DYNAMIC MODELLING RESEARCH ON THERMAL EFFICIENCY OF SUPERCRITICAL DOWN-FIRED BOILER

by

Chen HAN[a,b], Shaofan LI[c], Lianhong YANG[a,b], Yanqing LI[a,b], Xianyong PENG[c], and Zhi WANG[c*]

[a] Xinjiang Key Laboratory of High Value Green Utilization of Low-rank Coal, Changji, China
[b] School of Energy and Control Engineering, Changji University, Changji, China
[c] Jiangsu Smart Energy Technology and Equipment Engineering Research Center, School of Low-Carbon Energy and Power Engineering, China University of Mining and Technology, Xuzhou, China

Original scientific paper
https://doi.org/10.2298/TSCI250816179H

To accurately predict the coal-fired generating units' thermal efficiency under deep peak shaving conditions, a dynamic prediction method of boiler thermal efficiency is proposed. Determine the auxiliary variables affecting boiler thermal efficiency by analyzing the proportions of heat loss in the anti-balance method, and using the random forest algorithm to carry out supervised dimensionality reduction of auxiliary variables affecting the thermal efficiency. On this basis, the convolutional neural network (CNN) with dynamic modelling function is selected as the infrastructure, and to ensure the lightweight model, the cross-channel communication unit is inserted into the conventional CNN with only three convolutional layers to solve the problem of no interaction of feature maps in the same convolutional layer, and a boiler thermal efficiency predictive model based on the cross-channel communication CNN (C3-CNN) is constructed. Simulation experiments were conducted on the actual operation data for a 600 MW boiler, the results show that the developed method is equally suitable for both transient and steady-state conditions.

Key words: *coal-fired boiler, thermal efficiency, deep learning, CNN*

Introduction

In the practical operation of coal-fired power units, factors such as load fluctuations, variations in coal type, equipment aging, and system retrofitting often cause the boiler to deviate from its optimal operating conditions, thereby reducing thermal efficiency. Under deep peak-shaving conditions, boilers are frequently forced to work at low loads-substantially distant from their design points-for prolonged periods, which significantly deteriorates boiler thermal performance [1, 2]. Therefore, developing an accurate model to predict thermal efficiency is crucial for improving boiler operation under complex and variable conditions.

Modelling the boiler combustion process generally involves two primary approaches: mechanistic methods and data-driven methods. Mechanistic approaches utilize kinetic models, energy conservation laws, and other physical principles to establish mathematical representations of the underlying combustion mechanisms. Zhao *et al.* [3] developed a mathematical mod-

* Corresponding author, e-mail: zhi.wang@cumt.edu.cn

el for boiler thermal efficiency and NO_x emissions based on combustion mechanisms, taking into account in-furnace coal distribution and over-fire air (OFA). By simplifying the combustion system, they formulated a discrete-time linear state-space model. Similarly, earlier studies proposed a simplified dynamic model of a 600 MW coal-fired boiler-steam unit under low-load conditions, employing comparable simplifications of the boiler system [4]. While mechanistic modelling is advantageous for its interpretability, it demands a thorough understanding of the system's underlying mechanisms. However, in-furnace pulverized coal combustion involves coupled phenomena such as heat and mass transfer, fluid dynamics, and multi-phase flow. As a result, it remains challenging to achieve accurate thermal efficiency predictions using purely mechanistic models.

Compared with mechanistic approaches, data-driven methods leverage machine learning algorithms to achieve faster and often more accurate modelling of complex systems. In one study, Chen *et al.* [5] conducted 48 experiments on a co-firing boiler operating at a constant load of 300 MW, using 37 operational parameters to train a backpropagation neural network for predicting the impact of secondary air damper openings on NO_x emissions. However, due to the limited size of the training dataset, the model exhibited a maximum relative prediction error of up to 10%, underscoring the strong dependence of neural networks on sufficient data availability. Consequently, access to a rich and representative dataset is essential for the effective development of such models. In addition, ash fusion characteristics play a critical role in regulating boiler thermal efficiency. Ash fusion can lead to slagging and fouling within the boiler, which adversely affects heat transfer and combustion performance. Lawal *et al.* [6] developed predictive models for ash fusion temperatures utilizing various machine learning algorithms, including ANN, support vector regression (SVR), and Gaussian process regression. Among these approaches, the ANN model demonstrated the highest prediction accuracy. Laser-induced breakdown spectroscopy (LIBS) is an online detection technology capable of measuring the carbon content in fly ash, thereby enabling real-time assessment of boiler combustion efficiency. Integrating SVR as a predictive model for the unburned carbon content in fly ash can further enhance the quantitative accuracy of LIBS analysis [7]. However, the aforementioned modelling approaches do not account for the temporal dimension, their outputs depend solely on the current input, this static modelling method is only applicable to steady-state boiler conditions. In scenarios involving deep peak shaving or rapid load fluctuations, more advanced modelling methods are required to accurately capture the boiler's actual operating behavior. Furthermore, due to the heat storage effect of the working medium, boiler thermal efficiency is influenced by historical operating states. This is because heat storage reduces flue gas heat absorption, thereby providing more energy for steam generation. Consequently, capturing the temporal characteristics of boiler thermal efficiency necessitates the development of a dynamic modelling approach. Although some researchers have explored dynamic modelling approaches. Nevertheless, such models still fail to account for the influence of historical operating conditions on the boiler's current thermal performance [8]. With rapid advancements in computational power and the explosion of data, neural networks experienced a resurgence around 2010, propelled by the advent of deep learning techniques [9]. Among these, RNN are particularly well-suited for processing sequential data. The RNN-based models have been developed to predict NO_x emissions in boilers. Xie *et al.* [10] proposed a multi-step NO_x emission prediction model utilizing the long short-term memory (LSTM) and compared the performance with that of support vector machines and back propagation neural networks. Owing to its memory capability, the LSTM outperformed the other two in prediction accuracy and effectively captured the delay characteristics. Consequently, an enhanced LSTM was proposed to predict NO_x emission at the SCR

inlet under both steady-state and transient conditions by incorporating an attention mechanism to adaptively adjust the weights of the hidden layer output sequences [11]. Zhang *et al.* [12] developed model for heat losses based on the gated recurrent unit (GRU), using these outputs for real-time dynamic calculation of boiler efficiency. However, the GRU and LSTM memory units may lead to error accumulation, resulting in a gradual increase in predictive errors [13].

The CNN has been widely applied in industrial modelling. Employing 2-D CNN to construct NO_x emission prediction models presents a promising approach. At present, the substantial parameter usage represents the primary challenge hindering the widespread application of models in the power generation sector, particularly under extreme conditions involving rapid load variations [14]. The stability of the combustion-side flame exerts a disruptive influence on boiler thermal efficiency, making the lightweight design crucial to resolving this issue [15]. By adopting a lightweight architectural design, such models have been effectively adapted to transient load conditions in industrial boilers, achieving satisfactory prediction performance. To develop an accurate prediction model for boiler thermal efficiency, this study employs a CNN with dynamic modelling capabilities as the core architecture. A cross-channel communication unit (C3-unit) is integrated into a standard CNN with three convolutional layers to enhance interactions among feature maps. Building upon this enhancement, a dynamic prediction model for boiler thermal efficiency is constructed using the cross-channel communication CNN (C3-CNN). To evaluate the performance of the proposed C3-CNN, simulation experiments are conducted using historical data from a 600 MW coal-fired boiler.

Boiler introduction and auxiliary variables

Research objective

The down-fired boiler has become the predominant furnace type in China for burning anthracite and lean coal. However, its unique structural design renders it prone to poor combustion stability, severe slagging in the furnace, and significant temperature deviations along the water-cooled walls, factors that substantially reduce thermal efficiency. This study focuses on a 600 MW supercritical down-fired boiler. The boiler combusts low quality anthracite with a volatile matter content below 12%, leading to severe coking within the furnace. This coking behavior significantly disrupts the combustion environment and heat transfer processes, rendering accurate thermal efficiency prediction via mechanistic models particularly challenging. Data-driven modelling approaches offer a more suitable alternative.

Target boiler is equipped with twenty-four twin-swirl, direct-flow pulverized coal burners and six mills. Each coal mill supplies to four burners, which are symmetrically arranged on the front and rear arches of the furnace, with twelve burners on each side. As illustrated in fig. 1 a schematic of the matching relationship, each burner is equipped with an independently controlled F-layer secondary air damper that supplies the primary portion of secondary air for combustion. Accordingly, they are included as auxiliary variables.

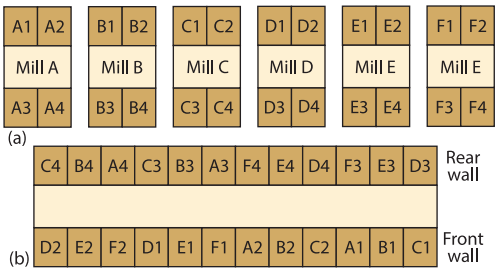

Figure 1. Schematic diagram of burner-to-mill matching relations; (a) correspondence between and burners and (b) burner layout

Auxiliary variables

Coal-fired power plant boilers primarily determine thermal efficiency using the anti-balance method (also known as the heat loss method). The thermal efficiency is calculated:

$$\eta = 100 - \sum_{i=2}^{6} q_i, \quad \sum_{i=1}^{6} q_i = 100 \tag{1}$$

where Q_f is the heat input and q_i – the expressed each heat loss as a percentage of heat input, [= (Q_i/Q_f) ×100], i = 2, 3,…, 6. By calculating the specific data of various heat losses, the operational status of the boiler can be assessed, and improve its thermal efficiency can be identified.

By systematically analyzing the variables that influence each type of heat loss, an auxiliary variable list reflecting factors that affect thermal efficiency can be constructed. The Q_2 is the exhaust heat loss, which is the largest item in the heat loss, accounting for 5%-6% in general, therefore, temperature in flue gas is selected as an auxiliary variable. The mechanical incomplete combustion heat loss Q_4 ranks second. Its key influencing factors include combustion mode, coal quality, air distribution strategy, and oxygen content. Accordingly, the auxiliary variable list includes pulverize capacity air-flow (coal blending), flue gas oxygen content, and secondary-air damper opening (air distribution). In summary, a total of 53 parameters is selected. Detailed information on all auxiliary variables is listed in tab. 1.

Dimension reduction method

With the in-depth study of boiler combustion process modelling, researchers have found that dimensionality reduction of input variables not only helps improve the learning efficiency of the model but also enhances its generalization performance [16]. Table 2 compares the characteristics of 3-Dity reduction methods. The reduction method (RF) is an extended variant of Bagging [17], particularly well-suited for evaluating the importance of high dimensional non-linear variables. The degree of performance degradation indirectly indicates the importance of a given variable [18]. The importance of an input variable before and after perturbation is measured:

$$\bar{D}_i = \frac{1}{B}\sum_{b=1}^{B}\left(R_b^{\mathrm{OOB}} - R_{b,i}^{\mathrm{OOB}}\right) \tag{2}$$

where R_b^{OOB} and $R_{b,i}^{\mathrm{OOB}}$ represent the errors of the out-of-bag data before and after perturbation.

Modelling method

In this study, a novel C3-CNN is proposed, which builds upon the conventional CNN by introducing the C3-unit. The C3-Unit is designed to facilitate interaction among feature maps, allowing the model to attain superior prediction accuracy using fewer layers. The structure of the C3-CNN is depicted in fig. 2.

Conventional convolution layer

The convolutional layer is the core component of a CNN, responsible for non-linear operations and feature extraction. Multiple filters are configured within each convolutional layer to slide systematically over the input tensor. The convolution operation is expressed:

$$\mathbf{X}_l^k = \mathbf{W}_l^k \mathbf{X}_{l-1} + b_l^k \tag{3}$$

where $\mathbf{X}_l^k$ is the feature matrix after the convolution operation and $\mathbf{W}_l^k$ – the weight matrix.

Table 1. List of auxiliary variables

Variables		Range	Number	Variables		Range
Load [MW]		[264.08, 605.05]	28	OFA-A2 [%]		[20, 70]
Total coal [tonne per hour]		[137.71, 400.72]	29	OFA-B2 [%]		[20, 75]
Capacity air of coal mill [%]	Mill-A-left	[0.15, 81.74]	30	F-layer secondary-air damper position [%]	F-A1	[20, 80]
	Mill-A-right	[0.05, 79.68]	31		F-B1	[20, 70]
	Mill-B-left	[0.09, 75.03]	32		F-C1	[20, 80]
	Mill-B-right	[0.92, 73.99]	33		F-D1	[20, 70]
	Mill-C-left	[0.18, 94.33]	34		F-E1	[20, 70]
	Mill-C-right	[0.46, 92.98]	35		F-F1	[20, 50]
	Mill-D-left	[0.25, 74.51]	36		F-A2	[20, 80]
	Mill-D-right	[0.38, 71.59]	37		F-B2	[20, 70]
	Mill-E-left	[0.25, 83.91]	38		F-C2	[20, 85]
	Mill-E-right	[0.67, 80.38]	39		F-D2	[20, 85]
	Mill-F-left	[0.85, 77.67]	40		F-E2	[20, 80]
	Mill-F-right	[0.71, 75.08]	41		F-F2	[20, 60]
Main steam temperature [℃]		[1024, 1888]	42		F-A3	[20, 100]
Main steam pressure [MPa]		[16.33, 25.59]	43		F-B3	[20, 100]
Main steam flow [tonne per hour]		[541.14, 574.95]	44		F-C3	[20, 100]
Flue gas temperature of SCR inlet [℃]		[336.67, 409.11]	45		F-D3	[20, 80]
Flue gas oxygen content [%]		[0.15, 9.06]	46		F-E3	[20, 100]
Feed-water flowrate [tonne per hour]		[1027.8, 1925.9]	47		F-F3	[20, 100]
Temperature in flue gas [℃]		[145.29, 177.98]	48		F-A4	[20, 100]
Total air volume [tonne per hour]		[1734.0, 2422.3]	49		F-B4	[20, 100]
Furnace pressure [Pa]		[-310.54, 94.92]	50		F-C4	[20, 100]
Differential pressure of air box [kPa]		[3.49, 29.11]	51		F-D4	[20, 100]
Highest temperature of the wall [℃]		[379.47, 496.25]	52		F-E4	[20, 100]
OFA-A1 [%]		[20, 60]	53		F-F4	[20, 100]
OFA-B1 [%]		[20, 65]	Target	Thermal efficiency [%]		[89.48, 93.28]

Table 2. Properties of the reduction methods

Methods	Computational complexity	Applicable data sets
PCA	High – Covariance matrix with large computation	Linear
Pearson correlation	Low	Linear
RF	Low – Parallel computation between decision trees	Linear/non-linear

As model training progresses, the feature extraction process in deeper convolutional layers often suffers from slow convergence or gradient vanishing due to uneven data distribution. To address this issue, the Batch normalization [19] was proposed, which not only effectively mitigates the problem of gradient vanishing but also enhances the robustness. To ensure the non-linearity of the CNN, a bias term is added to the output after the convolution operation,

which is then passed through an activation function generate the final feature map [20]. In this study, ReLU is adopted as the activation function, which is defined:

$$ReLU(x) = \max(0, x) = \begin{cases} 0, & \text{if } x < 0 \\ x, & \text{if } x \geq 0 \end{cases} \tag{4}$$

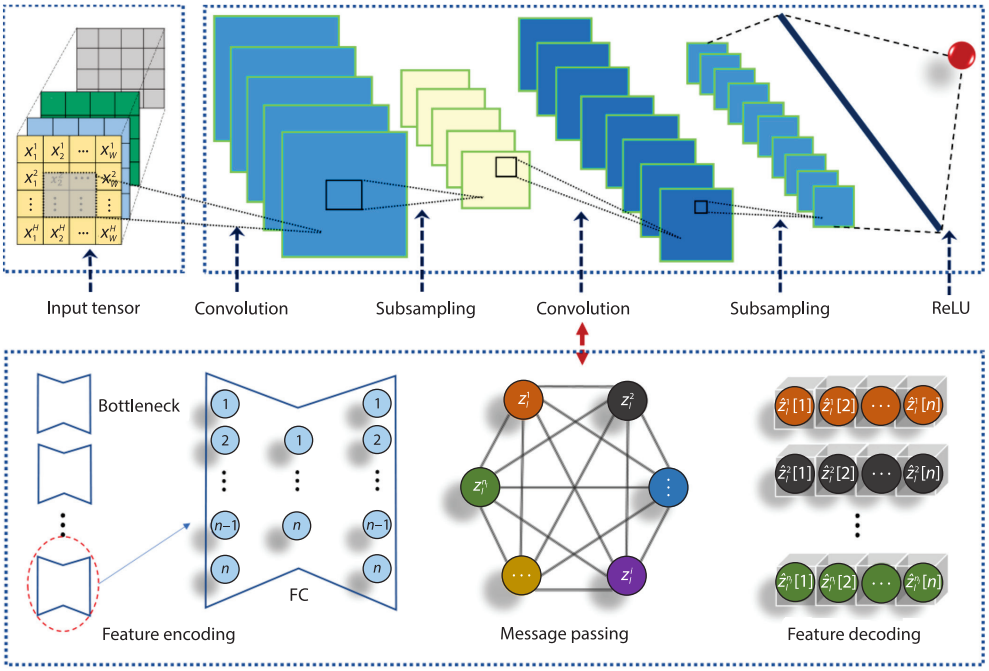

Figure 2. Structure of C3-CNN

The C3-unit

The C3-unit consists of three components: feature encoding, decoding, and message passing. All l^{th} layer feature maps are denoted as: $\mathbf{X}_l = \{\mathbf{X}_l^1, \mathbf{X}_l^2, \ldots, \mathbf{X}_l^{n_l}\}$. The update rule after completing information interaction is defined:

$$\hat{\mathbf{X}}_l^k = \mathbf{X}_l^k + f_l^k(\mathbf{X}_l^1, \mathbf{X}_l^2, \cdots, \mathbf{X}_l^{n_l}) \tag{5}$$

where $\mathbf{X}_l^k$ and $\hat{\mathbf{X}}_l^k$ are the feature map matrices of the k^{th} channel before and after interaction, respectively. All channels of size $C \times H_l \times W_l$ into a $C \times 1 \times 1$ vector by global average pooling were compressed in the SENet (squeeze and excitation networks) [21]. In contrast, the proposed C3-unit preserves the complete channel information, enabling richer and more comprehensive interaction:

– Feature encoding

To extract overall features from each feature map within the same convolutional layer, the feature encoding unit is developed, which are then send into the message passing unit. In the layer l, the k^{th} feature map $\mathbf{X}_l^k$ was flattened into a 1-D feature vector by the feature encoding unit, which is then passed through two fully-connected layers within a bottleneck structure. As shown in fig. 2, in this structure the two linear functions are defined:

$$\begin{aligned} \mathbf{y}_l^k &= f_{\text{in}}(\mathbf{X}_l^k, \mathbf{w}_1) \\ \mathbf{z}_l^k &= f_{\text{out}}\left[\sigma(\mathbf{y}_l^k, \mathbf{w}_2)\right] \end{aligned} \tag{6}$$

– Message passing

The message passing unit is responsible for interacting with the encoded features and subsequently updating after interaction. Graph convolutional networks [22] serve as the foundational frame for learning of this kind interactions. In graph convolutional networks, graph attention networks [23] were inserted to build an attention mechanism. A similar formulation was adopted to enable cross-channel communication. An encoded feature vectors are treated as nodes as shown in fig. 2. The $\mathbf{z} = \{\mathbf{z}_l\}$, $\mathbf{z}_l \in \mathbb{R}^{1\times H_l W_l}$. This equation shows the edge strength:

$$S_{k,q} = f_{att}(\mathbf{z}_l^k, \mathbf{z}_l^q) \tag{7}$$

This study uses the mean output to enhance the robustness during the message passing process. By utilizing a Softmax function, the nodes' distance is normalized into an attention score $a_{k,q}$. As shown in eqs. (8) and (9), the final output of the message passing unit, $\hat{\mathbf{z}} = \{\mathbf{z}_l^1, \mathbf{z}_l^2, \cdots, \mathbf{z}_l^{n_l}\}$, is computed:

$$a_{k,q} = \text{Softmax}_q(s_{k,q}) = \frac{\exp(s_{k,q})}{\sum_{i=1}^{n_l}\exp(s_{k,i})} \tag{8}$$

$$\hat{\mathbf{z}}_l^k = \sum_{q=1}^{n_l} a_{k,q}\mathbf{z}_l^q \tag{9}$$

– Feature decoding

After receiving the updated feature output, the feature decoding unit reshapes it from flattened vectors back to the original feature map shape, enabling further computation in the subsequent standard convolutional layers.

Results and discussion

Data preprocessing

Continuous operational data spanning approximately two weeks are extracted from the target boiler's DCS, resulting in 20160 samples. The raw data is preprocessed to meet the model's input requirements. Specifically, missing values in the sampled data are imputed using the median of non-missing values [24].

Feature selection

The importance of 53 auxiliary features affecting thermal efficiency was assessed using the RF, using 500 decision trees. The top 10 variables account for a cumulative weight of 0.933 and are chosen as input features for thermal efficiency prediction model. Figure 3 presents the RF evaluation results in descending order of importance. Based on the variable importance evaluation results, the necessity of each variable in thermal efficiency modelling is analyzed. As shown in fig. 3, the temperature in flue gas ranks highest in importance according to the RF computation, with a value of 0.838. This indicates a dominant contribution from a single variable, which aligns with the fact that

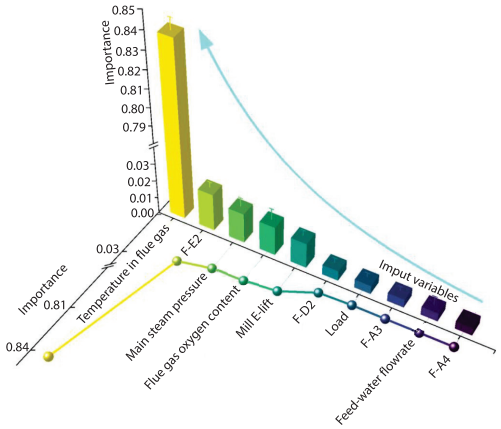

Figure 3. Results of the variables importance assessment related to thermal efficiency

exhaust heat loss is the largest component of total boiler heat loss. Several factors contribute to elevated flue gas temperatures, such as increased flue gas velocity, which leads to a higher excess air coefficient. Concurrently, the residence time of fuel within the furnace decreases, while the flue gas's ability to carry larger fuel particles increases. This results in greater mechanical incomplete combustion heat loss Q4, thereby reducing thermal efficiency. Moreover, upward displacement of the flame center further elevates flue gas temperature. Incomplete combustion elevates unburned carbon content, consequently diminishing thermal efficiency. The unit's steam parameters also significantly influence boiler thermal efficiency. Studies have shown that, for single-reheat units, increasing steam pressure by 2 MPa can improve boiler thermal efficiency by approximately 0.4%. Based on the standard coal consumption of 280 g/kWh for ultra-supercritical units, this translates to a savings of about 1.12 g of standard coal per kilowatt-hour of electricity generated. Main steam pressure ranks third in importance, further validating this finding. This outcome indicates that the variable importance rankings generated by the RF algorithm are consistent with established physical patterns governing boiler thermal efficiency. To balance model complexity and generalizability, the top 10 variables identified by the RF are selected as input data for the model.

Data reconstruction

To satisfy these input requirements of 2-D CNN, the input variable must be restructured. This study applies a sliding window approach to segment the time-series data along both the time and feature dimensions, following the data processing methods for sensor data outlined in [25]. The sliding step is 1, the feature dimension includes 10 variables ($W = 10$), and the time dimension includes 20 points ($H = 20$). The 2-D tensor's label corresponds to the thermal efficiency at the final time point of the time dimension.

Dataset segmentation

The original 20160 sampling data points are segmented into 20141 standardized input matrices. Each matrix is paired with a thermal efficiency label, originally indexed from sample 20 to sample 20160. These indices are reordered to range from 1-20141 to align with the matrix indices, ensuring that both the input matrices and labels are numbered starting from 1. The entire segmented dataset is split into training and testing sets at an 8:2 ratio. 20% of the entire dataset (indices 12086 to 16113) is designated as the validation set, which is extracted from the training set while maintaining the original sampling order. Figure 4 illustrates the data partitioning scheme using the thermal efficiency sampling sequence as an example.

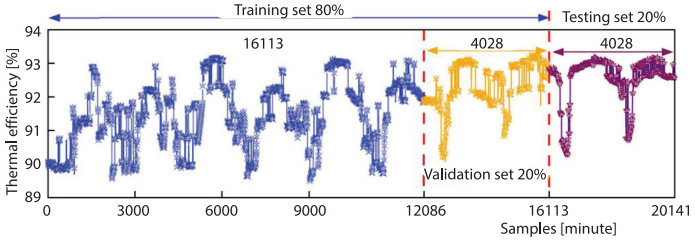

Figure 4. Schematic diagram of dataset segmentation ratio

Model parameter optimization

The model's weights and biases are determined using the training set, while the validation set is employed to optimize hyperparameters such as the number of convolutional layers,

number of filters, kernel size, learning rate, batch size. As shown in tab. 3, the hyperparameters of C3-CNN are determined through multiple rounds of validation. The batch size is set to 32, the compression factor, α, in the C3-Unit to 16, and the number of epochs to 200. This study employs the Adam optimizer with a learning rate of 0.001.

Table 3. Structure of C3-CNN when feature dimension is 10 variables

Layer name	Kernel size	Stride	Padding	Filter number	Input size
Input	–	–	–	–	20 × 10 × 1
Conv-1	2 × 2	1 × 1	1 × 1	64	20 × 10 × 1
Conv-2	3 × 3	1 × 1	1 × 1	64	21 × 11 × 64
Conv-3	–	–	–	–	13 × 13 × 64
C3-unit	3 × 3	2 × 2	1 × 1	128	21 × 11 × 64
Flatten	–	–	–	–	11 × 6 × 128
Fully-connected	–	–	–	–	8448 × 1

Comparative testing and analysis

To evaluate the performance of the C3-CNN thermal efficiency prediction model, this section compares it with multiple other models. The standard deviation (std) and mean predictive errors are derived from 20 experiments. The experimental models are run within the PyTorch framework, with all experimental code executed on a server equipped with a 24 GB NVIDIA GPU.

Different modelling methods

The C3-CNN model is compared with RNN-based models (GRU and LSTM) on the testing set. Both LSTM and GRU adopt one hidden layer. Given the input tensor of size 20 × 10, 20 LSTMCells and GRUCells in the hidden layer, respectively. the LSTM and GRU last cell as a whole prediction output. As shown in fig. 5, the distribution of prediction errors for LSTM, GRU, and C3-CNN across twenty independent experiments is presented. The C3-CNN achieves an RMSE of 0.101 ±0.008%, an MAE of 0.065 ±0.007%, and an R^2 of 0.978 ±0.003, outperforming both the LSTM and GRU models in all performance metrics.

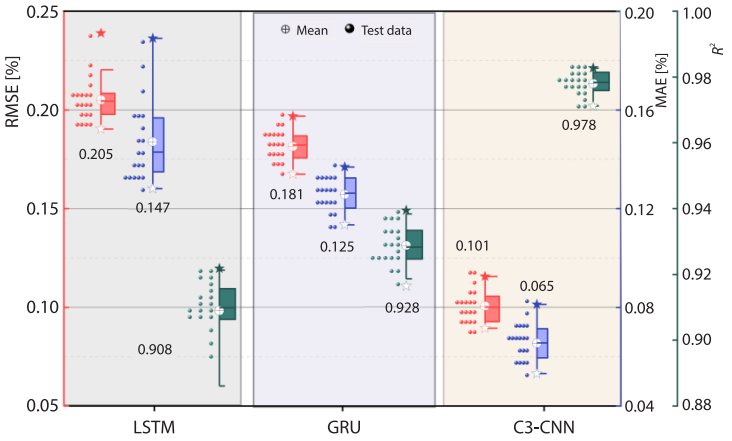

Figure 5. Prediction errors of three models

Figure 6 illustrates the prediction curve from one randomly selected experiment out of the twenty trials, along with the corresponding boiler load curve. Figure 6(a) shows the load reduction phase, where a strong positive correlation is observed between the boiler load and the variation in thermal efficiency. The higher the load, the greater the combustion efficiency within the furnace and the lower the boiler heat loss, so the higher the thermal efficiency. Conversely, at lower loads, combustion becomes incomplete, steam quality deteriorates, and work efficiency decreases, resulting in lower thermal efficiency. For C3-CNN, the error between the measured and predicted values is the smallest among the models in the load decrease interval, indicating that C3-CNN is capable of handling the load reduction conditions. The load increase interval is shown in fig. 6(b), where C3-CNN also exhibits the smallest deviation between the actual and predicted data among the strong models, suggesting that C3-CNN can handle the load increase conditions as well. Figure 6(c) shows the load stabilization interval, during which the boiler operates stably with minimal fluctuation in thermal efficiency, and the prediction performance of all three models is very similar. Above all, the C3-CNN is capable of handling both steady-state and transient conditions. The LSTM and GRU prediction performance is similar. As noted in reference [26], due to the highly similar architectures of GRU and LSTM, in many types of prediction tasks, GRU is not obviously better than LSTM.

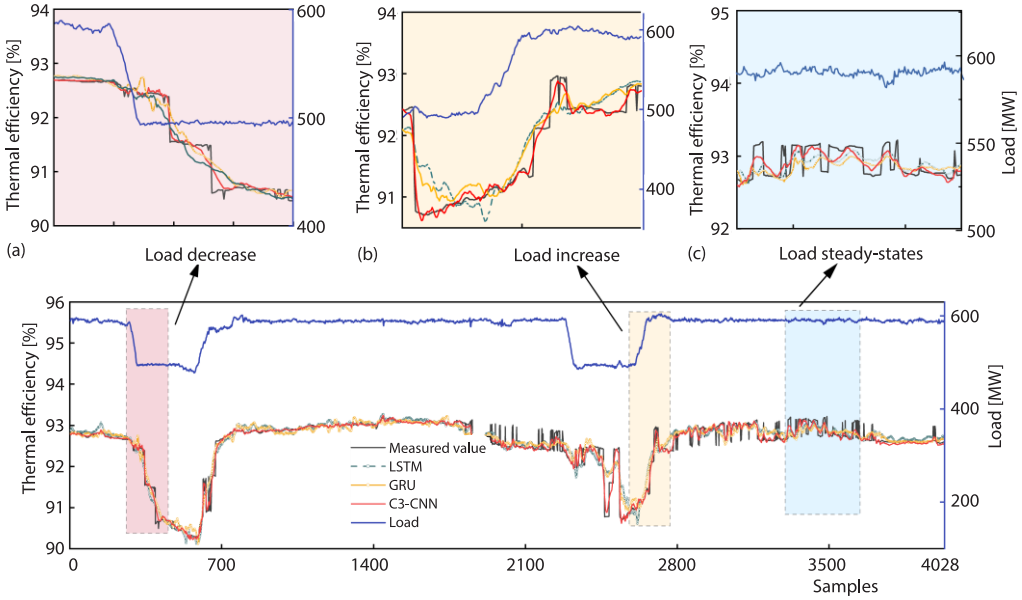

Figure 6. Prediction results of three models

Comparison of convolutional neural network architectures

To compare the model complexity and generalization performance of C3-CNN, two baseline models were designed: a 3-layer CNN and a 6-layer CNN. The 3-layer CNN is used to assess the impact of the C3-unit on the CNN. The 6-layer CNN aims to verify whether the three-layer C3-CNN can achieve the predictive accuracy of the deeper baseline CNN, it consists of six convolutional layers.

On the test set, the scatter plots of the prediction results of the three CNN models are present in fig. 7. The prediction results of the 3-layer CNN are shown in fig. 7(a), the fit-

ted curve deviates from the perfect line. The 95% confidence interval area is relatively large, indicating a considerable error between the measured and predicted data. These results indicate that only three convolutional layers struggles to achieve satisfactory thermal efficiency prediction performance. The prediction results of the 6-layer CNN and C3-CNN are shown in figs. 7(b) and 7(c), respectively. The two models' prediction results are distributed at relatively close distances on either side of the perfect line, with the fitted curves closely following the perfect line. The fitted slopes are 0.96 and 0.97, respectively. The C3-CNN shows a narrower 95% confidence interval, with fewer experimental points falling outside the prediction band. For the 6-layer CNN, its RMSE is 0.131 ±0.003%, the MAE is 0.085 ±0.005%, and the R^2 is 0.962 ±0.001, the prediction error was close to that of C3-CNN, indicating that the C3-CNN achieves, or even surpasses, the six-layer baseline model. Regarding the training times of the three models, the three convolutional layers model has the shortest training time at 181.249 ±4.279 seconds, the 6-layer CNN has the longest at 433.085 ±9.564 seconds, and the C3-CNN records a training time of 303.250 ±6.591 seconds.

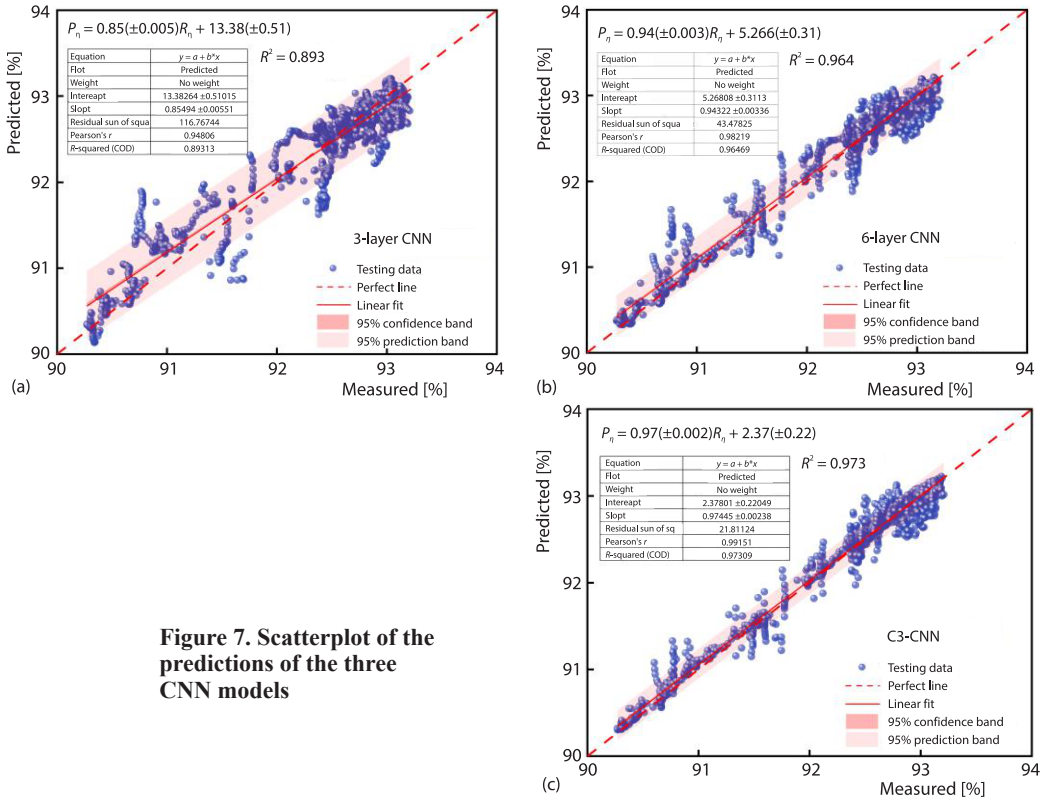

Figure 7. Scatterplot of the predictions of the three CNN models

Conclusions

To accurately predict boiler thermal efficiency, a lightweight dynamic modelling approach based on the proposed C3-CNN architecture is introduced. The model consists of a C3-unit and a baseline CNN. The C3-unit, as a plug-and-play lightweight component, innovatively interacts features from the same convolutional layer, effectively enhancing the model's predictive performance. Simulation experiments are conducted using historical data collected from a 600 MW down-fired coal-fired boiler. The conclusions are as follows.

- The input variables for thermal efficiency are obtained by supervised dimensionality reduction using RF, with flue gas temperature receiving a weight of 0.838 in the feature importance ranking. This result is consistent with the fact that flue gas heat loss represents the largest portion of total boiler heat loss.
- Comparative experiments with LSTM and GRU models demonstrate that, under steady-state conditions, the predictive curves of GRU, LSTM and C3-CNN are closely aligned, exhibiting minimal differences in performance. However, under transient conditions involving rapid load changes, C3-CNN maintains superior prediction accuracy, outperforming both LSTM and GRU.
- In the comparative experiment with an identical CNN architecture, the C3-Unit by promoting communication between feature mapping to improve the prediction performance of the model.

Multi-modal data input will provide models with rich feature information, representing the future direction of modelling.

Acknowledgment

This research was supported by 2024 Annual Xinjiang Key Laboratory of High-Value Green Utilization of Low-rank Coal Open Project [Grant No. XJDX2314-ZD202404], the Jiangsu Funding Program for Excellent Postdoctoral Talent [Grant No. 2024ZB049].

References

[1] Eslick, J. C., Zamarripa, M. A., *et al*., Predictive Modelling of a Subcritical Pulverized-Coal Power Plant for Optimization: Parameter Estimation, Validation, and Application, *Applied Energy*, *319* (2022), 119226
[2] Wang, H., *et al*., Experimental Study on Peak Shaving Operations for Efficient Pulverized Coal Combustion and Working Fluid Coordination, *Fuel*, *404* (2026), 136237
[3] Zhao, H., *et al*., Coal-Fired Utility Boiler Modelling for Advanced Economical Low-NO_x Combustion Controller Design, *Control Engineering Practice*, *58* (2017), Jan., pp. 127-141
[4] Niu, Y., et al., A Dynamic Non-Linear Model for a Once-Through Boiler-Turbine Unit in Low Load, *Applied Thermal Engineering*, *161* (2019), 113880
[5] Chen, H., Liang, Z., Damper Opening Optimization and Performance of A Co-Firing Boiler in a 300 MWe Plant, *Applied Thermal Engineering*, *123* (2017), Aug., pp. 865-873
[6] Lawal, A. I., *et al*., Prediction of Thermal Coal Ash Behavior of South African Coals: Comparative Applications of ANN, GPR, and SVR, *Natural Resources Research*, *32* (2023), 3, pp. 1399-1413
[7] Liu, R., *et al.*, Quantitative Analysis of Carbon Content in Fly Ash Using LIBS Based on Support Vector Regression, *Advanced Powder Technology*, *32* (2021), 8, pp. 2978-2987
[8] Lv, Y., *et al*., A Dynamic Model for the Bed Temperature Prediction of Circulating Fluidized Bed Boilers Based on Least Squares Support Vector Machine with Real Operational Data, *Energy*, *124* (2017), Apr., pp. 284-294
[9] Zhou, Z.-H., *Machine Learning*, Springer Nature, Berlin, Germany, 2021
[10] Xie, P., *et al*., Dynamic Modelling for NO_x Emission Sequence Prediction of SCR System Outlet Based on Sequence to Sequence Long Short-Term Memory Network, *Energy*, *190* (2020), 116482
[11] Wang, X., *et al.*, A hybrid NO_x Emission Prediction Model Based on CEEMDAN and AM-LSTM, *Fuel*, *310* (2022), 122486
[12] Zhang, Y., Combustion Control of Utility Boiler based on GRU Neural Network, *International Core Journal of Engineering*, *8* (2022), 1, pp. 665-667
[13] Wang, Z., Peng, X., *et al*., The NO_x Emission Prediction Using a Lightweight Convolutional Neural Network for Cleaner Production in a Down-Fired Boiler, *Journal of Cleaner Production*, *389* (2023), 136060
[14] Wang, Z., *et al.*, Dynamic Combustion Optimization of a Pulverized Coal Boiler Considering the Wall Temperature Constraints: A Deep Reinforcement Learning-Based Framework, *Applied Thermal Engineering*, *259* (2025), 124923
[15] Zhang, H., *et al*., Improving the Flexibility of Coal-Fired Power Plants Via a Pre-Gasification Burner with Ultra-Enhanced Flame Stability, *Engineering*, On-line first, https://dpi.org/10.1016/j.eng.2025.04.015, 2025

[16] Tang, Z., *et al*., Auto-Encoder-Extreme Learning Machine Model for Boiler NOx Emission Concentration Prediction, *Energy*, *256* (2022), 124552
[17] Breiman, L., Bagging Predictors, *Machine Learning*, *24* (1996), 2, pp. 123-140
[18] Verikas, A., *et al*., Mining Data with Random Forests: A Survey and Results of New Tests, *Pattern Recognition*, *44* (2011), 2, pp. 330-349
[19] Ioffe, S., Szegedy, C., Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, *Proceedings*, International Conference on Machine Learning, 2015, Lille, France, pp. 448-456
[20] Nair, V., Hinton, G. E., Rectified Linear Units Improve Restricted Boltzmann Machines, *Proceedings*, 27th International Conference on Machine Learning (ICML-10), Haifa, Israel, 2010, pp. 807-814
[21] Hu, J., *et al*., Squeeze-and-Excitation Networks, *Proceeding*s, IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, Ut., USA, 2018, pp. 7132-7141
[22] Jiang, B., *et al*., Semi-Supervised Learning with Graph Learning-Convolutional Networks, *Proceedings*, IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, Cal., USA, 2019, pp. 11313-11320
[23] Yang, J., et al., Graph r-CNN for Scene Graph Generation, *Proceeding*s, European Conference on Computer Vision (ECCV), Munich, Germany, 2018, pp. 670-685
[24] Dai, J., *et al.*. Deformable Convolutional Networks, *Proceedings*, IEEE International Conference on Computer Vision, Venice, Italy, 2017, pp. 764-773
[25] Jeon, Y., Kim, J., Active Convolution: Learning the Shape of Convolution for Image Classification, *Proceedings*, IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, Hi., USA, 2017, pp. 4201-4209
[26] Greff, K., et al., LSTM: A Search Space Odyssey, IEEE Transactions on Neural Networks and Learning Systems, *28* (2016), 10, pp. 2222-2232

Paper submitted: August 16, 2025
Paper revised: September 3, 2025
Paper accepted: September 4, 2025