A Neural-Network-Based Method for Ash Fouling Prediction of Heat Transfer Surface in Coal-Fired Power Plant Boiler

YUANHAO SHI1,2, MENGWEI LI1, FANGSHU CUI3, JIE WEN1, AND JIANCHAO ZENG3

1School of Electrical and Control Engineering, North University of China, Taiyuan 030051, China
2Department of Automation, Shanghai Jiao Tong University, Shanghai 200240, China
3School of Data Science and Technology, North University of China, Taiyuan 030051, China

ABSTRACT Soot blowing optimization and health management of coal-fired power plant boiler has received increasing attention in recent years. The ash fouling monitoring and prediction are the basis for achieve this goal. Nowadays, with the development of neural network technology, the new data-driven methodologies are provided for ash fouling monitoring and prediction. This paper presents a comprehensive method based on neural-network for ash fouling prediction. Firstly, the health factor-clearness factor of the heated surface was established from the actual heat transfer coefficient and the theoretical heat transfer coefficient. Wavelet threshold denoising algorithm will be used as data preprocessing method. Secondly, use Ensemble Empirical Mode Decomposition (EEMD) to obtain a series of frequency stable parts. Finally, Encoder-Decoder based Attention (EDA) is used to predict the ash deposit on the heat transfer surface. The EDA model consists of an encoder, a decoder and an attention mechanism. The encoder and decoder are composed of Bidirectional Long Short-Term Memory (BI-LSTM) and Long Short-term Memory Network (LSTM) respectively. The function of the attention mechanism is that the output of each time step of the encoder is given a different degree of attention, and it is sent to the decoder as an attention vector after a weighted average operation. Ash accumulation data on the heat transfer surface of various devices are used to verify the effectiveness of the proposed hybrid model. In addition, the experimental results show that this method has better prediction accuracy than other variant models.

INDEX TERMS Soot blowing optimization, ash fouling prediction, ensemble empirical mode decomposition (EEMD), long short-term memory (LSTM), encoder-decoder based attention (EDA).

I. INTRODUCTION

New energy generation, such as wind and solar energy, has a promising future, but thermal power generation is still the power supply mode in most countries at the present stage [1], [2]. As China’s efforts to save energy and reduce emissions increase, the requirements for controlling pollution emissions and reducing energy consumption are also increasing [3], [4]. As the main device of thermal power generation, coal-fired boilers shoulder the important task of converting thermal energy into electrical energy. Nowadays, the capacity and parameters of boilers are continuously increasing, and the problem of ash and slag in boilers is becoming more and more serious. After the pulverized coal is burned, the ash and slag will pass through [5] the high-temperature superheater, high-temperature reheater, low-temperature superheater, low-temperature reheater, economizer, etc. Ash will be deposited on the heated surface of almost all the devices mentioned above, due to the poor thermal conductivity of the ash, the increase in ash accumulation will cause a reduction in the overall heat transfer efficiency of the boiler [6]. In addition, if the ash is not cleaned up in time, it will increase the duct ventilation resistance and reduce the boiler output. What’s more, the deposition of ash and slag will cause corrosion of the pipeline and the heated surface, and the probability of industrial safety accidents will greatly increase.
Soot blowing is an effective method to maintain the health of the heated surface, which can ensure the safety and economy of boiler operation, but because it is difficult to estimate the level of ash and slag, so this operation has always been flawed. Nowadays, boiler soot blowing is still mostly based on the experience of regular soot-blowing [7]. However, there are many hidden dangers in this way: for example, low-frequency soot blowing will cause the heat loss and reduce heat transfer efficiency, while high-frequency soot blowing will cause waste of high-pressure steam and corrosion of pipelines, and even affect the life of power plant equipment and the occurrence of dangerous accidents. The occurrence of any of them is contrary to our desired goal of energy saving and emission reduction.

It is necessary to predict the cleaning status of the heated surface of the boiler in the future and make preparations in advance, but the soot blowing prediction of the heated surface has always been full of challenges. Its purpose is to let the soot-blower perform the soot blowing operation at the right time. With the continuous efforts of the majority of researchers, there has been some progress in the monitoring of ash accumulation. In the monitoring of ash accumulation, there are mainly device-based monitoring and model-based monitoring. Perez et al. [8] designed a new transient thermal fouling probe in order to accurately estimate the heat transfer coefficient and fouling thickness of convective heat exchangers. But this method usually requires more complicated instruments for measurement. Shi et al. [9] used a model-based indirect ash online monitoring method: the clearness factor was first proposed as an indicator of the health of the heated surface. Combined with the data available from the power station and the online monitoring model to calculate the ash deposition status. This monitoring method does not require special monitoring equipment and complicated calculation and aims at reducing steam consumption to design soot blowing optimization strategy [10]. Tong et al. [11] established the thermal resistance of the ash layer to reflect the degree of ash accumulation. In order to make the thermal resistance change of the ash layer more obvious, the thermal resistance curve was smoothed by wavelet threshold denoising. In addition, online monitoring is achieved through the support vector regression model, and the accuracy on the test set reached 98.5%.

In this paper, the Clearness Factor (CF) is used to characterize the health of the heated surface, which is composed of the ratio of the actual heat transfer coefficient and the theoretical heat transfer coefficient and has the characteristics of small extraction difficulty. In terms of the model, a pure data-driven method is proposed [12]. The clearness factor degradation curve is a highly nonlinear time series, which can be regarded as a mixed signal that includes low frequency and high frequency components. It may be difficult to adapt to these two features if predicting at the same time, so accurate prediction has always been a challenge. Therefore, the EEMD method is used to decompose the degradation curve into a series of clearness factor decomposition components, which include a residual component that represents the global degradation trend and multiple high-frequency components [13], [14]. Support vector regression (SVR) and Encoder-Decoder based Attention (EDA) are used to train and predict the low-frequency and high-frequency signals respectively, which greatly improves the accuracy of prediction. This data-driven method does not require complex, special instruments and computing systems, and can predict the health of the heated surface of coal-fired boilers only by using the readily available monitoring data.

In fact, the current research on soot blowing methods is generally a combination of real-time ash monitoring and soot-blowing threshold, but this prediction method cannot meet the requirements of ‘early warning’ [15]. This is because the future health condition of the heating surface cannot be obtained by real-time ash accumulation monitoring. If the ash accumulation state of the heating surface has reached the critical threshold, there will be no extra time to conduct preparation operation and personnel allocation of soot blowing, and the best soot blowing time will be missed. To solve such problems, rolling prediction is generally adopted. If the long-term rolling prediction is adopted, the long-term accumulation of errors will eventually make the experiment meaningless. In this paper, the finite-step rolling prediction method is used, which is part of the multi-step time series prediction task. The task of multi-step time series prediction is to require the model to predict the condition of the object for a long time in the future for maintenance and decision-making. In this paper, the time series of clearness factors available in the past will be used as input to train the deep learning model to mine the hidden information and long-term dependence of the time series. Finally, the trained model is used for rolling multi-step prediction. Compared with single-step prediction (only predict the state of the object at the next time in the future), multi-step prediction can know the time of failure of the system earlier. Although multi-step prediction will produce error accumulation and worse prediction results as the prediction step size increases, by adjusting the prediction step length reasonably, a more satisfactory prediction effect can still be obtained, and a certain preparation time is reserved for the soot-blowing operation. The impact of different rolling prediction steps on prediction performance will be described in detail in section IV. D. Predicting the future ash pollution status has laid the foundation for soot blowing decision-making and soot blowing optimization.

The remainder of the paper is organized as follows: Section II details the origin of the indicator that characterizes the health of the heated surface. Section III introduces basic theories of wavelet threshold denoising, EEMD, RNN, LSTM, encoder-decoder framework, and attention mechanism. In section IV, our model was tested on the fouling section dataset of four boiler devices. In addition, we analyzed the adaptability of the model on different datasets, the difference between our proposed model and the variant model, and the influence of different rolling prediction steps on the model. In order to verify the superiority of the model,
we conducted the comparative experiment with advanced models. Finally, conclusions are drawn in section V.

The main innovations and contributions are as follows:

(1) Aiming at the past fixed-time and fixed-frequency soot blowing operations, this paper proposes the ash prediction of boiler heated surface based on a deep learning model., using an encoder-decoder-attention mechanism architecture.

(2) For the ash accumulation curve of the heated area, wavelet threshold denoising is used to remove unnecessary noises that increase the difficulty of prediction. In order to complete the multi-scale analysis of the denoised curve, this paper uses the EEMD method to complete the multi-scale modeling and prediction of the ash on the premise of solving the modal aliasing problem of the EMD algorithm.

(3) Using ash accumulation datasets of the heated surface of the commonly used components in the boiler (economizer, low-temperature superheater, high-temperature superheater, reheater), the rolling prediction method is used to complete multi-step prediction. The experimental results verify the superiority and robustness of the model, as well as its good adaptability to a variety of datasets, and it provides a promising tool for the ash cleaning of coal-fired power stations.

II. 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, we combine the basic thermodynamic formula and real-time measured data from the boiler DCS system to obtain the indicator of 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.

\[ CF = \frac{k_r}{k_o} \]  

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.

\[ k_r = a_f + a_d \]  

In the formula, \( a_f \) represents the theoretical radiation heat transfer coefficient, and \( a_d \) is the theoretical convective heat transfer coefficient [7]. The following formula is the specific mechanism formula of the two heat transfer coefficients:

\[ a_f = 5.7 \times 10^{-8} a_{gb} \frac{1}{2} a_i T^2 K \]  

where \( a_{gb} \) and \( a_i \) are the blackness of the pipe wall and the flue gas respectively; \( T \) is the temperature of the flue gas and the pipe wall respectively, \( C_p \), \( C_z \) are the transverse and longitudinal directions of the heating surface, \( \lambda \) is the thermal conductivity of the flue gas, and \( d \) is 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.

\[ w = \frac{V_b}{A} \]  

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.

\[ V_b = \frac{\rho_v V_r}{(1 + \frac{t_r}{273.15})} \]  

where \( V_r \) is the measured flue gas flow through the heating surface, \( t_r \) is the flue gas temperature through the heating surface, \( \rho_v \) is the actual pressure of the flue gas, and \( \rho_b \) is the standard atmospheric pressure.

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

\[ K_r = \frac{Q_y}{F \Delta t_m} \]  

\[ \Delta t_m = (\Delta t_{max} - \Delta t_{min})/\ln \frac{\Delta t_{max}}{\Delta t_{min}} \]  

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

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, and 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:

\[ Q_y = Q_q \pm \Delta Q_j \pm \Delta Q_q \]
steam, and \( \Delta Q_q \) is the heat absorption change on the steam side.

Heat release on the flue gas side:

\[
Q_y = \varphi B_j (h_{in} - h_{out} + \beta h_f) \quad (11)
\]

\[
B_j = B \left(1 - \frac{q_4}{100}\right) \quad (12)
\]

\( \varphi \) is the heat retention coefficient, \( h_{in} \) and \( h_{out} \) are the flue gas enthalpy values at the inlet and outlet of the economizer, \( \beta \) is the air leakage coefficient of the flue section, and \( h_f \) 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.

\[
\Delta Q_j = C_j m_j \frac{\partial \theta_j}{\partial \tau} \quad (13)
\]

\[
\Delta Q_q = C_q m_q \frac{\partial \theta_q}{\partial \tau} \quad (14)
\]

\[
Q_q = D(H_{out} - H_{in}) \quad (15)
\]

In the formula, \( C_j \) 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. \( \theta_q \) and \( \theta_j \) are the metal pipe wall temperature and steam temperature, \( D \) is the mass flow of the working fluid of the economizer, \( H_{out} \) 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.

Fig. 1 shows the daily change curve of the economizer clearness factor obtained from the previously mentioned formula and the real-time data of the boiler DCS system and the corresponding load curve. It can be seen from Fig.1 that the value of the clearness factor decreases continuously from 11 am to 8 pm due to ash accumulation of the boiler. From 8 pm to 9 pm, the soot blowing operation makes the value of clearness factor increase continuously, which indicates that the condition of the heated surface of the boiler is getting better and better, the same is true from 9 am to 11 am. The cleanliness of the heated surface of the remaining devices of the boiler is basically similar to that of the economizer.

In addition, at S1 segment, we can find that the value of the clearness factor is significantly increased, but there is no soot blowing signal at this time. This is mainly because the flue gas flow rate rises sharply due to the sharp increase in unit load at this time. The quality of the ash taken away is far greater than the quality of the ash brought by the flue gas, which plays a role similar to soot blowing. After 11 am, the load of the unit stabilized, and the normal operation of coal-fired power stations caused increased ash accumulation. In fact, the main reason why the clearness factor still has strong nonlinear non-stationarity after denoising is the deposition and denudation of ash caused by the long-term uncertainty of flue gas rate changes, and this factor is not can be ignored, which creates difficulties for predicting.

### III. BASIC THEORY

In order to obtain the future changes of the clearness factor degradation curve, it is necessary to analyze and learn historical information. SVR and EDA fusion model can be better competent. In this section, we will briefly describe wavelet threshold denoising, EEMD, RNN, and EDA based on the attention mechanism.
A. WAVELET THRESHOLD DENOISING

The short-time Fourier transform solves the problem that the traditional Fourier transform cannot reflect the time domain information, but the characteristics of its fixed window will appear to be in conflict with time resolution and frequency resolution [16]. In the WT (Wavelet Transform), the adaptive ‘time-frequency window’ that varies with frequency can observe any details of the signal, which is usually wide in the low-frequency part and narrow in the high-frequency part. WT starts from the perspective of the basis function, which has scale parameter $a$ and translation parameter $\tau$. These characteristics make the WT have a good time-frequency analysis ability and solve the problem of sudden change and nonlinear signal analysis. In fact, in engineering applications, since the analyzed signal is discrete, it is usually necessary to discretize the wavelet transform. The DWT (Discrete Wavelet Transform) can be calculated following (16):

$$w_f(2^j, b) = \langle f(t) , \psi_{j,b}(t) \rangle = 2^{-\frac{j}{2}} \int_R f(t) * \psi(\frac{t-b}{2^j})dt$$

(16)

The basic principle of DWT decomposition is as follows: the original signal is continuously decomposed through high-pass and low-pass filters [17]. 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 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 repeat the process continuously. The decomposition figure of DWT is shown in Fig. 2.

![Wavelet decomposition structure](image)

**FIGURE 2. Wavelet decomposition structure.**

After multi-level wavelet decomposition of the original signal containing noise, a series of high-frequency and low-frequency wavelet decomposition coefficients will be obtained. In general, the effective part of the signal is stored in the low-frequency component, and the noise is stored in the high-frequency component. The wavelet decomposition coefficients corresponding to noise are generally small, on the premise of retaining the wavelet coefficients of all low-frequency components, the wavelet coefficients of high-frequency components containing noise are processed. The threshold function generally has a soft threshold and a hard threshold function. The hard threshold function compares the absolute value of the signal with the threshold, the coefficients smaller than the threshold are set to zero, and the others remain unchanged; The soft threshold function sets the coefficient less than the threshold to zero, and points greater than or equal to the threshold will not be reduced to zero, but will become the difference between the point value and the threshold. The mathematical representation is shown in (17-18):

$$w_\lambda = \begin{cases} w & |w| \geq \lambda \\ 0 & |w| < \lambda \end{cases}$$

(17)

$$w_\lambda = \begin{cases} \text{sgn}(w)(|w| - \lambda) & |w| \geq \lambda \\ 0 & |w| < \lambda \end{cases}$$

(18)

where $\lambda$ is the fixed threshold.

The reconstructed signal obtained by the hard threshold processing method may cause oscillation and local distortion, and the data after the soft threshold function denoising is relatively smooth. The denoised signal can be obtained by performing wavelet inverse transform on the processed wavelet decomposition coefficients. Fig. 3 shows the dataset of the economizer’s all-day clearness factor denoising and the extracted ash accumulation coefficients. It is not difficult to find that the denoised data more clearly reflects the deterioration trend and nonlinear fluctuations, which lays the foundation for the following multi-scale prediction.

B. ENSEMBLE EMPIRICAL MODE DECOMPOSITION

Empirical Mode Decomposition (EMD) is a new signal processing method that has powerful processing capabilities in multi-frequency complex signals. It is not only suitable for linear analysis, but also for nonlinear, non-stationary signals. The key to the EMD algorithm is that this algorithm will decompose the multi-feature fusion signal into a single feature signal, which is a method of analyzing non-stationary signals [18]. It is now widely used in the PHM field. 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 $h_k(k)$ 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):

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

Ensemble Empirical Mode Decomposition (EEMD) has made improvements on the basis of EMD. Considering that when the EMD method is used to adaptively decompose signals [19], [20], it is easy to cause end effects and mode confusion problems, these problems will cause the prediction accuracy to decrease. By adding appropriate noise and ensemble averaging, the deficiencies of EMD can be improved to avoid. In addition, the significant advantage of EEMD decomposition compared with wavelet decomposition is that it is a signal adaptive analysis method, which completely relies on the signal itself to adaptively determine the number of modal components obtained by decomposition, and does not require custom basis functions and decomposition layers [21].

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 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.

Fig.4 shows the decomposition of the clearness factor degradation curve of the economizer after denoising (We only show four high-frequency components). If the degradation sequence of clearness factors is directly predicted, the aliasing of multiple modes will cause difficult prediction and large errors. The EEMD decomposition decomposes the multi-frequency aliasing signal into the overall degradation trend and multiple high-frequency components to predict separately. The relatively constant frequency of multiple high-frequency components can bring better prediction results. In addition, the time spent is an important part of evaluating the performance of the model. It is worth noting that the introduction of the decomposition algorithm brings about a huge improvement in the prediction effect, while the time spent in the whole multi-step prediction process is almost negligible (the running time of EMD and EEMD is about 2s). This part of the discussion will be given in detail later.

C. SUPPORT VECTOR REGRESSION

SVM (Support vector machine) show certain advantages in non-linear high-dimensional pattern recognition problems, as a statistical learning theory of machine learning methods.
Proposed by Corinna and Vapnik in 1995, SVR [22], [23] is proposed on the basis of SVM, and its sample target value is the continuous value when doing regression analysis. As a branch of SVM, SVR has good stable predictive ability for nonlinear time series. Its main function is to map training samples to high-dimensional space through nonlinear transformation, and establish a linear regression function in the high-dimensional feature space as shown in formula (20):

\[ f(x) = w^T \phi(x) + b \]  

where \( w, x, \) and \( b \) are weight matrices, input sample features, respectively, \( f(\cdot) \) represents the input-output mapping function, \( \phi(x) \) is the nonlinear mapping function.

The basic idea of SVM is to separate the two types of samples by finding an optimal classification surface, while the basic idea of SVR is to obtain an optimal classification surface with the smallest error from all training samples. The SVR problem is (introducing slack variables):

\[ \min_{w, b, \xi, \hat{\xi}} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{m} (\xi_i, \hat{\xi}_i) \]  

\[ \begin{align*}
    y_i - w^T x_i - b & \leq \varepsilon + \xi_i \\
    w^T x_i - b & \leq \varepsilon + \hat{\xi}_i \\
    \xi_i & \geq 0 \\
    \hat{\xi}_i & \geq 0, \quad i = 1, 2, \ldots n
\end{align*} \]  

where \( \xi_i \) and \( \hat{\xi}_i \) represent slack variables, \( \{x_i, y_i\} \) is the input and output combination of the \( i \)th training sample, \( \varepsilon \) is the loss coefficient and \( C \) represents a regularization constant. After introducing the Lagrange multiplier and the kernel function, the nonlinear mapping SVR expression is obtained:

\[ f(x) = \sum_{i=1}^{m} (a_i - a_i^*) k(x, x_i) + b \]  

The KKT condition needs to be met during the whole solution process. In addition, \( a_i \) and \( a_i^* \) are called Lagrange multipliers, and \( k(x, x_i) \) is called a kernel function. The function of the kernel function can transform low-dimensional nonlinear problems into high-dimensional linear problems. The commonly used kernel functions in SVR generally include linear kernels, polynomial kernels, and Gaussian kernels. The linear kernel function is selected in this article.

D. RECURRENT NEURAL NETWORK

As an important branch of deep learning, Recurrent Neural Network (RNN) can study variable length sequences due to its fixed loop structure. General RNN has certain advantages in processing time series because its current hidden state contains all previous input information and the output will be affected by historical information. In addition, it has the characteristics of hidden layer parameter sharing [24]. RNN structure diagram as shown in Fig. 5. The mathematical expression of RNN is shown in (24-25):

\[ h_t = \tanh(w_{xh}x_t + w_{hh}h_{t-1} + b) \]  

\[ y_t = f(w_{hy}h_t + c) \]  

where \( w_{xh}, w_{hh} \) and \( w_{hy} \) are the training weights respectively for input-to-hidden, hidden-to-hidden and hidden-to-output connections, while \( b \) and \( c \) are the bias vectors which allow each node to learn an offset. \( x_t, h_t, y_t \) are the input, hidden state, and output of RNN, respectively. In general, the activation function in the process of obtaining the output \( h_t \) of the current hidden layer is tanh, which can limit the output of the hidden layer to \([-1, 1]\). However, the activation function of the output \( y_t \) is not fixed, and it is selected according to different situations. Due to the problem of gradient explosion and gradient disappearance, RNN is still very difficult to deal with long-term dependency problems. Long Short-term Memory Network (LSTM) and Gated Recurrent Unit (GRU) add a threshold mechanism on the basis of traditional RNN, which solves this problem well.

Different from RNN, LSTM was proposed by Hochreiter and Schmidhuber [25], [26]. It expands three gates on the basis of the RNN framework. Each LSTM unit can determine the information that needs to be discarded or retained, so even if it is important information in the early stage will be retained by long-distance transmission, rather than forgotten. This allows LSTM to mine long-term dependencies between time series. More literature shows that LSTM can handle time series with a length of hundreds, while RNN is very limited. It can be seen from the following formula that the input, forget gate, and output gate can be calculated from the current output and the hidden state at the previous moment. In each step, the forgetting gate determines what information is forgotten in the past, and the input gate determines what information is newly added to the cell state (from the current input information and the hidden state at the previous moment). The hidden state at the moment will be determined by the output gate and the cell state.

\[ i_t = \delta(w_i [h_{t-1}, x_t] + b_i) \]  

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

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

\[ h_t = o_t \cdot \tanh(c_t) \]  

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

where \( i_t, f_t, \) and \( o_t \) are the state values of the input gate, forget gate, and output gate respectively. \( c_t, h_t \) are the cell state and hidden state value at the current moment, respectively.
$w_i, w_f, w_o, w_c$ are the weight matrix of input gate, forget gate, output gate, and cell state respectively. $h_t, b_i, b_f, b_o, b_c$ are the input gate, forget gate, output gate, and cell state bias vector respectively. $\delta(\cdot)$ is the gate function. It can be seen from the above formula that LSTM will not cause overwriting and rewriting of previous content due to its clever threshold mechanism setting. In addition, under the action of this special mechanism, even early important information will not be lost due to long-distance transmission, which means that LSTM is completely different from general neural networks in capturing long-distance dependence.

In fact, traditional LSTM still has weaknesses in time series prediction, and it can only mine information from a single direction. The Bidirectional Long Short-Term Memory (BI-LSTM) is composed of two LSTMs. When all input sequences are available, it receives the input sequence from the forward direction (1 to T) and the reverse direction (T to 1) and learns in the opposite direction. This allows us to obtain additional hidden states for more adequate learning.

$$i_t = \delta(w_i [h_{t-1}, x_t] + b_i) \quad (31)$$
$$f_t = \delta(w_f [h_{t-1}, x_t] + b_f) \quad (32)$$
$$o_t = \delta(w_o [h_{t-1}, x_t] + b_o) \quad (33)$$
$$\tilde{h}_t = o_t \cdot \tanh(c_t) \quad (34)$$
$$\tilde{c}_t = f_t \cdot c_{t-1} + i_t \cdot \tanh(w_c [h_{t-1}, x_t] + b) \quad (35)$$
$$h_t = \tilde{h}_t \odot \tilde{c}_t \quad (36)$$

The forward learning formula is basically the same as the reverse learning formula. We give different direction arrows to distinguish them, but only the reverse learning formula is given here, the meaning of each variable is the same as LSTM. The output of the hidden state at each time of forwarding learning and reverse learning will be added to each other to get the final hidden state at each time. Fig. 6, Fig. 7 shows the structure of LSTM and BI-LSTM.

**E. ENCODER-DECODER FRAMEWORK BASED ON ATTENTION MECHANISM**

The Encoder-Decoder Framework (ED) model was proposed by Cho et al. It was originally derived from machine translation [27]. It will be used when the input and output lengths are not equal, which solves the problem of fixed-length output such as LSTM and GRU.

The framework is mainly composed of two parts, the encoder part and the decoder part. In the encoder part, we encode the input sequence $x$ into a fixed-length context vector for use by the decoder. The context vector contains input information (generally determined by the hidden state of the RNN at the last moment). Then the decoder will decode the context vector to generate the target value step by step. This framework allows the output sequence length of the decoder to be different from the input sequence length of the encoder, which improves the flexibility of use. When the length of the input sequence increases, the context vector can only carry partial information of all inputs due to dependency problems, but this is far from enough for practical engineering use.

Many scholars now generally add an attention mechanism to the ED architecture [28], [29], the attention mechanism is inspired by the biological vision system, which allows...
us to pay attention to certain things in certain scenes. It is making its mark in machine translation and image analysis.

In the attention model, the output of each step of the decoder will perform a weighted average operation on the output of all steps of the encoder. Therefore, the ED model based on the attention mechanism removes the fixed-length encoding bottleneck of the general ED, and the input information is transferred from the encoder to the decoder without loss [30]. The Encoder-Decoder based on Attention (EDA) structure is shown in Fig. 8.

\[
C_i = \sum_{k=1}^{T} a_{ij} h_{ij} \tag{37}
\]

\[
a_{ij} = \frac{\exp(\beta(s_j, h_i))}{\sum_{k=1}^{T} \exp(\beta(s_j, h_i))} \tag{38}
\]

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

**F. OVERALL STRUCTURE**

The EEMD-SVR-EDA (ESE) model can be used to train and predict the clearness factor degradation curve. The addition of SVR to EDA (Encoder-Decoder based Attention) not only improves the accuracy of EDA but also reduces the amount of calculation. The overall algorithm framework is shown in Fig. 9:

The main steps of the prediction are as follows:

Step1: The original degradation data is denoised and smoothed using wavelet threshold denoising theory.

Step2: Use EEMD to decouple the preprocessed clearness factor degradation signal to obtain a residual component and

Step3: Use SVR and EDA to train the low-frequency and high-frequency components respectively and perform rolling prediction with limited steps.

Step4: Add the prediction results of each component.

The multi-step time series prediction task is to predict future data change by learning historical data. The training set is generally constructed by the method of ‘sliding time window’ [31], as shown in Fig. 10, this method can be better capture sudden increases and dips in time series. Embedding this method into the ESE model can effectively improve prediction accuracy. \(\alpha\) is the total length of input and output, \(\lambda\) is the size of each window sliding along the time axis. In fact, the different lengths of ‘sliding window’ will affect the final prediction results, longer or shorter window lengths may not get the best prediction results. In this paper, through a lot of experimental tests, we set \(\alpha = 25, \lambda = 20\) (When in the five-step predictions), \(\alpha = 23, \lambda = 20\) (three-step prediction).

**IV. EXPERIMENTAL VERIFICATION**

In order to verify the feasibility and effectiveness of the proposed model, we use economizer, low-temperature superheater, high-temperature superheater, and reheater data sets. Then set two kinds of rolling prediction steps to verify the matching problem between rolling prediction steps and soot blowing preparation operation time. In addition, the attention heat map is used to show the rationality of introducing the attention mechanism, and the role of the EEMD algorithm in prediction is analyzed. Finally, the proposed model is compared with more advanced models to verify its superiority.

**A. DATASET DESCRIPTION AND PREPROCESSING**

The dataset in this article comes from the economizer, low-temperature superheater, high-temperature superheater, and
reheater in the coal-fired boiler of Guizhou thermal power station as examples (the daily monitoring value of the clearness factor of the four heated surfaces). The daily change data of the clearness factor comes from the unit’s DCS system. It is necessary to denoise the data because the original clearness factor degradation curve has strong noise, which not only adversely affects our prediction results but also increases the difficulty of prediction. The clearness factor degradation curve of each device denoised by wavelet threshold is shown in Fig. 11. Since this article is mainly about the prediction of the ash accumulation on the heated surface in the future, the ash accumulation section of each device should be extracted.

After denoising the data, feature extraction helps increase the prediction accuracy. After using the EEMD algorithm, the residual component generally represents the low-frequency component, and the IMF component represents the high-frequency component. Due to the long-existing ash deposition and erosion, the economizer clearness factor time series still has strong nonlinear and nonstationary characteristics after denoising, and it can be regarded as a mixed signal with multi-scale components. The way to solve this problem is to use EEMD for multi-scale prediction. EEMD decomposes the denoised signal into an overall degradation curve and multiple frequency-domain stable components. Low-frequency components and high-frequency components have different prediction difficulties. Generally, low-frequency components have the characteristics of lower prediction difficulty than high-frequency components due to their fixed trend changes and low volatility. However, many proposed methods use the same method to predict them, which may cause problems of low prediction accuracy and poor prediction stability. In this article, we give different prediction models to the decomposed high and low-frequency components, which not only reduces the amount of calculation but also improves the prediction accuracy.

B. EVALUATION METRICS

The quality of a model is generally judged by model evaluation indicators. For the dataset used in this article,
we introduce Root Mean Square Error (RMSE), Mean Square Error (MAPE), and Mean Absolute Error (MAE) as model evaluation indicators. These three indicators are widely used in the evaluation of classification and regression effects in machine learning and deep learning, and their mathematical expressions are shown by (39-41):

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (N_{\text{true},i} - N_{\text{predicted},i})^2}$$  \hspace{1cm} (39)

$$\text{MAPE} = \sum_{i=1}^{n} \left| \frac{N_{\text{true},i} - N_{\text{predicted},i}}{N_{\text{true},i}} \right| \ast \frac{100}{n}$$  \hspace{1cm} (40)

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |N_{\text{true},i} - N_{\text{predicted},i}|$$  \hspace{1cm} (41)

where $N_{\text{true},i}$ and $N_{\text{predicted},i}$ respectively represent the true value and predicted value of the clearness factor at the $i$th moment.

C. VARIOUS HEATED SURFACE ANALYSIS RESULTS

The experimental environment is as follows: We use the open-source machine learning library Scikit-learn and the deep learning library Keras to implement the model. The programming language is python 3.7. The experimental hardware environment is CPU Intel(R) Core (TM) i7-9752H CPU @ 2.60GHz 2.59 GHz.
In this experiment, the parameter configurations of SVR and EDA models are shown in Tab 1. The first 60% of the experimental data is used as the training set, and the last 40% is used as the test set. The prediction method is limited-step rolling prediction, and the number of rolling prediction steps is three steps. Use RMSE and MAPE as model error evaluation indicators. The training set is established by a sliding window method to extract historical data information and features. In addition, SVR not only has a better prediction effect in predicting low-frequency components but also has a small amount of calculation and greatly improves the running speed.

Some variant methods are used to compare the performance effects of multi-step rolling prediction. Including EEMD-EDA, EEMD-ED, EDA, ED. In the entire comparison experiment process, it is necessary to keep the network parameters and structure configuration the same. Both EDA and ED models use BI-LSTM and LSTM as the encoder and decoder models. The loss function, optimizer, epoch, and batch size are all consistent with the parameters of the proposed method (same as Tab 1).

The economizer, low-temperature superheater, high-temperature superheater, and reheater’s heated surface cleanliness predictions are shown in Fig. 12, Fig. 13, Fig. 14,
Fig. 15. It can be seen from the figure that the proposed hybrid model (ESE) has good adaptability in the ash prediction of the heated surface of the four boiler installations.

In addition, unlike the proposed model, the EEMD-ED model shows different prediction effects under different datasets. For example, the prediction effect of the EEMD-ED model applied to the economizer and reheater is seriously deviated from the real situation, but the opposite is true for the low-temperature superheater. For the EEMD-EDA model, its prediction effect in the economizer data set has many burrs, and the prediction curve is not smooth, which may be due to the randomness and instability of the prediction, but the overall effect is good.

In order to further explore the role of the attention mechanism in the test set, the attention weight is generally visualized. Fig. 16, Fig. 17, Fig. 18, Fig. 19 is the visualization result pictures (take five-step prediction as an example), the abscissa is the number of hidden layer outputs of the input encoder, and the ordinate is the number of multi-step prediction steps. Combined with the calculation formula of the attention mechanism (see formula 37-38), the predicted value at the next moment will be obtained from the BI-LSTM hidden state of the previous $k$ steps. Under the action of the attention mechanism, when mining the potential information and long-term dependence of the time series, the $k$ step hidden state will be given different weights. The attention heat
map visualizes the weights assigned to the $k$ step hidden state when calculating the attention vector. It can be found in the visualization figure that in the prediction stage of the residual component, more attention is paid to the latest moment, and less attention is paid to the older historical information, which is also in line with intuitive understanding. However, in the IMF1 (The highest frequency part in IMF) prediction stage, the information provided by the attention distribution has no regularity, which may be caused by the irregularity of high-frequency components. However, it can be seen from the above experimental results that the attention mechanism can still mine long-term dependencies in information.

**D. FINITE STEP PREDICTION**

In this part, we will analyze the prediction effect of various variant models and different rolling prediction steps. Tab 2. a and Tab 2. b shows the prediction errors (RMSE and MAPE) of the five models in two multi-step predictions (3 steps and 5 steps). It can be seen from the tables that the prediction error of our proposed model is smaller than that of any comparison model, which significantly improves the prediction accuracy. In addition, it is found that some variant methods have different adaptability and errors on different datasets, such as the EEMD-ED model. Among them, we give five-step prediction figures for the low-temperature superheater.
TABLE 2. (a) A comparison of the average prediction error (RMSE and MAPE) of our model (ESE) with other variant models on four datasets (three-step). (b) A comparison of the average prediction error (RMSE and MAPE) of our model (ESE) with other variant models on four datasets (five-step).

| Models  | Three-step prediction error table for each device | Five-step prediction error table for each device |
|---------|-----------------------------------------------|---------------------------------------------|
|         | Economizer RMSE MAPE LTS RMSE MAPE HTS RMSE MAPE Reheater RMSE MAPE | Economizer RMSE MAPE LTS RMSE MAPE HTS RMSE MAPE Reheater RMSE MAPE |
| ESE     | 0.00298 0.00340 0.00123 0.00161 0.00111 0.00113 0.00078 0.00147 | |
| EEMD-EDA| 0.00308 0.00374 0.00176 0.00256 0.00108 0.00110 0.00083 0.00158 | |
| EEMD-ED | 0.00422 0.00540 0.00226 0.00332 0.00147 0.00169 0.00214 0.00359 | |
| EDA     | 0.00856 0.01114 0.00481 0.00746 0.00245 0.00274 0.00519 0.00794 | |
| ED      | 0.00568 0.00691 0.00805 0.01300 0.00314 0.00365 0.00160 0.00205 | |

Note: LTS and HTS are the abbreviations of low-temperature superheater, and high-temperature superheater respectively.

FIGURE 16. Attention heat map of low-frequency and high-frequency components (IMF1) of the economizer.

FIGURE 17. Attention heat map of low-frequency and high-frequency components (IMF1) of low-temperature superheater.

FIGURE 18. Attention heat map of low-frequency and high-frequency components (IMF1) of high temperature superheater.

FIGURE 19. Attention heat map of the low-frequency and high-frequency components (IMF1) of the reheater.

and reheater, as shown in Fig. 20, Fig. 21. Under the 3-step prediction, the problem caused by this model is not significant due to the small number of prediction steps. But there is a big deviation from the true value under the 5-step prediction. This problem is particularly evident in the reheater (Fig. 21(b)). Furthermore, if the low-frequency components are predicted

by SVR instead of EDA, although the accuracy of the model is not greatly improved, the complicated calculation process and calculation time of deep learning can be avoided. In addition, this multi-step predicting method can reserve 5 minutes for the soot-blowing preparation operation, which is sufficient.

Based on the finite-step predicting method, we can find from the predicting figures that all the predicting trends except for the economizer-based three-step prediction all
have the situation of ‘floating upward’, which is actually good. If the soot blowing threshold is introduced as the starting point of soot blowing, then the ‘floating downward’ is similar to an over-soot blowing situation, causing the soot blowing operation to proceed ahead of schedule. If the heated surface of the boiler is in this state for a long time, it will cause high-temperature corrosion of the heated surface or pipeline, which will make unnecessary economic waste. The ‘floating upward’ is similar to the under-blowing situation. Although it may cause some heat transfer losses, it is much better than the previous case.

The EEMD algorithm ensures the smoothness and accuracy of the prediction curve. Fig. 22 shows that the two comparative experiments are EEMD-EDA, EDA, and EEMD-ED, ED (three-step prediction). It can be seen from the figure that the prediction error increases significantly without the participation of the EEMD algorithm. It is worth noting that in the error analysis table (Tab 2) when analyzing the situation.
of the reheater (five-step prediction), the prediction error even appears ‘chaotic’, as if the attention mechanism has a negative effect in the prediction, which also verifies the ability of the EEMD algorithm to predict nonlinear and non-stationary time series. In fact, this situation also occurs on other datasets.

In order to further verify the superiority of the model, different methods were implemented for comparison. RNN and LSTM, as a major branch of deep learning, are widely used in time series prediction due to the cyclic mechanism. LSTM adds a gate mechanism on the basis of RNN to make it more advantageous in mining long-term dependencies (GRU as a neural network with functions similar to LSTM, it reduces the computational complexity, so it is not within the scope of this article). In order to maintain consistency with the proposed method, we also use the EEMD algorithm for comparison methods, and the Epoch and Batch-size is given by Tab 1 are both applied to the model. In addition, we use a single-layer RNN or LSTM network and add two fully connected layers (the number of
neurons is 100 and 1 respectively). In addition, in terms of data sets, we use economizers and low-temperature superheaters to verify. The prediction result graph is given by Fig. 23 and Fig. 24.

It can be clearly seen from the figures that our proposed model has the best effect compared to other comparison methods, which also verifies the advantages of the EDA model in ash prediction of various heating areas of coal-fired power plant boilers. Under the effect of the decomposition algorithm of EEMD, both RNN and LSTM models are better for predicting the volatility of the curve, but the proposed model is closer to the real curve. In all the result figures, both EEMD-LSTM and EEMD-RNN will produce a “floating downward” as the prediction time continues to advance, and this situation is similar to the “overblowing” mentioned above. The situation has an adverse effect on the long-term operation of the heated surface.

In Tab 3, we can see that, whether in the economizer or in the low-temperature superheater, the proposed model has the smallest error in all evaluation indicators relative to other models. Compared with EEMD-LSTM and EEMD-RNN under the three-step prediction of the economizer dataset, the proposed model has increased by 40.7% and 66.4% respectively, and increased by 30.1% and 63.29% under the five-step prediction (in RMSE), the situation is similar in MAPE and MAE. In summary, the superiority of the ESE model is verified, and the adaptability and accuracy of the proposed ESE model in fouling prediction under different devices are proved.

As the number of prediction steps continues to increase, although long-term prediction methods can reserve more time for soot blowing operations compared with short-term prediction methods, their final prediction results will lose practical engineering significance due to their serious accumulation of...
errors. From the previous analysis, it can be seen that the prediction results are very consistent with the early real value of the prediction, and the prediction error becomes larger and larger as time goes by. In fact, by reasonably adjusting the number of rolling prediction steps, the prediction accuracy and soot blowing retention time can reach the best state.

V. SUMMARY

In order to solve the problems of energy waste and industrial safety accidents caused by unreasonable soot blowing time in coal-fired power stations at this stage, we propose a hybrid model to predict the future health of the boiler heated surface. First, the wavelet threshold denoising method can reduce the unnecessary glitch and noise of the original ash signal. Secondly, the implantation of EEMD decomposition can solve the problem of high prediction difficulty of nonlinear and non-stationary time series. Finally, a hybrid learning model of SVR and EDA is used to predict the future ash accumulation situation. According to the prediction results of four common heated surfaces, the proposed model has the lowest prediction error. Furthermore, we use the variant types of the proposed model and two advanced methods as comparative experiments in this paper. The experimental results verify the accuracy and superiority of the proposed model. In addition, because the prediction model involves in deep learning models, its prediction accuracy is closely related to the adaptability of the dataset itself and parameter adjustment. How to obtain a higher and more stable prediction ability to meet the requirements of ‘early warning’ will be the focus of future research. On the basis of this paper, for the energy loss problem, a soot blowing optimization model can be established (heat transfer of the heating surface and the steam required for soot blowing are the optimization targets). The prediction of ash accumulation will play an important role in soot blowing optimization decision.
REFERENCES

[1] M. Aldrich, “The rise and decline of the kerosene kitchen: A neglected energy transition in rural America,” *Agricult. Hist.*, vol. 94, pp. 24–60, Jan. 2020.

[2] B. Peña, E. Teruel, and L. Díez, “Soft-computing models for soot-blowing optimization in coal-fired utility boilers,” *Appl. Soft Comput.*, vol. 11, no. 2, pp. 1657–1668, Mar. 2011.

[3] S. Tang, M. Li, F. Wang, Y. He, and W. Tao, “Fouling potential prediction and multi-objective optimization of a flue gas heat exchanger using neural networks and genetic algorithms,” *Int. J. Heat Mass Transf.*, vol. 152, p. 15, May 2020.

[4] B. Wu, X. Bai, W. Liu, S. Lin, S. Liu, L. Luo, Z. Guo, S. Zhao, Y. Lv, C. Zhu, Y. Hao, Y. Liu, J. Hao, L. Duan, and H. Tian, “Non-negligible stack emissions of noncriteria air pollutants from coal-fired power plants in China: Condensable particulate matter and sulfur trioxide,” *Environ. Sci. Technol.*, vol. 54, no. 11, pp. 6540–6550, Jun. 2020.

[5] K. M. Zierold and C. Odoh, “A review on fly ash from coal-fired power plants: Chemical composition, regulations, and health evidence,” *Rev. Environ. Health*, vol. 35, no. 4, pp. 401–418, Nov. 2020.

[6] D. K. Mohanty and P. M. Singru, “Fouling analysis of a shell and tube heat exchanger using local linear wavelet neural network,” *Int. J. Heat Mass Transf.*, vol. 77, pp. 946–955, Oct. 2014.

[7] Y. Shi, J. Wang, and Z. Liu, “On-line monitoring of ash fouling and soot-blowing optimization for convective heat exchanger in coal-fired power plant boiler,” *Appl. Thermal Eng.*, vol. 78, pp. 39–50, May 2015.

[8] L. Perez, B. Ladevie, P. Tochin, and J. C. Batsale, “A new transient thermal fouling probe for cross flow tubular heat exchangers,” *Int. J. Heat Mass Transf.*, vol. 52, nos. 1–2, pp. 407–414, Jan. 2009.

[9] Y. Shi, J. Wen, F. Cui, and J. Wang, “An optimization study on soot-blowing of air preheaters in coal-fired power plant boilers,” *Energies*, vol. 12, p. 15, Jan. 2019.

[10] Y. Shi, Q. Li, J. Wen, F. Cui, X. Pang, J. Jia, J. Zeng, and J. Wang, “Soot blowing optimization for frequency in economizers to improve boiler performance in coal-fired power plant,” *Energies*, vol. 12, p. 19, Jan. 2019.

[11] S. Tong, X. Zhang, Z. Tong, Y. Wu, N. Tang, and W. Zhong, “Online ash fouling prediction for boiler heating surfaces based on wavelet analysis and support vector regression,” *Energies*, vol. 12, p. 20, Jan. 2020.

[12] L. Mao, J. Xu, J. Chen, J. Zhao, Y. Wu, and F. Yao, “A LSTM-STW and GS-LM fusion method for lithium-ion battery RUL prediction based on EEMD,” *Energies*, vol. 13, p. 13, Jan. 2020.

[13] N. Bokde, A. Feijóo, and K. Kulat, “Analysis of differencing and decomposition preprocessing methods for wind speed prediction,” *Appl. Soft Comput.*, vol. 71, pp. 926–938, Oct. 2018.

[14] G. Zhang, H. Liu, J. Zhang, Y. Yan, L. Zhang, C. Wu, X. Hua, and Y. Wang, “Wind power prediction based on variational mode decomposition multi-frequency combinations,” *J. Mod. Power Syst. Clean Energy*, vol. 7, no. 2, pp. 281–288, Mar. 2019.

[15] Q. Li, P. Yan, J. Liu, Y. Shi, and D. Yu, “Prediction of pollution state of heating surface in coal-fired utility boilers,” *IEEE Access*, vol. 8, pp. 206132–206145, 2020.

[16] C. Kumar and R. Prakash, “Ultrasound medical image denoising using threshold based wavelet transformation method,” *J. Med. Imag. Health Informat.*, vol. 10, no. 8, pp. 1825–1830, Aug. 2020.

[17] B. Xie, Z. Xiong, Z. Wang, L. Zhang, D. Zhang, and F. Li, “Gamma spectrum denoising method based on improved wavelet threshold,” *Nucl. Eng. Technol.*, vol. 52, no. 8, pp. 1771–1776, Aug. 2020.

[18] Z. Qian, Y. Pei, H. Zareipour, and N. Chen, “A review and discussion of decomposition-based hybrid models for wind energy forecasting applications,” *Appl. Energy*, vol. 235, pp. 939–953, Feb. 2019.

[19] Y. Ji, X. Wang, Z. Liu, Z. Yan, L. Jiao, D. Wang, and J. Wang, “EEMD-based online milling chatter detection by fractal dimension and power spectral entropy,” *Int. J. Adv. Manuf. Technol.*, vol. 92, p. 1185, Sep. 2017.

[20] X. Zhang, J. Li, X. Liu, and Z. Chen, “Improved EEMD-based standardization method for developing long tree-ring chronologies,” *J. Forestry Res.*, vol. 31, no. 6, pp. 2217–2224, Dec. 2020.

[21] N. Bokde, A. Feijóo, D. Villanueva, and K. Kulat, “A review on hybrid empirical mode decomposition models for wind speed and wind power prediction,” *Energies*, vol. 12, no. 2, p. 254, 2019.

[22] L. Cai, J. Meng, D.-I. Stroe, J. Peng, G. Luo, and R. Teodorescu, “Multiobjective optimization of data-driven model for lithium-ion battery SOH estimation with short-term feature,” *IEEE Trans. Power Electron.*, vol. 35, no. 11, pp. 11855–11864, Nov. 2020.

[23] I. A. Naguib, E. A. Abdelaleem, E. S. Hassan, N. W. Ali, and M. Gamal, “Partial least squares and linear support vector regression chemometric models for analysis of Norfloxacin and Tinidazole with Tinidazole impurity,” *Spectrochim. Acta A, Mol. Biomolecular Spectrosc.*, vol. 239, p. 8, 2020.

[24] Q. Wang, S. Bu, and Z. He, “Achieving predictive and proactive maintenance for high-speed railway power equipment with LSTM-RNN,” *IEEE Trans. Ind. Informat.*, vol. 16, no. 10, pp. 6509–6517, Oct. 2020.

[25] P. Arora, H. Kumar, and B. K. Panigrahi, “Prediction and analysis of COVID-19 positive cases using deep learning models: A descriptive case study of India,” *Chaos, Solitons Fractals*, vol. 139, Oct. 2020, Art. no. 110017.

[26] F. Zhao, Y. Li, X. Wang, L. Bai, and T. Liu, “Lithium-ion batteries state of charge prediction of electric vehicles using RNN-CNNs neural networks,” *IEEE Access*, vol. 8, pp. 98168–98180, 2020.

[27] Y. Chen, G. Peng, Z. Zhu, and S. Li, “A novel deep learning method based on attention mechanism for bearing remaining useful life prediction,” *Appl. Soft Comput.*, vol. 86, p. 11, Jan. 2020.

[28] J. Qu, F. Liu, Y. Ma, and J. Fan, “A neural-network-based method for RUL prediction and SOH monitoring of lithium-ion battery,” *IEEE Access*, vol. 7, pp. 87191–97178, 2019.

[29] F. Li, Z. Gui, Z. Zhang, D. Peng, S. Tian, K. Yuan, Y. Sun, H. Wu, J. Gong, and Y. Lei, “A hierarchical temporal attention-based LSTM encoder-decoder model for individual mobility prediction,” *Neurocomputing*, vol. 403, pp. 153–166, Aug. 2020.

[30] F. Liu, Y. Lu, and M. Cai, “A hybrid method with adaptive sub-series clustering and attention-based stacked residual LSTMs for multivariate time series forecasting,” *IEEE Access*, vol. 8, pp. 62423–62438, 2020.

[31] S. F. Naqi, S. A. A. Ali, N. Yahya, M. A. Yasin, Y. Hafeez, A. R. Subhani, S. H. Adil, U. M. Al Saggaf, and M. Moinuddin, “Real-time stress assessment using sliding window based convolutional neural network,” *Sensors*, vol. 20, no. 16, p. 4400, 2020.
FANGSHU CUI received the B.Sc. degree from Ludong University, Yantai, China, in 2006, the M.Sc. degree from Taiyuan University of Science and Technology, Taiyuan, China, in 2009, and the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China, in 2018.

She is currently a Lecturer with the School of Data Science and Technology, North University of China, Taiyuan. Her current research interests include evolutionary optimization and machine learning.

JIE WEN received the B.Sc. and Ph.D. degrees from the University of Science and Technology of China, Hefei, China, in 2010 and 2015, respectively.

He is currently a Lecturer with the School of Electrical and Control Engineering, North University of China, Taiyuan, China. His current research interests include optimization of boiler soot blowing, intelligent control, and deep learning.

JIANCHAO ZENG received the B.Sc. degree from Taiyuan Heavy Machinery Institute, Taiyuan, China, in 1982, and the M.Sc. and Ph.D. degrees from Xi’an Jiao Tong University, Xi’an, China, in 1985 and 1990, respectively.

He is currently a Professor with the School of Electrical and Control Engineering, North University of China, Taiyuan, where he is the Vice President. He has published over 200 international journals and conference papers. His current research interests include prognostics and health management, intelligent computation, swarm intelligence, and swarm robotics.

Dr. Zeng is the Vice President of the Technical Committee on System Simulation of the Chinese Association of Automation. He is the Director of China Simulation Federation.

***