Prediction of NOx Emissions from a Coal-Fired Boiler Based on Convolutional Neural Networks with a Channel Attention Mechanism

Nan Li 1,*, You Lv 2 and Yong Hu 3

1 School of Information and Electrical Engineering, Lu Dong University, Yantai 264025, China
2 School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China
3 State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources, North China Electric Power University, Beijing 102206, China
* Correspondence: nan_li_ldu@outlook.com; Tel.: +86-17853546260

Abstract: This paper presents a small and efficient model for predicting NOx emissions from coal-fired boilers. The raw data collected are processed by the min–max scale method and converted into a multivariate time series. The overall model’s architecture is mainly based on building blocks consisting of separable convolutional neural networks and efficient channel attention (ECA) modules. The experimental results show that the model can learn good representations from sufficient data covering different operation conditions. These results also suggest that ECA modules can improve the model’s performance. The comparative study shows our model’s strong performance compared to other NOx prediction models. Then, we demonstrate the effectiveness of the model proposed in this paper in terms of predicting NOx emissions.

Keywords: coal-fired boiler; separable convolutional neural network; channel attention mechanism; NOx emissions; prediction

1. Introduction

Nitrogen oxides (NOx) are an unavoidable pollutant in coal-fired power generation. To reduce the harmful effects of NOx on the environment, several low-NOx technologies have been introduced into coal-fired boilers, as they can control NOx emissions more economically [1]. For example, ammonia selective catalytic reduction is an industrialized flue gas-based De-NOx technology that can convert NOx emissions into nitrogen and water [2]. Combustion optimization can reduce NOx emissions by carefully setting the operating parameters of the boiler [3]. These technologies are based on the accurate measurement of NOx emissions. Continuous emission-monitoring systems (CEMS) are routinely used to measure the NOx emissions from coal-fired boilers. However, the analyzer of CEMS is frequently off-line for maintenance due to the harsh environment. Therefore, it is essential to develop an accurate NOx prediction model.

Numerous studies have shown that machine learning (ML) algorithms can be considered as an effective method for modeling NOx emissions [4,5]. ML algorithms are data-driven algorithms that do not require specific knowledge regarding coal-fired boilers when building predictive models. Early researchers often built NOx prediction models based on traditional ML algorithms and small data sets. These studies have demonstrated that these models can develop the relationship between the related operational parameters and NOx emissions. For example, Zhou et al. also built an NOx prediction model based on historical operational parameters, support vector regression (SVR), and ant colony optimization [6]. Lv et al. built models based on historical operational parameters,
least-squares support vector regression (LSSVR), and a partial least-squares method [7]. However, these traditional ML algorithms trained on small data sets have difficulty achieving results with good generalization, which limits the use of predictive models in practice [8].

Recently, several studies focused on NOx prediction models based on deep-learning (DL) algorithms [9]. Some results show that the NOx prediction models based on DL algorithms outperform those based on traditional ML algorithms. For example, Tan et al. showed that a DL-based NOx prediction model based on a long short-term memory (LSTM) network outperforms an SVR-based NOx prediction model [10]. Yang et al. showed that a DL-based NOx prediction model consisting of two LSTM networks is superior to an LSSVR-based NOx prediction model [11]. However, an LSTM network is a time-consuming computational unit with a complex internal structure, making it challenging to learn effective representations from complex data [12]. In addition, pre-processing data methods were often introduced to reduce the difficulty in terms of learning representations in LSTM networks. For example, principal component analysis is used to reduce the dimensions of samples, and complete ensemble empirical mode decomposition adaptive noise is used to provide time-domain dynamics information regarding the samples [13]. These methods reduce the dimensions of the samples based on global knowledge of the dataset, making it difficult to deal with a single sample or a small batch of samples effectively. These drawbacks limit the practical application of NOx prediction models trained using reduced dimensional data. An alternative approach is to design advanced architecture for DL-based models without reducing the dimensions of the samples.

Compared to an LSTM network, a convolutional neural network (CNN) is considered a simple and efficient computational unit that has emerged as the master algorithm in the ML community [14]. In recent years, CNNs have become the dominant algorithm in computer vision and developing efficient network architectures for deep CNN-based models has been a topic of great interest. Simonyan et al. developed a very deep convolutional network, consisting of a stack of convolutional layers followed by three Fully-Connected (FC) layers, which was used for an image classification task [15]. He et al. utilized the residual learning architecture of CNNs to achieve impressive generalization performance on the image recognition task [16]. Sandler et al. built a deep CNN-based model primarily from depth-wise separable convolutions to reduce the model’s computational cost [17]. Ma et al. introduced the channel shuffle operation to boost the performance of depthwise-separable convolutions [18]. These efficient CNN-based models can learn to recognize good representations from complex image data and have been successful in image and video recognition applications. However, little attention has been paid to the use of a CNN-based model for modeling NOx emissions [19]. Therefore, it is worth exploring methods with which to design efficient CNN-based models to accurately predict NOx emissions from coal-fired boilers.

The primary purpose of this paper is to create a novel architectural design of a CNN-based model for NOx prediction. The proposed CNN-based models can efficiently learn good representants from multivariate time series that originate from the distributed control system of a 330 MW subcritical, tangential, pulverized coal-fired boiler. Numerical experiments have shown that the developed model can provide accurate predictions of NOx emissions at the outlet of the furnace. Detailed comparisons with other DL-based NOx prediction models have also been carried out. The remainder of this paper is organized as follows. Section 2 describes the architectural design of the CNN-based model. Section 3 describes the detailed model comparisons and the application of the CNN-based model. Section 4 closes this paper with the conclusions of the study.
2. Methods

2.1. Brief Description of the Boiler

The raw data points for the present research were extracted from a database attached to a 330 MW subcritical, tangential, pulverized coal-fired utility boiler. The boiler has a 14.022 × 14.022 m² cross-section furnace and a height of 65.1 m, and it belongs to a unit of the Dong Sheng power plant in the province of Inner Mongolia, China. The schematic diagram of the boiler is shown in Figure 1. Five layers of primary air nozzles (A, B, C, D, and E) and seven layers of secondary air nozzles (AA, AB, BC, CC, DD, DE, and EE) are located alternately in the vertical direction. Coal–air mixtures are fed to the burners on levels A–D.

Figure 1. Schematic diagram of the furnace.

2.2. Data Preparation

NOx emissions from the combustion process consist mainly of thermal and fuel NOx. The NOx formed by the reaction of nitrogen and oxygen in the air at high temperatures is known as thermal NOx. The formation of thermal NOx is affected by the combustion temperature and oxygen concentration. The fuel NOx originates from the nitrogen content in coal and is affected by the coal’s properties, combustion temperature, and oxygen concentration. Furthermore, the boiler’s operating parameters also have a significant impact on NOx emissions. Firstly, the furnace temperature will gradually reduce when the boiler load is reduced. In addition, when boilers are operated at low loads, the oxygen concentration gradually increases, eventually increasing NOx emissions. Secondly, the primary air will create a recirculation area where the reduction reaction will take place, reducing NOx emissions. Thirdly, the secondary air will reduce the oxygen concentration in the primary combustion zone, resulting in gradual temperature reduction, and ultimately reducing NOx emissions. Based on the theoretical mechanisms of NOx formation, 55 measures affecting the NOx emissions were determined for NOx modeling, as shown in Table 1. Although no real-time data concerning coal quality parameters were recorded in the operational database of the investigated boiler, coal quality parameters can be reflected by the historical sequence of operational variables and NOx emissions. Therefore, coal quality parameters are not included in the input variables.

Table 1. Variables used in the modeling.

<table>
<thead>
<tr>
<th>Parameter Name</th>
<th>Identity</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boiler load</td>
<td>(m_1)</td>
<td>MW</td>
</tr>
<tr>
<td>Total airflow</td>
<td>(m_2)</td>
<td>t/h</td>
</tr>
</tbody>
</table>
Total fuel flow \( m_3 \) t/h
Main steam pressure \( m_4 \) Mpa
Main steam temperature \( m_5 \) °C
Main steam flow \( m_6 \) t/h
Coal-feeder rate \( m_{17} \sim m_{11} \) t/h
Primary air temperature \( m_{17} \sim m_{21} \) °C
Primary airflow \( m_{17} \sim m_{21} \) %
Secondary air temperature \( m_{22} \sim m_{23} \) °C
total secondary airflow \( m_{24} \sim m_{25} \) %
Secondary airflow \( m_{26} \sim m_{49} \) %
OFA airflow \( m_{50} \sim m_{53} \) %
Oxygen concentration before the selective catalytic reduction inlet \( m_{54} \sim m_{55} \) %

In this study, we collected 86,400 raw data points covering 5 days from the distributed control system with a time resolution of 5 s. Three steps were taken sequentially regarding processing the raw data points to construct the dataset for modeling NOx emissions. Firstly, extreme outliers were removed from raw data points to enhance dataset quality. Secondly, the remaining data points were processed using the min–max scaling method to eliminate scale differences between data points based on formula (1) [20]

\[
x_i = \frac{m_i - \min(m_i)}{\max(m_i) - \min(m_i)}
\]

where \( m_i \) denotes the measurement of the raw data points, and \( x_i \) denotes the measurement of the processed data points. The values of \( \max(m_i) \) and \( \min(m_i) \) can be determined due to the collected raw data points. In addition, a processed data point is of the same dimensions as a raw data point.

Thirdly, each sample is formatted into a multivariate time series with timestep of 30 in the following manner [19]

\[
\begin{pmatrix}
x_1(k) & \cdots & x_{55}(k) \\
\vdots & \ddots & \vdots \\
x_1(k+30) & \cdots & x_{55}(k+30)
\end{pmatrix}
\]

where \( x_i(k) \) denotes the \( x_i \) at timestep \( k \), and \( e_A \) and \( e_\eta \) denote the true values of the NOx emissions from CEMS.

The dataset was divided into a training subset, a validation subset, and a test subset. The three subsets contain sufficient samples to account for different operation conditions of the boiler. The training subset consists of 51,810 samples covering the first three days of five consecutive days; the validation subset consists of 17,250 samples covering the fourth of five consecutive days; and the test subset consists of 17,250 samples covering the last five consecutive days. The optimal parameters of the model were determined based on the training subset and validation subset. The test subset is used to evaluate the model’s performance using the root-mean-square-error (RMSE) criterion [13], the mean absolute error (MAE) criterion [13], and the \( R^2 \) score [20]. These three criteria are defined as follows

\[
RMSE = \frac{\sqrt{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}}{N}
\]

\[
MAE = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i|
\]

and

\[
R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \sum_{i=1}^{N}y_i/N)^2}
\]
where $N$ denotes the number of samples, $y_i$ denotes the measured values, and $\hat{y}_i$ denotes the corresponding predicted values.

2.3. A Brief Introduction to CNN

CNN is a neural network that uses a standard convolution instead of general matrix multiplication in its layers. For the standard convolution, a patch of size $N \times N$ is extracted from the input, and a scalar value is computed by dot products of the patch and a convolution kernel of the size $N \times N$. The effect of the standard convolution is to filter the input according to the convolution kernel and combine the features to produce a new representation.

To improve the computational efficiency of CNN, the standard convolutional has been replaced by separable convolution, which consists of two successive standard convolutions. The first standard convolution is performed independently over each channel of input. The second standard convolution is used to create a linear combination of the output channel of the first convolution operation. Compared to a standard convolution, a separable convolution has a lower computational cost. For example, if the standard convolution and the separable convolution have the same input with the dimensions $D_p \times D_p \times M$ and an output with dimensions $D_p \times D_p \times N$, then the computational cost of the standard convolution is $D_p^2 D_p^2 MN$ and the computational cost of the separable convolution is $D_p^2 D_p^2 M + D_p^2 MN$, where $D_p$, $M$, and $N$ are positive integers greater than 2. Note that $(D_p^2 D_p^2 M + D_p^2 MN) / (D_p^2 D_p^2 MN) = 1/N + 1/D_p^2 < 1$, which implies that a separable convolution has a lower computational cost than a standard convolution. A CNN that uses the separable convolution is called “separable CNN”.

2.4. A Brief Introduction to the Channel Attention Mechanism

Recently, the channel attention mechanism has been shown to have the potential to improve the performance of CNN-based models. In this study, we introduce the efficient channel attention (ECA) module [21]. As shown in Figure 2, the ECA module first performs a global-average-pooling operation and then performs a standard convolution with a convolution kernel size determined based on formula (6) [21]

$$k = \left\lfloor \log_2 C + 1 \right\rfloor_{\text{odd}}$$

where $k$ denotes the size of the convolution kernel, $C$ denotes the given channel dimension, and $\lfloor t \rfloor_{\text{odd}}$ denotes the nearest odd number of $t$.

Subsequently, the ECA module performs a sigmoid activation function on the output of the CNN to determine the channel attention weight. Finally, the input of the ECA module is multiplied by the channel attention weight.

**Figure 2.** The overview of the ECA module.

2.5. The Architectural Design of the NOx Prediction Model

Figure 3 shows a schematic representation of our model. The model has a streamlined architecture that uses multiple building blocks. In our model, the parameters of the initial CNN and all of the separable CNNs need to be set manually. The parameters of CNNs in the ECA modules also need to be set manually, aside from the size of the convolution kernel. Table 2 shows the detailed configuration parameters of our model.
Table 2. The detailed configuration of the components in our model (shown in columns).

<table>
<thead>
<tr>
<th>Components</th>
<th>Filter</th>
<th>Kernel Size</th>
<th>Stride</th>
</tr>
</thead>
<tbody>
<tr>
<td>The initial CNN</td>
<td>32</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Building block 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Separable CNN</td>
<td>32</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Separable CNN</td>
<td>64</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>CNN in the ECA module</td>
<td>1</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Building block 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Separable CNN</td>
<td>64</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Separable CNN</td>
<td>128</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>CNN in the ECA module</td>
<td>1</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Building block 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Separable CNN</td>
<td>128</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Separable CNN</td>
<td>256</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>CNN in the ECA module</td>
<td>1</td>
<td>4</td>
<td>0</td>
</tr>
</tbody>
</table>

Firstly, the initial CNN with stride two is used to downsample the input and transform the input into different channels for the following building blocks.

Secondly, three building blocks with varying parameter settings are repeatedly placed in a cascade to learn the intermediate representations. Building deep neural networks by stacking building blocks is necessary for learning good data representations. These building blocks are based on the practical guidelines proposed for lightweight CNN architectural design [22]. The practical guidelines contain two points. Firstly, the separable CNNs should be used to build a lightweight deep neural network with a streamlined architecture. Secondly, an efficient building block should consist of separable CNNs followed by batch normalization operation and the ReLU activation function. However, the attention mechanism needs to be included in the practical guidelines to improve the performance of the developed model. Therefore, in addition to using separable CNNs in our model, we have added a channel attention mechanism to enhance the model’s performance. The separable CNN, which is more computationally efficient than CNN, constitutes the main body of the building block [23]. Subsequently, the ECA module is implemented after each separable CNN to refine the intermediate representations. The following batch normalization (BN) operation was used to accelerate the model’s training [24], and the ReLU activation function can effectively improve the performance of CNN-based models.

Thirdly, global average pooling is used to learn the final representations. The input of the global-average-pooling operation is a tensor consisting of feature maps, which are the output of the third building block. These feature maps can be considered as numerical matrices of equal dimensionality. Global average pooling calculates each feature map’s average and forms the one-dimensional features that can be regarded as the final representations learned by our model. The final component is a regular, fully connected (FC) layer that predicts NOx emissions at sides A and B based on the final representations.

Figure 3. The architectural overview of our model.
2.6. Training Infrastructure

In this study, all models were implemented using the DL library Keras with TensorFlow as the backend and trained on a single NVIDIA GeForce RTX 2080 with 16 GB of memory. All models were trained using the Adam optimizer and a batch size of 256 based on the RMSE criterion. The early-stopping strategy and the checkpoint procedure were applied to the validation subset to combat the over-fitting of the model. When the model’s performance on the validation subset could not be improved, the early-stopping strategy terminated the model’s training, and the checkpoint procedure saves the optimal model parameters.

3. Results and Discussion

3.1. Effect of the ECA Module

We removed all the ECA modules from our model to analyze the effectiveness of the ECA module in our model. Still, we kept the parameter settings of the initial CNN, the separable CNNs in the three building blocks, and the fully connected layer unchanged. To eliminate the effect of algorithmic randomness on the model’s evaluation, our model was calculated 30 times with and without the ECA module. Therefore, each case has thirty RMSE values, thirty MAE values, and thirty $R^2$ scores. Table 3 shows the summary statistics of our model with and without the ECA module for 30 runs. The overall performance of our model without ECA modules is worse than our model with ECA modules: firstly, the means of the RMSE values and MAE values for the model without the ECA modules are larger than those of the models with the ECA module. Secondly, the $R^2$ score values for the model without the ECA module were smaller than those for the model with the ECA module. Thirdly, the standard deviations of the RMSE values, MAE values, and $R^2$ scores for the model without the ECA modules are larger than those of the models with the ECA module. Although the introduction of the ECA module adds an additional computational burden, the ECA module can retain the intermediate representations that are more effective in improving the model’s predictive performance during the training process while discarding those that are less effective. Therefore, the inclusion of the ECA module in the model enhances the ability of the model to learn representations, thus improving the model’s performance.

Table 3. The comparison results of our model with and without the ECA module.

<table>
<thead>
<tr>
<th></th>
<th>RMSE (mg/Nm$^3$)</th>
<th>MAE (mg/Nm$^3$)</th>
<th>$R^2$ Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean Standard Deviation</td>
<td>Mean Standard Deviation</td>
<td>Mean Standard Deviation</td>
</tr>
<tr>
<td>With the ECA module</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Side A</td>
<td>3.03 0.81</td>
<td>1.93 0.54</td>
<td>99.23% 0.39%</td>
</tr>
<tr>
<td>Side B</td>
<td>3.24 0.77</td>
<td>2.22 0.65</td>
<td>99.01% 0.42%</td>
</tr>
<tr>
<td>Without the ECA module</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Side A</td>
<td>4.59 1.63</td>
<td>2.98 1.17</td>
<td>98.14% 1.38%</td>
</tr>
<tr>
<td>Side B</td>
<td>4.71 1.80</td>
<td>3.33 1.23</td>
<td>97.74% 1.91%</td>
</tr>
</tbody>
</table>

3.2. NOx Prediction Results

Table 4 shows the results of the 13th run, which can be considered as the best results of our model among the 30 runs. Figure 4 shows the boiler load and our model’s detailed predictive results at the test subset for the 13th run. Our model’s predicted values of NOx emissions are very close to the reference values of NOx emissions obtained by CMES. The relative errors of almost all the predicted values fall into the interval $[-5\%, 5\%]$ on both sides. The excellent relative error distribution shows that the predicted values are very close to the reference values. These results indicated that our models could accurately predict NOx emissions from the studied boiler.
Table 4. The results of the 13th run for our model.

<table>
<thead>
<tr>
<th></th>
<th>RMSE (mg/Nm$^3$)</th>
<th>MAE (mg/Nm$^3$)</th>
<th>R$^2$ Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our model</td>
<td>Side A</td>
<td>Side B</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.03</td>
<td>3.24</td>
<td>99.23%</td>
</tr>
<tr>
<td></td>
<td>1.93</td>
<td>2.22</td>
<td>99.01%</td>
</tr>
</tbody>
</table>

Figure 4. The variations of unit load and detailed predictive results of our model at the test subset for the 13th run.

3.3. Model Comparison and Discussion

This section conducts a comparative analysis of different DL-based models for modeling NOx emissions to further assess the proposed CNN-based model with ECA modules. The outlines of the DL-based NOx prediction models used for comparison are listed below:

(1) The DNN-based NOx prediction model in [9] is a deep neural network that contains five FC layers. The 5 layers have—in order—57, 32, 16, 4, and 2 neurons.

(2) The LSTM-based model in [11] is a stack of two LSTM layers. The first LSTM layer contains 400 units, and the second LSTM layer includes 100 units. Principal component analysis is used to reduce the dimensions of the samples in the dataset.

(3) VGGNet is a CNN-based architecture that is a stack of CNNs followed by an FC layer with two neurons. This model follows most of the hyper-parameters used in [15].

(4) ResNet is a CNN-based architecture composed of efficient residual learning structures. This model follows most of the hyper-parameters used in [16].

(5) MobileNetV2 is a lightweight CNN architecture that depends on separable CNNs. This model follows most of the hyper-parameters used in [17].

(6) The CNN-based model [19] can be considered as a lightweight CNN architecture based on ShuffleNetV2. The building blocks of the model are composed of separable CNNs and CNNs.
All models were run 30 times in the same hardware and software environment. Figure 5 shows the compared results among the seven models based on the RMSE, MAE, and $R^2$ scores. For each subgraph of Figure 5, the height of the bar indicates the mean of RMSE values, MAE values, or $R^2$ scores, and the black line on the bar indicates the standard deviation corresponding to these three terms. Our model achieved the best results among the seven DL-based models: first, the means of our model’s RMSE values and MAE values are smaller than those of the other six models, indicating that our model is more capable of providing accurate predictions than the other models. Second, the higher values of the $R^2$ scores for our model compared to the other six models indicate that our model establishes a more vital link between a boiler’s operating parameters and NOx emission values. Third, the standard deviations of our model’s RMSE values, MAE values, and $R^2$ scores are smaller than the other six models, indicating that our model has better numerical stability than the other models.

![Figure 5](image.png)

**Figure 5.** Prediction performance of the seven NOx prediction models on the test subset for 30 runs. (a) The mean of RMSE values with the standard deviation of RMSE values. (b) The mean of MAE values with the standard deviation of MAE values. (c) The mean of $R^2$ values with the standard deviation of $R^2$ values.
values with the standard deviation of RMSE values. (c) The mean of R² scores with the standard deviation of R² scores.

Table 5 shows the simulation time taken for developing the prediction models. Although our model takes more time to train than the LSTM-based model, the DNN-based model, and VGG, the predictive performance of our model is much better than these three models. The overall performance of the CNN-based model is slightly inferior to our model. However, it takes approximately twice as long to train the CNN-based model as it does to train our model. Considering the model’s predictive accuracy, the time taken to train our model is acceptable.

Our model achieves stable and accurate prediction results. This is mainly due to two factors: firstly, the use of separable CNNs for learning intermediate representations reduces the computational cost of the model and thus suppresses the overfitting problem to a certain extent. Secondly, the learned intermediate representations are further refined using a channel attention mechanism after each separable CNN, retaining only those features that are valid for predicting NOx emissions, thus improving the predictive accuracy of our model.

Table 5. The simulation time among different prediction models.

<table>
<thead>
<tr>
<th>Time (Second)</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our model</td>
<td>598</td>
<td>169</td>
</tr>
<tr>
<td>LSTM-based Model</td>
<td>233</td>
<td>110</td>
</tr>
<tr>
<td>DNN-based Model</td>
<td>143</td>
<td>54</td>
</tr>
<tr>
<td>CNN-based Model</td>
<td>1267</td>
<td>287</td>
</tr>
<tr>
<td>VGG</td>
<td>526</td>
<td>123</td>
</tr>
<tr>
<td>ResNet</td>
<td>2128</td>
<td>777</td>
</tr>
<tr>
<td>MobileNetV2</td>
<td>791</td>
<td>229</td>
</tr>
</tbody>
</table>

4. Conclusions

We have presented a novel CNN-based model for predicting NOx emissions from a 330 MW tangentially coal-fired power plant boiler. We found that a model consisting of separable CNNs and ECA modules can predict NOx emissions accurately and consistently. In addition, the performance of the model developed herein was improved by the channel attention mechanism. This study, therefore, indicates that a well-designed model architecture is a significant factor in building an accurate NOx prediction model. In particular, the best RMSE values for side A and side B were 3.03 mg/Nm³ and 3.24 mg/Nm³, respectively; the best MAE values for side A and side B were 1.93 mg/Nm³ and 2.22 mg/Nm³, respectively; and the best R² score values for side A and side B were 99.23% and 99.01%, respectively. Our results suggest that the developed model could be applied to yield accurate predictions for NOx emissions from similar pulverized coal-fired boilers. Future work should focus on reducing the simulation time of CNN-based models while maintaining the predictive accuracy of the model. Although our model achieves good prediction results with respect to the current dataset, more parameters need to be set and adjusted by humans. This will result in the model parameters needing to be changed again in practice due to differences in boiler types. Future work should be designed so as to determine the model’s key parameters adaptively.

Author Contributions: Conceptualization, N.L. and Y.L.; methodology, N.L.; software, N.L.; validation, N.L., Y.L. and Y.H.; formal analysis, N.L. and Y.L.; investigation, Y.H.; data curation, Y.H.; writing—original draft preparation, N.L.; writing—review and editing, N.L., Y.L. and Y.H.; visualization, N.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.
References


Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.