surface. Under the premise of ignoring the thermal resistance of the working fluid and the tube wall and the internal resistance of the metal, it is usually the sum

of the theoretical radiation heat transfer coefficient and the theoretical convective heat transfer coefficient.

$$k_r = a_f + a_d \tag{2}$$

In formula 2, a_f represents the theoretical radiation heat transfer coefficient, and a_d is the theoretical convective heat transfer coefficient. The following formula is the specific mechanism formula of the two heat transfer coefficients:

$$a_f = 5.7 \times 10^{-8} \frac{a_{gb} + 1}{2} a_h T^3 K \tag{3}$$

$$a_d = 0.65 \, C_s C_z \frac{\lambda}{d} \left(\frac{wd}{v} \right)^{0.64} Pr^{0.33} \tag{4}$$

$$K = \left\{ \left(1 - \left(\frac{T_{gb}}{T} \right)^4 \right) \Big/ \left(1 - \frac{T_{gb}}{T} \right) \right\} \tag{5}$$

In formulas 3−5, a_{gb} and a_h are the blackness of the pipe wall and the flue gas respectively; T and T_{gb} are the temperature of the flue gas and the pipe wall respectively, C_s and C_z are the transverse and longitudinal directions of the heating surface, λ is the thermal conductivity of the flue gas, and d is the pipe diameter, w is the flue gas flow rate, v is the dynamic viscosity of the flue gas, and Pr is the Reynolds number.

The flue gas flow rate w is the ratio of the flue gas flow rate to the area of the tube section of the heating surface.

$$w = \frac{V_b}{A} \tag{6}$$

where V_b is the standard flue gas volume passing through the heating surface, A is the official cross-sectional area of the heating surface, and the standard flue gas flow rate is obtained by Avogadro's law.

$$V_b = \frac{\rho_r V_r}{\rho_b} \Big/ \left(1 + \frac{t_r}{273.15} \right) \tag{7}$$

In formula 7, V_r is the measured flue gas flow through the heating surface, t_r is the flue gas temperature through the heating surface, ρ_r is the actual pressure of the flue gas, and ρ_b is the standard atmospheric pressure.

The actual heat transfer coefficient is obtained by the dynamic energy balance and iterative method.

$$K_r = \frac{Q_y}{F \Delta t_m} \tag{8}$$

$$\Delta t_m = (\Delta t_{max} - \Delta t_{min})/\ln\frac{\Delta t_{max}}{\Delta t_{min}} \tag{9}$$

where Q_y is the energy released on the flue gas side, F is the heat transfer area of the heating surface, Δt_m is the average heat exchange temperature difference between the flue gas side and the working fluid side, and Δt_{max} and Δt_{min} are the maximum and minimum temperature differences of heat exchange on both sides.

Considering that during the operation of the boiler, as the load changes, the boiler's coal feed, air supply, and other variables are dynamically changing, the corresponding temperature of each heating surface is also changing, and the specific heat capacity of the working fluid will also change with the change of temperature. Therefore, the energy released by the flue gas side in the dynamic process is not completely equal to the heat absorbed by the working fluid. At this time, the change in the heat storage of the working fluid needs to be considered. Therefore, the energy conservation on the flue gas side and the working fluid side in this dynamic process can be expressed as

$$Q_y = Q_q \pm \Delta Q_j \pm \Delta Q_q \tag{10}$$

where Q_q is the heat absorption of the working fluid on the working fluid side, ΔQ_j is the change in the heat storage of steam, and ΔQ_q is the heat absorption change on the steam side.

Heat release on the flue gas side

$$Q_y = \varphi B_j(h_{in} - h_{out} + \beta h_{lf}) \tag{11}$$

$$B_j = B\left(1 - \frac{q_4}{100}\right) \tag{12}$$

φ is the heat retention coefficient, h_{in} and h_{out} are the flue gas enthalpy values at the inlet and outlet of the economizer, β is the air leakage coefficient of the flue section, and h_{lf} is the cold air enthalpy of the air leakage. B_j is the calculated fuel quantity, B is the actual measured fuel quantity entering the furnace, and q_4 is the heat loss of the mechanical incomplete combustion of the boiler.

The metal heat storage change of the pipe wall, the steam heat storage change, and the heat absorption of the steam side are as shown in the formulas.

$$\Delta Q_j = C_j m_j \frac{\partial \theta_j}{\partial \tau} \tag{13}$$

$$\Delta Q_q = C_q m_q \frac{\partial \theta_q}{\partial \tau} \tag{14}$$

$$Q_q = D(H_{out} - H_{in}) \tag{15}$$

In formulas 13−15, C_j and C_q are the average specific heat capacity of metal and working fluid respectively. m_j and m_q are the metal quality of the tube wall on the heated surface and the

quality of the working fluid inside. θ_q and θ_j are the metal pipe wall temperature and steam temperature, D is the mass flow of the working fluid of the economizer, and H_{out} and H_{in} are the side enthalpy values of the working fluid in and out of the economizer. The enthalpy value of the working fluid can be obtained by the international general industrial water and water vapor property calculation formula.

2.2. Data Preprocessing. CF is used as the indicator of the health condition of the heated surfaces to reflect the real-time ash condition well. In fact, the daily change of the CF has a strong non-linearity. It is challenging and unnecessary for datasets to be directly used for ash deposit prediction and soot blowing optimization. Generally speaking, the noise of the CF curve is generally divided into two types: one is the on-site environmental change. The other is that when the flue gas carrying ash is used for heat exchange, the flow of the flue gas causes the ash in the flue gas to deposit on the heated surface or take away part of the ash from the heating surface (it has a relationship with the flow rate of the flue gas). The former situation is what we do not want to appear, and the latter one, as the physical change inside the boiler, occurs almost all the time and cannot be ignored, which is also one of the difficulties in the ash prediction. Among many denoising algorithms, the combination of wavelet analysis and threshold denoising is an advanced data smoothing method, which has the characteristics of a high signal-to-noise ratio and strong adaptability after denoising.

As a bridge between the time domain and frequency domain, Fourier transform plays an extremely important role in early signal analysis and processing.[25] Wavelet transform can obtain not only the frequency component of the signal but also the occurrence time of each frequency signal. From a mathematical point of view, the wavelet transform is composed of a set of wavelet basis functions, which can be obtained by the translation and scaling of the wavelet basis functions. Its formula is shown as 16:

$$WT_f(\alpha, \tau) = \langle f(t), \psi_{\alpha,\tau}(t)\rangle = a^{-1/2} \int_R f(t)\overline{\psi\left(\frac{t-\tau}{\alpha}\right)} \tag{16}$$

The original signal $f(t) \in L^2(R)$. $\psi_{\alpha,\tau}(t)$ is the wavelet basis function, and a and τ are the translation and scaling coefficients, respectively. The inner product of x and x completes the continuous wavelet transform.

Due to practical engineering needs, binary discrete wavelet transform (DWT) (discretization of translation coefficient and scaling coefficient) is commonly used when dealing with time series problems, as shown below:

$$WT_f(2_j, b) = <f(t), \varphi_{j,b}(t)$$
$$> = 2^{-j/2} \int_R f(t) \times \varphi\left(\frac{t-b}{2^j}\right) dt \tag{17}$$

The basic principle of DWT decomposition is as follows: the original signal is continuously decomposed through high-pass and low-pass filters. First, the original signal is passed through high-pass and low-pass filters to obtain high-frequency components (H1) and low-frequency components (L1). Then, we let the low-frequency component (L1) pass through the high-pass and low-pass filters to obtain the new high-frequency component (L2) and the new low-frequency component (H2). Then, we repeat the process continuously, until the specified number of decomposition layers is reached.

The decomposition figure of DWT is shown in Figure 2. Choosing the appropriate wavelet basis function and the
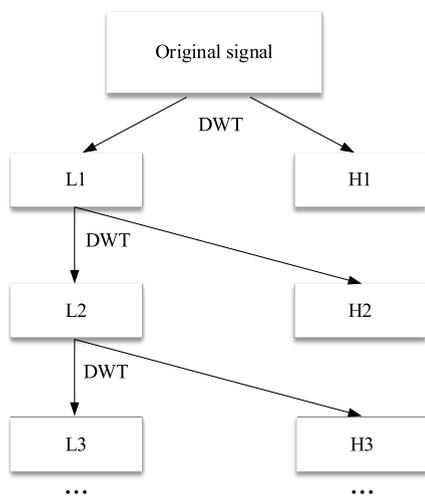
Figure 2. Wavelet decomposition structure.

number of decomposition layers is one of the keys to denoising. Generally speaking, after obtaining the wavelet decomposition coefficients of various levels, the final low-frequency coefficients of the wavelet decomposition coefficients are retained, and the high-frequency coefficients of each level are quantized. Because the noise part of the signal is usually located in the high-frequency segment, and the wavelet coefficient of the noise is generally smaller than the effective signal. The hard threshold function allows the signal points whose absolute value is less than the threshold value to be directly set to 0, while the soft threshold value shrinks the points with discontinuous boundaries to 0 on its basis. The soft threshold function is used to obtain a smoother denoising signal under the premise of ensuring the signal-to-noise ratio of the denoised signal, thereby solving the problem that the reconstructed signal may oscillate at some points. The wavelet threshold denoising algorithm strengthens the adaptability of the subsequent prediction algorithm to the time series of ash accumulation.[26] After quantifying the wavelet decomposition coefficients at all levels, the pure ash signal can be reconstructed by inverse wavelet transform.

2.3. Deep Feature Extraction. *2.3.1. Decomposition Algorithm.* In order to extract the high-dimensional details of the ash segment, this article will introduce EMD and its derivative algorithms, such as EEMD and CEEMDAN. Considering that the importance of the training time of the deep learning model in the entire ash deposit prediction and soot blowing optimization process and the high-frequency imfs obtained by the decomposition algorithm have certain redundant characteristics, this paper uses the KPCA algorithm to reduce the dimensionality of the high-frequency feature components obtained by decomposition. Therefore, under the premise of ensuring the minimum loss of effective information, a lot of time is reduced for the training of the deep learning model in the future.

Compared with other commonly used decomposition algorithms, the EMD algorithm has strong analysis and processing ability in both linear and nonlinear signal processing and can adaptively select the decomposition basis function and decomposition layer number according to the signal.

The EMD algorithm is based on the following assumptions:

(1) The original signal extreme point and the number of zero points must be equal or at most.
(2) The upper enveloped line defined by the maximum value point and the average value defined by the minor value point is zero, that is, the upper and lower envelope of the signal respects the time axis symmetry.

The EMD decomposition process is as follows:

Step 1: Connect all local extremum points in $x(t)$ with three spline interpolation curves to form up and down envelopes and m_{low}.

Step 2: The mean curve $m_1(t) = [m_{up} + m_{low}]/2$ of the envelope.

Step 3: Calculate the difference $h_1(t) = x(t) - m_1(t)$, if it does not satisfy the two sufficient conditions of the intrinsic mode function (IMF) component, use $h_1(t)$ instead of $x(t)$, repeat step 1 and step 2 until the k may be given to $h_1(k)$ satisfying two conditions.

Step 4: The IMF1 component is $c_1(t) = h_{1k}(t)$, and the remaining component is $r_1(t) = x(t) - c_1(t)$.

Step 5: Repeat the remaining component $r_1(t)$ as the original sequence to decompose, and finally obtain an n IMF component and a residual component $r_n(t)$, where the residual component is a monotonic sequence or a regular value sequence.

Step 6: Finally, the EMD decomposition formula is shown in the formula (x):

$$x(t) = \sum_{i=1}^{n} c_i(t) + r_n(t) \tag{18}$$

However, the conventional EMD algorithm has a poor effect on ash accumulation analysis, and the main problem is modal aliasing, that is, the single imf has the problem of feature coupling.

As a noise-assisted decomposition algorithm, EEMD reduces the problem of mode aliasing. The principle is to use the characteristic of a uniform distribution of a white noise spectrum to add white noise to the signal to be analyzed. In this way, the signals of different time scales can be automatically separated into the corresponding reference scales. However, the signal reconstruction error of such a method is large, and if the decomposition algorithm is added in the prediction, it is inevitable to reconstruct to obtain the final prediction result, so the EEMD algorithm still needs to be improved.

The EEMD algorithm steps are as follows:

1) Add the normally distributed white noise to the original signal.
2) Take the signal with white noise as a whole, and then perform EMD decomposition to obtain each IMF (intrinsic mode function) component.
3) Repeat steps 1 and 2, adding a new normal distribution white noise sequence each time.
4) The IMF obtained each time is integrated and averaged as the final result.

CEEMDAN adds the adaptive white noises on the basis of the EEMD algorithm, which not only reduces the reconstruction error, but also effectively reduces the calculation cost (refer to the introduction to the requirements and decomposition process of EMD, the operation process of CEEMDAN will not be elaborated). In addition, the weight of white noise (δ)

and the number of times of adding white noise (T) need to be determined in advance.[27]

Compared with the general shallow model, the deep learning model shows its accuracy and superiority in time series prediction. However, due to the depth of the deep learning model and many hyper-parameters, it takes too long to train the model. In this case, a hybrid prediction model is formed by combining the decomposition algorithm, sacrificing a long training time to obtain a greater improvement in prediction accuracy, the overall effect may be very low, but fortunately, the decomposition algorithm itself takes less time, which is almost negligible. Therefore, a large number of imfs representing various features is the main problem that model training takes a long time. Taking the research object of this paper as an example, the forward prediction time provided by the multistep-ahead prediction can be broadly understood as the preparation time for the soot blowing operation. In general time series prediction, the model training time is generally not counted, but in practice, if the training time exceeds a certain proportion of the multistep-ahead prediction time, then the preparation time reserved for the soot blower operation may be much less than the theoretical result to be insufficient to complete the soot blower preparation and staffing. In addition, there is a lot of information redundancy among the various imfs of the CEEMDAN algorithm. The method of data dimensionality reduction and reasonable adjustment of dimensionality reduction can not only retain most of the effective information and save the overall training time of the model but also ensure the high efficiency of the model.

2.3.2. Kernel Principal Component Analysis (KPCA). As mentioned above, without the function of dimensionality reduction algorithm, the datasets decomposed by CEEMDAN are successively put into the model for training, which consumes a lot of time and loses its significance in practical problems. In the multiscale modeling prediction by the decomposition algorithm, not all features of the object are required, that is, many features are redundant. Such characteristics not only do not reflect the nature of the object but also cause a lot of unnecessary trouble for subsequent operations.

As a widely used data preprocessing method, dimensionality reduction preserves some of the most important features of high-dimensional data and removes noise and unimportant features, so as to improve the data processing speed. The dimensionality reduction of data can save a lot of time and calculation costs within a certain range of information loss.[28,29]

The main function of the principal component analysis (PCA) algorithm is to reduce the dimensionality of the data. The linear correlation between the data is removed through the diagonalized covariance matrix. The data correlation here is considered as redundant noise; at the same time, the small variance dimension in the diagonal matrix is discarded, and the large variance dimension is retained to achieve data dimensionality reduction. KPCA is one step more than PCA, that is, the dimensionality is increased first (both RBF and polynomial kernel are increased to infinite dimensionality) and then the projection is performed because some non-linearly separable datasets are only linearly separable from the perspective of ascending dimensions.[30]

PCA operation process:

(1) Standardize the original input variable matrix. As shown in formula 19: where X is the standardized matrix, k is the sample length and in this experiment is the length of the ash accumulation time series. n is the number of features.

$$X = \begin{matrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{k1} & x_{k2} & \cdots & x_{kn} \end{matrix}$$

(19)

(2) Find the correlation coefficient matrix of X, that is, the covariance matrix, as formula 20:

$$\sum = \frac{1}{n} X^T X$$

(20)

(3) Calculate the eigenvalue λ of $\sum$, rearrange the order according to the rule from large to small, and calculate the standardized eigenvector.

(4) Finally, the cumulative contribution rate C_i and the actual contribution rate C of all the feature roots are obtained.

$$C_i = \frac{\lambda_i}{\sum_1^k \lambda_i} (i = 1, 2, .., k)$$

(21)

$$C = \frac{\sum_1^p \lambda_i}{\sum_1^k \lambda_i} (p = 1, 2, .., k)$$

(22)

The kernel method is a method of transforming the nonlinear separable problem in low-dimensional space into linear separable problem in high-dimensional space. In detail:

Let χ be the input space (that is, $x_i \in \chi$, χ is a subset or discrete set of R^n), and H is the feature space (**H** is the Hilbert space), if there is a mapping from χ to **H**.

$$\Phi(x): \chi \rightarrow \mathbf{H}$$

(23)

Such that for all $x, z \in \chi$ the function $K(x, z)$ satisfies the condition:

$$K(x, z) = \langle \Phi(x), \Phi(z) \rangle$$

(24)

then we call K the kernel function, where $\Phi(x)$ is the mapping function and $\langle ., . \rangle$ is the inner product.

The kernel inputs two vectors, and it returns the same value as if you took the Φ mapping of each of these vectors and then took the dot product. In addition, commonly used kernel functions generally include linear kernels, polynomial kernels, and Gaussian kernels. The Gaussian kernel function is selected in this article.

KPCA replaces the original n features with a smaller number of m features. Also, it maximizes the sample variance and makes the new m features as uncorrelated as possible. The mapping from old features to new features captures the inherent variability in the data. KPCA reduces high-dimensional features to low-dimensional uncorrelated principal components. In addition, the extracted low-dimensional features also ensure the integrity of the effective information in the original data. KPCA reduces the training time of deep learning model, saves time cost, and improves operational efficiency. High-frequency imfs obtained by the CEEMDAN algorithm are reconstructed by KPCA and become the final input of the deep learning network.

2.4. Adaptive Sliding Window. Time series prediction is the prediction of future development trend through the statistical analysis of the past time series. The sliding window

is generally used to construct the prediction model. Normal sliding window strategy and multistep time series prediction tasks are as follows (one-step-ahead).

$$V_t = (CF_t, \ CF_{t-1}, \ CF_{t-2} \dots CF_{t-d+1}) \tag{25}$$

$$\overline{CF_{t+1}} = f(CF_t, \ CF_{t-1}, \ CF_{t-2} \dots CF_{t-d+1}) \tag{26}$$

Assuming that t represents time, d represents the length of the sliding window, and CF_t represents the ash accumulation on the heating surface corresponding to the time t. CF_{t+1} represents the predicted future dust accumulation situation at $t + 1$. The vector V_t was constructed according to the corresponding time relation to represent the heated surface pollution degree at the past. In addition, the input–output mapping relationship f represents the constructed deep learning model.

As the new CF data is updated, the window is constantly shifted back by a fixed unit to be updated. Figure 3 shows a
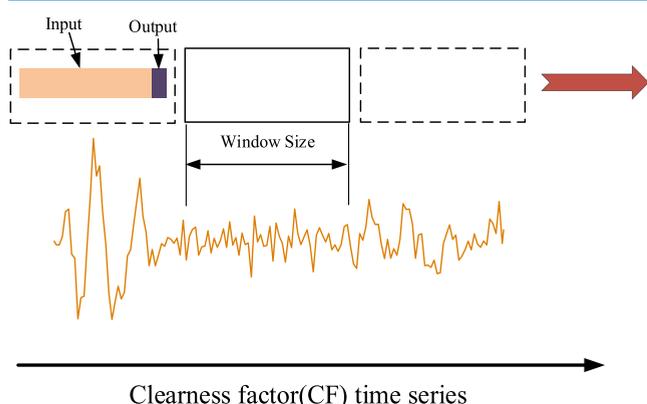
Figure 3. Time-based sliding window.

specific graphical representation of a sliding window. The sliding window contains $d + 1$ data points, among which the first d is used to build the deep learning model (when it is a single-step prediction). The multistep-ahead prediction has a similar principle to one-step-ahead prediction.

Although this method can deeply excavate the degradation and oscillation state of the time series in the ash accumulation period, since the predictive maintenance of heating surface requires the high precision ash accumulation prediction as to the support, a sliding window with an adaptive width is inserted into the whole prediction algorithm framework. As the sliding window moves forward, the length of the window is recalculated, depending on how the data in the adjacent window changes. Compared with the fixed window method, the advantage of this algorithm is that when the window width is small, the deep learning model trained by narrow window data can easily capture the mutation of CF, and the wide window can more easily cover the degradation trend of the health condition of the whole heating surface. The size of the window depends on the recent changes in the health condition of the heating surface. When the health condition changes significantly, the window size will shrink sharply, and vice versa. To illustrate the effectiveness of sliding s, the following strategies for adaptive window adjustment are presented.

$$VS_i = \left| \frac{var(Z_i) - var(Z_{i-1})}{max(Z) - min(Z)} \right| \tag{27}$$

$$DS_i = \left| \frac{std(Z_i) - std(Z_{i-1})}{max(Z) - min(Z)} \right| \tag{28}$$

$$Dif_i = max\left(\frac{|DS_i - DS_i|}{DS_i} \times \frac{|VS_i - VS_{i-1}|}{VS_i} \right) \tag{29}$$

VS_i and DS_i represent the mean fluctuation and difference fluctuation of the data distribution of the ith ($i > 1$, which is an integer) window. Z_i represents the data sequence of the ith window, and Z represents all the data sequences required to calculate the size of the new window this time. Var and std, respectively, represent the variance and standard deviation of the calculated sequence. Finally, the variable Dif_i is defined to characterize recent data changes.

The main idea of this method is to slice data segment to obtain multiple local informational pieces. The adaptive sliding window updating strategy determines the width of the new slidng window based on the distribution of previous windows, which is a strategy for adjusting for local distribution differences between data slices. According to formula X, the window width can be shrunk or enlarged by setting a reasonable threshold during operation. When the calculated Dif_i is smaller than the threshold, it is considered that the distribution difference of the nearest window data is small, and the window width should be expanded to improve the training and prediction speed. If it is bigger than the threshold value, it indicates that the recent data segment has entered the oscillating region, and the window width should be reduced so that the deep learning model can better remember these abrupt situations. Compared with the fixed-size sliding window, this strategy improves the detection accuracy of the mutation and operation efficiency and enables the model to remember the overall deterioration trend and local mutation status of the CF more quickly.

2.5. Encoder–Decoder Based on Attention Mechanism (EDA).

$$h_t = \tanh(w_{xh}x_t + w_{hh}h_{t-1} + b) \tag{30}$$

$$y_t = f(w_{hy}h_t + c) \tag{31}$$

As the earliest form of neural network, the recurrent neural network (RNN) is generally composed of a recursive architecture, and the hidden state of each time step depends on the previous input. This characteristic gives it a great advantage in processing serialized data compared with other neural networks. Mathematically, given a time series $X(t)$, the hidden state h_t and output y_t can be updated as follows:

The problems of gradient vanishing and gradient explosion[31] (due to the chain rule of derivatives and the use of nonlinear functions) make it difficult for the input with a long distance to establish an effective connection when adjusting parameters in the reverse error propagation. Therefore, there are challenges in capturing the long-term dependence of time series. Different from the simple recursive method, Long Short-Term Memory (LSTM) cell on the basis of the RNN can selectively memorize and forget information through the gate mechanism composed of an input gate, output gate, and forget gate to further avoid the problems of gradient disappearance and gradient explosion. This dynamic learning method makes it easy to remember even the early useful information.

$$i_t = \delta(w_i \cdot [h_{t-1}, \ x_t] + b_i) \tag{32}$$
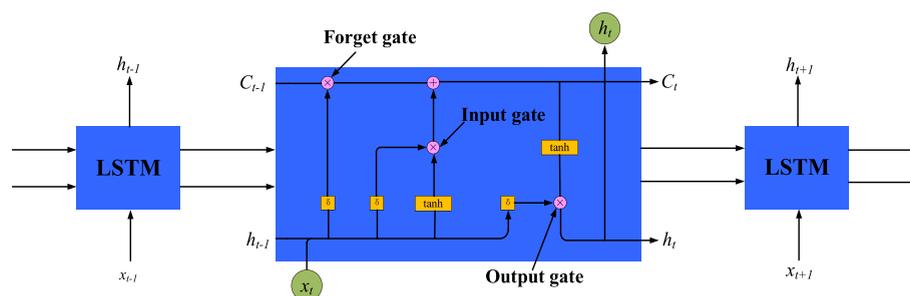
Figure 4. LSTM structure.
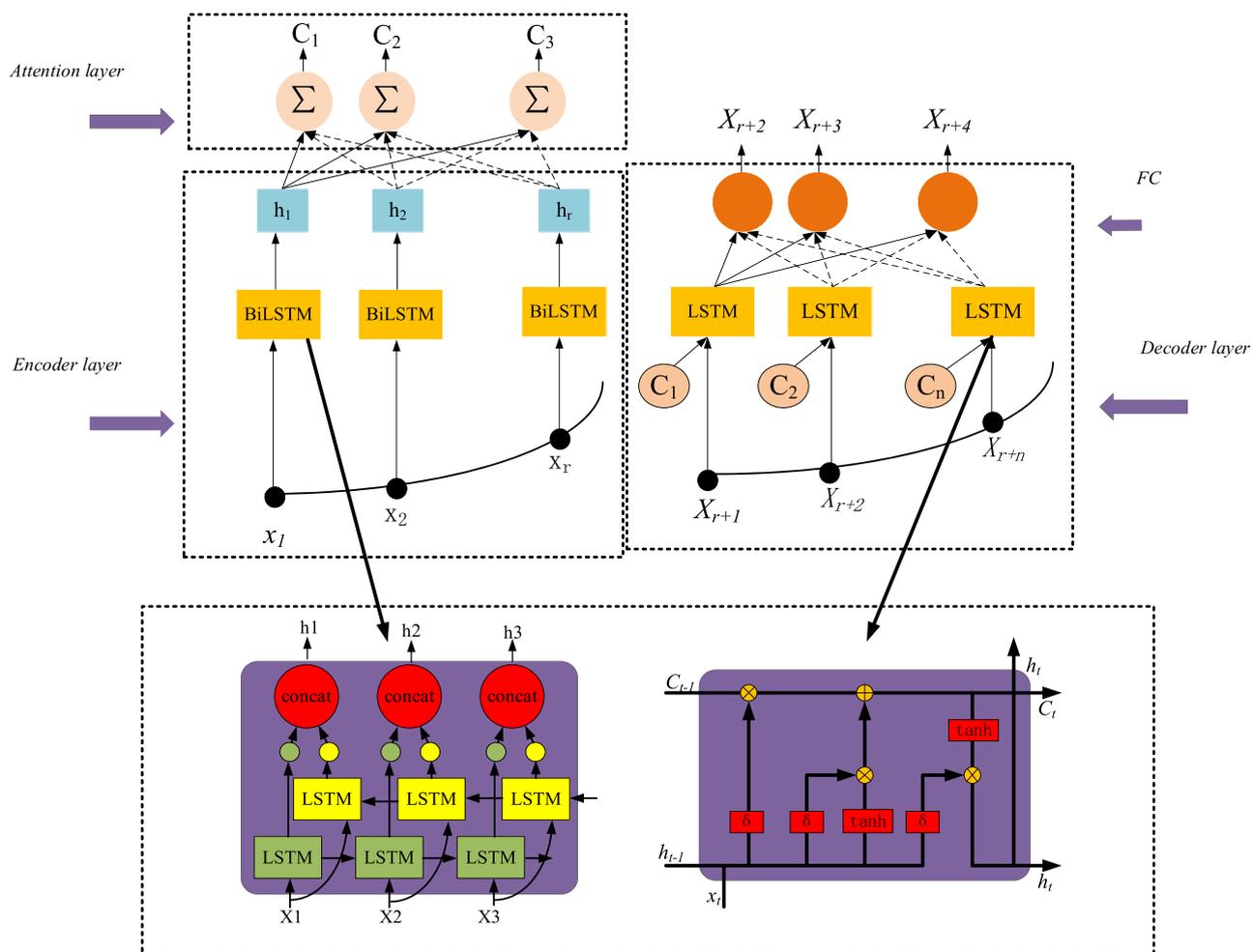
Figure 5. Encoder−decoder based on attention structure.

$$f_t = \delta(w_f \cdot [h_{t-1}, x_t] + b_f) \tag{33}$$

$$o_t = \delta(w_o \cdot [h_{t-1}, x_t] + b_o) \tag{34}$$

$$h_t = o_t \times \tanh(c_t) \tag{35}$$

$$c_t = f_t \times c_{t-1} + i_t \times \tanh(w_c \cdot [h_{t-1}, x_t] + b_c) \tag{36}$$

f_t, i_t, and o_t are the output vectors of the three gates, which are mainly calculated from the input x_t at the current moment and the hidden state h_{t-1}. Sigmad and tanh are used as the activation functions of the gate mechanism and the output activation functions of the LSTM cell C_t, respectively. Sigmoid and hyperbolic tangent functions are used to realize the non-linearization of LSTM. w_i, w_f, w_o, w_c and b_i, b_f, b_o, b_c are

respectively used as weight matrix and bias vectors, which can be updated by the error back propagation algorithm during training. From the internal structure of the LSTM (see Figure 4), it can be seen that the status of the old internal cell state C_{t-1} of the LSTM is mainly updated through the forget gate and the input gate. The new cell state C_t has two main functions: one is to complete the self-renewal with new input and hidden state, so as to further complete the long-distance transmission of information and long-term memory. Second, the information flow is outputted to complete the update of the hidden state h_t, and finally the output y_t of the current moment LSTM is established.

Bidirectional Long Short-Term Memory (BILSTM) contains LSTM networks in both positive and negative directions.[32]
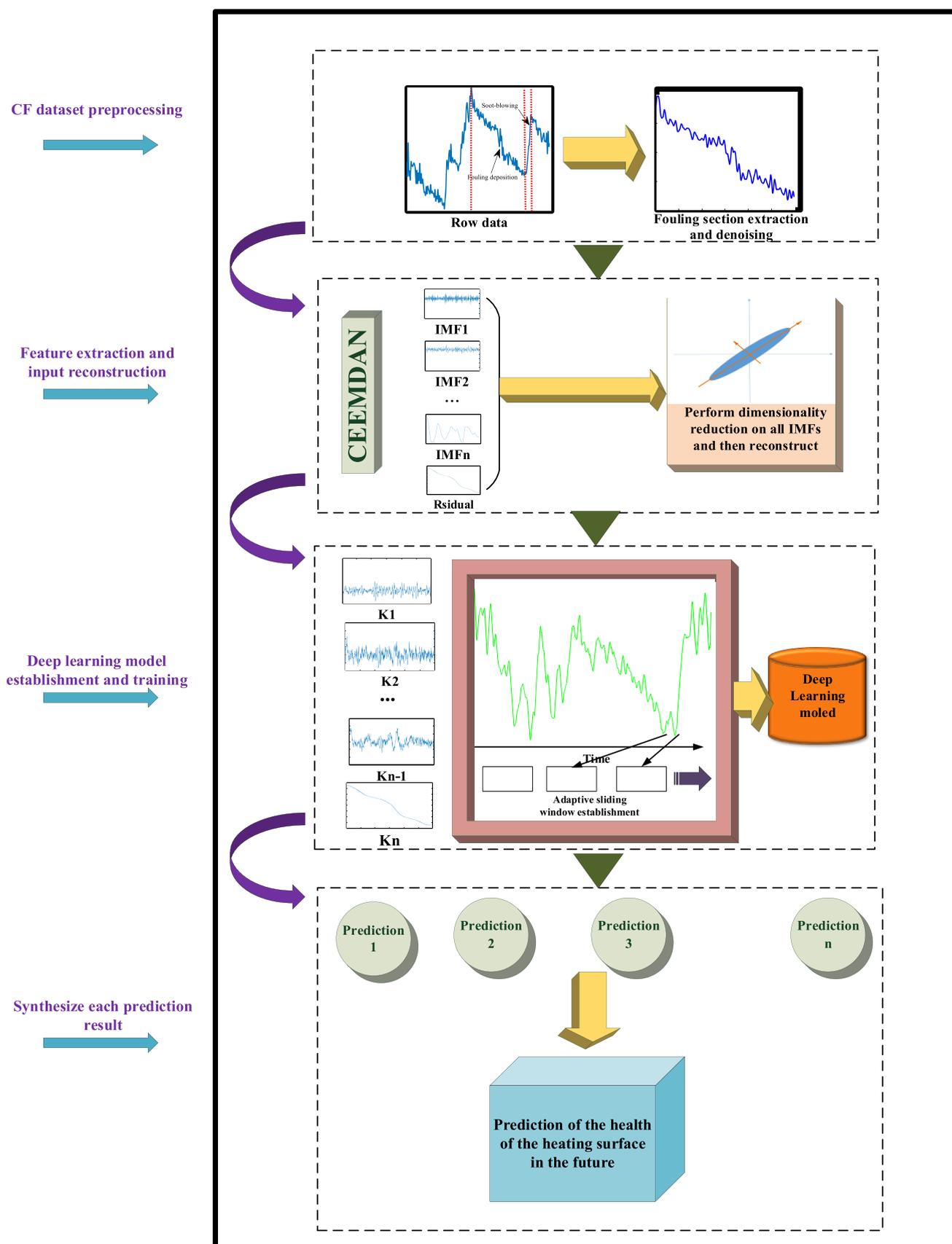
Figure 6. Theoretical framework of ash accumulation prediction.

When input information is available, BILSTM can receive sequences from both forward and reverse directions for learning, so that more characteristic hidden conditions can be obtained and more complete time series feature mining can be completed.
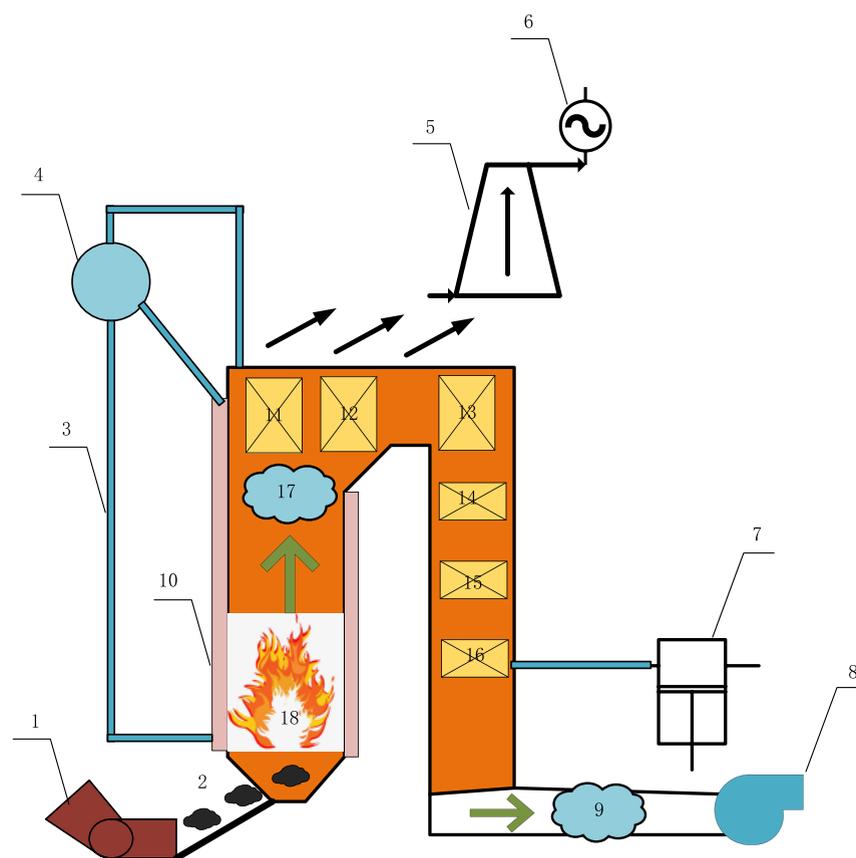
Figure 7. Boiler schematic: (1) pulverizers, (2) coal powder, (3) downcomer, (4) steam drum, (5) turbine, (6) generator, (7) air preheater, (8) supply air fan, (9) high-temperature flue gas, (10) water wall, (11) platen superheater, (12) high-temperature superheater, (13) high-temperature reheater, (14) low-temperature superheater, (15) low-temperature reheater, (16) economizer, (17) low-temperature flue gas, (18) furnace combustion.

The single-direction learning is the same as regular LSTM, but the final hidden layer output is a linear superposition of two hidden layer outputs in opposite directions. Its mathematical expression is as follows:

$$\overleftarrow{i_t} = \delta(\overleftarrow{w_i} \cdot [h\overleftarrow{_{t-1}}, x\overleftarrow{_t}] + b\overleftarrow{_i}) \tag{37}$$

$$f\overleftarrow{_t} = \delta(w\overleftarrow{_f} \cdot [h\overleftarrow{_{t-1}}, x\overleftarrow{_t}] + b\overleftarrow{_f}) \tag{38}$$

$$o\overleftarrow{_t} = \delta(w\overleftarrow{_o} \cdot [h\overleftarrow{_{t-1}}, x\overleftarrow{_t}] + b\overleftarrow{_f}) \tag{39}$$

$$h\overleftarrow{_t} = o\overleftarrow{_t} \times \tanh(c\overleftarrow{_t}) \tag{40}$$

$$c\overleftarrow{_t} = f\overleftarrow{_t} \times c\overleftarrow{_{t-1}} + i\overleftarrow{_t} \times \tanh(w\overleftarrow{_c} \cdot [h\overleftarrow{_{t-1}}, x\overleftarrow{_t}] + b\overleftarrow{}) \tag{41}$$

$$h_t = \vec{h_t} \odot h\overleftarrow{_t} \tag{42}$$

In fact, due to the inherent shortcoming of the RNN structure for long sequence processing and the fact that a large amount of input information is only represented by a fixed-length vector B_i, which may lead to the loss of information, the actual use has great limitations. The researchers then developed the attention mechanism by providing an intuitive interpretation of the human visual mechanism. As an intuitive explanation of the human visual mechanism, it allows the decoder to directly access all the hidden output of the encoder when generating each time-step output. Furthermore, this article introduces the attention module to the encoder−decoder network structure to complete

the hidden state of the automatic learning encoder and decoder of hidden state correlation to calculate attention weights. Finally, all the hidden layer outputs of the encoder are weighted by the calculated attention weights to complete the final representation vector B_i and make it participate in the output of the decoder. It can be seen that the attention module will produce an attention representation vector B_i, which is obtained by the weighted sum of the hidden states of the decoder and all the encoders at the last moment before the decoder obtains the output of each step. This is also the essence of attention operation.[33] The Encoder−Decoder based on Attention (EDA) structure is shown in Figure 5.

$$B_i = \sum_{k=1}^{T} a_{ij} h_{ij} \tag{43}$$

$$a_{ij} = \frac{\exp(\beta(s_{j-1}, h_i))}{\sum_{k=1}^{T} \exp(\beta(s_{j-1}, h_i))} \tag{44}$$

where a_{ij} is the attention vector, h_{ij} is the attention weight, B_i is the final result of the attention mechanism after the weighting operation, β is a correlation operator (such as dot multiplication operation), and s_j is the output of the hidden layer of the decoder at time j.

2.6. CF Prediction Based on the Hybrid Model. Based on the above models and algorithms, we proposed a hybrid model based on deep feature extraction and deep learning model for multistep prediction and predictive maintenance of heated surface health conditions. The overall detailed prediction

process framework is shown as follows (see Figure 6), which is mainly divided into four parts. (1) Based on the establishment of health factors reflecting the ash accumulation condition of the heated surface, the construction of the datasets on the change of health condition of the heated surface throughout a day was completed. After that, the ash accumulation segment for various heated surface datasets was extracted and denoised to complete the data preprocessing operation. (2) In the part of feature extraction and input reconstruction of the deep learning model, CEEMDAN, an improved model of EMD, was adopted to complete multiscale analysis of the dust accumulation segment after denoising, and it was decomposed into the overall deterioration component and several high-frequency components. In addition, the KPCA algorithm was used to complete the input reconstruction in order to solve the problems of feature redundancy after decomposition and operation efficiency in the deep learning model. (3) We improved the shortcomings of the traditional sliding window, such as low efficiency and poor ability in learning the mutation of the *CF* value, and then proposed the adaptive sliding window method, which combined with the deep learning prediction model of the encoder−decoder model based on the attention mechanism to complete the accurate prediction of each component of the reconstructed input. (4) We integrated all the prediction results to complete the final heating surface health condition prediction task.

3. EXPERIMENT VERIFICATION

3.1. Dataset Description and *CF* Data Smoothing. The dataset used in this paper to verify the performance of the proposed model comes from a 300 MW coal-fired boiler in a thermal power station in Guizhou, China, where the schematic diagram of the boiler is shown in Figure 7. The main design parameters of the boiler are shown in Table 1. The boiler type is

Table 1. Boiler Internal Parameters

parameter	unit	value of number
rated condition	MW	300
fuel flow	kg/s	35.4
rated evaporation	t/h	909.6
rated main steam pressure	MPa	17.25
rated main steam temperature	°C	540
reheat steam flow	t/h	732.2
reheat steam pressure	MPa	3.18
reheat steam temperature	°C	540
feed water temperature	°C	278
air volume	kg/s	295

HG-1025/17.3-WM18. The boiler features subcritical, natural circulation, intermediate reheating, double arch single furnace, "W" flame combustion method, dual flue at the tail, and flue gas baffle temperature adjustment, balanced ventilation, etc.

This article selects three types of heat-receiving surface datasets of boiler components: economizer, low-temperature superheater, and reheater. Each dataset uses the clearness factor as a health indicator and records the ash on the heated surface of the boiler for a day (under the same working conditions). In addition, they are in the same working conditions. It is necessary to denoise and smooth the clean factor dataset obtained from the DCS online monitoring data because a large amount of noise and burrs increase unnecessary prediction difficulty and damage the stability and accuracy of the prediction results. The abscissa

is time, the unit is hours, and the ordinate is *CF* reflecting the health status of the heated surface.

The *CF* curve of the economizer before denoising and its corresponding load are shown in Figure 9a. The *CF* curve obtained by combining the DCS online monitoring data and the thermodynamic model has strong nonlinearity. There are two general reasons: random noise caused by the normal operation of the economizer and the worksite. Such noise can be eliminated by a reasonable denoising algorithm. However, the 'noise' caused by normal physical phenomena inside the economizer is worthy of our attention. These ″noises″ are inevitable and cannot be ignored in the entire forecasting process. In order to understand this non-negligible noise, we conduct a detailed analysis: When the flue gas passes through the convective heated surface, the ash in the flue gas will be deposited on the heated surface, resulting in a decrease in heat transfer efficiency, and the passing of the flue gas will take away part of the ash on the heated surface, resulting in an increase in heat transfer efficiency. In addition, the flow rate of the flue gas will also greatly affect the degree of fouling on the heated surface.

It is worth noting that S1 is not an effective soot blowing point, while S2 is (the descending section before S2 is an effective soot accumulation section, and the ascending section after S2 is an effective soot blowing section). This is due to the surge in the boiler load during this time period, resulting in an effect similar to soot blowing. The dust accumulation section used in this paper to verify the proposed model is D1 because D1 is in a stable load state, and *CF* has a more obvious trend of change (large load changes will not reflect normal ash accumulation changes). The analysis of other heated surfaces is basically similar to the economizer heating surface. In order to ensure a better denoising effect, this paper adopts the wavelet threshold denoising method. The Daubechies wavelet is used as the basis function, wavelet order is designated as 4, and the soft threshold is used to quantify the wavelet coefficients. Figure 8 shows the denoising results under 5, 6, and 7 wavelet decomposition layers respectively. The denoised signal under 5 still retains more noise, while signal 7 has filtered more effective signals, and the denoising signal with a decomposition level of 6 is finally used as the result of data preprocessing. Further, similar to the above discussion, Figure 9b,c shows the all-day *CF* datasets of the economizer, low-temperature superheater, and reheater after denoising and extracting only the ash accumulation section. It is more obvious from the figure that the *CF* dataset after denoising still has strong nonlinearity and non-stationarity and can be regarded as a multifeature fusion signal, so even if advanced algorithms are used, it is difficult to obtain key information through direct prediction and adapt to multiple features at the same time.

The datasets given above is for the heated surface of multiple pieces of equipment under the same working conditions, but in fact, the working conditions of the boiler may be different. Figure 10 shows the ash accumulation dataset of 20 sets of economizers (all under stable load), which belong to the complete ash accumulation dataset of the economizer from health to complete failure under the same working conditions. Similarly, the same preprocessing operation is also required.

3.2. Evaluation Index. In this paper, in order to visually represent the model performance, appropriate evaluation indicators are needed to verify the prediction performance, and the overall evaluation indicators RMSE and MAPE, MAE are introduced. These evaluation indicators are widely used to measure the accuracy of the results of classification and
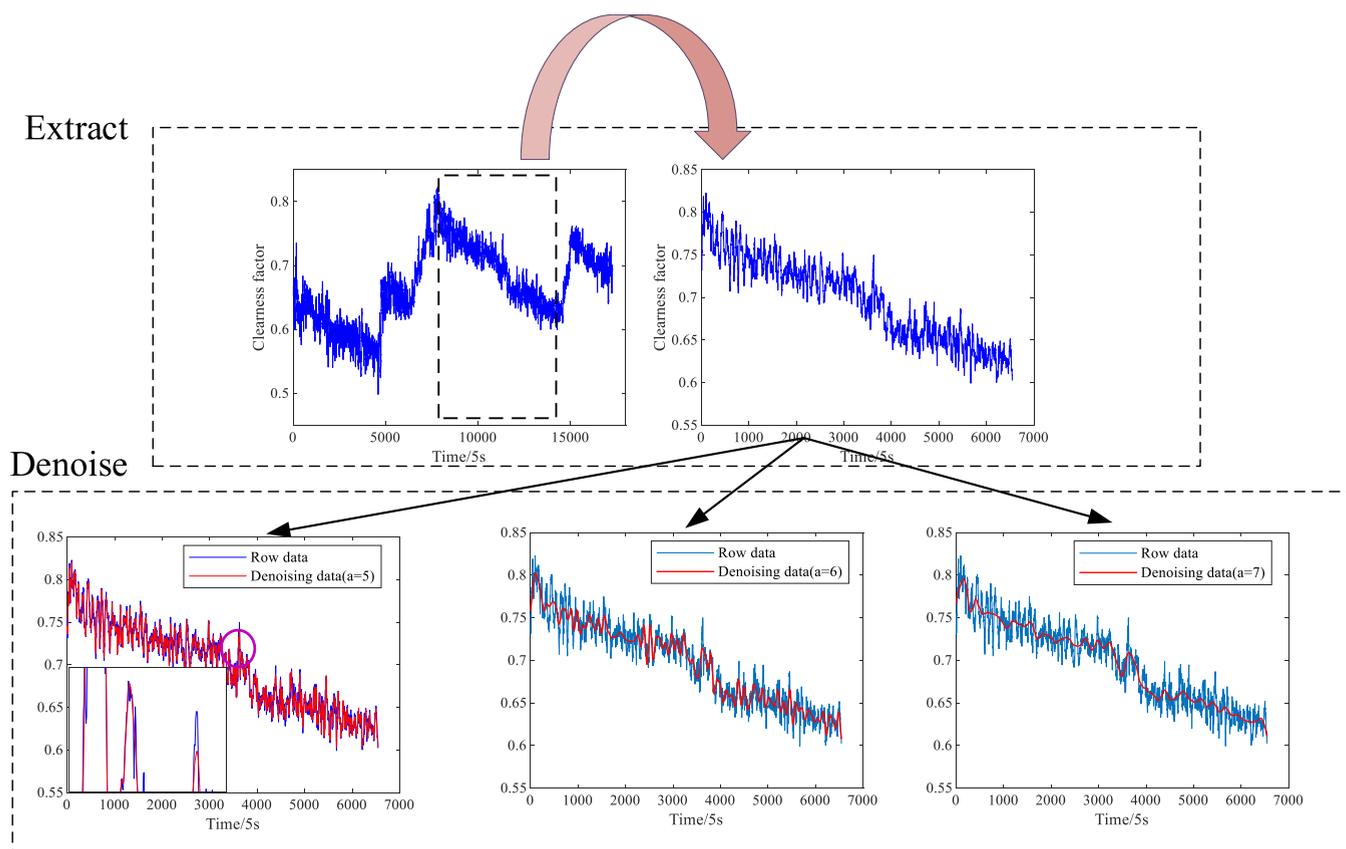
Figure 8. Wavelet threshold denoising results of economizer datasets under different decomposition levels.
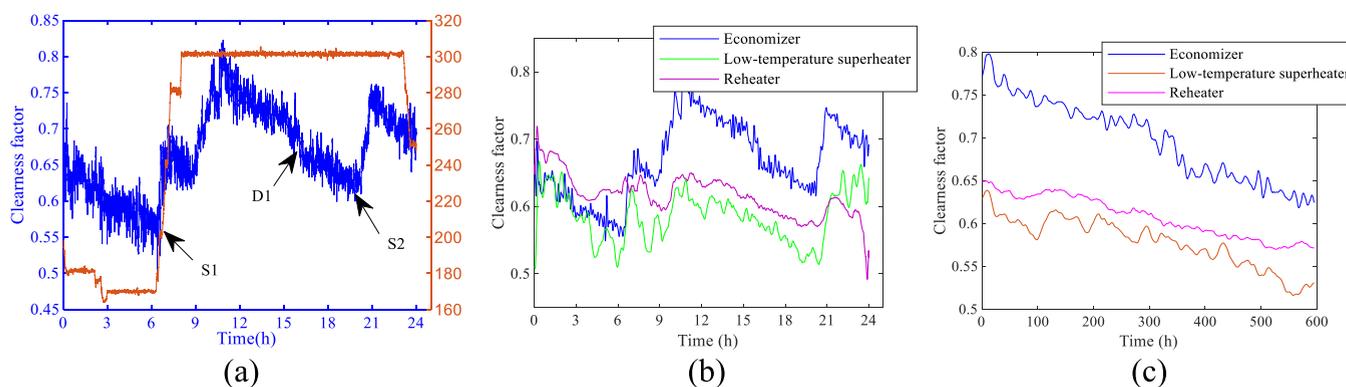
Figure 9. (a) CF data and load of the original economizer, (b) economizer, low-temperature superheater, reheater after denoising, and (c) extract only the ash section.
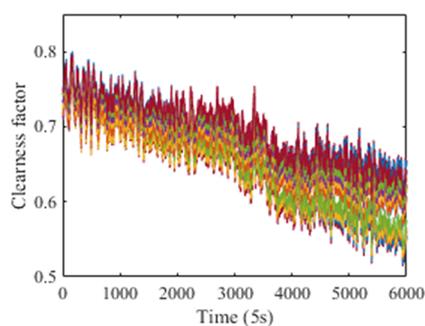
Figure 10. Variable working condition economizer dataset (20 groups).

regression algorithms. The specific mathematical expression is shown in formulas eq 45−47.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(N_{true_{-i}} - N_{predicted\_i}\right)^2} \tag{45}$$

$$\text{MAPE} = \sum_{i=1}^{n}\left|\frac{N_{true_{-i}} - N_{predicted\_i}}{N_{true_{-i}}}\right| \times \frac{100}{n} \tag{46}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|N_{true_{-i}} - N_{predicted\_i}| \tag{47}$$

where $N_{true_{-i}}$ and $N_{predicted\_i}$, respectively, represent the true value and predicted value of the CF at the ith moment.

3.3. Implementation Details. In this article, the superiority of the proposed model will be reflected in comparison with many commonly used models and variant models of the proposed model. The prediction model introduced below will be used in the comparative experiment of this paper: proposed model (M1), EDA deep learning model without adaptive sliding window (M2), replacement of the EDA deep learning model with LSTM (M3), and LSTM model without adaptive sliding window (M4) (see Table 2). After obtaining multiple sets of

Table 2. Model Details

model	model details
M1	the proposed model
M2	without adaptive sliding window
M3	replace EDA with LSTM
M4	without adaptive sliding window and replace EDA with LSTM

samples from the historical heating surface cleaning factor data after feature extraction through an adaptive sliding window, various parameter configurations of the deep learning model are necessary, which is related to the final prediction performance. For deep learning models built on the basis of recurrent neural networks, time-based backpropagation methods are generally used to correct parameters. In addition, MSE is used as a loss function to measure the difference between the predicted value and the actual value, and the Adam optimizer is used to make it approach and minimize such differences. Epoch and batch size and the internal parameters of the EDA model need to be properly configured to avoid under-fitting and over-fitting of the prediction model. The learning rate step size is an important hyper-parameter for supervised learning, and its reasonable setting ensures that the network can be quickly and correctly find the optimal solution. Finally, the initialization parameters for the adaptive sliding window should also be taken into consideration. We give the final values of the parameters in Tables 3 and 4.

Table 3. Hyperparameter Settings of the Experimental Model

parameter	value
encoder hidden layer number	1
decoder hidden layer number	1
bidirectional LSTM merge mode	Sum
activation function of the attention layer	Softmax
encoder neurons	150
decoder neurons	100
loss function	MSE
optimizer	Adam
epoch	100
batch Size	20

Table 4. Adaptive Sliding Window Parameters

parameter	value
maximum window length	30
minimum window length	2
update window length each time	1
w traverse length	1
initial window size of the training set	15
initial window size of the testing set	15

3.4. Ash Feature Extraction and Model Input Establishment. *3.4.1. CEEMDAN Result Analysis of CF Data.* In order to extract the deep abstract features hidden in the ash accumulation section, Figures 11 and 12 show the result of the economizer dataset after data smoothing after being decomposed by EMD and CEEMDAN. A series of imfs and a residual are obtained. It can be seen from the figure that the residual gives a better indication of the overall deterioration trend of the fouling section of the economizer under steady load. It is worth noting that the number of components obtained by EMD decomposition is small, which is caused by the problem of modal aliasing.

Imfs with high-frequency characteristics represent the non-stationary and non-linear part of the fouling, and each decomposed mode is called imf_i ($i = 1, 2, 3, ...$). The frequency decreases from top to bottom in the figure. The reason for the multiscale analysis of the ash accumulation time series is that the direct use of the deep learning model for prediction may not be able to adapt to all frequency features at the same time, and the accuracy, stability, and robustness of the prediction will be poor. Through experiments, there are 9 decomposition components for both the low-temperature superheater and the reheater. Before the experiment, two parameters need to be determined: the noise weight (δ) and the number of times of adding noise (T), and we finally set them to 100 and 0.05 after many experiments.

3.4.2. Deep Feature Extraction and Input Reconstruction. KPCA first raises the dimensionality of the original features through the kernel function and then reduces the dimensionality according to the maximization of variance. This method extracts deep abstract features from imfs and converts high-dimensional related features into low-dimensional irrelevant features. All principal components can almost cover all the effective information of all original features, ensuring the completeness and validity of the input data for the next stage of predicting. In addition, KPCA, as a dimensionality reduction algorithm, greatly alleviates the training time consumption problem of the deep learning model, so that the training time is greatly reduced in the forward prediction time.

In this paper, we only perform dimensionality reduction operations on CEEMDAN components except for residual. Figure 13 shows the relationship between the number of reconstructed features of the KPCA and the loss of ash information. When the reconstructed input increases to a certain amount, the loss of the amount of ash information caused by the reconstruction of KPCA is almost no longer reduced. In other words, continuing to increase the number of reconstructed inputs will only increase the computational load and cause unnecessary time consumption. Considering the prediction performance of the model, the economizer, low-temperature superheater, and reheater time series groups obtained by CEEMDAN are reconstructed into 4, 4, and 3 sub-sequences, respectively. In this paper, the optimal number of dimensionality reduction layers of KPCA is selected to select the minimum information loss, and the cumulative contribution rate of economizer, low temperature superheater, and reheater can reach 95.3%, 97.6%, and 95.1%.

3.5. Deep Learning Model Establishment and Online Prediction. *3.5.1. Multistep-Ahead Prediction of Ash Deposition on Heating Surfaces.* In this section, the performance of the short-term prediction (five-step-ahead prediction) of the proposed model will first be verified under the CF datasets of the three heated surfaces. In the fouling time series after deep
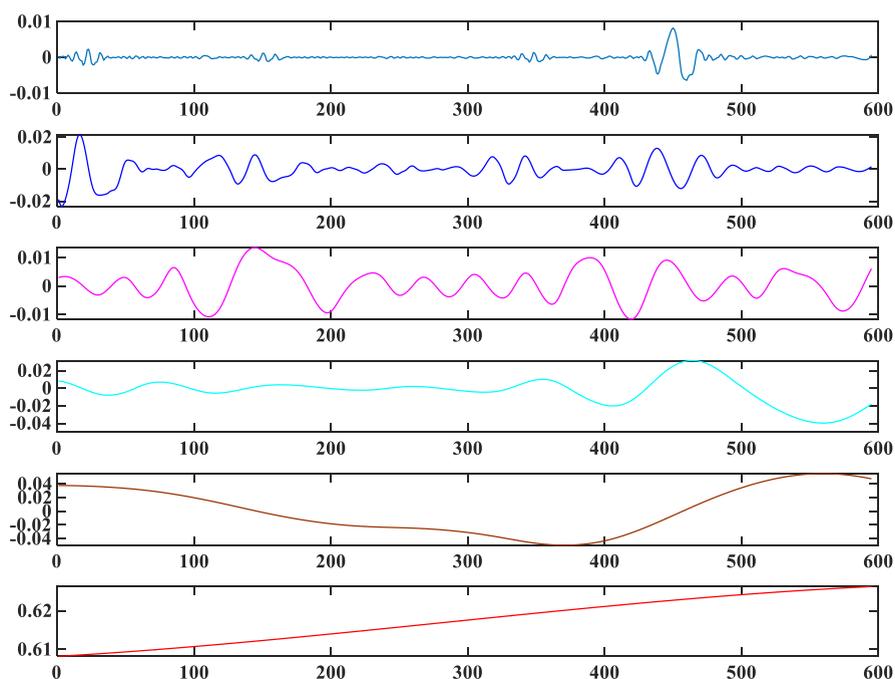
Figure 11. EMD decomposition of economizer fouling dataset.
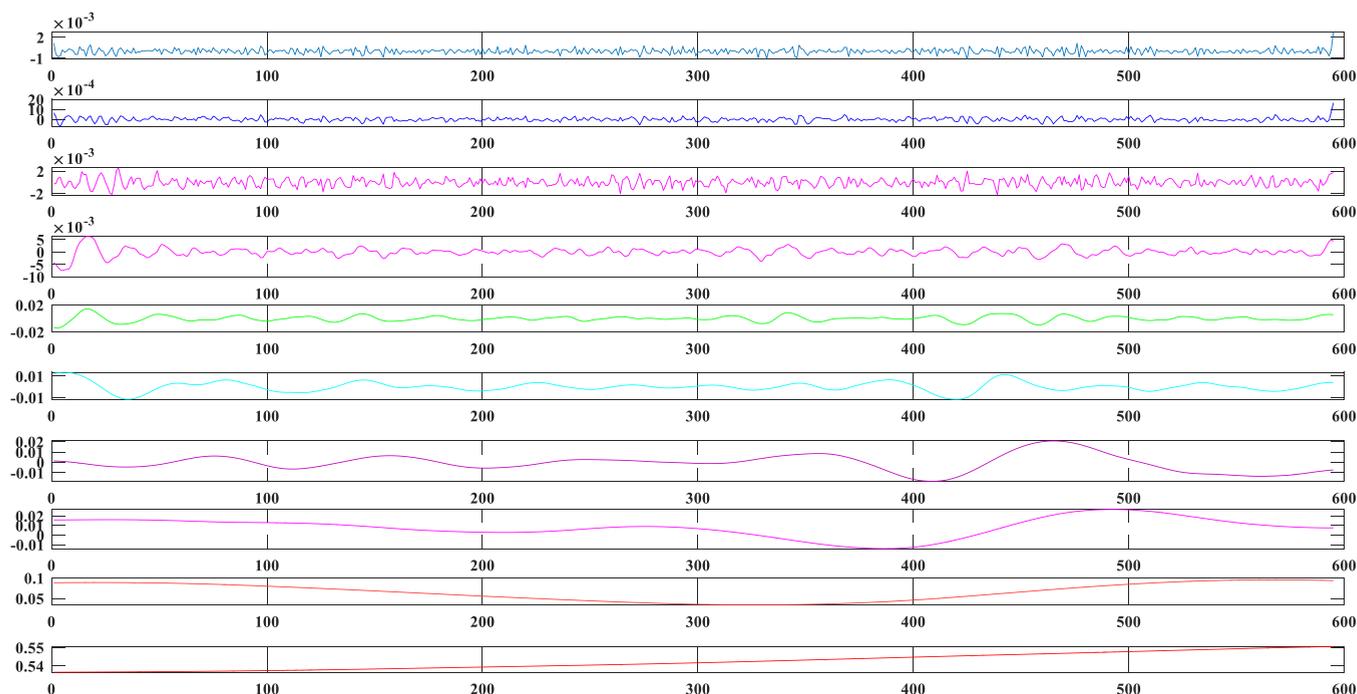
Figure 12. CEEMAD decomposition of economizer fouling dataset.

feature extraction, the magnitude of each component is quite different. We will perform maximum and minimum standardization processing on each subcomponent that has undergone input reconstruction operations to improve model efficiency and convergence speed. After initializing the maximum and minimum widths, thresholds, and adjustments of the sliding window, *CF* data is sent to the adaptive sliding window to obtain multiple time series sample sets to complete the deep learning model training on historical data. Finally, the adaptive sliding window is initialized again with the same sliding window parameter configuration to complete the short-term

prediction of ash of the heated area. Figures 14— are the short-term prediction results of the proposed model and the comparison model on the three heating surfaces. In the comparison, the prediction of M1 model has the best prediction accuracy and can almost reproduce the original volatility and overall deterioration trend. In contrast, M3 has a slightly inferior effect but still shows a better volatility prediction ability under the action of an adaptive sliding window. The corresponding prediction errors are presented in Figures 17—. M1 has the smallest RMSE of 0.0085, 0.00400, and 0.001960 under three heated surface datasets, and it is significantly lower than the
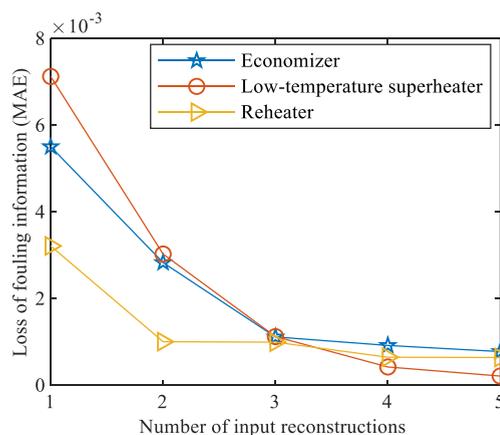
Figure 13. Amount of information loss in different reconstruction input variables.
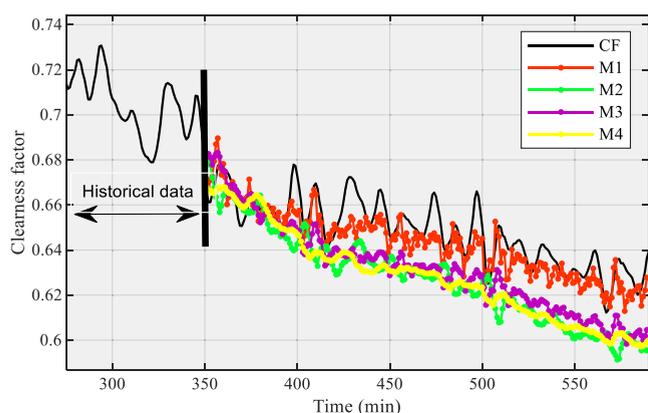
Figure 14. Five-step-ahead prediction of M1−M4 on the economizer.

other three models. MAPE and MAE also have a similar situation for M1.

In addition, the prediction result of M1 is still close to the true value near the end of the life of the heated surface. In fact, predicting the performance in the late stage is more important than the early stage because the failure threshold (soot blowing threshold) is often distributed in the last 10 to 20% of the overall degradation curve, which is the key time period for making predictive soot blowing decisions, and M1 can meet this requirement in the short-term prediction.

In order to further explore the superiority of the proposed method for the adaptive sliding window and the EDA deep learning model, we give the short-term prediction results of the low-temperature superheater in Figure 15. In the top line of the figure, we only change the use of the adaptive sliding window, while the bottom line controls the selection of the deep learning model.

When comparing Figure 15a,b, without the adaptive sliding window, the prediction curve has more spikes and burrs. Comparing Figure 15c,d, EDA has better stability and ability to track the mutation of the heated surface degradation curve than the LSTM. Therefore, the EDA framework and adaptive sliding window are of great significance in ash prediction of the heated area of the heat exchanger. Similarly, the short-term prediction result of the proposed model in the reheater also highlights its superiority as shown in Figure 16.

In fact, the forward prediction time provided by the short-term forecast (five-step-ahead prediction) often cannot meet the

complex equipment configuration and personnel arrangement of the soot blowing operation. A larger number of prediction steps can be improved to effectively solve such problems. In general, the prediction effect should be similar to the short-term prediction, but the accumulation of errors in the long-term prediction causes the prediction accuracy to decrease as the number of forwarding steps increases.

Figures 20− show the ten-step-ahead prediction. The fouling prediction in the future is more deviated from the true value than the short-term prediction, but the proposed model still achieves the best prediction effect on different heat exchanger heated surfaces. Tables 5− also reflect this point in three evaluation indicators. Without adding an adaptive sliding window, the prediction result of the low-temperature superheater still has a lot of glitch noise, which is similar to the prediction performance reflected in the five-step-ahead prediction. In addition, from the economizer error table (see Table 5), the prediction error of M4 seems to be smaller than that of M3, but M4 is already relatively poor in predicting the non-linear part of the ash accumulation degradation curve. However, the function of the deep feature extraction module based on multiscale analysis can still maintain a good overall prediction effect. There is a similar situation under the reheater dataset.

When the number of forwarding prediction steps is increased to twenty-five steps (Figures 23−), each model has a large deviation from the true value in the dataset except for the low-temperature superheater, especially M3 and M4 under the economizer data set and M2 and M4 under the reheater have greatly deviated from reality. In detail, for the twenty-five-step-ahead predictions, the RMSE of the adaptive sliding window and the EDA deep learning model in M1 are improved by 44.2% and 60.8% in the economizer, 26.54% and 14.54% in the low-temperature superheater, 7.4% and 18.0% in the reheater, respectively. Similarly, compared to M2, the MAPE of M1 increased by 94%, 26.3%, and 8.1%, respectively, and compared with M3, increased by 64.5%, 14.2%, and 25.6%. M1 also has the smallest value under MAE. For the twenty-five-step-ahead prediction, the RMSE of M1 under the economizer is increased by 12.9%, 60.4%, and 88.6%, respectively, compared with the others. For the other two heated surface CF datasets, the situation is similar in the three evaluation indicators. Time consumption under the model under twenty-five-step predic-tion: (1) proposed model: 4 min 25 s; (2) model without deep feature extraction: 5 min 58 s; (3) model without sliding adaptive window: 4 min 8 s; (3) to replace EDA with LSTM: 3 min 49 s. It can be seen that the deep feature extraction model can significantly reduce the training time of deep learning and can obtain good prediction results.

In summary, it is shown that the combination of deep feature extraction and adaptive sliding window is evaluated by RMSE, MAPE, and MAE for the effectiveness of the proposed hybrid method in multistep ahead prediction.

In addition, the deep feature extraction method of combining CEEMDAN and KPCA is further discussed. Figures 26 and 27 shows the prediction results of the non-depth feature extraction algorithm under different forward prediction steps from the economizer and the low-temperature superheater. It can be seen that without the effect of multiscale analysis and dimensionality reduction, due to the characteristics of short-term prediction (five-step-ahead prediction), it can still give satisfactory results under the two datasets. (But for the economizer dataset, the non-linear and non-stationary part of its degradation curve can hardly be responded to, but fortunately, it can still get the basic
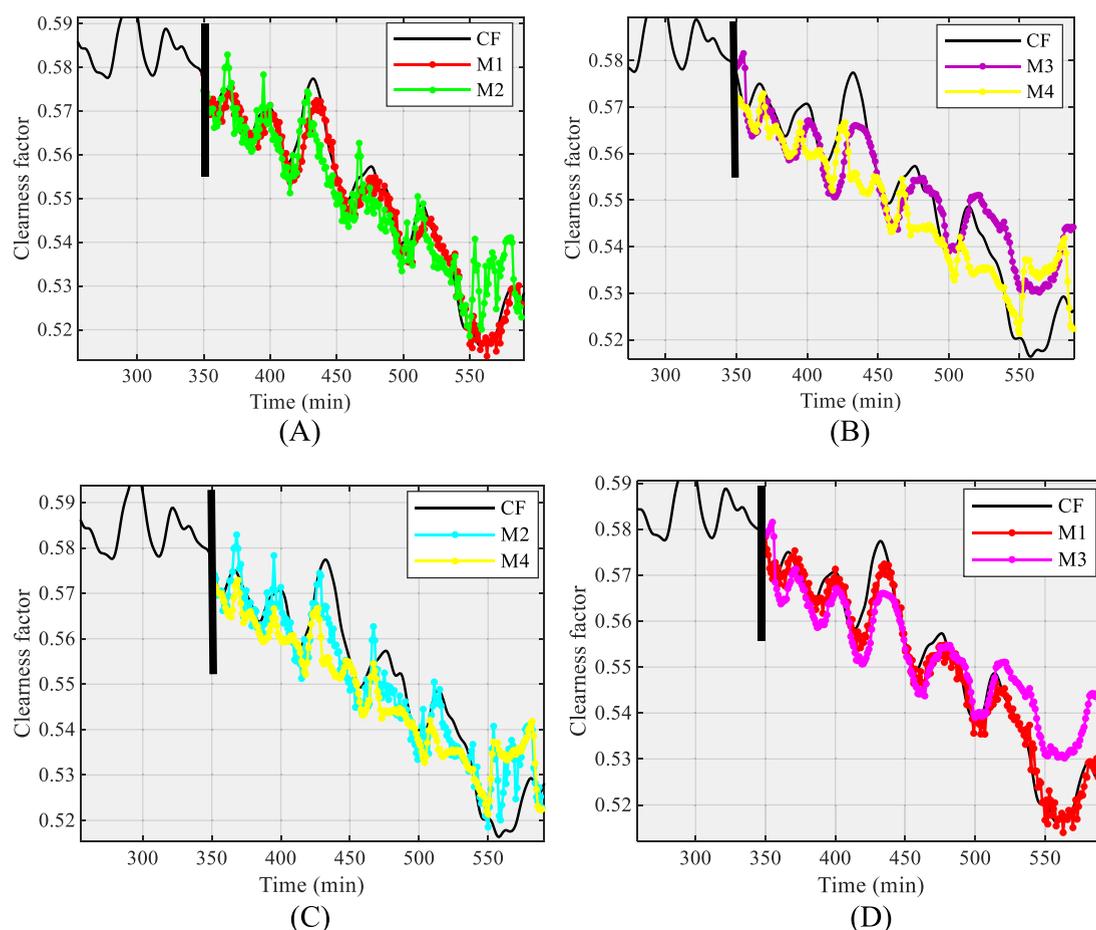
Figure 15. Five-step-ahead prediction of M1−M4 on the low-temperature superheater (from left to right, top to bottom, the graphs are numbered A, B, C, and D).

degradation trend under the combined action of the adaptive sliding window and the deep learning model.) As the number of forwarding prediction steps increases, prediction performance drops rapidly, and it can hardly reflect the real ash accumulation on the heated surface. The accumulation of errors in long-term predictions and the coupling of complex features in the fouling curve have resulted in large prediction errors and randomness in multiple experiments. Therefore, this deep feature extraction algorithm plays an important role in the estimation of ash deposition and the predictive maintenance of the heated surface of the heat exchanger.

In order to verify the robust performance of the proposed model, we tested the clean factor prediction results of the economizer dataset from 250, 350, and 450 min for different prediction starting points (see Figures 28 and 29) (respectively given from the forward prediction steps of five and ten). In this regard, we observe that when the starting point is in the later stage, the effect is always better than that in the early stage. This is because more historical information can be obtained to provide deep learning network training, and in the early stage, it will be limited by the amount of available data. In addition, we usually hope that the prediction result can complete the task of high-precision prediction regardless of whether the starting point is forward or backward. However, early prediction is faced with the problems of the lack of historical data, serious error accumulation, and the coupling of various characteristics of ash accumulation, which all bring about the problems of high prediction error and low accuracy. In the proposed model, the

prediction error brought by the earlier starting point is completely within the acceptable range, so this method has broad application prospects in the ash prediction task of the heated area. Such prediction characteristics have important value for the predictive maintenance of the heating surface in the later stage, that is, the soot blowing and cleaning work can be carried out in a timely and accurate manner by reasonably judging the results of the later prediction, avoiding the problems of over-blowing and under-blowing.

3.5.2. Predictive Soot Blowing Strategy. In the heated surface maintenance task, the most important decision faced is when performing preventive soot blowing operations. The strategy of soot blowing operation depends on the real-time health status (ash accumulation status) $X(t)$ of the heated surface of the system and the predicted health status. This part is based on the above detailed explanation, which ensures the reliability of the proposed model under the multistep-ahead prediction task of the ash condition of the heated area.

We define the failure soot threshold L and the actual soot blowing threshold ε $(\varepsilon > L)$, with S as the starting point of fouling. When to perform the soot blowing operation depends on the value of the current *CF* predicted. The multistep-ahead prediction time will be used as the preparation time reserved for the soot blowing operation. There are generally three situations in the relationship between the predicted CF value $X(KT)$ and the threshold ε, L:
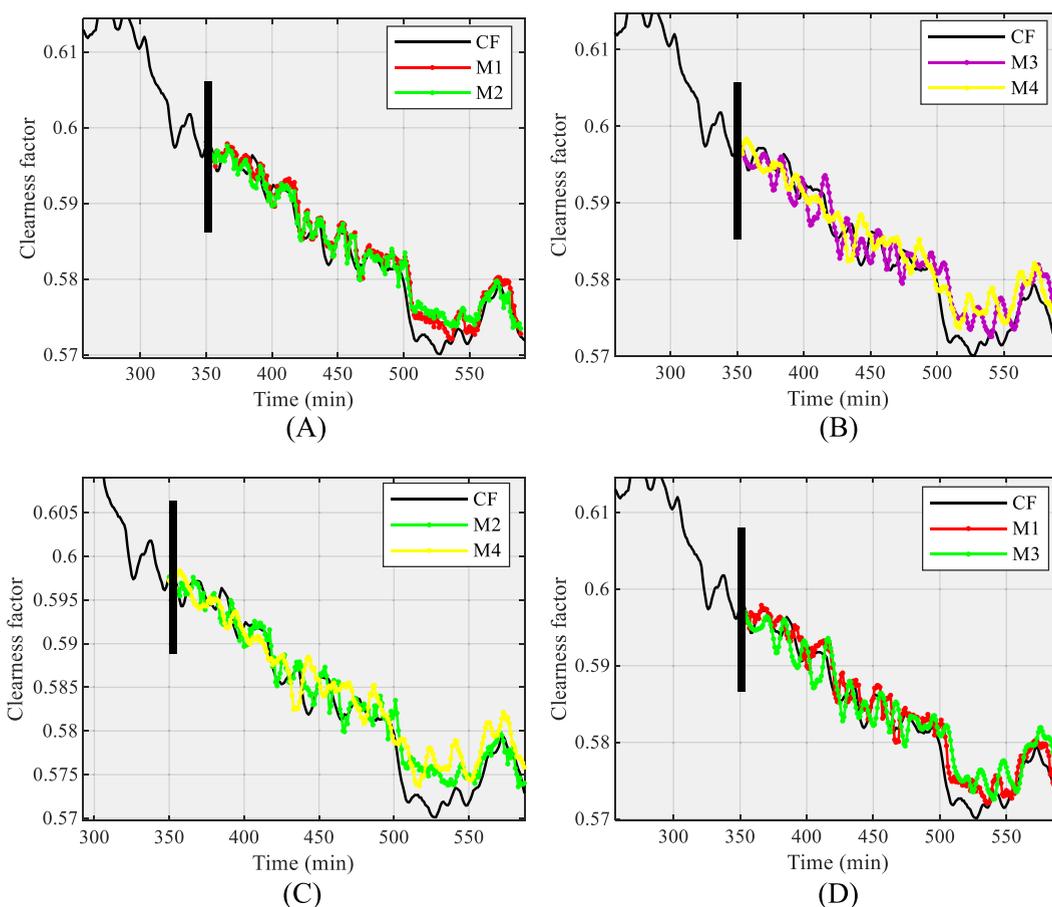
Figure 16. Five-step-ahead prediction of M1−M4 on the reheater. (from left to right, top to bottom, the graphs are numbered A, B, C, and D).
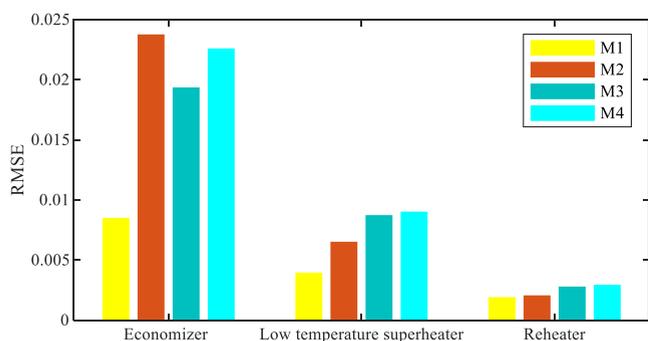
Figure 17. RMSE comparison of five-step-ahead prediction.
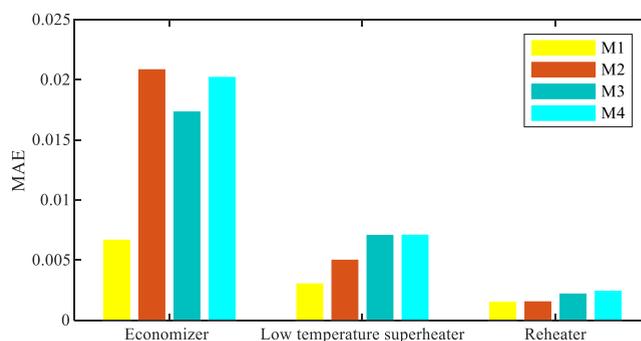
Figure 19. MAE comparison of five-step-ahead prediction.
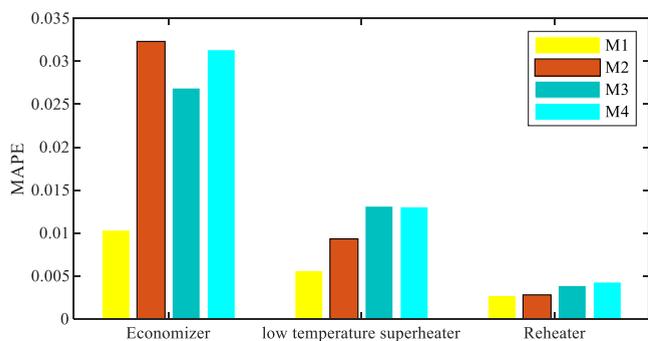
Figure 18. MAPE comparison of five-step-ahead prediction.

(1) When $S > X(KT) > \varepsilon$, the ash degradation state of the heated area of the system is within the normal range, and no soot blowing operation is required. The area of $[\varepsilon, S]$ in Figure 30 is the normal range of the ash degradation state of the heated area of the system. This corresponds to 0-T1.

(2) When $\varepsilon \geq X(KT) > L$, the ash degradation state of the heated area of the system is within the threshold of near failure ash accumulation. The system can continue to run without soot blowing, but there is a high risk of system failure or shutdown. As shown in Figure 30, if the ash degradation state of heated surface is detected in the area of $[\varepsilon, L]$, preventive soot blowing operations are required. Corresponding to T2-T3, T3-T4.
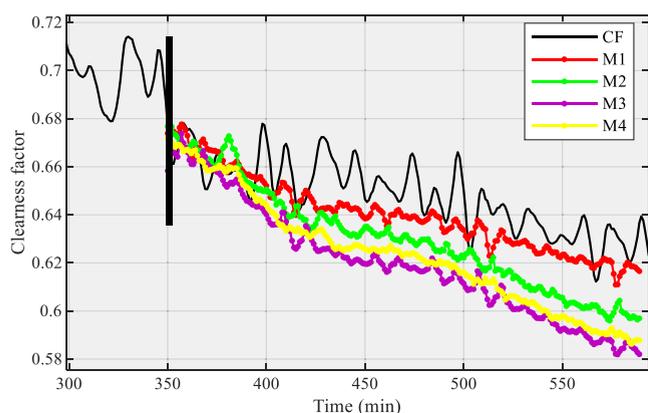
Figure 20. Ten-step-ahead prediction of M1−M4 on the economizer.
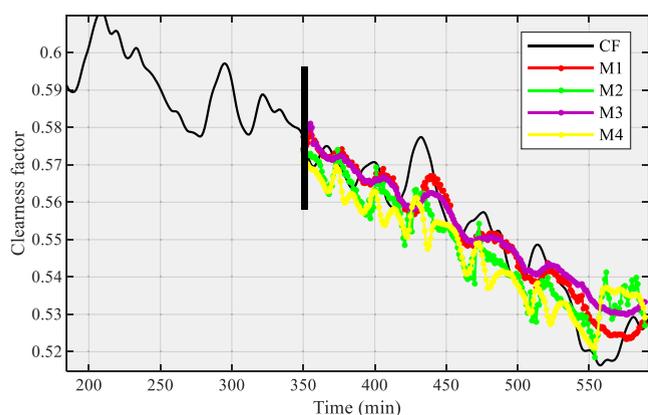
Figure 21. Ten-step-ahead prediction of M1−M4 on the low-temperature superheater.
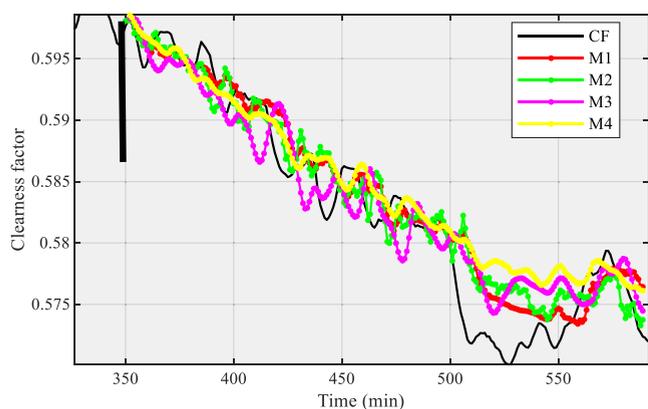
Figure 22. Ten-step-ahead prediction of M1−M4 on the reheater.

Table 5. RMSE Comparison of Multistep-Ahead Prediction

prediction step	model	economizer	low-temperature superheater	reheater
10	M1	0.01243	0.00664	0.002522
	M2	0.02230	0.00904	0.002725
	M3	0.03167	0.00777	0.003078
	M4	0.02832	0.01015	0.003233
25	M1	0.02018	0.00771	0.002704
	M2	0.03252	0.00923	0.003946
	M3	0.05107	0.01111	0.003276
	M4	0.17688	0.009583	0.004947

Table 6. MAPE Comparison of Multistep-Ahead Prediction

prediction step	model	economizer	low-temperature superheater	reheater
10	M1	0.015735	0.009739	0.0032989
	M2	0.03058	0.013232	0.003593
	M3	0.04437	0.011352	0.0044350
	M4	0.039125	0.015139	0.0040909
25	M1	0.027282	0.011078	0.002704
	M2	0.045979	0.013750	0.003846
	M3	0.079429	0.015488	0.0032764
	M4	0.250527	0.013597	0.0049473

Table 7. MAE Comparison of Multistep-Ahead Prediction

prediction step	model	economizer	low-temperature superheater	reheater
10	M1	0.010247	0.005351	0.001911
	M2	0.019779	0.007266	0.002081
	M3	0.028644	0.0061884	0.002575
	M4	0.02526	0.0083328	0.002364
25	M1	0.0176892	0.006081	0.002164
	M2	0.029722	0.0075966	0.003022
	M3	0.051079	0.006188	0.002567
	M4	0.0161152	0.0083328	0.003934
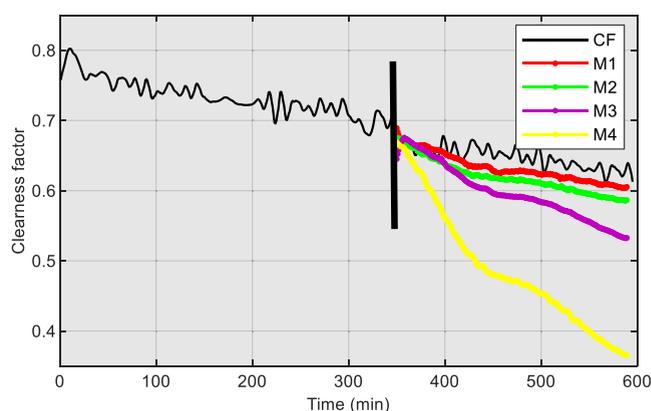
Figure 23. Twenty-five-step-ahead prediction of M1−M4 on the economizer.
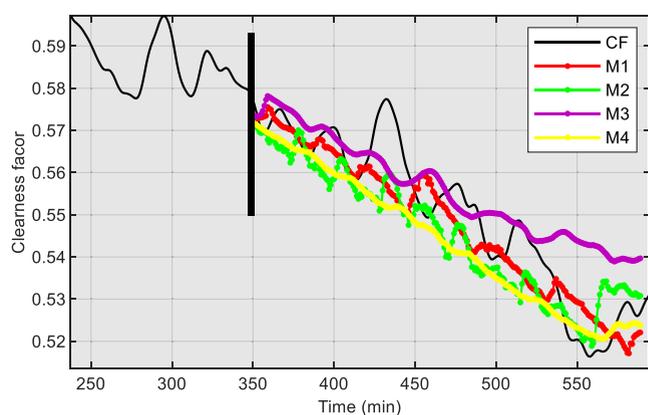
Figure 24. Twenty-five-step-ahead prediction of M1−M4 on the low-temperature superheater.

(3) When $X(KT) < L$, the ash degradation state of the heated area of the system reaches the failure ash accumulation threshold. This means that the system efficiency decreases
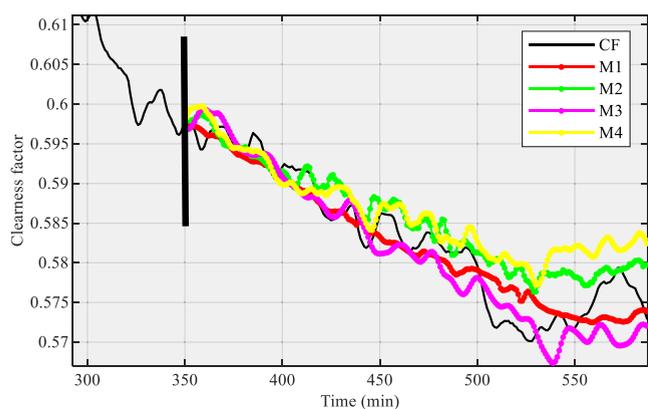
Figure 25. Twenty-five-step-ahead prediction of M1—M4 on the reheater.
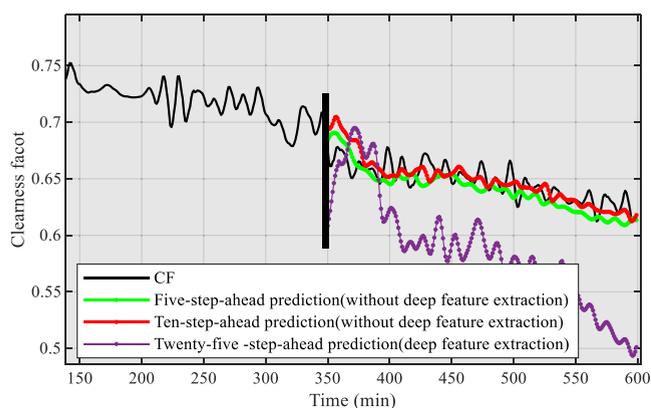
Figure 26. Multistep-ahead prediction without deep feature extraction in the economizer.
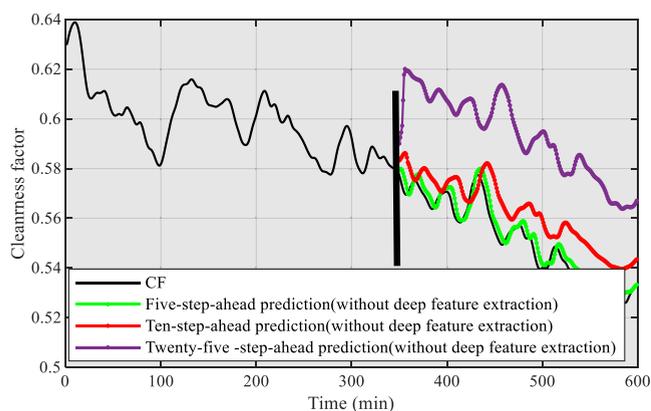
Figure 27. Multistep-ahead prediction without deep feature extraction in the low-temperature superheater.

or fails seriously due to the serious accumulation of ash. It is necessary to immediately carry out strong soot blowing operation. Corresponds to T5-T6 in Figure 30.

It is easy to cause safety accidents after long-term failure (in situation 3). Also, in the soot blowing optimization task based on heat transfer efficiency and cost rate,[12] by optimizing the start and end time of soot blowing, a larger net profit can be obtained (which is generally the difference between the heat transfer amount obtained by the soot blowing operation and the corresponding amount of high-pressure steam lost). According to experience, the starting point is within the predictive

maintenance period of the heating surface to obtain the maximum net profit. Therefore, accurate soot deposit prediction can lay the foundation for soot blowing optimization.

The main task of predictive maintenance in the soot blowing strategy is actually the prediction of the end of life (Eol). In this article, we set Eol as the soot blowing threshold. In other words, it is necessary to predict the time point when the *CF* value reaches the preset soot blowing threshold.

In order to further illustrate the effectiveness of the proposed model in predictive maintenance tasks, we used 20 economizer fouling datasets under various working conditions. In order to ensure consistency, the model hyper-parameters of all data are given the same, and the starting points are predicted from 250, 350, and 450 min, respectively. Figure 31 shows the predictive maintenance schematic diagram and the PDF diagram of the first set of data after 20 repeated experiments (five-step-ahead prediction), where the results follow a normal distribution. The actual Eol of this set of data is 487 min, and the predicted average value of the normal distribution is 486 min under starting point 450 min, 475 min under 350 min, and 436 min under 250 min. The average values of the normal distribution of the results at the predicted starting points of 450 min and 350 min are similar to the actual Eol, but the error is larger at 250 min, which is also in line with the multistep-ahead prediction results of the different prediction starting points in the previous chapter. In addition, the prediction results of the remaining 19 sets of data are shown in Figure 32, where the label 'predicted Eol' is the mean of the normal distribution. The five-step-ahead and ten-step-ahead prediction have small prediction errors, and the final results verify that all improve the effectiveness and credibility of the model in predictive maintenance.

4. CONCLUSIONS

Aiming at the new direction of energy saving, emission reduction, and environmental protection, a fusion model (CEEMDAN-KPCA-EDA) was proposed to predict the health condition of the heated surface and complete the predictive maintenance task in order to maintain the health condition of the heating surface and the efficient heat transfer. This method integrates multiscale analysis of nonlinear and non-stationary fouling time series to obtain IMFs of various frequencies. Then, the global degradation component is retained, and the characteristic dimension reduction is carried out for the IMFS components of different scales to eliminate redundant information, improve the training speed, complete the input reconstruction, and solve the decline of other indicators caused by the training speed of the deep learning model. The adaptive sliding window can adjust the window width adaptively according to the mutation of time series, complete more detailed feature extraction, and improve the prediction performance. In the selection of the prediction models, traditional recurrent neural networks such as LSTM and GRU are abandoned. It uses the framework of EDA to complete the deep extraction of deterioration information. This decomposition—reconstruction—aggregation approach makes it possible to model and predict the time series of nonlinear, non-stationary heating surface degradation with high accuracy. Finally, in the experimental part, the effectiveness and superiority of feature extraction and dimension reduction, adaptive sliding window, and deep learning model in this experiment are analyzed and verified from the perspectives of various models and heating surfaces. In addition, the robustness of the model is proven by experiments from different starting
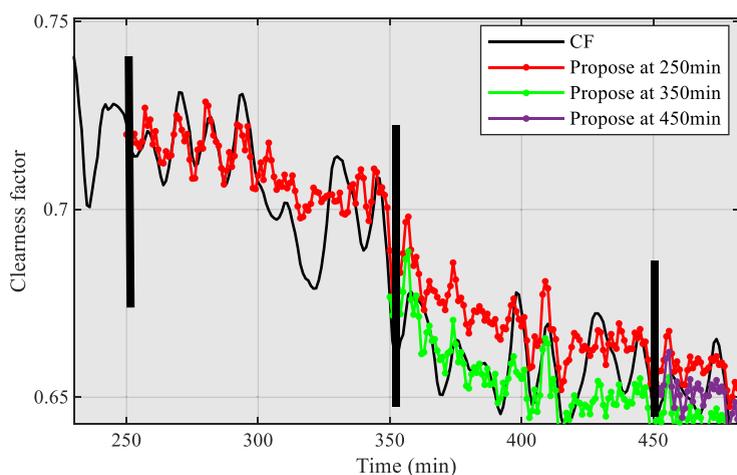
(a) Five-step prediction with different prediction starting points in economizer.
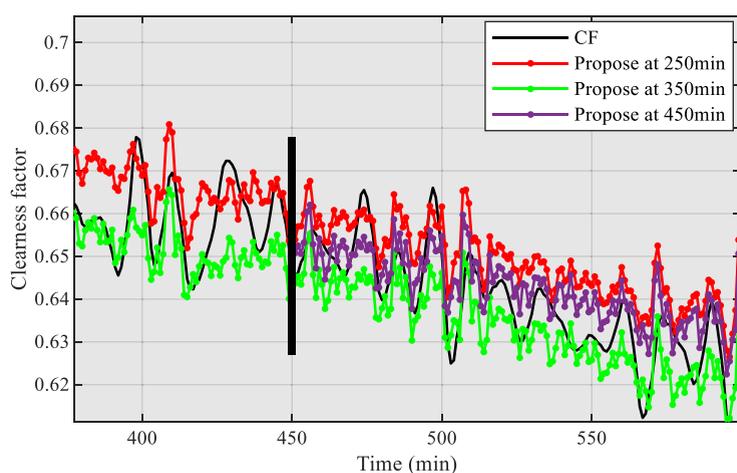
(b) Five-step prediction with different prediction starting points in economizer.

Figure 28. (a) Five-step prediction with different prediction starting points in economizer. (b) Five-step prediction with different prediction starting points in economizer.
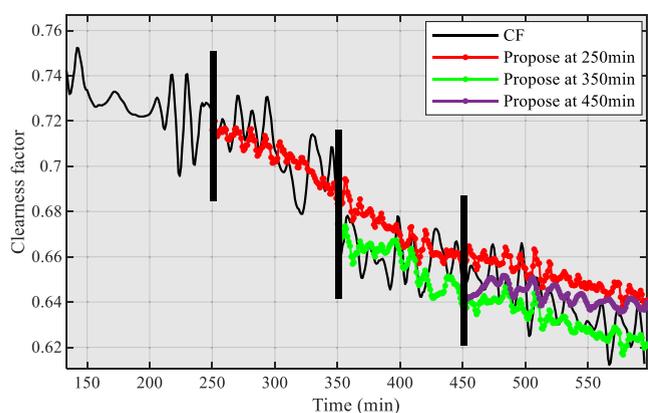
Figure 29. Ten-step prediction with different prediction starting points in economizer.

Figure 30. Three types of ash accumulation.

points of prediction. The predictive maintenance of the heating surface was completed with the data of the economizer under variable working conditions, and the feasibility of the proposed model under this task was verified.

For the numerous hyperparameters inherent in deep learning, this paper only selects moderate and identical hyperparameter

groups for experiments. Therefore, adding a reasonable hyperparameter configuration method is an effective method to optimize this experiment and is also the focus of future work. In addition, the health factor-clearness factor is a time series composed of many salient characteristics (such as flue gas side and working medium side heat transfer temperature difference on average, entrance exit flue gas enthalpy), if these features are further integrated into the deep learning model, the prediction error will be greatly reduced. Finally, in future work, we will further study and discuss how to integrate the high-precision heating surface ash pollution prediction model into the soot
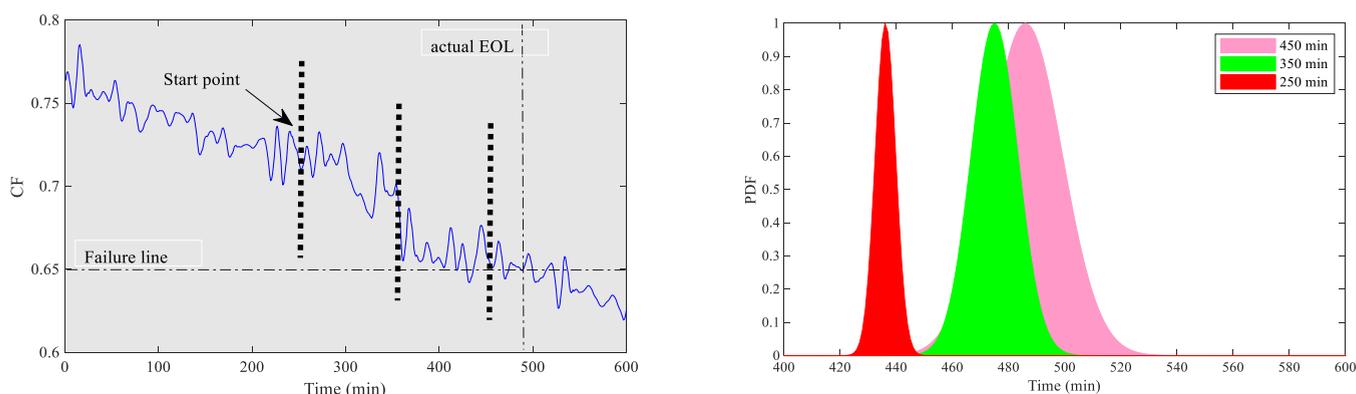
Figure 31. Predictive maintenance principle (left) and predicted uncertainty of Eol at different prediction starting points (right).
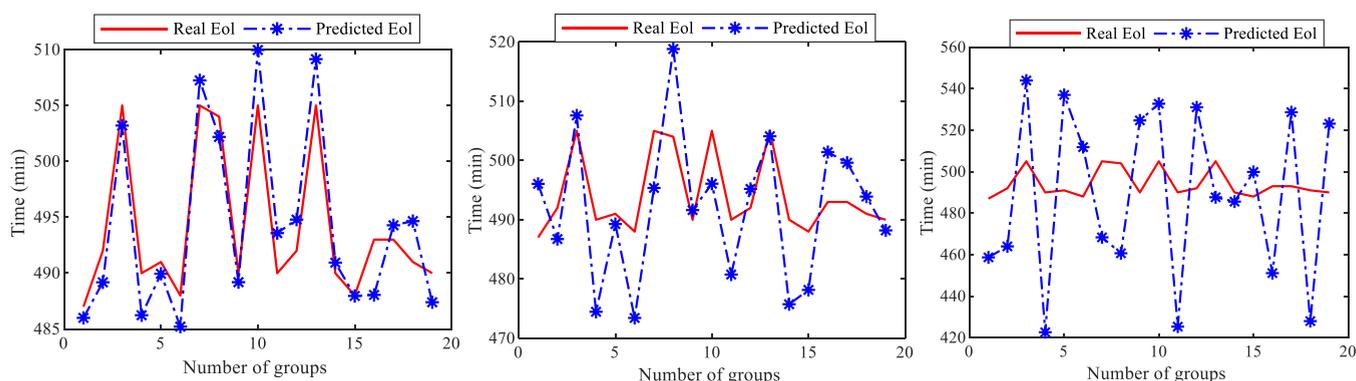
Figure 32. Results of predictive maintenance of 20 sets of data at different starting points (from left to right: 450, 350, 250 min).

blowing optimization, so that a reasonable and economical soot blowing optimization model becomes possible.

■ AUTHOR INFORMATION

Corresponding Author

Yuanhao Shi − *School of Electrical and Control Engineering, North University of China, Taiyuan 030051, China;* ⊙ orcid.org/0000-0001-5680-5658; Email: yhshi@ nuc.edu.cn

Authors

Mengwei Li − *School of Electrical and Control Engineering, North University of China, Taiyuan 030051, China*

Jie Wen − *School of Electrical and Control Engineering, North University of China, Taiyuan 030051, China*

Yanru Yang − *School of Electrical and Control Engineering, North University of China, Taiyuan 030051, China*

Jianchao Zeng − *School of Electrical and Control Engineering, North University of China, Taiyuan 030051, China*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acsomega.2c03052

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Wu, B.; Bai, X.; Liu, W.; Lin, S.; Liu, S.; Luo, L.; Guo, Z.; Zhao, S.; Lv, Y.; Zhu, C.; Hao, Y.; Liu, Y.; Hao, J.; Duan, L.; Tian, H. Non-Negligible Stack Emissions of Noncriteria Air Pollutants from Coal-Fired Power Plants in China: Condensable Particulate Matter and Sulfur Trioxide. *Environ. Sci. Technol.* **2020**, *54*, 6540−6550.

(2) Aldrich, M. The Rise and Decline of the Kerosene Kitchen: A Neglected Energy Transition in Rural America. *Agric. Hist.* **2020**, *94*, 24−60.

(3) Mohanty, D. K.; Singru, P. M. Fouling analysis of a shell and tube heat exchanger using local linear wavelet neural network. *Int. J. Heat Mass Tranf.* **2014**, *77*, 946−955.

(4) Yu, Z.; Jin, J.; Hou, F.; Zhang, Y.; Wang, G.; Liu, B.; Zhai, Z. Understanding effect of phosphorus-based additive on ash deposition characteristics during high-sodium and high-calcium Zhundong coal combustion in drop-tube furnace. *Fuel* **2021**, *287*, 119462.

(5) Kalisz, S.; Pronobis, M. Investigations on fouling rate in convective bundles of coal-fired boilers in relation to optimization of sootblower operation. *Fuel* **2005**, *84*, 927−937.

(6) Romeo, L. M.; Gareta, R. Neural network for evaluating boiler behaviour. *Appl. Therm. Eng.* **2006**, *26*, 1530−1536.

(7) Valero, A.; Cortés, C. Ash fouling in coal-fired utility boilers. Monitoring and optimization of on-load cleaning. *Prog. Energy Combust. Sci.* **1996**, *22*, 189−200.

(8) Qin, Y.; Wang, J.; Liu, Y. W.; Duan, Y.; Liu, M.; Yan, J. Experimental research on convection-condensation heat transfer characteristics and sootblowing operation strategy of waste heat recovery heat exchanger for dryer exhaust. *Drying Technol.* **2019**, *37*, 1633−1646.

(9) Xu, L.; Huang, Y.; Yue, J.; Dong, L.; Liu, L.; Zha, J.; Chen, B.; Zhu, Z.; Liu, H. Improvement of slagging monitoring and soot-blowing of

waterwall in a 650MWe coal-fired utility boiler. *J. Energy Inst.* **2021**, *96*, 106−120.

(10) Shi, Y. H.; Wang, J. C.; Wang, B. Ash fouling monitoring dynamic models of coal-fired utility boiler's heat transfer surface. *Chin. Control Conf.* **2011**.

(11) Perez, L.; Ladevie, B.; Tochon, P.; Batsale, J. C. A new transient thermal fouling probe for cross flow tubular heat exchangers. *Int. J. Heat. Mass Trans.* **2009**, *52*, 407−414.

(12) Shi, Y.; Wang, J.; Liu, Z. On-line monitoring of ash fouling and soot-blowing optimization for convective heat exchanger in coal-fired power plant boiler. *Appl. Therm. Eng.* **2015**, *78*, 39−50.

(13) Zhang, S.; Shen, G.; An, L.; Li, G. Ash fouling monitoring based on acoustic pyrometry in boiler furnaces. *Appl. Therm. Eng.* **2015**, *84*, 74−81.

(14) Ma, Z. H.; Iman, F.; Pisi, L.; Rod, S.; Kong, L. B.; Rokanuzzaman, A. S.; McCollor, D. P.; Steven, A. B. A comprehensive slagging and fouling prediction tool for coal-fired boilers and its validation/application. *Fuel Process. Technol.* **2007**, *88*, 1035−1043.

(15) L., Sun; Y., Zhang; X., Zheng; S., Yang; Y., Qin, Research on the Fouling Prediction of Heat Exchanger Based on Support Vector Machine, *International Conference on Intelligent Computation Technology & Automation*; IEEE, 2008.

(16) Tong, S.; Zhang, X.; Tong, Z.; Wu, Y.; Tang, N.; Zhong, W. Online Ash Fouling Prediction for Boiler Heating Surfaces based on Wavelet Analysis and Support Vector Regression. *Energies* **2020**, *13*, 59.

(17) Shi, Y.; Wang, J. Ash fouling monitoring and key variables analysis for coal fired power plant boiler. *Therm. Sci.* **2015**, *19*, 253−265.

(18) Sivathanu, A. K.; Subramanian, S. Extended Kalman filter for fouling detection in thermal power plant reheater. *Control Eng. Pract.* **2018**, *73*, 91−99.

(19) Teruel, E.; Cortés, C.; Díez, L. I.; Arauzo, I. Monitoring and prediction of fouling in coal-fired utility boilers using neural networks. *Chem. Eng. Sci.* **2005**, *60*, 6035−5048.

(20) Shi, Y.; Li, Q.; Wen, J.; Cui, F.; Pang, X.; Jia, J.; Zeng, J.; Wang, J. Soot Blowing Optimization for Frequency in Economizers to Improve Boiler Performance in Coal-Fired Power Plant. *Energies.* **2019**, *12*, 2901.

(21) Li, Q.; Yan, P.; Liu, J.; Shi, Y.; Yu, D. Prediction of Pollution State of Heating Surface in Coal-Fired Utility Boilers. *IEEE Access.* **2020**, *8*, 206132−206145.

(22) M., Li; Y., Shi; F., Cui; J., Wen; J., Zeng, Research on Gray Prediction of Heated Surface Combining Empirical Mode Decomposition and Long Short-term Memory Network, *International Conference on Control, Automation, Robotics and Vision (ICARCV)*; IEEE 2020.

(23) Pena, B.; Teruel, E.; Diez, L. Soft-computing models for soot-blowing optimization in coal-fired utility boilers. *Appl. Soft Comput.* **2011**, *11*, 1657−1668.

(24) Yang, L.; Wang, F.; Zhang, J.; Ren, W. Remaining useful life prediction of ultrasonic motor based on Elman neural network with improved particle swarm optimization. *Measurement* **2019**, *143*, 27−38.

(25) Li, W.; Zhu, X.; Chen, Z.; Dai, Z.; Li, J.; Ran, C. Code multipath error extraction based on the wavelet and empirical mode decomposition for Android smart devices. *Gps Solut.* **2021**, *25*, 1.

(26) Upadhyay, P.; Upadhyay, S. K.; Shukla, K. K. Magnetic resonance images denoising using a wavelet solution to laplace equation associated with a new variational model. *Appl. Math Comput.* **2021**, *400*, 126083.

(27) Zhang, Y.; Yan, B.; Aasma, M. A novel deep learning framework: Prediction and analysis of financial time series using CEEMD and LSTM. *Expert Syst. Appl.* **2020**, *159*, 113609.

(28) Chen, Z.; Wang, L.; Zhang, J.; Guo, G.; Fu, S.; Wang, C.; Li, X. Framework and development of data-driven physics based model with application in dimensional accuracy prediction in pocket milling. *Chin. J. Aeronaut.* **2021**, *34*, 162−177.

(29) Wang, X.; Zhang, Y.; Yu, B.; Salhi, A.; Chen, R.; Wang, L.; Liu, Z. F. Prediction of protein-protein interaction sites through eXtreme gradient boosting with kernel principal component analysis. *Comput. Biol. Med.* **2021**, *134*, 104516.

(30) Kong, D.; Zhu, J.; Duan, C.; Lu, L.; Chen, D. Bayesian linear regression for surface roughness prediction. *Mech. Syst. Signal Pr.* **2021**, *142*, 106770.

(31) Wu, J.; Hu, K.; Cheng, Y.; Zhu, H.; Shao, X.; Wang, Y. Data-driven remaining useful life prediction via multiple sensor signals and deep long short-term memory neural network. *ISA Trans.* **2020**, *97*, 241−250.

(32) Zheng, J.; Zhang, H.; Dai, Y.; Wang, B.; Zheng, T.; Liao, Q.; Liang, Y.; Zhang, F.; Song, X. Time series prediction for output of multi-region solar power plants. *Appl. Energy* **2020**, *257*, 114001.

(33) Han, Y.; Fan, C.; Xu, M.; Geng, Z.; Zhong, Y. Production capacity analysis and energy saving of complex chemical processes using LSTM based on attention mechanism. *Appl. Therm. Eng.* **2019**, *160*, 114072.