A Deep Learning Model for Measuring Oxygen Content of Boiler Flue Gas

ZHENHAO TANG[1], **YANYAN LI**[1], **AND ANDREW KUSIAK**[2], **(Life Member, IEEE)**

[1]School of Automation Engineering, Northeast Electric Power University, Jilin 132012, China
[2]College of Engineering, The University of Iowa, Iowa City, IA 52242, USA

Corresponding author: Andrew Kusiak (andrew-kusiak@uiowa.edu)

ABSTRACT The oxygen content of boiler flue gas is a valid indicator of boiler efficiency and emissions. Measuring the oxygen content of boiler flue gas is time consuming and costly. To overcome the latter shortcomings, a novel deep belief network algorithm based hybrid prediction model for the oxygen content of boiler flue gas is proposed. First, the algorithm is used to build a model based on the historical data collected from the distribution control system. The variables are divided into control variables and state variables to meet the needs of advanced control requirement. Then, a lasso algorithm is used to select variables highly related to the oxygen content as the inputs of the prediction model. Two basic models based on the deep-belief network are established, one using control variables, and the other, state variables. Finally, the two basic models are combined with a least square support vector machine to improve prediction accuracy of the oxygen content of boiler flue gas. To test the accuracy of the proposed algorithm, experiments based on three industrial datasets are performed. Performance of the comparison of the proposed deep belief algorithm is compared with five machine learning algorithms. Computational experience has shown that the model derived with the deep-belief algorithm produced better accuracy than the models generated by the other algorithms.

INDEX TERMS Boiler production, deep belief network, feature selection, oxygen content of flue gas.

I. INTRODUCTION

With growing concerns about environmental protection, combustion optimization has become an important issue in the operation of coal-fired boilers [1]. The oxygen content of flue gas [2] is an important metric of coal-fired boiler combustion operation, and is closely related to boiler combustion efficiency [3]–[5] and NOx emissions [6]. Accurate measurement of the oxygen content of flue gas can help improve boiler combustion efficiency and reduce coal consumption [7].

Many approaches have been suggested to measure the oxygen content of flue gas. They can be mainly grouped into two categories: direct measurement methods [8] and soft measurement methods. The direct methods measure the oxygen content of flue gas by utilizing oxygen sensors, such as thermomagnetic oxygen sensors and ZrO_2 oxygen sensors. The manufacture of these oxygen sensors is based on different principles. They achieve good performance under

The associate editor coordinating the review of this manuscript and approving it for publication was Qichun Zhang.

various applications. However, they suffer from limitations with regard to the balance between measurement precision and manufacturing cost. Furthermore, dysfunctions due to the high temperatures and loud noises characteristic of power plant boilers are common.

Compared with direct measurement methods, soft measurement methods [9] [10], [11] utilize data to train a forecasting model. They are easy to use. Therefore, soft measurement methods have increasingly established themselves as effective and popular methods. Currently, more advanced soft measurement methods are based on data-driven algorithms such as least squares support vector machine (LSSVM) [12], Gaussian process (GP) [13], neural networks [14], radial basis function (RBF) [15], deep belief network (DBN) [16], and so on. For instance, a quantitative transformation fusion model based on a cloud model have been proposed in [17]. They thus used the advantages of the existing sensors and adapted the idea of data fusion to improve the accuracy of oxygen content measurements, thereby improving the combustion efficiency of the boiler. In [18], the authors established a combustion process model based on a data-driven method and

proposed a multiple-model-based fuzzy predictive control algorithm to predict the oxygen content of flue gas. A data-driven method with fuzzy c-means clustering and subspace identification was applied to identify the model parameters. In [19], the authors proposed a hypergraph (HG) feature extraction technique to describe the relations among the data containing feature vectors and developed a PNN model to classify the frequency responses of the sensors which can correspond to various mass concentrations of flue gas by detecting the nonlinearity. The capability of HGPNN has been validated with datasets. The authors in [13] used GP to build a NOx emissions model for a 330 MW tangentially fired boiler. They compared a 13-input GP model and a 21-input GP model using 670 sets of production data from fired boilers. This 13-input GP model could achieve reasonable accuracy and provide the optimum operation parameters for reducing NOx emissions. These approaches achieved good experimental results, indicating the feasibility of soft measurement methods. However, they can hardly extract deep information because of the limitations posed by their shallow network structures.

Consequently, deep learning of has attracted much interest. In recent years, deep learning have been successfully applied in numerous fields, such as fault diagnosis [20], pattern recognition [21], and image processing [22]. Compared with other modeling methods, deep learning has several advantages; it can realize network self-learning, obtain deep feature information contained in input data, facilitate faster convergence speeds, and improve prediction accuracies [23]. Therefore, this study used a DBN to model the oxygen content of boiler flue gas.

Besides the network type, the input selection has a significant impact in the modeling process [24]. To reduce the complexity of the prediction model and improve its accuracy and efficiency, several input variable selection methods have been applied. Examples include classification and regression trees (CARTs) [25], kernel principal component analysis (KPCA) [26], and the lasso [27], However, selection of the parameters of CART and KPCA requires much experimentation and practical experience, which is not conducive to the application of the algorithm. Lasso offers an advantage here; it can function with fewer parameters, accurately select variables that are strongly related to the target variable, and reduce the dimensions of the input variables. Therefore, the lasso algorithm is utilized for features selection in this study.

Combining all these improvements, a nonlinear combined DBN method is proposed in this paper. Firstly, the input variables are divided into control variables and state variables, and the most relevant state variables that are selected as the input variables using the lasso method. Then, the control variables and selected state variables are modeled as inputs to obtain the control prediction and the state prediction model, respectively. Finally, the two models are nonlinearly combined to obtain the nonlinear combined deep belief network (NCDBN) model.

To implement the proposed approach, the remainder of this paper is organized as follows. Section 2 describes the whole framework of the proposed approach and its novelties. Section 3 presents the steps required to build the nonlinear combined model. Section 4 presents the proposed model forecasting experiments and a related discussion. Section 5 concludes this study.
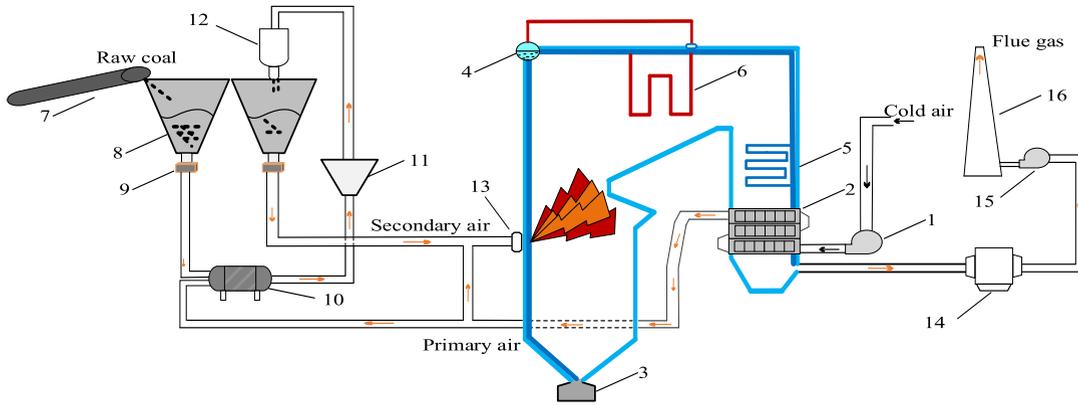
II. ANALYSIS OF BOILER PROCESS VARIABLES

The oxygen content of flue gas is closely related to the production process of the boiler, which has many characteristics, including multiple parameters and the fact that the process is nonlinear. In this study, process variables are divided into two categories according to the actual process characteristics and the causes of variation. The first category involves control variables, which can be controlled by technical operation such as changes in fuel quantity and valve opening. This type of variable can be directly operated by workers to control the variations in the combustion process. The second category involves state variables, which reflect the state characteristics of the coal-fired boiler, such as furnace pressure, furnace temperature, and superheated steam temperature. This type of variable cannot be directly controlled by workers; the variations in the state variables can only be realized indirectly by changing the control variables. In addition, this division of process variables facilitates the application of advanced control algorithms. The change in the combustion process is caused by controllable factors such as modifications to fuel quantity and valve opening alterations, which can be quickly relayed to workers to ensure their safety during the combustion process. This aspect can also improve the combustion efficiency of coal-fired boilers. The oxygen content of flue gas reflects the combustion characteristics of the boiler. As the input data changes, the model changes accordingly, so the oxygen content of flue gas is modelled as a dynamic model.

Fig. 1 depicts the production process of coal-fired boilers. There are 24 process variables, such as unit load, fuel quantity, exhaust temperature, induced draft valve opening, blower baffle opening, and furnace temperature. Among them, the control variables include fuel quantity, primary air flow, secondary air flow, blower flow, total air flow, water supply flow, induced draft valve opening, blower baffle opening, primary fan valve opening, and secondary fan valve opening. The state variables include main steam flow, main steam pressure, main steam temperature, unit load, feed water temperature, flue gas pressure, exhaust temperature, furnace temperature, furnace negative pressure, feed water pressure, primary air temperature, secondary air temperature, reheat steam temperature, and superheated steam temperature.

III. OXYGEN CONTENT IN FLUE GAS PREDICTION MODEL
A. DATA PRE-PROCESSING
The operation process data of the coal-fired boiler affect the prediction accuracy of the model. As the ranges of the variations of the boiler operating parameters are large,

1 - blower, 2 - air preheater, 3 - ash bucket, 4 - drum, 5 - fuel economizer, 6 - superheater, 7 - coal handing system, 8 - coal bunker, 9 - coal feeder, 10 - coal mill, 11 - coarse separator, 12 - cyclone separator, 13 - burner, 14 - ash separator, 15 - induced draft fan, 16 – chimney

FIGURE 1. Production process of coal-fired boilers.

the magnitudes of the operating parameters can be very different, leading to a reduction in the accuracy of the model. Thus, the model will not be able to accurately reflect the relationships among the variables. Therefore, the original process data need to be normalized so as to eliminate the influence of magnitude variations within the target parameters before beginning the modeling. The data pre-processing is performed with the Z-score method, as shown in formula (1):

$$z_i^* = \frac{z_i - \mu}{\sigma} \tag{1}$$

where z_i^* represents the value of the parameter after normalization by the Z score, and z_i is the original process data, μ is the mean of the sampled data, σ is the standard deviation, and i is the number of samples.

B. SELECTION OF INPUT VARIABLES

The input variables directly affect the computational cost and prediction accuracy of the model. Too many input variables will decrease the prediction accuracy and increase the computational time. Therefore, it is necessary to remove redundant variables before beginning the modeling. The lasso algorithm [28] can achieve dimension reduction and improve modeling accuracy. The lasso-based feature selection method aims to eliminate the redundant variables by minimizing the sum of the squared residuals, which is described using formula (2).

$$\arg\min\{\sum_{i=1}^{n}(y_i - \sum_{j=1}^{m}x_{ij}\beta_j)^2\}$$

$$s.t. \sum_{j=1}^{m}|\beta_j| \leq t \tag{2}$$

where x_{ij} denotes the independent variables, y_i is the dependent variable, and β_j is the regression coefficient of the jth variable. The threshold t is a paradigm penalty for the regression coefficient, and the value of t ranges from 0 to $+\infty$. When t is small, the coefficients of some redundant variables are compressed to 0, resulting in a more compact model.

TABLE 1. Process variables.

Variable	Name	Unit
State variable	Furnace negative pressure	MPa
	Reheat steam temperature	°C
	Furnace temperature	°C
	Main steam flow	t/h
	Main steam temperature	°C
	Unit load	MW
	Exhaust temperature	°C
	Fuel quantity	t/h
	Primary air flow	t/h
	Secondary air flow	t/h
Control variable	water supply flow	t/h
	blower flow	t/h
	induced draft valve opening	%
	blower baffle opening	%

Lasso is a variable selection method based on a linear model, which can reflect the correlation between several input variables and response variables. Compared with other feature selection methods, lasso-based feature selection can accurately select input variables with strong correlation, and also has the stability of variable selection. Therefore, lasso have significance for variable selection. The control variables and state variables are selected according to a lasso-based feature selection algorithm, respectively. 7 control variables and 7 state variables were chosen as input variables for establishing the control prediction model and the state prediction model. Detailed information about the variables is presented in Table 1. In addition, a time lags analysis was performed on the actual production data of multiple research objects. The correlation between the variables with different time lags and the prediction variable showed no significantly difference.

C. MODELING PROCESS BASED ON THE DEEP BELIEF NETWORK

1) DEEP BELIEF NETWORK

DBN [29] is a neural network based on brain neuron in which multiple restricted Boltzmann machines (RBMs) are stacked one by one to realize a back propagation (BP) neural network. The training method of the DBN firstly adopts unsupervised pre-training to initialize the parameters of the DBN model layer by layer. The data are inputted into the bottom layer of the DBN, which is the first visible RBM layer. Then, supervised fine-tuning is used to optimize the network structure. The steps of DBN algorithm modeling are as follows:

Step 1: Divide the processed data into training and testing sets, and input the training sets at the bottom of the DBN.

Step 2: The DBN performs unsupervised pre-training. Randomly initialize the parameters of the network and set the layer node number of the DBN network, the maximum number of layers being m. The probability distribution of a single RBM can be considered as an energy function E(v, h), which is expressed as (3):

$$E(v, h|\theta) = -\sum_i^m a_i v_i - \sum_j^n b_j h_j - \sum_i^m \sum_j^n v_i \omega_{ij} h_j \quad (3)$$

where $\theta = (\omega, a, b)$ denotes the model parameters, and vi and h_j represent the states of the ith visible layer neuron and the jth hidden layer neuron. ω_{ij} denotes the weight vector between the ith visible layer neuron and the jth hidden layer neuron, and a and b are the biases.

Step 3: Calculate the activation probabilities of the visible layer neuron and the hidden layer neuron as follows:

$$P(h_j = 1/v, \theta) = f(b_j + \sum_i \omega_{ij} v_i) \quad (4)$$

$$P(v_i = 1/h, \theta) = f(a_i + \sum_j \omega_{ij} h_j) \quad (5)$$

where $f(x) = 1/(1 + \exp(-x))$ is the sigmoid function, and θ can be calculated by training the RBM with the contrast divergence algorithm.

Step 4: Update the weight vector and bias of the network between the visible layer and the hidden layer. The formula for updation can be inferred from (6)–(8).

$$[\omega_{ij}]_{n+1} = \lambda[\omega_{ij}]_n + \eta(<v_i h_j>_{data} - <v_i h_j>_{\text{mod } el}) \quad (6)$$

$$[a_i]_{n+1} = \lambda[a_i]_n + \eta(<h_j>_{data} - <h_j>_{\text{mod } el}) \quad (7)$$

$$[b_j]_{n+1} = \lambda[b_j]_n + \eta(<v_i>_{data} - <v_i>_{\text{mod } el}) \quad (8)$$

where $<\cdot>_{data}$ represents the expectation of data distribution, and $<\cdot>_{\text{mod } el}$ represents the expectation of model distribution. λ is momentum, and η is the learning rate.

Step 5: Each RBM must be fully and successively trained and stacked till the maximum number of layers of the DBN is obtained. Then, apply supervised fine-tuning to modify the weight vector between the visible layer and the hidden layer using the BP method. Finally, the prediction model is established.

2) NONLINEAR COMBINED PREDICTION MODEL BASED ON THE DEEP BELIEF NETWORK

During the operation of the coal-fired boiler, the oxygen content of the flue gas can be influenced by adjusting the total air flow, blower baffle opening, and so on. State variables, such as exhaust temperature and furnace temperature, also affect or reflect the oxygen content of the flue gas. The inputs of the control prediction model are fuel quantity, primary air flow, secondary air flow, blower flow, water supply flow, induced draft valve opening, and blower baffle opening, and the inputs of the state prediction model are furnace negative pressure, reheat steam temperature, furnace temperature, main steam flow, main steam temperature, unit load, exhaust temperature. The outputs of the control prediction model and the state prediction model are the oxygen content of flue gas. The DBN algorithm is applied to establish the control prediction model and state prediction model. The nonlinear combination of the control and state prediction models can provide the final combined prediction model, which reflects the influence of different operation parameters on the oxygen content of the flue gas, and allows us to obtain a more accurate prediction model.

In this study, the LSSVM is implemented to construct the final NCDBN prediction model.

First, the original process datasets are divided into training datasets and testing datasets. The training datasets are utilized for training the control and state prediction models. The control prediction model and the state prediction model are trained separately by using the training datasets.

Then, supposing that the predicted results of the two sub-models are fc_g and fs_g, respectively. fc_g, fs_g and the real values of oxygen content of flue gas f_r are combined into a new training datasets for the nonlinear combination. The control prediction model and the state prediction model are combined by using the new training datasets.

Finally, the LSSVM is applied to construct the nonlinear combined forecasting model. The testing datasets are used to verify the prediction accuracy of the nonlinear combined forecasting model and to store the model parameters.

$$fn_g = f(fc_g, fs_g, w, b) \quad (9)$$

where fn_g represents the predicted values of nonlinear combined prediction model, fc_g denotes the predicted values of the control prediction model based on the DBN, fs_g refers to the values predicted by the state prediction model based on the DBN, w is the weight vector of the nonlinear combined prediction model, and b is the bias of the nonlinear combination prediction model.

D. ERROR METRICS AND ALGORITHM FLOW

To evaluate the performance of the prediction model, three error metrics, namely mean square error (*MSE*), mean relative error (*MRE*), and mean absolute error (*MAE*), are used to measure the performance of each prediction model. The three
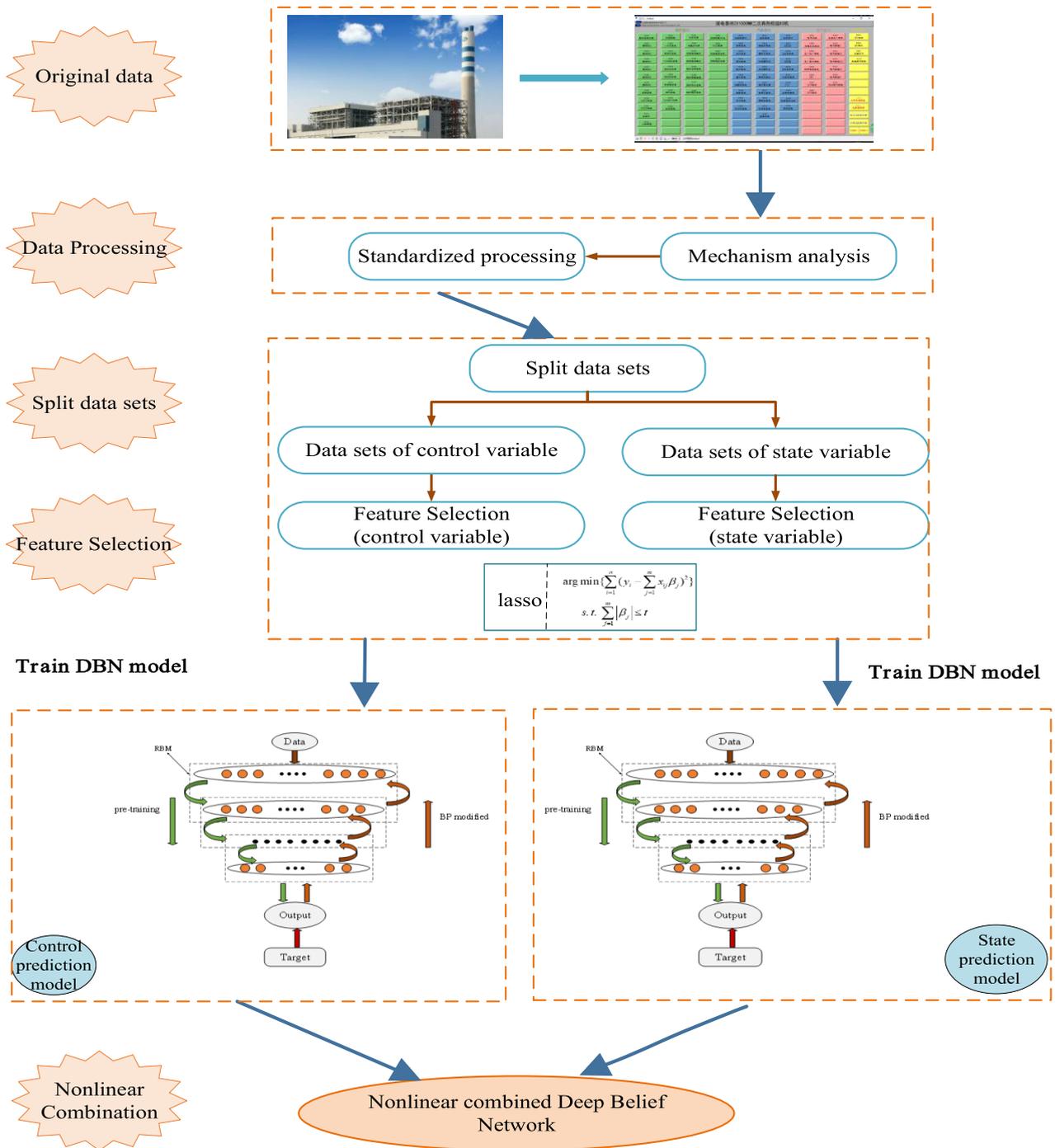
FIGURE 2. Flow chart of the NCDBN algorithm.

error metrics are shown in (10)–(12).

$$MSE = \frac{1}{N}\sum_{i}^{N}(Y_i - \hat{Y}_i)^2 \qquad (10)$$

$$MRE = \frac{1}{N}[\sum_{i=1}^{N}\frac{\left|Y_i - \hat{Y}_i\right|}{Y_i}] \times 100\% \qquad (11)$$

$$MAE = \frac{1}{N}\sum_{i}^{N}\left|Y_i - \hat{Y}_i\right| \qquad (12)$$

where N is the sample number of the testing datasets, Y_i is the measured value of the oxygen content of the flue gas, and $\hat{Y}_i$ is the predicted value of the oxygen content of the flue gas.

The overall flow of the algorithm is shown in Fig. 2. The algorithm comprises the following four main steps.
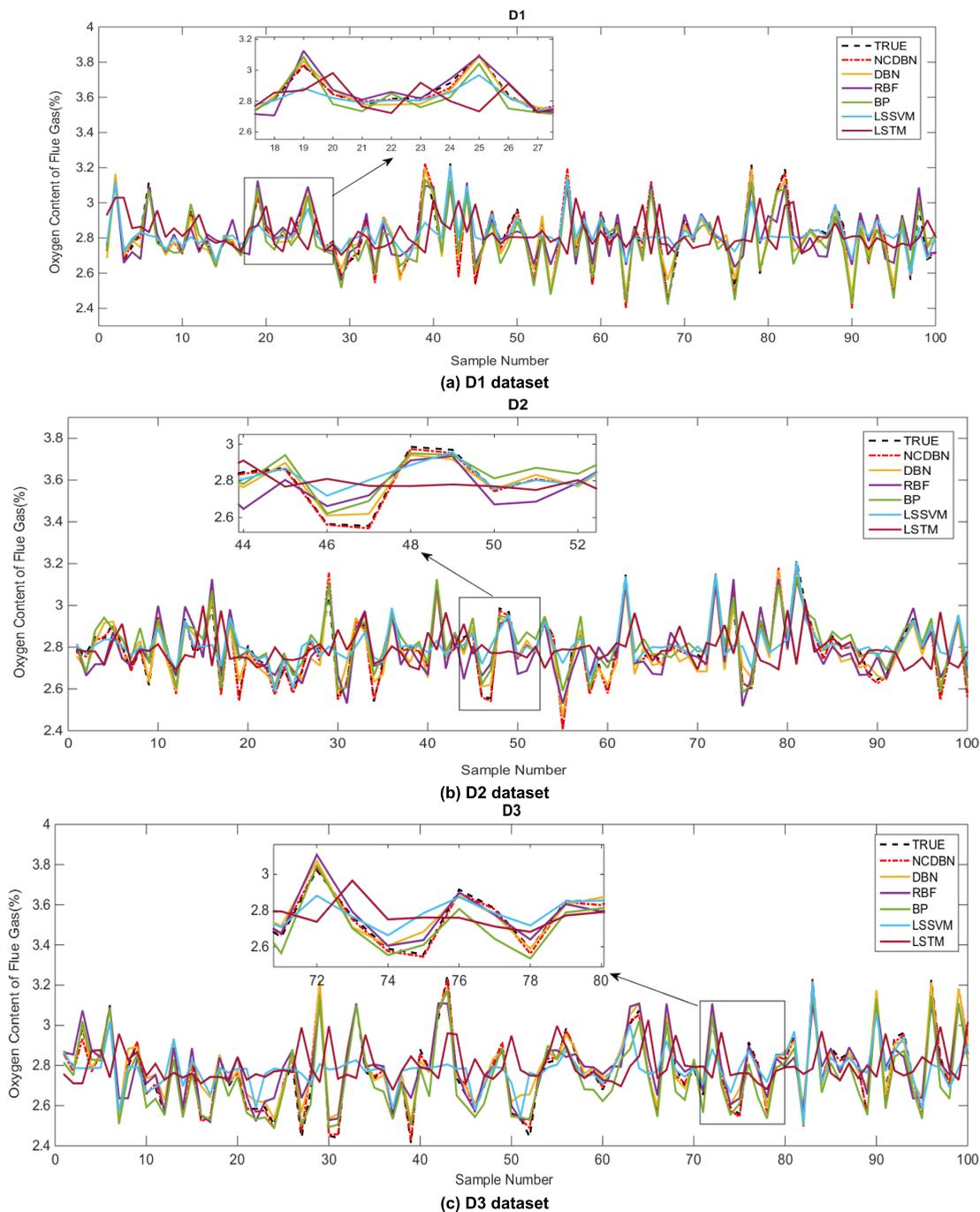
FIGURE 3. Predicted results of different models.

Step 1: This step involves data pre-processing. Standardize the original process data using the *Z*-score method. The details of the same are described in Section A.

Step 2: Divide the input variables into the control variables and state variables. The features selection of the control variables and the state variables using the lasso method are described in Section B.

Step 3: Train the DBN model using the training datasets of the control variables and state variables, and obtain the control prediction model and the state prediction model, respectively. The specific methods are explained in Section C.

Step 4: Obtain the NCDBN model by nonlinear combination of the control and state prediction models. For the nonlinear combination method, see Section 2).

IV. CASE STUDY AND DISCUSSION

To assess the validity of the proposed model, experiments based on the original process data were carried out.

TABLE 2. Information about the experimental data.

Dataset	D1	D2	D3
Training datasets	350	350	350
Training datasets for nonlinear combination	150	150	150
Testing datasets	100	100	100
Data acquisition times	01.10.2016 00:50–10:50	01.15.2016 06:00–16:00	02.16.2016 10:30–20:30

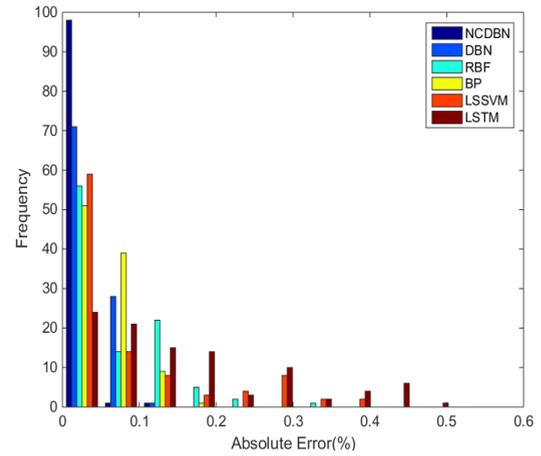**TABLE 3.** Results of feature selection.

Error metrics	NCDBN model	
	With lasso	Without lasso
MAE	0.0146	0.0377
MRE	0.5151	1.3741
MSE	0.0005	0.0023
Time (s)	25.8	40.1

The results of the NCDBN model were compared with those of the LSSVM, radial basis function neural network (RBFNN), long short time memory (LSTM) [30], and BP neural network (BPNN) models. The structure of 3 hidden layer is determined for DBN, The control prediction model is found of has the structure of 7-15-15-15-1. Namely, the control prediction model has 7 input nodes, 15 hidden nodes in the first hidden layer, 15 hidden nodes in the second hidden layer, 15 hidden nodes in the third hidden layer and 1 output layer node. The state prediction model has the structure of 7-15-10-15-1. Similarly, the state prediction model has 7 input nodes, 15 hidden nodes in the first hidden layer, 10 hidden nodes in the second hidden layer, 15 hidden nodes in the third hidden layer and 1 output layer node. The structure of the above two models are determined by considering a range of neuron numbers and the performance on the training datasets. The learning rate of RBM is selected as 0.1. the learning rate of supervised training for DBN is 0.1.
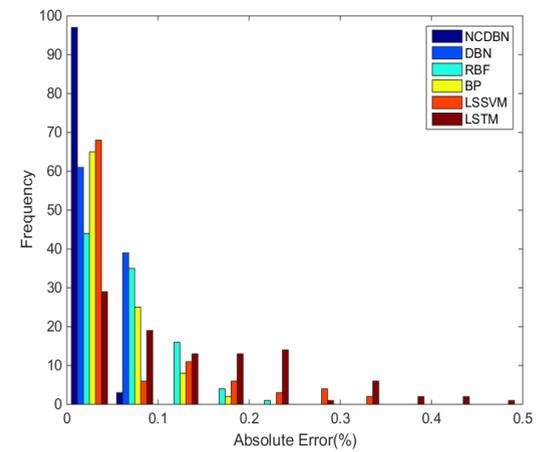
Three experimental datasets were randomly collected from the distribution control system of a 1000MW unit by the Guodian Taizhou power plant. To facilitate the subsequent description, the three datasets are represented by D1, D2, and D3. To verify the generalization of the proposed algorithm, different data were tested separately. The specific information of datasets is shown in Table 2. All the experiments were performed on a Core i5 processor with a 4 GB RAM and a Microsoft Windows 10 operating system. The algorithms in the experiments have been trained using python.
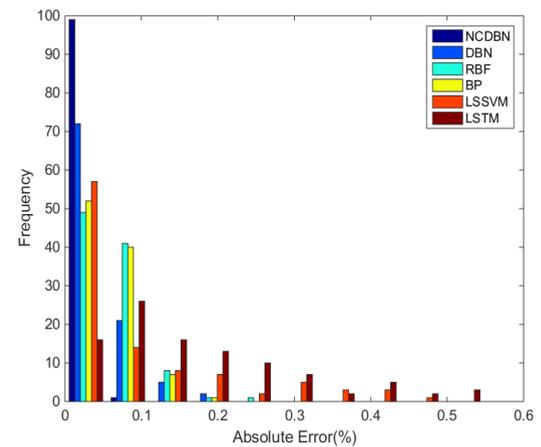
A. INPUT SELECTION ANALYSIS

This section mainly analyzes the influence of the lasso method on the accuracy and computational efficiency of the prediction model. Table 3 shows the results with and without the features selection using the lasso method. From Table 3, it can be concluded that the NCDBN model training time is

(a) D1 dataset

(b) D2 dataset

(c) D3 dataset

FIGURE 4. Absolute errors of different models.

shortened by 14.3 s, resulting in a reduction of 36% in the modeling time. It can be concluded that the lasso method can improve the efficiency and accuracy of the prediction model by reducing its input dimensions and eliminating some of the variables with coupling correlations.

TABLE 4. Comparison of the predicted results of the different models.

Data sets	Error metrics	NCDBN		DBN		RBFNN		BPNN		LSTM		LSSVM	
		Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
D1	*MAE*	0.0138	0.0146	0.0374	0.0380	0.0170	0.0680	0.0292	0.052	0.1601	0.1515	0.034	0.0818
	MRE	0.4951	0.5151	1.3448	1.3813	0.6077	2.4870	1.0487	1.8426	5.6420	5.433	1.2242	2.9826
	MSE	0.0005	0.0005	0.0134	0.0022	0.0005	0.0083	0.0014	0.0041	0.0410	0.0382	0.0019	0.0167
D2	*MAE*	0.0031	0.0141	0.0321	0.0420	0.0145	0.0621	0.029	0.0465	0.1073	0.1327	0.0299	0.0595
	MRE	0.3082	0.4981	1.1593	1.5177	0.5166	2.2594	1.0403	1.6961	3.8597	4.7535	1.0854	2.1821
	MSE	0.0001	0.0004	0.0116	0.0024	0.0004	0.0059	0.0013	0.0034	0.0161	0.0305	0.0017	0.0100
D3	*MAE*	0.0057	0.0161	0.0359	0.0410	0.0175	0.0565	0.0306	0.0568	0.1023	0.1713	0.0318	0.0894
	MRE	0.5743	0.5846	1.2994	1.5286	0.6348	2.0490	1.1079	2.0410	3.7118	6.1948	1.1423	3.2383
	MSE	0.0005	0.0004	0.0137	0.0031	0.0005	0.0051	0.0015	0.0044	0.0153	0.0457	0.0015	0.0324

B. THE NONLINEAR COMBINED DEEP BELIEF NETWORK MODEL VS OTHER MODELS

Fig. 3 shows the comparison between the measured and predicted values for different models. To clarify the differences among the six models, a partial curve is drawn for each (Figs. 3 (a)–(c)). It can be clearly seen that the NCDBN prediction curves follow the direction of the real data (marked as ''TRUE'' in the figures), indicating that this model can effectively predict the oxygen content of the flue gas. The worst results are provided by the LSTM algorithm as the predicted curve does not reflect the true data. This result may be attributed to the fact that the LSTM model is not suitable for solving non-time series problems.

To further verify the performance of the NCDBN algorithm, the absolute errors between the measured and predicted values of several models were compared in Fig. 4. The absolute error of the NCDBN model is distributed within the minimum interval [0, 0.05] as per Fig. 4(a). With the increase of absolute error, its frequency gradually decreases (Figs. 4 (b) and (c)), similar to the distribution law. The frequency distributions of the DBN, BP, LSTM, RBF, and LSSVM models were similar to that of the NCDBN algorithm, but the absolute error frequency of the latter decreased rapidly, and none of the distributions existed in the higher absolute error interval. This result may be attributed to the fact that the nonlinear combination improve two submodels prediction accuracy. Therefore, the absolute error of the NCDBN model is the smallest, and the NCDBN shows the best prediction accuracy.

Table 4 shows the error metrics of the different prediction algorithms on training data and testing data. Taking the testing data results of dataset D1 as an example, the *MAE* and *MRE* of the NCDBN model showed reductions of 62% and 63%, respectively, whereas the *MSE* decreased by 80%. The error metrics of the NCDBN model also declined for

TABLE 5. Experimental results of the nonlinear combination strategy.

Model	Error metrics	D1	D2	D3
Control model	*MAE*	0.0606	0.0756	0.0736
	MRE	2.1944	2.7671	2.6786
	MSE	0.0055	0.0087	0.0076
	R^2	0.8113	0.7599	0.7424
State model	*MAE*	0.0383	0.0417	0.0383
	MRE	1.3902	1.5223	1.3971
	MSE	0.0024	0.0026	0.0024
	R^2	0.916	0.9100	0.9189
Combined model	*MAE*	0.0146	0.0141	0.0161
	MRE	0.5151	0.4981	0.5846
	MSE	0.0005	0.0004	0.0004
	R^2	0.9859	0.9825	0.9879

the testing data of datasets D2 and D3. In addition, the error metrics of the NCDBN model were lower than those of the LSSVM, RBF, LSTM, and BPs model for all three datasets. The error metrics of the NCDBN model are even better than those of the single DBN model for all the datasets. As seen from Table 4, the error metrics of BP and RBF models with training data were lower than results on testing data. This result may be attributed to the fact that the BP and RBF models suffered from overfitting problem.

C. ASSESSMENT OF THE NONLINEAR COMBINATION STRATEGY

Table 5 shows the comparison of the error metrics of the elements of the nonlinear combination strategy. Taking D1 as an example, the comparison between the combined and control
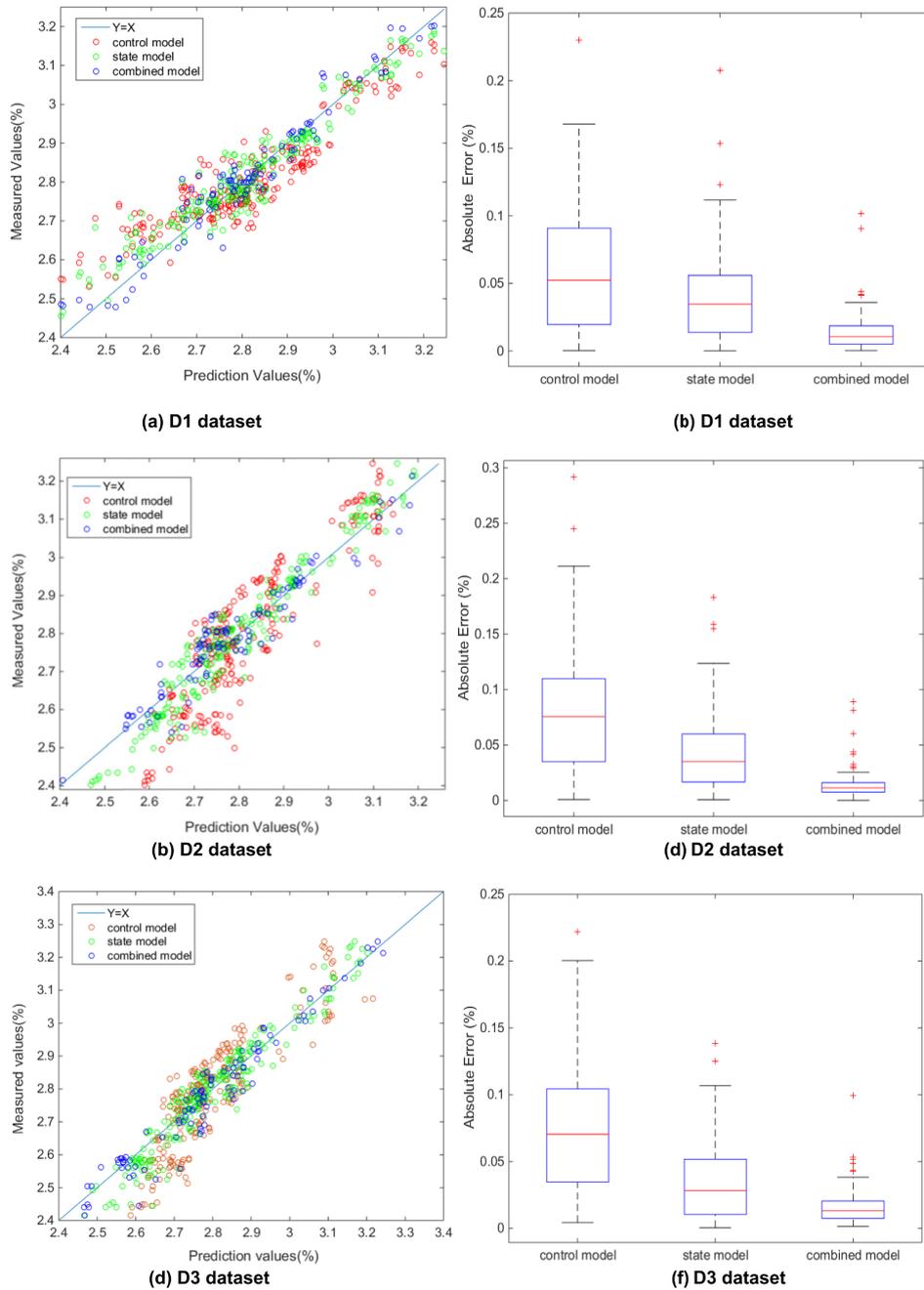
(a) D1 dataset

(b) D1 dataset

(b) D2 dataset

(d) D2 dataset

(d) D3 dataset

(f) D3 dataset

FIGURE 5. Measured and predicted values of oxygen content of flue gas with and without nonlinear combination.

prediction models shows that the *MAE*, *MRE*, and *MSE* of the former were reduced by 76%, 77%, and 92%, respectively. Compared with the state prediction model, the corresponding values decreased by 62%, 63%, and 81%, respectively. It can be seen that the nonlinear combined model showed the best performance, which verifies that it has superior prediction accuracy. Furthermore, the NCDBN model showed the lowest errors among all three models for all the datasets.

Fig. 5 shows the comparisons between the measured and predicted values for the oxygen content of the flue gas before

and after the nonlinear combination. Fig. 5 (a), (c), and (e) show the regression analyses of the measured and predicted values for the datasets. The datasets clearly show a good fit, with perfect lines indicating that the predicted value is equal to the measured one. As per Fig. 5 (a), the nonlinear combined prediction model shows promising performance. Fig. 5 (c) and (e) have similar distributions. R^2 represents the relevant index. If R^2 is closer to 1, it means that the relationship between the measured and predicted values is almost linear. From Table 5, the NCDBN model shows the

best performance, with $R^2 = 0.9859$, 0.9825, and 0.9879 for the different datasets. Fig. 5 (b), (d), and (f) show the box plots of the absolute errors between the measured and predicted values for all three datasets, with the red lines representing the median errors. It is clear that the errors of the NCDBN algorithm vary within a small range. The NCDBN model presents the lowest errors compared with the control prediction and state prediction models. Thus, the nonlinear combination can improve the prediction accuracy of the oxygen content of flue gas.

V. CONCLUSION

The process of coal-fired boiler combustion is complicated and the production environment is harsh. These factors lead to serious and frequent equipment dysfunctions and loss when attempting direct measurements of the oxygen content of flue gas. Thus, it is difficult to maintain a high level of measurement accuracy. To solve this problem, a nonlinear combined deep learning approach is proposed in this paper to predict the oxygen content of flue gas. The whole algorithm is analyzed in three parts: data pre-processing, feature selection, and data analysis modeling. One unique feature of the proposed approach is that the process variables are divided into control and state variables to facilitate the application of an advanced control algorithm. Another major feature is that a nonlinear combined scheme based on DBN is proposed to predict the oxygen content of flue gas. Experiments based on actual production data are carried out to evaluate the proposed approach. The results show that the proposed combination modeling approach and feature selection strategies are effective and promising. In the future, we plan to continue develop an application of predictive control using the proposed algorithm.

REFERENCES

[1] W. Zhang, Y. Zhang, X. Bai, J. Liu, D. Zeng, and T. Qiu, "A robust fuzzy tree method with outlier detection for combustion models and optimization," *Chemometrics Intell. Lab. Syst.*, vol. 158, pp. 130–137, Nov. 2016.

[2] J. Luo, L. Wu, and W. Wan, "Optimization of the exhaust gas oxygen content for coal-fired power plant boiler," *Energy Procedia*, vol. 105, pp. 3262–3268, May 2017.

[3] Q. Li and G. Yao, "Improved coal combustion optimization model based on load balance and coal qualities," *Energy*, vol. 132, pp. 204–212, Aug. 2017.

[4] F. Wang, S. Ma, H. Wang, Y. Li, and J. Zhang, "Prediction of NO_X emission for coal-fired boilers based on deep belief network," *Control Eng. Pract.*, vol. 80, pp. 26–35, Nov. 2018.

[5] H. Satyavada and S. Baldi, "Monitoring energy efficiency of condensing boilers via hybrid first-principle modelling and estimation," *Energy*, vol. 142, pp. 121–129, Jan. 2018.

[6] S. Baldi, T. L. Quang, O. Holub, and P. Endel, "Real-time monitoring energy efficiency and performance degradation of condensing boilers," *Energy Convers. Manage.*, vol. 136, pp. 329–339, Mar. 2017.

[7] Y. Jin, N. Gao, and T. Zhu, "Techno-economic analysis on a new conceptual design of waste heat recovery for boiler exhaust flue gas of coal-fired power plants," *Energy Convers. Manage.*, vol. 200, Nov. 2019, Art. no. 112097.

[8] M. Willett, "Oxygen sensing for industrial safety—Evolution and new approaches," *Sensors*, vol. 14, no. 4, pp. 6084–6103, Mar. 2014.

[9] W. Yan, H. Shao, and X. Wang, "Soft sensing modeling based on support vector machine and Bayesian model selection," *Comput. Chem. Eng.*, vol. 28, no. 8, pp. 1489–1498, Jul. 2004.

[10] R. Grbić, D. Slišković, and P. Kadlec, "Adaptive soft sensor for online prediction and process monitoring based on a mixture of Gaussian process models," *Comput. Chem. Eng.*, vol. 58, pp. 84–97, Nov. 2013.

[11] C. Shang, F. Yang, D. Huang, and W. Lyu, "Data-driven soft sensor development based on deep learning technique," *J. Process Control*, vol. 24, no. 3, pp. 223–233, Mar. 2014.

[12] Y. Lv, J. Liu, T. Yang, and D. Zeng, "A novel least squares support vector machine ensemble model for NO_x emission prediction of a coal-fired boiler," *Energy*, vol. 55, pp. 319–329, Jun. 2013.

[13] C. Wang, Y. Liu, S. Zheng, and A. Jiang, "Optimizing combustion of coal fired boilers for reducing NO_X emission using Gaussian Process," *Energy*, vol. 153, pp. 149–158, Jun. 2018.

[14] F. Yu and X. Xu, "A short-term load forecasting model of natural gas based on optimized genetic algorithm and improved BP neural network," *Appl. Energy*, vol. 134, pp. 102–113, Dec. 2014.

[15] C. Cecati, J. Kolbusz, P. Rozycki, P. Siano, and B. M. Wilamowski, "A novel RBF training algorithm for short-term electric load forecasting and comparative studies," *IEEE Trans. Ind. Electron.*, vol. 62, no. 10, pp. 6519–6529, Oct. 2015.

[16] J. Qiao, G. Wang, W. Li, and X. Li, "A deep belief network with PLSR for nonlinear system modeling," *Neural Netw.*, vol. 104, pp. 68–79, Aug. 2018.

[17] X. Han, Y. Yan, C. Cheng, Y. Chen, and Y. Zhu, "Monitoring of oxygen content in the flue gas at a coal-fired power plant using cloud modeling techniques," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 4, pp. 953–963, Apr. 2014.

[18] H. Xiaoying, W. Jingcheng, Z. Langwen, and W. Bohui, "Data-driven modelling and fuzzy multiple-model predictive control of oxygen content in coal-fired power plant," *Trans. Inst. Meas. Control*, vol. 39, no. 11, pp. 1631–1642, Nov. 2017.

[19] R. S. Lakshmi, A. Sivakumar, G. Rajaram, V. Swaminathan, and K. Kannan, "A novel hypergraph-based feature extraction technique for boiler flue gas components classification using PNN—A computational model for boiler flue gas analysis," *J. Ind. Inf. Integr.*, vol. 9, pp. 35–44, Mar. 2018.

[20] W. Zhang, C. Li, G. Peng, Y. Chen, and Z. Zhang, "A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load," *Mech. Syst. Signal Process.*, vol. 100, pp. 439–453, Feb. 2018.

[21] J. Yu, X. Zheng, and S. Wang, "A deep autoencoder feature learning method for process pattern recognition," *J. Process Control*, vol. 79, pp. 1–15, Jul. 2019.

[22] Y. Wang, X. Wang, and W. Liu, "Unsupervised local deep feature for image recognition," *Inf. Sci.*, vol. 351, pp. 67–75, Jul. 2016.

[23] F. Li, J. Zhang, C. Shang, D. Huang, E. Oko, and M. Wang, "Modelling of a post-combustion CO_2 capture process using deep belief network," *Appl. Therm. Eng.*, vol. 130, pp. 997–1003, Feb. 2018.

[24] S. Salcedo-Sanz, L. Cornejo-Bueno, L. Prieto, D. Paredes, and R. García-Herrera, "Feature selection in machine learning prediction systems for renewable energy applications," *Renew. Sustain. Energy Rev.*, vol. 90, pp. 728–741, Jul. 2018.

[25] W.-Y. Loh, "Classification and regression trees," *WIREs Data Mining Knowl. Discovery*, vol. 1, no. 1, pp. 14–23, Jan. 2011.

[26] Z. Fan, J. Wang, B. Xu, and P. Tang, "An efficient KPCA algorithm based on feature correlation evaluation," *Neural Comput. Appl.*, vol. 24, nos. 7–8, pp. 1795–1806, Jun. 2014.

[27] Y. Kim, J. Hao, T. Mallavarapu, J. Park, and M. Kang, "Hi-LASSO: High-dimensional LASSO," *IEEE Access*, vol. 7, pp. 44562–44573, 2019.

[28] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Stat. Soc., B, Methodol.*, vol. 58, no. 1, pp. 267–288, Jan. 1996.

[29] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, Jul. 2006.

[30] X. Qing and Y. Niu, "Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM," *Energy*, vol. 148, pp. 461–468, Apr. 2018.

ZHENHAO TANG received the B.S. degree from Qingdao University, Qingdao, China, in 2007, and the M.S. and Ph.D. degrees from the College of Information Science and Engineering, Northeast University, Shenyang, China, in 2009 and 2014, respectively.

From 2014 to 2017, he was an Assistant Professor with the School of Automation Engineering, Northeastern Electric Power University, Jilin, China, where he is currently an Associate Professor. His research interests include data mining and computational intelligence with applications in modeling, monitoring, optimization, and operations of systems in the renewable energy and conventional energy.

YANYAN LI was born in Heze, Shandong, China, in 1994. She received the M.E. degree from Shandong First Medical University, China. From 2013 to 2017, she was a Student with the School of Information Engineering, Shandong First Medical University. She is currently a Graduate Student with the School of Automation Engineering, Northeastern Electric Power University.

Her research interests include deep learning methods and thermal power prediction algorithms.

ANDREW KUSIAK (Life Member, IEEE) received the B.S. and M.S. degrees in precision engineering from the Warsaw University of Technology, Warsaw, Poland, in 1972 and 1974, respectively, and the Ph.D. degree in operations research from the Polish Academy of Sciences, Warsaw, in 1979.

He is currently a Professor with the Intelligent Systems Laboratory, Department of Mechanical and Industrial Engineering, The University of Iowa, USA. He has authored or coauthored numerous books and technical articles in journals sponsored by professional societies, such as the Association for the Advancement of Artificial Intelligence, the American Society of Mechanical Engineers, the Institute of Industrial and Systems Engineers, and other societies. His current research interests include the application of computational intelligence in automation, renewable energy, smart manufacturing, product development, and healthcare.

Dr. Kusiak is a Fellow of the Institute of Industrial and Systems Engineers and the Editor-in-Chief of the *Journal of Intelligent Manufacturing*. He has served on the Editorial Board of over 50 journals.

• • •