Title: Prediction of NOx Emissions from a Biomass Fired Combustion Process based on Flame Radical Imaging and Deep Learning Techniques

Authors: Nan Li [1]

Gang Lu [2] (Corresponding author)

Xinli Li [1]

Yong Yan [2]

Address: [1] School of Control and Computer Engineering

North China Electric Power University

Beijing, 102206, China

[2] School of Engineering and Digital Arts

University of Kent

Canterbury

Kent CT2 7NT

UK

Tel: 00-44-1227823015 (Dr Gang Lu)

Fax: 00-44-1227456084

Email: smile_mokou@163.com, g.lu@kent.ac.uk, lixinli@ncepu.edu.cn and y.yan@kent.ac.uk

Abstract

This paper presents a methodology for predicting NOx emissions from a biomass combustion process through flame radical imaging and DL (deep learning). The dataset was established experimentally from flame radical images captured on a biomass-gas fired test rig. MCA (Morphological component analysis) is undertaken to improve the quality of the dataset, and the region-of-interest extraction is introduced to extract the flame radical part and rescale the image size. The developed DL-based prediction model contains three successive stages for implementing the feature extraction, feature fusion and emission prediction. The fine-tuning based on the prediction is introduced to adjust the process of the feature fusion. The effects of the feature fusion and fine-tuning are discussed in detail. A comparison between various image- and machine learning-based prediction models shows that the proposed DL prediction model outperforms other models in terms of RMSE (Root Mean Square Error) criteria. The predicted NOx emissions are in good agreement with the measurement results.

Keywords: Flame radical imaging, biomass, NOx emission, image processing, deep learning, de-noising auto-encoder

1 **Introduction**

In the foreseeable future, biomass is one of the most common renewable energy resources for power generation due to their wide availability and environmental friendliness (Saidur et al., 2001). Similar to fossil fuel combustion, however, biomass firing produces harmful emissions such as NOx (Skalska et al., 2010). The control of NOx emissions from biomass combustion is therefore still a concern in power generation industry (Demirbas, 2005). Techniques for the monitoring and on-line predication of NOx emissions from biomass combustion are thus highly desirable to meet increasingly stringent environmental regulations.

Studies on NOx emission production and control in fossil fuel combustion have been undertaken computationally and experimentally for many years. The development of an accurate model for NOx emission prediction is crucial. Considerable efforts have been paid in this area but most focused on basic learning algorithms (such as neural networks, support vector machine and their variants) which attempt to use historical and current plant data to establish the relationships between the NOx emissions and the combustion process variables (Zhou et al., 2001, Le Bris et al., 2007, Si et al., 2009). In practical applications, however, some process parameters such as the concentration of pulverized coal and the mass flow rate of coal particles are not available or difficult to be measured by hardware-based sensors due to technical and economic limitations (Zhang et al., 2008). In addition, it is very difficult, if not impossible, to build an accurate analytical model for on-line NOx prediction due to the lack of available combustion parameters (Zhou et al., 2001). It is therefore necessary to develop techniques for NOx prediction using more effective inputs such as characteristic information of the flame which relates directly to the combustion process. Some research has been devoted to establish CFD models for NOx estimation (Li et al., 2004, Stopford, 2009). CFD modelling is, however, generally very time consuming and not applicable to real-time processes (Li et al., 2004).

Flame radicals (i.e., OH*, CN*, CH* and C_2*) act as intermediates in fuel thermal and chemical reactions. Early studies have suggested that the characteristics of flame (including radicals) are closely linked to furnace conditions such as burner structure, fuel type, air-to-fuel ratio, etc. For instance, the information of the flame radicals has proven to be a very good indicator of the air-to-fuel ratio (Sandrowitz et al., 1998, Hardalupas et al.,2004, Krabicka et al., 2011). Previous studies have also shown that there is a close correlation between the characteristics of flame radicals and NOx emissions (Demory, 2007, Love et al., 2011, Hernández et al., 2013). Attempts have been made to establish NOx emission predication models based on hand-crafted features extracted from flame radical images and software computing algorithms. Li et al. (2014) proposed a prediction model based on sparse coding and neural network techniques. Further attempts were made to develop a variable image feature structure based on Contourlet transform and Zernike moments for improved prediction results (Li et al., 2015a), and a prediction model based on NMF (non-negative matrix factorization) texture analysis and fast sparse regression (Li et al. (2015b). The NMF algorithm produces different decomposition modes, each containing independent components. The accuracy of the prediction result can be improved by adjusting the number of the independent components. Another attempt was made where both the radical image features and flame temperature were used to build a neural network for predicting NOx emissions (Li et al., 2012). The above studies have achieved some success on the prediction of NOx emissions, however, some issues remain unresolved. This is mainly due to the fact that fuel combustion is a complex and time-varying process, which would introduce significant noises into the images (such as variations in size, shape, luminosity, temperature, etc.). Such variations can have a great impact on the accuracy in the prediction model. Therefore, the extracted features of the flame image should be representative of and robust to the corresponding operation condition. This is particularly referred to the ability of robustness to exhibit classification invariance to a diverse range of transformations and distortions (Arel and Karnowski, 2010). In the algorithm (Li e al., 2014), the designed feature extractor cannot be optimally tuned to the prediction task. The number of the independent components can be considered as a parameter

adjusted based on the prediction task (Li et al., 2015b). However, the adjustment process cannot directly act on the image feature. Therefore, the feature extractor based on the prediction task requires further research. In addition, the multi-model fusion for the different types (up to four) of radical images has not been studied in detail, leading to the inefficient prediction and additional computational burden. Finally, the existing learning algorithms have limited modeling and representational power which may affect the accuracy of prediction results. Therefore, an effective prediction model of NOx emissions which should rely more on automatic learning and less on hand-designed heuristics is highly desirable.

In the present work, we address these intractable problems by presenting a prediction model based on DL (deep learning) algorithms. The DL algorithms work on a layered, hierarchical architecture for extracting complex structure and building internal representation from rich sensory inputs (Schmidhuber, 2015). In comparison with existing learning algorithms, a key benefit of the DL is that the data (such as signals and images) representations (features) are extracted automatically from massive amounts of unsupervised data. In this study, the FRI (flame radical image) dataset is established experimentally. Image preprocessing techniques such as MCA (morphological component analysis) (Starck et al., 2004) and the ROI (region-of-interest) extraction are employed to improve the quality of the FRI dataset and reduce the computational load of the DL model. The discreetly designed numerical experimentation shows the effectiveness of the proposed techniques in the prediction model.

2 **METHODOLOGY**

2.1 Brief introduction to deep learning

Deep learning has become a hot topic in various fields of research such as classification tasks (Bengio et al., 2007), modeling textures (Osindero, and Hinton, 2008), modeling motion (Taylor, and Hinton, 2009), and object segmentation (Levner, 2009). This is due to recent advances in machine learning and

information processing research, particularly the significant increases of chip processing abilities and training data size (Yu, and Deng, 2011). In comparison with other machine learning algorithms, DL algorithms can effectively exploit complex, compositional nonlinear functions, which make it possible to learn distributed and hierarchical feature representations for both labeled and unlabeled data. Most DL algorithms overcome the limitations of the traditional machine learning architectures which have generally an insufficient depth, resulting in a poor generalization ability in classification and regression tasks. In DL algorithms, the hierarchical architecture with a sufficient depth offers an excellent learning ability on the data representation, particularly for some pattern recognition applications.

In the deep learning, basic structures are usually stacked to construct the hierarchical architecture of DL models (Bengio, 2009). An auto-encoder is a neural network based on the concept of sparse coding (Olshausen, 1996). The aim of the auto-encoder is to learn a compressed, distributed representation for a set of data, typically for the purpose of dimensionality reduction (Hinton, and Salakhutdinov, 2006). The auto-encoder is a discrimination model which learns an approximation to the identity function and discovers the internal structure of the data by limiting the number of hidden units. The DAE (de-noising auto-encoder) adds noise to the original radical image to acquire the corrupted image from which the DAE intends to reconstruct the uncorrupted-image. Recent work has suggested that the DAE is effective in learning the underlying data generating distribution (Vincent et al., 2008), which would be very useful in the construction of the DL model.

2.2 Deep learning model for NOx prediction

Figure 1 shows the block diagram of the prediction model of NOx emissions based on images of flame radicals (i.e., OH*, CN*, CH* and C_2*). The model has three successive stages. The first stage is a learning unit for extracting the image feature of four flame radicals. The resulted image features are taken as the input to the second stage to learn the fusion feature. Some DAEs are stacked to construct

the first two stages. The number of the DAE depends on the image size and the practical problem. In comparison to the feature extraction method referring to hand-crafted features, the designed learning units rely more on automatic learning. In the third stage, the prediction unit establishes the relationships between the fusion features and NOx emissions. In addition, the interaction of the last two stages can improve the accuracy of the prediction results. A detailed description of each stage is given as follows.

The first stage of the prediction model aims to generate image features which are extracted by the four independent channels corresponding to the four flame radicals, as shown in Figure 2. The pixel values of the radical images are given to the corresponding channel to acquire the compacted image feature. Each channel is in fact a DDAE (deep DAE) network. A DDAE network is basically a neural network which has multiple hidden layers. The outputs of each layer are connected to the inputs of the successive layer. The greedy layer-wise approach for the DDAE pre-training is used which trains each layer in turn based on the reconstruction error. The image feature is the output of the activation function of the deepest layer of hidden units. The number of the hidden units reduces layer by layer for the compacted image feature. The training procedure of the n-layer DDAE network can be formulated as following,

$$n_i = c_j(a_i), i = 0,1, \dots, n; j = 1,2, \dots, n \quad (1)$$

$$a_{i+1} = f_j(n_i), i = 0,1, \dots, n; j = 1,2, \dots, n \quad (2)$$

$$r_i = g_j(a_i), i = 1, \dots n; j = 1,2, \dots, n \quad (3)$$

$$a_i^* = \operatorname{argmin}(\|a_i - r_i\|_2), i = 0,1, \dots, n \quad (4)$$

where a_i is the input of the DDAE network, $c_j(\cdot)$ is the corrupted function which adds the noise to input a_i, n_i is the corrupted version of input a_i, and $f_j(\cdot)$ is the encoder function mapping n_i to the a_i; $g(\cdot)$ is the decoder function mapping a_i to the reconstruction r_i. Term $\|a_i - r_i\|_2$ denotes the reconstruction error in a single layer. The greedy layer-wise approach can be expressed by solving equation (4) in every layer rather than among the all DDAE network noted by $\sum_{i=1}^{n}\|a_i - r_i\|_2$. Therefore, the equation (4) is solved in every hidden layer in the DDAE network, and a_i^* is the output of the activation function

of the i^{th} hidden layer. For the n-layer DDAE network, a_n^* is the output of the activation function of the deepest layer of hidden units, and can be considered as the image feature.

The second stage is also a DDAE network which is designed to learn the image features generated in the first stage. In comparison with the DDAE in the first stage, this DDAE network interacts with the prediction unit as shown in Figure 3. The third stage is a three-layer feedforward backpropagation neural network considering as the prediction unit. Supposed the m pairs sample (a_{nk}, y_k), where a_{nk} is the fusion feature computed by the DDAE network, y_k is the measured value of NOx, the pre-training procedure of the n-layer DDAE network in the second stage also follows the equation (1)-(4). , The difference is, however, that the backpropagation algorithm is used to solve the following optimization problem,

$$\{a_{nk}^*\}_{k=1}^m = \operatorname{argmin}\left\{\frac{1}{m}\sum_{i=1}^m\left(\frac{1}{2}\|h(a_{nk}) - y_k\|_2^2\right)\right\}, \quad (5)$$

where $h(\cdot)$ is the activation function in the prediction unit, the term $\frac{1}{m}\sum_{i=1}^m\left(\frac{1}{2}\left\|h_{W,b}(a_{nk}) - y_k\right\|_2^2\right)$ denotes the prediction error. To solve the equation (5), the backpropagation algorithm is used to perform fine-tuning on the DDAE network in the second stage: the prediction error is back propagated through the whole DDAE network in the second stage, and used to improve all the weights of the DDAE network.

3 **Results and Discussion**

3.1 Experimental generation of FRI (flame radical image) dataset and NOx emissions

The FRI dataset is a collection of the real-world images of flame radials captured on a biomass-gas fired combustion test rig using a spectroscopic imaging system developed in early studies (Figure 4(a), Krabicka et al., 2011). The imaging system combines optical splitting and filtering and intensified digital imaging techniques, capable of taking simultaneously the emission images of four flame radicals. A full

description of the imaging system and test rig is given in (Krabicka et al., 2011). In the experiments, the primary fuel is biomass (willow) and propane gas is used to support the combustion. The willow was ground to fine particles and pneumatically conveyed to the burner. The proximate and ultimate analyses of biomass fuel are summarized in Table 1 (Williams et al, 2012). A turntable feeder was used controlled the flow rate of the air and thus the willow. A total of seven different biomass mass flow rates were created by changing the speed of the feeder from 10Hz to 40Hz with an interval of 5Hz under a fixed propane flow rate of 0.6 liter per minute. For each test condition, images of four flame radicals were concurrently captured using the spectroscopic imaging system. A total of 100 images were collected under each test condition. The image sets were collected to reflect the full range of conditions of the combustion process.

The actual values of NOx emissions in the gas flue were measured using a portable gas analyzer (KANE 900 PLUS) during the image acquisition (Figure 4). Three samples were taken under each condition and the average of the three sample data was used as the reference values for the NOx prediction. These reference values are attached to the FRI dataset to form the complete label dataset. Note that, due to the fact that the temperature of the biomass combustion could not reach 1600°C or above under the experimental conditions, the main source of NOx produced was considered to be fuel NOx in this study.

The dataset has a total of 2800 examples which consists of four of image sequences of flame radicals, i.e., OH* (308 nm), CH* (432 nm), CN* (387 nm), and C_2* (514 nm), and is divided into the training set (1680 examples) and testing set (1120 examples). All images have a size of 256×256 pixels, each pixel is represented by a value between 0 and 1, where 0 is black, 1 is white and any value, g, in between represents a different grey-level. Figure 4(b) illustrates typical examples of flame radical images captured in the experiments.

It should be mentioned that, in practical furnaces, whether the images of flame radicals can be acquired or not by an imaging system is dependent up on the accessibility of the furnace. However, considering the fact that many modern power plants have been equipped vision-based monitoring systems such as CCTVs, it is feasible to install the flame imaging system on practical furnaces without a significant modification of the furnace.

3.2 Image Preprocessing

Although the cooling system of the EMCCD camera allows the efficient reduction of dark noise, the images acquired still contain unwanted variances. The MCA algorithm (Starck et al., 2004) was applied over the FRI data to decompose the raw flame radical image into the smoothed image part and the noisy part (noisy texture). Compared with traditional de-noising methods, the noise model is able to learn via the MCA algorithm rather than based on assumptions on the noise source. The selected dictionaries in this study were curvelet dictionary (Candès and Donoho, 1999) and local DCT (discrete cosine transform) dictionary (Starck et al., 2005). The selected dictionaries are mutually incoherent so that each has sparse representations for its intended content type, while giving non-sparse representations on the other content type (Starck et al., 2004). Figure 5 illustrates an example of the decomposition results. It is clear that the smoothed image (Figure 5(b)) and the noisy part (Figure 5(c)) are successfully decomposed. The irregular impurities in Figure 5(c) suggest that the MCA algorithm can appropriately model the irregular noise content by the selected dictionaries whilst traditional noise models based on the probabilistic assumption (such as Gaussian noise) cannot.

In addition, it is recognized that a 256×256 pixel image results in a 65536-dimension vector which induces computationally expensive to train the constructed model. As shown in Figure 4, there is a large area which is the background of the image to be processed, resulting in too much irrelevant information in the 65536-dimension vector. Therefore, the ROI of radical images are extracted and rescaled to 64×64

pixels. The extraction is done by computing first the minimum area of the rectangle containing the flame radical part, and then scaling the segmented image area a square. The square is finally rescaled to 64×64 pixels.

3.3 Impact of the feature fusion and fine-tuning on the prediction model

The assessment of the proposed prediction model focuses on two aspects, i.e., the necessity of the learning unit for the fusion feature; the effect of the fine-tuning done by the prediction unit. Three subsequent model architectures are constructed to facilitate comparative analysis using the RMSE (root mean square error),

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}}(\mathrm{y_i} - \widehat{y_\mathrm{i}})^2}\,, \tag{6}$$

where $\mathrm{y_i}$ is the measured value of the NOx emission, $\widehat{y_\mathrm{i}}$ is the corresponding predicted value, and N is the number of samples. The three model architectures examined are,

MA1 (Model architecture 1), the DL based prediction model as described in section 2. The DDAE network's topology was fixed to five layers with 4096-3000-2000-1500-1000-500 (the numbers in the sequence mean the number of the hidden units in every layer) in the first stage, six layers with 2000-1500-1000-500-300-80 in the second stage. The resulting fusion feature is of an 80-dimension vector.

MA2 (Model architecture 2), a modified model architecture which has the same first a and second stages as MA1. Its third stage was replaced by the LSSVR model (Li et al., 2015a).

MA3 (Model architecture 3), a modified model architecture which has the same first stage as MA1. Its second and third stages were replaced by the LSSVR model. The DDAE network's topology was fixed to 10 layers with 4096-3000-2000-1500-1000-500-300-150-100-50-20. The four kinds of image feature are simply concatenated the four types of radical image features.

It is hoped that, by comparing the RMSE values of MA1 and MA2, we can assess the impact of the fine-tuning on the accuracy of the prediction results; and by comparing the RMSE values of MA2 and MA3, we can assess the impact of the fusion features on the prediction results.

The same training and testing sets of the FRI dataset were used to train and test the three models, respectively. The RMSE values of the three model architectures are illustrated in Figure 6. The RMSE value of the MA2 is lower than that of the MA3, which suggests that the fine-tuning can improve the prediction results. For all three models, the MA1 gives the best prediction accuracy for the testing set with RMSE 1.5 ppm. It shows that the fusion features go simply beyond concatenating the four kinds of radical image features, and the designed prediction model is completeness. It can therefore be concluded the proposed DL based prediction model (MA1) has the most appropriate architecture.

3.4 NOx prediction based on the deep prediction model

Figure 7 shows the predicted values and the standard deviations. The error bars represent the standard deviation of the mean computed on the testing dataset. It is clear that the predicted values are in a good agreement with the reference values which were obtained using the gas analyzer. The maximum standard deviation is 3 ppm occurred for Test 7 and the minimum standard deviation is 0.6 ppm occurred for Test 1. Figure 8 shows the relative errors between the predicted and reference values of the NOx emissions. The maximum relative errors for Tests 1 to 7 are 1.8%, 2%, 1.8%, 1.7%, 1.8%, 1.9%, and 2.6%, respectively. As can be seen from the figure, the proposed DL based prediction model has a good accuracy in the predication of the NOx emissions in comparison to the reference values.

As previously addressed, the major source of NOx emission in this study is considered to fuel NOx. In the case where the flame temperature reaches 1600°C or above such as on a full-scale coal/biomass

boiler, the flame temperature may have to be measured and used as an input to the prediction model to reflect the contribution of the thermal NOx to the total NOx emission. It should also stressed that the combustion process in a full-scale furnace is far more complicated than that in a combustion test rig. The variation of the furnace accessibility by an imaging system may also result in the geometric and luminous distortions of the images acquired. Therefore, other techniques, such as image de-noising, image segmentation, and geometric and luminous corrections, may have to be introduced for preprocessing the images, making them suitable for the proposed DL-based prediction model.

3.5 Comparative analysis of different techniques for NOx emission prediction

There are prediction models which have been developed based on the hand-crafted feature extraction and basic learning algorithms have achieved some success in the NOx prediction for different biomass types and combustion conditions (Li et al., 2015a, Li et al., 2015b, Li et al., 2012). A comparative analysis of different prediction models for NOx emissions was conducted so as to have a further assessment of the proposed DL based prediction model. The outlines of these prediction methods are described as follows: IR (Intensity ratio) and BP (Backpropagation) (IR+BP) neural network (Li et al., 2012) where the hand-crafted feature extraction is based on the ratios of flame radical intensities (OH*, CH*, CN*, and C_2*) together with the flame temperature. A standard three-layer BP neural network is used as the predict unit; CTZM+LSSVR (Li et al., 2015a) where the hand-crafted feature extraction is based on the CTZM (contourlet transform and zernike moment) algorithm, and the LSSVR is used as the predict unit; and NMF-TA+ FSR (fast sparse regression) (Li et al., 2015b) where the hand-crafted feature extraction based on the non-negative matrix factorization and texture analysis, and the FSR is used as the predict unit.

The same training set was used to train these prediction methods, the RMSE values on the testing set was used to evaluate the results. The RMSE values of the different prediction methods are illustrated in

Figure 9. As can be seen, the RMSE value of the DL based prediction model is significantly lower than that of other methods. It attributes to the advantages of the fusion features and the fine-tuning strategy in the DL model. The IR+BP network (Li et al., 2012) has the highest RMSE (32.6 ppm). This may be due to the fact that the ratio of flame radical intensities is consider to be the incomplete statistical description which is not robust enough to represent image variations especially in a real combustion environment. The other two prediction methods in (Li et al., 2015a, Li et al., 2015b) exhibit the improved prediction results and the prediction results are still inferior to the proposed DL based prediction model.

The results presented have proven that there is a possibility of using the radical information for predicating the NOx emissions so as to shorten the control loops for more effective furnace control and emissions reduction. Data obtained under laboratory conditions have given us some confidence in applying such a technique to practical furnaces. It has been recognized, however, that, due to the complexity of the combustion process in a full-scale furnace, other factors such as flame temperature should be considered and used to extend the developed DL-based prediction method for applications on a full-scale furnace.

4. **Conclusions**

In this study, a novel model architecture based on flame radical imaging and DL (deep learning) techniques has been developed for predicating NOx emissions from a biomass combustion process. The FRI (flame radical image) dataset acquired on a biomass-gas combustion test rig has been used to train and test the prediction model. The morphological component analysis and ROI extraction algorithms have been employed to improve the quality of the FRI dataset and reduce the computational load in the data processing. The experimental results have shown that the DL based prediction model, which

integrates well the feature learning, multi-feature fusion and the prediction units, can automatically learn and fuse the radical image features, and thus establish the correlations between the characteristics of flame radials and NOx emissions. Furthermore, the feature fusion and fine-tuning strategies have been shown to be efficient to improve the accuracy of the prediction results. The comparisons among the different prediction model have suggested that the DL based prediction model has the best performance among the prediction models which are based on the hand-crafted feature extraction and basic learning algorithms.

Acknowledgments

The authors wish to acknowledge the Chinese Ministry of Science and Technology (MOST) and the Chinese Ministry of Education for providing financial support for this research as part of the 973 Project (2012CB215203) and part of the 111 Talent Introduction Projects (B13009) at North China Electric Power University.

References

Arel, I., Rose, D. C., and Karnowski, T. P. 2010. Deep machine learning-a new frontier in artificial intelligence research. *Computational Intelligence Magazine, IEEE*, **5(4)**, 13-18.

Bengio, Y., Lamblin, P., Popovici, D., and Larochelle, H. 2007. Greedy layer-wise training of deep networks. *Advances in neural information processing systems 19 (NIPS'06)*, 153-160, Edited Schölkopf, B., Platt, J., Hofmann ,T., MIT Press (1st edition), ISBN-13: 978-0262195683, ISBN-10: 0262195682.

Bengio, Y. 2009. Learning deep architectures for AI. *Foundations and Trends® in Machine Learning*, **2(1)**, 1-127, doi.org/10.1561/2200000006.

Candès, E. J., and Donoho, D. L. 1999. Curvelets: A surprisingly effective nonadaptive representation for objects with edges. *Curves and Surfaces Fitting*, 105–120, Edited by Rabut, C., Cohen, A., and Schumaker, L. L., Vanderbilt University Press, Nashville, TN.

Demirbas, A. 2005. Potential applications of renewable energy sources, biomass combustion problems in boiler power systems and combustion related environmental issues. *Progress in Energy and Combustion Science*, **31(2)**, 171-192.

Demory, R. 2007. Optical measurement of nitric oxide and hydroxyl radicals distributions in combusting diesel sprays. University of Brighton.

Hardalupas, Y., Orain, M., Panoutsos, C. S., Taylor, A. M. K. P., et al. 2004. Chemiluminescence sensor for local equivalence ratio of reacting mixtures of fuel and air (FLAMESEEK). *Applied thermal engineering*, **24(11)**, 1619-1632.

Hernández, J.J., Lapuerta, M., Barba, J. 2013. Flame stability and OH and CH radical emissions from mixtures of natural gas with biomass gasification gas. *Applied Thermal Engineering*, **55(1)**, 133-139.

Hinton, G. E., and Salakhutdinov, R. R. 2006. Reducing the dimensionality of data with neural networks. *Science*, **313(5786)**, 504-507.

Krabicka, J., Lu, G., and Yan, Y. 2011. Profiling and characterization of flame radicals by combining spectroscopic imaging and neural network techniques. *Instrumentation and Measurement, IEEE Transactions on*, **60(5)**, 1854-1860.

Le Bris, T., Cadavid, F., Caillat, S., Pietrzyk, S., Blondin, J., and Baudoin, B. 2007. Coal combustion modelling of large power plant, for NOx abatement. *Fuel*, **86(14)**, 2213-2220.

Levner, I. 2009. Data driven object segmentation, *PhD thesis*, Department of Computer Science, University of Alberta.

Li, K., Thompson, S., and Peng, J. 2004. Modelling and prediction of NOx emission in a coal-fired power generation plant. *Control Engineering Practice*, **12(6)**, 707-723.

Li, N., Lu, G., Li, X. L., and Yan, Y. 2014. Prediction of pollutant emissions of biomass flames using digital imaging, contourlet transform and Radial Basis Function network techniques. *Proceedings of the IEEE International Instrumentation and Measurement Technology Conference (I2MTC'2014)*, 697-700, Montevideo, Uruguray.

Li, N., Lu, G., Li, X. L., and Yan, Y. 2015a. Prediction of Pollutant Emissions of Biomass Flames Through Digital Imaging, Contourlet Transform, and Support Vector Regression Modeling. *IEEE Transactions on Instrumentation and Measurement* (In press, 8 pages), doi:10.1109/TIM.2015.2411999.

Li, N., Lu, G., Li, X. L., and Yan, Y. 2015b. Prediction of NOx emissions from a biomass fired combustion process through digital imaging, nonnegative matrix factorization and fast sparse. *Proceedings of the IEEE International Instrumentation and Measurement Technology Conference (I2MTC' 2015)*, Pisa, Italy, 176-180.

Li, X. L., Sun, D., Lu, G., Krabicka, J., and Yan, Y. 2012. Prediction of NOx emissions based on flame radical imaging and neural network techniques. *Proceedings of the 12th IEEE Imaging Science and Techniques (IST'2012)*, Manchester, UK, 502-505.

Love, N., Parthasarathy, R. N., Gollahalli, S. R. 2011. Concentration measurements of CH and OH radicals in laminar biofuel flames. *International Journal of Green Energy*, **8(1)**, 113-120.

Olshausen, B. A. 1996. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, **381(6583)**, 607-609.

Osindero, S., and Hinton, G. E. 2008. Modeling image patches with a directed hierarchy of Markov random fields. *Advances in neural information processing systems 20 (NIPS' 07)*, 1121-1128, Curran Associates Incorporated, ISBN 160560352X, 9781605603520.

Saidur, R., Abdelaziz, E. A., Demirbas, A., Hossain, M. S., and Mekhilef, S. 2011. A review on biomass as a fuel for boilers. *Renewable and Sustainable Energy Reviews*, **15(15)**, 2262-2289.

Sandrowitz, A. K., Cooke, J. M., and Glumac, N. G. 1998. Flame emission spectroscopy for equivalence ratio monitoring. *Applied Spectroscopy*, **52(5)**, 658-662.

Schmidhuber, J. 2015. Deep learning in neural networks: An overview. *Neural Networks*, **61**, 85-117.

Si, F. Q., Romero, C. E., Yao, Z., Schuster, E., Xu, Z. G., Morey, R. L., and Liebowitz, B. N. 2009. Optimization of coal-fired boiler SCRs based on modified support vector machine models and genetic algorithms. *Fuel*, **88(5)**, 806-816.

Skalska, K., Miller, J. S., and Ledakowicz, S. 2010. Trends in NOx abatement: A review. *Science of the total environment*, **408(19)**, 3976-3989.

Starck, J. L., Elad, M., and Donoho, D. 2004. Redundant multiscale transforms and their application for morphological component separation. *Advances in Imaging and Electron Physics*, **132**, 287-348, doi: 10.1016/S1076-5670(04)32006-9.

Starck, J. L., Elad, M., and Donoho, D. L. 2005. Image decomposition via the combination of sparse representations and a variational approach. *IEEE Transactions on Image Processing*, **14(10)**, 1570-1582.

Stopford, P. J. 2002. Recent applications of CFD modelling in the power generation and combustion industries. *Applied Mathematical Modelling*, **26(2)**, 351-374.

Taylor, G. W., and Hinton, G. E. 2009. Factored conditional restricted Boltzmann machines for modeling motion style. *Proceedings of the 26th Annual International Conference on Machine Learning*, 1025-1032, Montreal, Canada.

Vincent, P., Larochelle, H., Bengio, Y., and Manzagol, P. A. 2008. Extracting and composing robust features with denoising autoencoders. *Proceedings of the 25th International Conference on Machine Learning*, 1096-1103, Helsinki, Finland.

Williams, A., Jones, J. M., Ma, L., and Pourkashanian, M. 2012, Pollutants from the combustion of solid biomass fuels. *Progress in Energy and Combustion Science*, **38(2)**, 113-137.

Yu, D., and Deng, L. 2011. Deep learning and its applications to signal and information. Processing of *Signal Processing Magazine, IEEE*, **28(1)**, 145-154.

Zhou, H., Cen, K. F., and Mao, J. B. 2001. Combining neural network and genetic algorithms to optimize low NOx pulverized coal combustion. *Fuel*, **80(15)**, 2163-2169.

Zhang, X., Yan, W. W., and Shao, H. H. 2008. Nonlinear multivariate quality estimation and prediction based on kernel partial least squares. *Industrial & Engineering Chemistry Research*, **47(4)**, 1120-1131.

List of Figure's Captions

List of Table’s Title

List of Figures

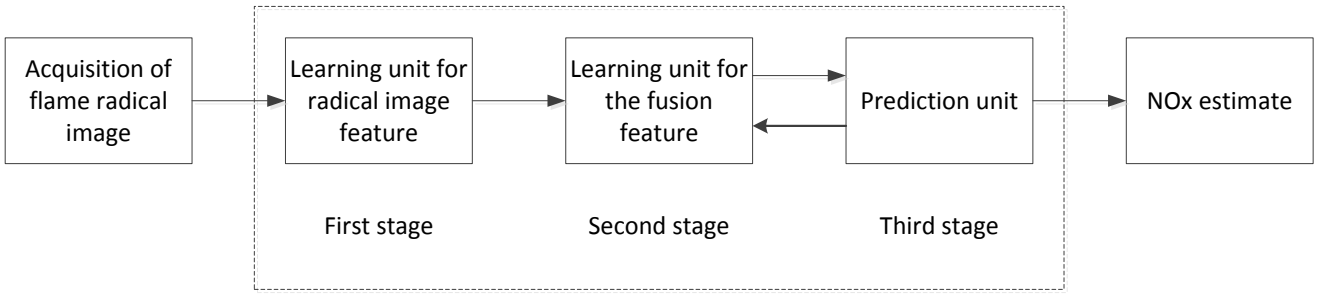


Figure 1. Block diagram of the deep learning model for NOx emissions prediction.

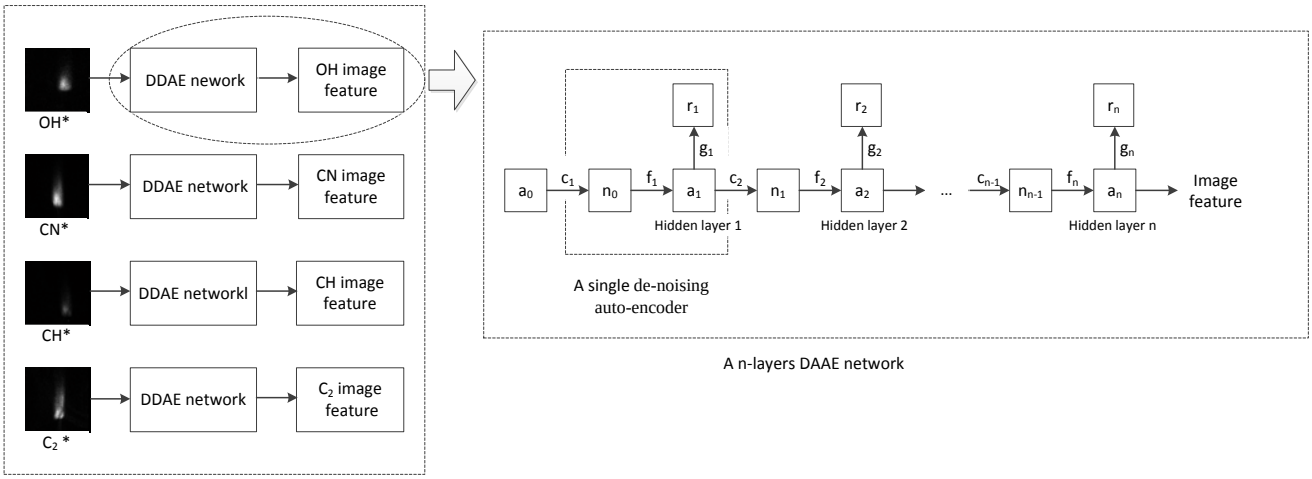


Figure 2. Block diagram of the first stage.

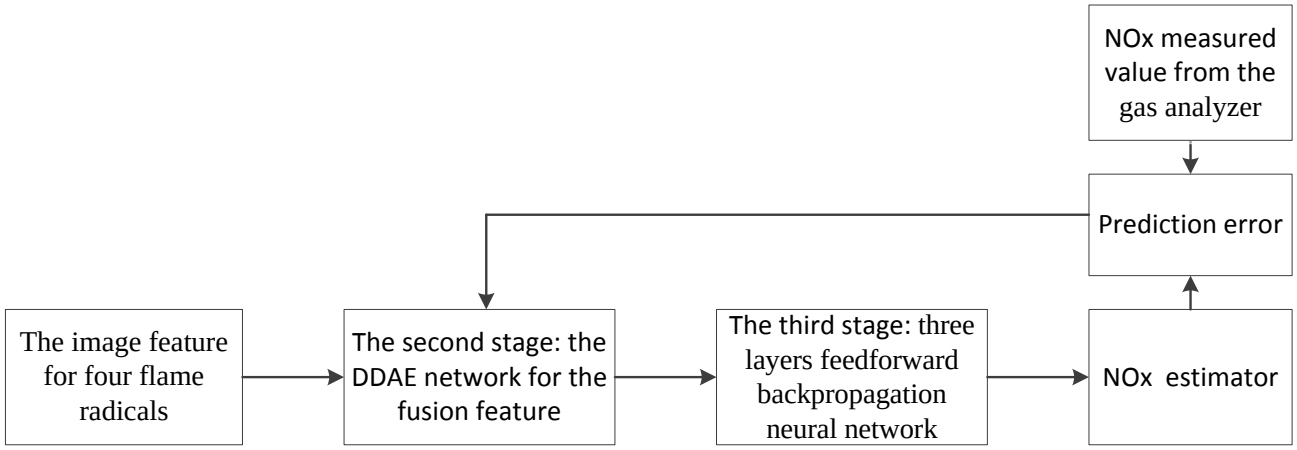
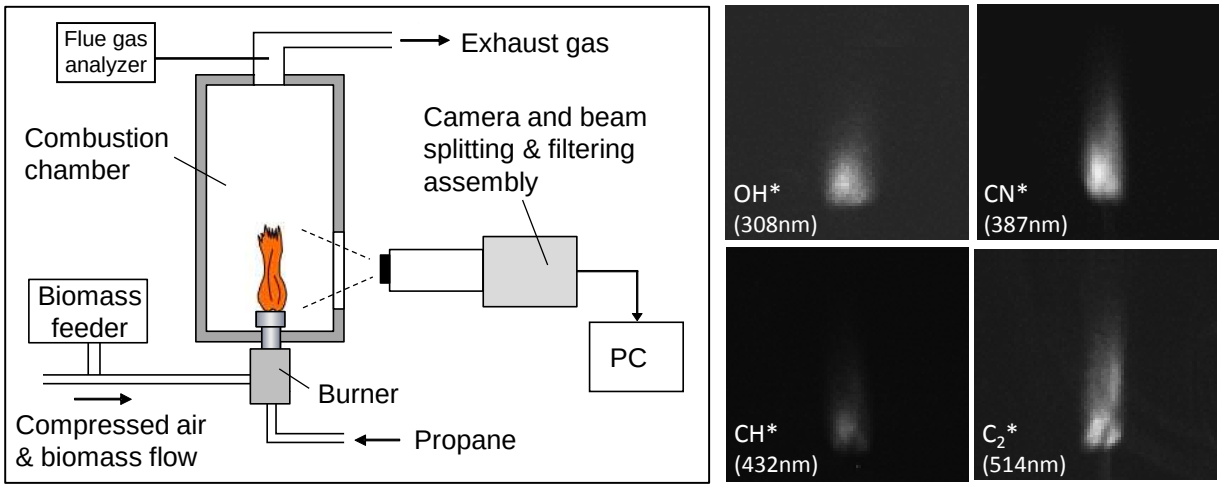
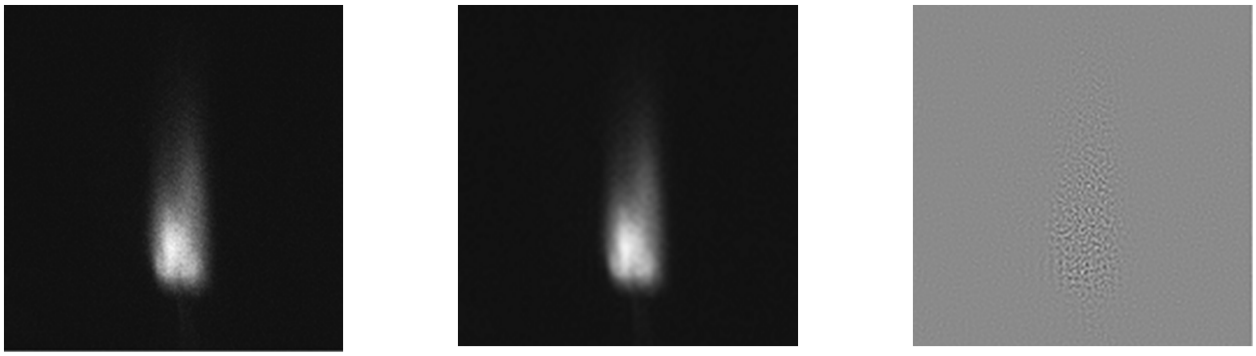


Figure 3. Block diagram of the second and third stages.



(a) Schematic of experimental setup (b) Example images of flame radicals

Figure 4. Schematic of experimental setup and example images of flame radicals.



(a) Original radical image (b) smoothed image (c) Noisy texture

Figure 5. Example results after the MCA.

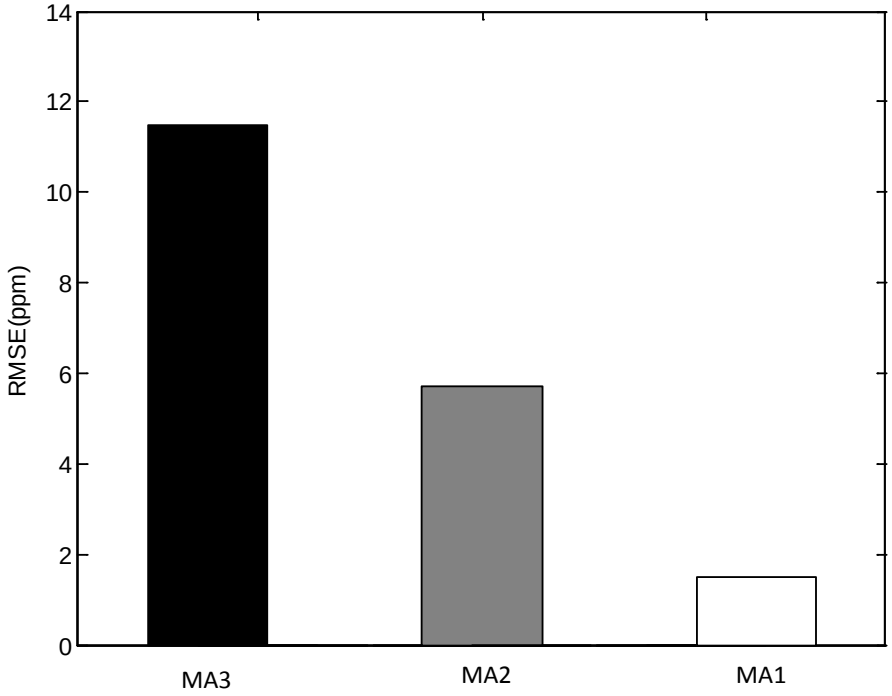


Figure 6. Comparison of RMSE values among three models.

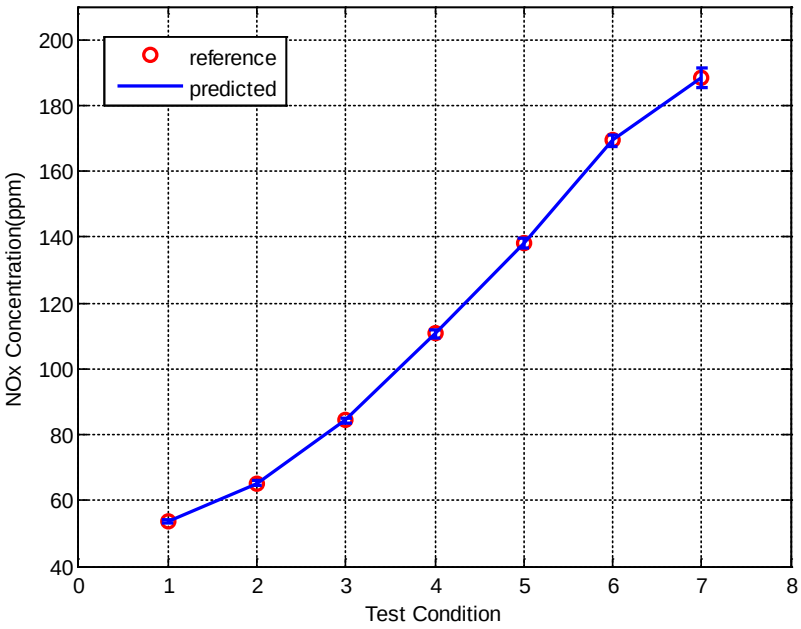


Figure 7. Estimation of NOx emissions.

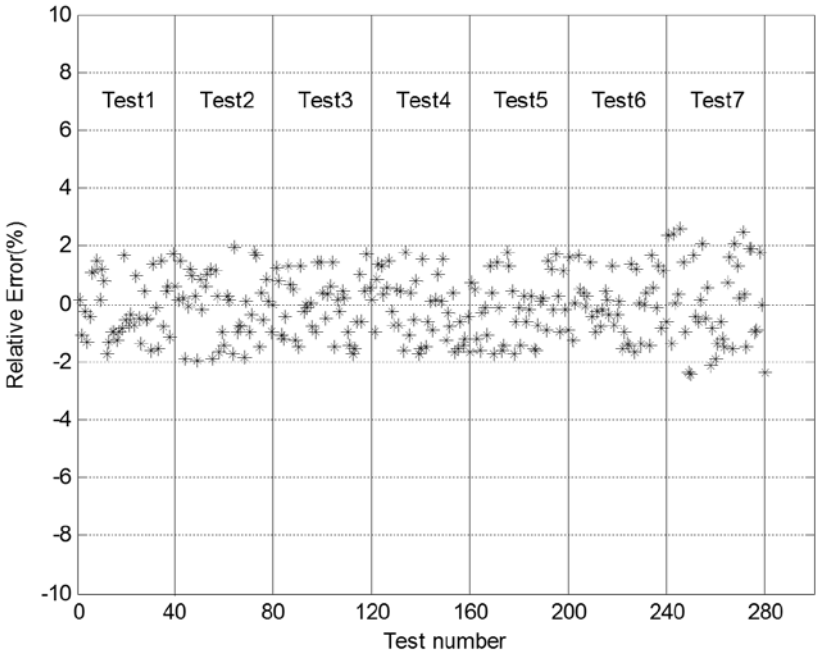


Figure 8. Variations of relative error for NOx emission prediction.

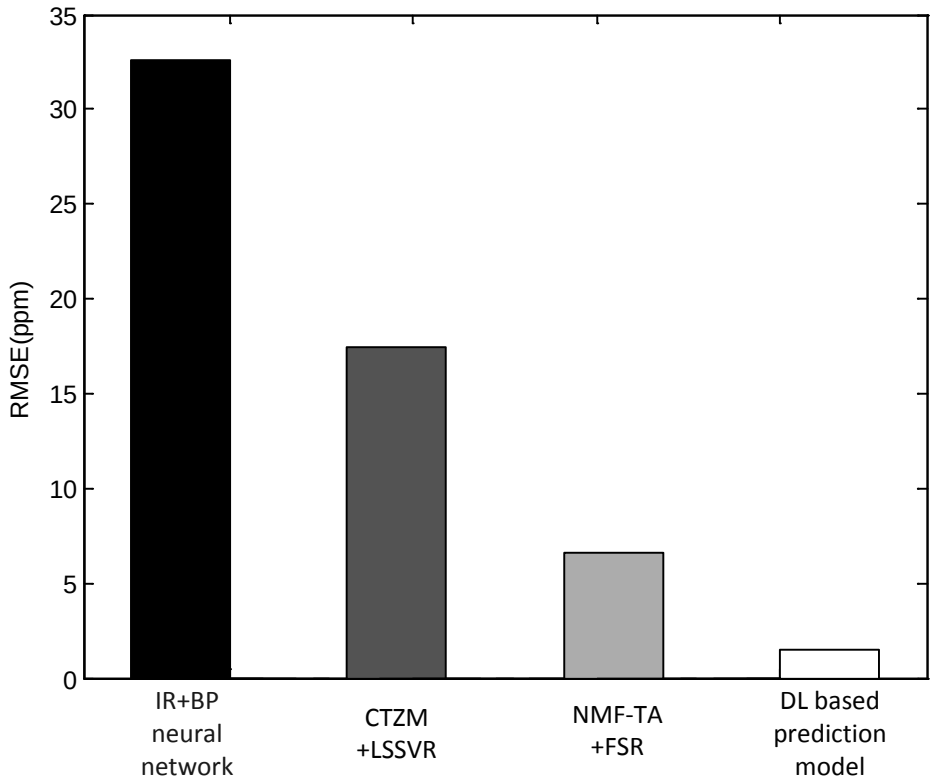


Figure 9. Comparison of RMSE values among the four prediction models.

List of tables

Table 1 Proximate and ultimate analyses of the willow sawdust (Note: *ar- as received, **daf- dry-ash free basis) (Williams et al, 2012)

Content	Value (%)
Moisture (ar*)	6.96
Volatile Matter (ar)	75.7
Fixed Carbon (ar)	16.31
Ash (ar)	1.03
C (daf**)	51.62
H (daf)	5.54
O (daf)	42.42
N (daf)	0.38
S (daf)	0.03
Cl (daf)	0.01