

Article

Multi-Gas Regression from High-Speed Image Sequences Using 3D CNN and 3DResNet Architectures in Biomass Co-Combustion: A Feasibility Case Study

Andrzej Kotyra 

Department of Electronics and Information Technology, Lublin University of Technology, 20-618 Lublin, Poland; a.kotyra@pollub.pl

Abstract

This study explored a spatio-temporal deep learning approach for optical soft sensing of combustion emissions in a coal–biomass co-firing scenario. High-speed RGB flame sequences from a 0.5 MW test rig co-firing hard coal with 10% straw were synchronized with extractive measurements of O₂, CO₂, and NO. These sequences were used to train three shallow 3D CNNs and three 3D ResNet-50 architectures with squeeze-and-excitation attention. The proposed 3D CNN/ResNet models performed simultaneous regression of all three gas species from flame image volumes. The best configuration achieves R² values of 0.975, 0.987, and 0.980, accompanied by mean absolute errors of 0.23% by volume, 13.15 mg/m³, and 0.19% by volume for O₂, NO, and CO₂, respectively, at a resolution of 128 × 96 × 96 pixels. Within the scope of the available dataset, comprising a single measurement run and a single fuel mixture, the results indicate that a comprehensive spatio-temporal analysis of flame images can yield accurate estimates of multiple gas concentrations, thereby providing a promising foundation for the future development of soft optical sensors. At the same time, the study is limited to a single combustion experiment, a single biomass fraction, and a single borescope orientation, and the inference delay and hardware requirements were not quantified; therefore, issues regarding the generalizability of the proposed approach to different conditions and its implementation remain open for further work.

Keywords: biomass–coal co-firing; combustion diagnostics; flame imaging; deep learning regression; multi-gas estimation



Academic Editor: Georgios Mavropoulos

Received: 25 April 2026

Revised: 20 June 2026

Accepted: 25 June 2026

Published: 27 June 2026

Copyright: © 2026 by the author.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

The co-firing of biomass with hard coal in existing power and heating plants is considered a transitional step towards reducing net greenhouse gas emissions in coal-based energy systems [1,2]. In Poland, where electricity and district heating production remain heavily dependent on hard coal and lignite, biomass co-firing is a key strategy for reducing the CO₂ intensity associated with fossil fuels while still using existing industrial facilities [3,4]. Notwithstanding the environmental benefits, co-firing biomass with coal causes several technological issues, including increased slagging, combustion instability, and highly variable pollutant emissions [5]. It has been established that biomass ash, noted for its high alkali metal and chlorine contents, can reduce the ash melting point. This, in turn, encourages the deposition of pollutants and corrosion, processes that have been shown to have detrimental effects on the plant's efficiency and service life. Concurrently,

substantial disparities in moisture content, volatile matter, and calorific value between coal and biomass exert a pronounced influence on flame dynamics, ignition characteristics, and the formation of nitrogen oxides (NO_x), sulfur dioxide (SO_2), and particulate matter [6,7].

Flame instability is a particularly salient issue in biomass–coal co-firing systems due to the simultaneous interaction of multiple fuel-related factors [7,8]. It is well established that biomass typically has a higher moisture content, which absorbs heat during the process of evaporation [9,10]. This, in turn, results in a local decrease in flame temperature, which, in turn, can delay the process of ignition and potentially create cold spots. The presence of elevated levels of volatile matter, coupled with a reduced ignition temperature, has the potential to trigger the unanticipated release of combustible gases. Such releases can lead to alterations in the composition of the flame and the local heat, which may thus become more unpredictable. Furthermore, the varied size, shape, and composition of biomass particles, combined with the variable chemical composition of the resulting ash, lead to uneven combustion and spatially heterogeneous flame propagation [11,12]. The abovementioned factors are contingent upon the mixture ratio, fuel feed rate, and air distribution. The consequence of these factors is increased NO_x , SO_2 , and particulate matter emissions [13]. In this context, effective real-time monitoring and control mechanisms are of key importance. However, it has been demonstrated that traditional combustion sensors often fail to provide the precision and response speed necessary for such rapidly changing combustion conditions [14–16].

From the perspective of combustion efficiency, the CO_2 and O_2 contents in the flue gas are the primary diagnostic indicators [17,18]. In the context of coal-fired boilers, elevated levels of carbon dioxide (CO_2) accompanied by low residual oxygen (O_2) are often indicative of optimal combustion and near-optimal excess air, while high levels of oxygen (O_2) accompanied by relatively lower CO_2 levels suggest excessive aeration, reduced flame temperature, and higher flue gas losses. In the context of biomass co-firing, the previously outlined alterations in ignition behavior, volatile matter release, and local temperature modify the correlations between the air-to-fuel ratio, flame stability, and degree of combustion. Consequently, air and fuel feeding strategies that are optimal for pure coal no longer ensure low levels of CO and unburned carbon under co-firing conditions [19]. A series of experimental and modeling studies on the co-firing of coal and biomass in grate and fluidized bed boilers have confirmed that continuous, accurate measurement of carbon dioxide (CO_2) and oxygen (O_2) is essential for the effective monitoring of combustion efficiency under varying biomass fractions. This is imperative for the regulation of air distribution and timing in real-time, as well as for limiting energy losses and the formation of pollutants. In addition, reliable CO_2 and O_2 data are necessary to determine CO_2 emission factors for individual installations; to quantify the biogenic fraction of CO_2 , which is treated as zero-rated under the EU Emissions Trading System; and for designing or validating post-combustion CO_2 capture systems, whose performance is highly dependent on inlet CO_2 and O_2 levels [20,21].

Nitrogen oxides (NO_x), mainly nitrogen monoxide (NO) and nitrogen dioxide (NO_2), are among the most strictly regulated pollutants for power plants in the EU [22]. The formation of NO_x in coal and coal–biomass blend flames depends heavily on the local temperature, oxygen availability, and nitrogen bound in the fuel, which are influenced by the combustion instabilities and heterogeneous combustion patterns described above. Numerous laboratory and full-scale studies have shown that co-firing biomass with coal can cause significant changes in NO_x emissions—often reductions but, in some cases, increases—depending on the type of biomass, the blend ratio, the air staging configuration, and the overall excess air. Understanding and controlling these trends requires systematic measurement of NO_x alongside O_2 in the flue gas, such that NO_x formation can be linked to

the actual combustion conditions and normalized relative to the reference oxygen content used in emissions regulations [23,24].

Flame image processing has been successfully adapted as a source of information for combustion processes [25–27]. The main advantage of image-based approaches is that they are non-intrusive, provide controlled spatial resolution, and offer substantially lower latency relative to extractive gas analyzers, which are constrained by sampling lines and transport delays. High-speed imaging enables the tracking of rapid transient changes during combustion and, consequently, in gas composition, which is highly valuable in unstable biomass–coal co-firing scenarios, where the local equivalence ratio and temperature fluctuate strongly.

Many researchers have successfully implemented image-based neural networks to estimate exhaust gas concentrations. One of the earliest and most important contributions was the application of deep belief networks (DBNs) to CCD flame images for oxygen prediction. In this approach, stacked Restricted Boltzmann Machines are trained as a multilayer DBN to extract features from raw flame images [28]. The resulting high-level features are then fine-tuned in a supervised regression stage to predict continuous flue-gas O₂ content. It was demonstrated that DBN features more accurately identify nonlinear combustion trends than linear PCA features, and that the DBN-based soft sensor significantly improved the oxygen prediction accuracy relative to the PCA + SVR baseline, thus establishing “DBN + regression” as a strong baseline for flame-image-based O₂ estimation.

Convolutional neural networks (CNNs) trained on RGB images have been successfully applied for O₂ content prediction [29]. More than 4 million high-resolution flame images from a gas-fired boiler were labeled with oxygen levels and used to train a custom multi-channel CNN that stacks several convolution and pooling layers to learn spatial features, followed by fully connected layers that output either oxygen classes or the continuous oxygen percentage, within a semi-supervised “deep entity” framework. In direct comparison with a neural-cascade DBN, this CNN-based model achieved significantly lower prediction error, demonstrating that end-to-end convolutional feature extraction is highly effective for oxygen estimation from flame imagery.

Another important development was the construction of an estimation framework for online oxygen estimation using Wasserstein loss-based generative models and a CNN, in which a variational WGAN is first trained to generate synthetic flame images for under-represented oxygen conditions, following which a CNN regressor was trained on the augmented dataset to predict the continuous O₂ [30]. This combination of generative augmentation and CNN regression yields very high accuracy, exceeding traditional chemometric methods such as PLS and SVR built on hand-crafted image features, underlining the advantage of deep convolutional models for image-based oxygen soft sensing.

The most advanced image-based models for nitrogen oxides generally use convolutional neural networks (CNNs) on physically meaningful image-derived fields, rather than raw color alone. An online deep learning framework that uses flame images to predict NO_x emissions first reconstructs two-dimensional (2D) flame temperature maps using two-color pyrometry. These temperature images are then fed into a convolutional neural network (CNN), which learns the spatial correlations between temperature patterns and emitted NO_x. The network is trained using a combined regression and knowledge distillation loss to improve its generalization ability across operating conditions. This CNN-based soft sensor was shown to accurately predict NO_x emissions not only for conditions seen during training, but also for “unseen” operating conditions [31]. This indicates that deep spatial feature extraction from combustion images (or temperature fields derived from them) can encode sufficient information about NO_x formation. Meanwhile, classical feed-forward or dynamic neural networks have been used to estimate NO_x, CO, and O₂ based on image-derived

flame features, such as size, shape, color, oscillation characteristics, and chemiluminescence intensities. A digital imaging study of industrial flames used engineered descriptors from camera images and chemiluminescence signals as inputs to a feed-forward artificial neural network, achieving NO_x prediction errors of only a few percent under industrial furnace conditions [32]. Another study on flame image analysis applied principal component analysis (PCA) to reduce flame image features. Then, these features were fed into dynamic neural networks and genetic algorithm-optimized neural networks to monitor NO_x and O₂ based on “flame status.” The obtained results demonstrated that even relatively shallow architectures can perform multi-output gas estimation when supplied with informative image features [26].

Information about the state of the ongoing combustion process obtained through flame image processing is more comprehensive when the spatio-temporal nature of the acquired data is taken into account [33,34]. The concentrations of gases are not determined from a single snapshot of a flame; rather, they result from the evolution of ignition, mixing, and heat release over time. For this reason, three-dimensional convolutional neural network (3D CNN) architectures, which process image sequences as x–y–t volumes and learn filters that span space and time, were employed in this study. Such a design was engineered to capture and analyze flicker, motion, and evolving textures, which were synchronized with changes in the equivalence ratio and temperature. This approach is expected to provide a superior inductive bias for unstable biomass–coal co-firing, where feed fluctuations and local quenching lead to transient changes in flame length, intermittency, and brightness that affect the flue gas composition.

The range of architectures considered in this study was limited to 3D CNN and 3D ResNet frameworks to assess relatively shallow 3D CNNs and deeper models, such as ResNet-50, which incorporate residual connections across different spatial resolutions. An additional objective was to verify whether the addition of extra branches containing information on brightness variation statistics and the frequency signature of flame flickering, as a well-known approach for assessing combustion processes [35,36], affects the behavior of the considered models.

2. Materials and Methods

The co-firing experiments were conducted using a 0.5 MW laboratory test rig. The rig was equipped with a combustion chamber 2.5 m long and 0.7 m in internal diameter. The combustion chamber was equipped with a rotating low-emission pulverized-fuel burner, an auxiliary oil burner, and a gas igniter. Observation of the flame was facilitated by viewing windows on both sides of the chamber and a water-cooled Cesyco endoscope (40° lateral field of view) (CESYCO, Bordeaux, France) inserted at a 60° angle to the axis, aimed towards the near-burner zone. The coal–biomass mixture was premixed, stored, and fed to the burner via a screw feeder along with primary air. Secondary air was introduced through a separate inlet to adjust the overall stoichiometry, with the total combustion air volume resulting from the sum of the primary and secondary air streams. A more detailed description of the test setup and its geometric layout can be found in [37].

Flame images were recorded using a Mikrottron MC1311 (Mikrottron GmbH, Unterschleißheim, Germany) high-speed color camera (1280 × 1024 pixels, dynamic range of 59 dB, peak spectral sensitivity at 550 nm) and saved in a lossless 24-bit RGB (8 bits per channel) format. The original resolution of the acquired images was 800 × 800 pixels. The image data were transmitted for storage via the CameraLink interface, ensuring negligible latency. The concentrations of O₂, CO₂, and NO in the exhaust gases were measured simultaneously using an Ultramat 23 analyzer (Siemens AG, Karlsruhe, Germany) every 1 s and recorded using a computerized data acquisition system.

All the data were collected during a single combustion test, in which a mixture containing 10% straw biomass was co-fired with hard coal. Image series were captured at 150 frames per second. None of the image acquisition parameters, including white balance, analog gain, and exposure time, were altered during the test.

Although gas concentrations were recorded throughout the experiment, flame images were captured only during 22 sessions of varying lengths, ranging from 11.1 s to 3 min and 13.6 s. The readouts of the three gas concentrations (O_2 , CO_2 , and NO) are shown in Figure 1, which reveals that the combustion process ran through different operating regimes. During the combustion test, the fuel feed and air flow were altered, resulting in changes in gas concentration. O_2 roughly varied from 0 to 9% vol, CO_2 from about 11 to 18% vol, and NO from ~ 350 to 1200 mg/m^3 , with clear transient spikes and drops rather than smooth trends. The data cover several different regimes (high- O_2 /low- CO_2 , low- O_2 /high- CO_2 , varying NO) with fast transitions, demonstrating the nonlinear and dynamic properties of co-combustion.

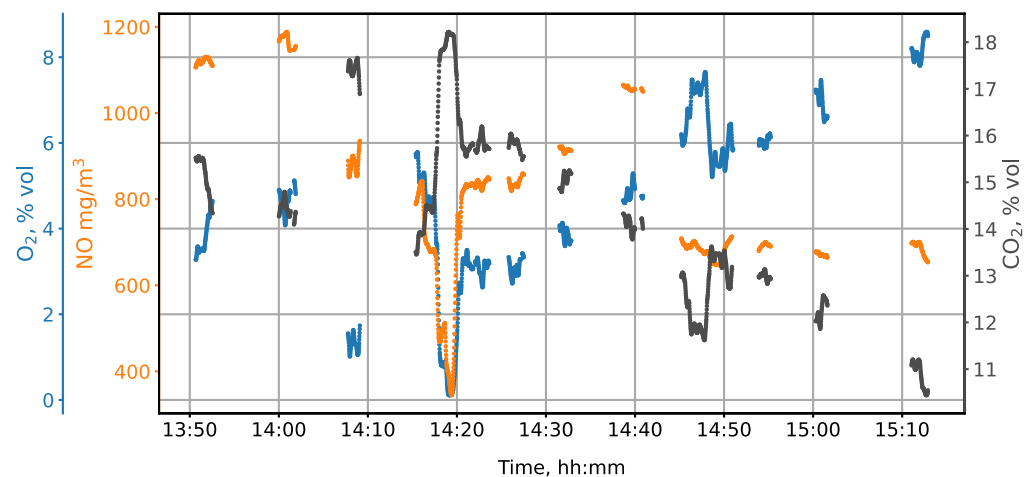


Figure 1. Timeline of the co-combustion experiment providing readouts of gas concentrations inside the combustion chamber: O_2 (left axis, blue), NO (left axis, orange), and CO_2 (right axis, gray).

Prior to the experiment, a synchronization process was implemented between the test bench's data acquisition system clocks and the computer recording the high-speed images. This procedure was undertaken to establish a common time base. To generate the measurements, the time series from the gas analyzer was shifted by the estimated transport delay due to the approximately 7 m sampling line and the analyzer system, such that each image sequence was aligned with the delay-corrected gas concentration reading reflecting the combustion state responsible for that measurement. Consequently, any residual uncertainty in the precise delay manifests as minor fluctuations in the measurements rather than a systematic time deviation.

The gas concentration histograms, presented in Figure 2, confirm that the dataset covers relatively wide, multimodal ranges for all three gases, which is favorable for training and testing a data-driven model. For O_2 , the distribution is clearly not unimodal: there is substantial probability mass between 2 and 6%, with local peaks around 3–4% and additional weight at higher O_2 levels (6–7%). The NO histogram is narrower in absolute terms but also exhibits a structured, multi-peaked shape. A dominant peak appears around $650\text{--}750 \text{ mg/m}^3$, with secondary peaks at higher values ($\approx 900\text{--}1100 \text{ mg/m}^3$), consistent with the distinct combustion states characterized by different temperatures and air-staging conditions. For CO_2 , the kernel estimate shows a broad maximum around 14–16% and

noticeable probability at lower and higher CO_2 , indicating multiple regimes rather than a single operating point.

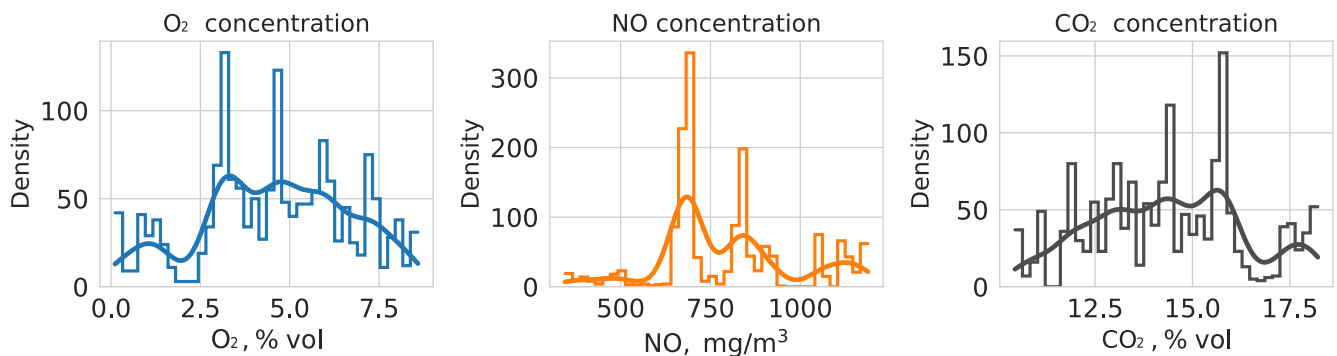


Figure 2. Probability density histograms with kernel density estimates for O_2 , NO , and CO_2 concentrations. The area under all bars sums to 1.

All the histograms show that the data do not concentrate around one narrow value but instead sample a range of combustion conditions with overlapping clusters in the O_2 , NO , and CO_2 concentrations.

2.1. The Shallow 3D CNN Models

The first three models are characterized as shallow 3D CNN architectures, with three 3D convolutional blocks forming the basic framework. These models include a regression head consisting of a dense layer, followed by three parallel paths whose outputs are concatenated to form a vector of gas concentrations.

In Model #1 (Figure 3), the sole input is the image sequence ($T \times H \times W \times C$). A three-block 3D CNN with increasing channel depth ($32 \rightarrow 64 \rightarrow 128$), interleaved with spatial max-pooling and a few $3 \times 1 \times 1$ “bottleneck” convolutions, produces a sequence-level feature vector via global average pooling. This shared feature is passed through a dense layer (128 units), then split into three identical heads (Dense-64 $\rightarrow$ Dropout $\rightarrow$ Dense-1), each predicting a single gas. The three scalars are concatenated to form the output vector.

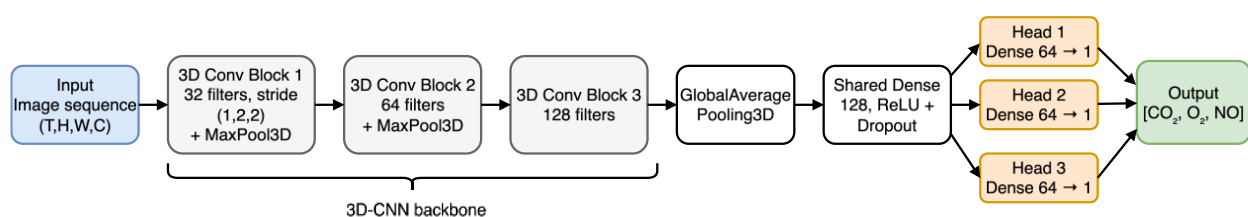


Figure 3. Model #1's architecture. The 3D CNN backbone is used in Models #2 and #3.

In Model #2, the 3D CNN backbone for the image sequence remains unaltered; however, it is augmented with an auxiliary branch that incorporates simple brightness statistics (mean, standard deviation, minimum, and maximum) as an additional input. These scalar features are processed by a small two-layer multi-layer perceptron (MLP) and concatenated with the image features. The fused vector is subsequently fed into a deeper shared multi-layer perceptron (MLP) comprising a Dense-256 layer, a Dropout layer, a Dense-128 layer, and another Dropout layer. This configuration enables the direct output of all gas concentrations in a single linear Dense layer (Figure 4).

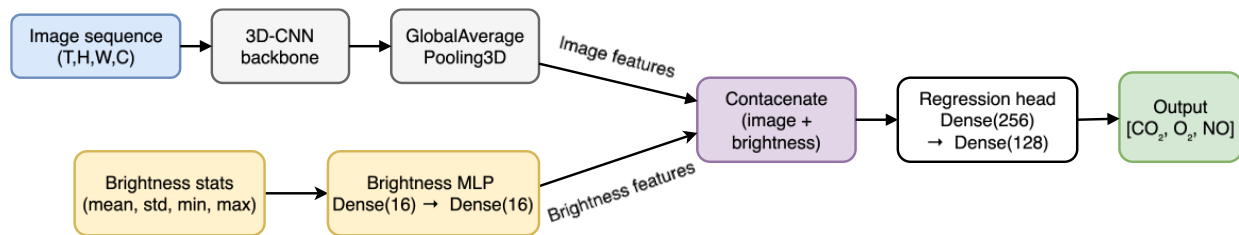


Figure 4. Model #2's architecture.

The third model (Model #3, Figure 5) expands the concept of multimodal fusion through the incorporation of a secondary 3D CNN branch that processes a precomputed short-time Fourier transform (STFT) magnitude volume of the flame. This approach introduces knowledge on the distribution of temporal oscillations at varying frequencies across the flame image, thereby enabling the network to discern gas-related patterns within the time–frequency domain [27,38].

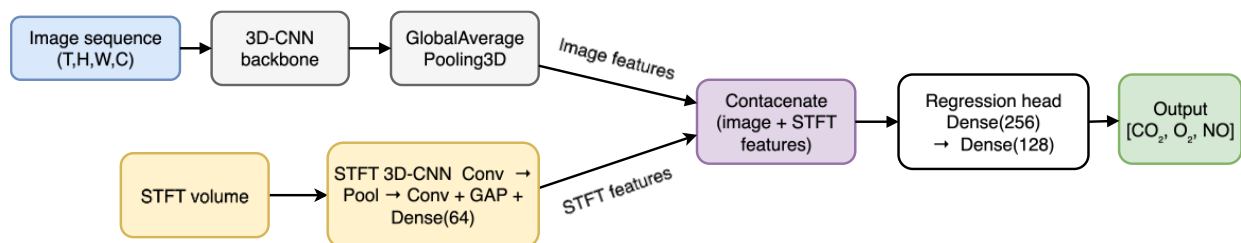


Figure 5. Model #3's architecture.

For each image sequence, a compact 3D time–frequency representation is constructed and used as input to a dedicated 3D CNN branch. After converting to greyscale, each sequence is reduced to a stack of down sampled frames of size (T, 32, 32, 1). For a given sequence, the spatial resolutions of every frame are flattened so that each pixel position (over all 32×32 locations) yields its own 1D intensity-versus-time signal. The short-time Fourier transform (STFT) is then applied to every such temporal signal using a Hann window with a frame length of 32 time steps, a hop of 4, and an FFT size of 64. This produces a complex STFT tensor whose magnitude encodes the energy distribution across the frequency dimension for each pixel. Only the lowest-frequency bins (here, 16) are retained to concentrate on the most relevant temporal frequencies, and the STFT magnitudes are averaged over the time axis, yielding a frequency energy vector per pixel. These per-pixel spectra are reshaped back to form a 3D volume, where the first dimension indexes frequency bins and the remaining two are spatial coordinates; finally, a singleton channel dimension is added to obtain an STFT volume.

These volumes are stacked over all training and test sequences, which provide the 3D-CNN with a complementary view of the flame instead of raw intensities over time. This provides information on how temporal oscillations at different frequencies are distributed across the flame image, allowing the network to learn gas-related patterns in the time–frequency domain.

The image branch has the same three-block 3D CNN backbone, followed by a Dense-256 layer. In contrast, the STFT branch uses a lightweight 3D CNN architecture: Conv-16, Pool, Conv-32, and global average pooling (GAP), culminating in Dense-64. Their feature vectors are concatenated and passed through a two-layer regression head (Dense-256 → Dropout → Dense-128 → Dropout → Dense-3), yielding simultaneous estimates of the three gas concentrations from fused spatio-temporal image information and its time–frequency representation.

2.2. The Models with 3D ResNet-50 Backbone

All subsequent models (Models #4–#6) are based on a 3D ResNet-50-style backbone that extends the canonical 2D ResNet-50 bottleneck design to spatio-temporal inputs using 3D convolutions and squeeze-and-excitation (SE) channel attention [39].

The backbone architecture is a 4-stage 3D ResNet-50 layout with bottleneck residual blocks of the form $1 \times 1 \times 1 \rightarrow 3 \times 3 \times 3 \rightarrow 1 \times 1 \times 1$, analogous to the 2D ResNet-50 bottleneck design but operating over time, height, and width. The network begins with a spatio-temporal stem consisting of a 3D convolution (64 filters, kernel $3 \times 7 \times 7$, stride $1 \times 2 \times 2$) followed by batch normalization, ReLU, and a 3D max-pooling layer (pool $1 \times 3 \times 3$, stride $1 \times 2 \times 2$), which reduces spatial resolution while preserving temporal sampling. The subsequent four stages follow the classical ResNet-50 depth configuration, with bottleneck blocks, channel widths {64, 128, 256, 512}, and a $4\times$ expansion in the last $1 \times 1 \times 1$ convolution of each block, as in the standard ResNet-50 architecture [40].

- Stage 1 (conv2_x): 3 bottleneck blocks with 64 base filters, all with unit stride, operating at the stem resolution.
- Stage 2 (conv3_x): 4 bottleneck blocks with 128 base filters; the first block performs spatial downsampling using stride (1, 2, 2), while subsequent blocks use stride 1.
- Stage 3 (conv4_x): 6 bottleneck blocks with 256 base filters; again, only the first block downsamples spatially via stride (1, 2, 2), keeping the temporal resolution fixed.
- Stage 4 (conv5_x): 3 bottleneck blocks with 512 base filters; the first block uses stride (1, 2, 2) for further spatial downsampling, the others keep a stride of 1.

Within each bottleneck, the $1 \times 1 \times 1$ reduce convolution adjusts the channel width, the $3 \times 3 \times 3$ convolution performs spatio-temporal feature extraction, and the $1 \times 1 \times 1$ expand convolution restores the channels to $4\times$ the base filter count, forming the residual branch. If the spatial/temporal stride or the number of channels changes, a $1 \times 1 \times 1$ projection on the shortcut path is applied (with a matching stride) to ensure shape compatibility, which is standard practice in deep residual network architectures [41].

To enhance channel-wise feature recalibration, a 3D SE module is integrated at the end of the residual branch in each bottleneck block (before addition with the shortcut). The SE block performs global average pooling throughout the full spatio-temporal volume to produce a channel descriptor, followed by a two-layer MLP with reduction ratio $r = 16$ (ReLU $\rightarrow$ sigmoid) and a $1 \times 1 \times 1$ re-expansion to per-channel weights, which are applied multiplicatively to the residual features. This encourages the model to emphasize channels that are informative for gas-related structures while limiting less relevant responses, similarly to 2D SE-ResNet and more recent channel attention variants.

After the final residual stage, global average pooling along the time and space dimensions is applied to obtain a compact feature vector. This is fed into a small MLP regression head composed of two fully connected layers (256 and 128 units, ReLU activations, L2 weight decay) with dropout (rate 0.2) for regularization, followed by a linear output layer producing the target gas concentrations.

The idea behind the structure of models #4–#6 is the same as for models #1–#3. The models share the same 3DResNet-50 + SE backbone described above but differ in their input branches and late fusion strategy. Model #4 is a single 3D ResNet-50 backbone, which uses only the main spatio-temporal input (flame image sequence) processed by the backbone and regression head as described above. Model #5 is a dual-stream architecture including the brightness branch, which augments Model #4 with an additional branch that takes a scalar or low-dimensional brightness descriptor sequence processed by a shallow MLP as input. The brightness embedding is concatenated with the pooled 3DResNet backbone features before the final regression layers, permitting the model to fuse global intensity context with deep spatio-temporal features. Model #6 contains an STFT branch, as

described for Model #3. The STFT embedding is fused with the 3DResNet backbone output (by concatenation) prior to the MLP head, enabling joint exploitation of spatial, temporal, and frequency information for gas concentration regression.

2.3. Data Preparation

The creation of each training sample entails synchronizing frame data and gas concentration measurements via their timestamps, then packaging these data elements into fixed-length sequences. Initially, the image data and the gas analyzer records (stored in a separate file) were matched in time, ensuring that each frame path is associated with CO₂, O₂, and NO concentrations derived from the nearest analyzer timestamp. These aligned records were then used to create non-overlapping windows of consecutive frames with a fixed length for each clip. The list of frame paths forms the image sequence for each window, and the gas concentration vector from the first frame is used as the label for the entire window. After windowing, the actual image data were loaded for each frame path from the HDF5 file, thereby transforming each sequence of paths into a tensor of shape (T, H, W, C), where T denotes the number of time steps; H and W are the image height and width, respectively; and the final quantity, C, is defined as the number of channels (for RGB images, C is set to 3). Each training example consists of a temporally consistent 128-frame flame sequence and a synchronized label vector of CO₂, O₂, and NO concentrations derived from the time-aligned analyzer data.

After the cleaning process, the dataset totaled 1890 entries containing image series tensors (T, H, W, C) and gas concentration vectors. The dataset was split into training and test sets at a 0.8:0.2 ratio, with this split applied to each of the 22 recording sessions shown in Figure 1. Consequently, the measured gas concentration values were comparable across the training and test sets. It should be underlined that the train–test split was performed within each recording session, such that temporally adjacent windows from the same session could appear in both sets. As the gas concentrations are not uniformly distributed but form several clusters (Figure 2), this strategy mainly tests interpolation within these clustered operating regimes rather than generalization to entirely unseen readouts.

The dimensions of the recorded images were recalibrated to match the specified H and W values in the image series tensor. Given that the recorded images were 8-bit per color, the RGB values were normalized to the interval [0, 1].

Gas concentration targets were standardized prior to model training to place all variables on a comparable numerical scale and improve the conditioning of the optimization problem. This transformation lessens the dominance of high-magnitude variables, supports steady convergence of gradient-based learning algorithms, and ensures that regularization acts equally across parameters [42]. In addition, standardization reduces sensitivity to gas sensor-specific offsets and drift, thereby enhancing the stability and generalization capabilities of the resulting predictive models [43]. Standardization was applied exclusively during the training phase, and predictions were subsequently converted back to their original physical units (vol% for O₂ and CO₂, mg/m³ for NO) prior to evaluation.

Training was performed for a maximum of 300 epochs, with early stopping applied if no improvement in validation performance was observed over 20 consecutive epochs. The initial learning rate was set to 0.0005 and was reduced by a factor of 2 if the training loss did not decrease over 8 successive epochs. All models used the Adam optimizer and the Huber loss, which is often more robust than MSE for regression tasks. The Huber loss is an alternative to the standard squared loss used for the RMSE, as it interpolates between quadratic and linear behavior and thus reduces the influence of outliers while retaining high efficiency under near-Gaussian noise. After each epoch, the model was saved in HDF5 format, provided that an improvement in validation loss was achieved.

All models were trained on a system equipped with A100 GPUs, operating under Ubuntu 22.02. The software environment included Python 3.12.3, CUDA 12.8, TensorFlow 2.17, Scikit-learn 1.5.2, and Jupyter 4.3.5.

3. Results

The six models were trained for three different image resolutions (height $\times$ width: 64×64 , 96×96 , and 128×128), with the same sequence length of 128. The assessment results for all model configurations are presented in Table 1. The evaluation was based on standard metrics, namely, the coefficient of determination and the mean absolute error (MAE). The MAE was calculated for each gas individually and for all gases collectively.

Table 1. Coefficients of determination (R^2) and mean absolute error (MAE) for O_2 , NO, and CO_2 concentration regression for all CNN architectures (Models #1–#6) and input resolutions. All metrics were computed on de-standardized predictions and targets.

Model	Gas	R^2	MAE	Global MAE
Model #1 $128 \times 64 \times 64$	O_2	0.9715	0.2550	5.6920
	NO	0.9763	16.5922	
	CO_2	0.9742	0.2289	
Model #1 $128 \times 96 \times 96$	O_2	0.9671	0.2656	5.8272
	NO	0.9705	16.9840	
	CO_2	0.9704	0.2321	
Model #1 $128 \times 128 \times 128$	O_2	0.9554	0.3086	8.3382
	NO	0.9249	24.4344	
	CO_2	0.9614	0.2714	
Model #2 $128 \times 64 \times 64$	O_2	0.9745	0.2408	5.6512
	NO	0.9750	16.4979	
	CO_2	0.9775	0.2149	
Model #2 $128 \times 96 \times 96$	O_2	0.9796	0.2240	5.0468
	NO	0.9865	14.7260	
	CO_2	0.9823	0.1903	
Model #2 $128 \times 128 \times 128$	O_2	0.9776	0.2328	5.3468
	NO	0.9830	15.6130	
	CO_2	0.9812	0.1948	
Model #3 $128 \times 64 \times 64$	O_2	0.9764	0.2397	5.3009
	NO	0.9850	15.4497	
	CO_2	0.9784	0.2134	
Model #3 $128 \times 96 \times 96$	O_2	0.9458	0.3555	8.9085
	NO	0.9275	26.0499	
	CO_2	0.9522	0.3202	
Model #3 $128 \times 128 \times 128$	O_2	0.9660	0.2882	6.8596
	NO	0.9731	20.0289	
	CO_2	0.9682	0.2617	
Model #4 $128 \times 64 \times 64$	O_2	0.9704	0.2625	4.9618
	NO	0.9872	14.4061	
	CO_2	0.9753	0.2169	
Model #4 $128 \times 96 \times 96$	O_2	0.9754	0.2258	4.5220
	NO	0.9873	13.1510	
	CO_2	0.9800	0.1892	
Model #4 $128 \times 128 \times 128$	O_2	0.9754	0.2315	4.9193
	NO	0.9841	14.3343	
	CO_2	0.9801	0.1920	

Table 1. Cont.

Model	Gas	R ²	MAE	Global MAE
Model #5 128 × 64 × 64	O ₂	0.9725	0.2480	5.0577
	NO	0.9860	14.7135	
	CO ₂	0.9765	0.2116	
Model #5 128 × 96 × 96	O ₂	0.9746	0.2388	5.3857
	NO	0.9834	15.7109	
	CO ₂	0.9778	0.2074	
Model #6 128 × 64 × 64	O ₂	0.9681	0.2546	5.9421
	NO	0.9730	17.3503	
	CO ₂	0.9723	0.2215	
Model #6 128 × 96 × 96	O ₂	0.9666	0.2596	5.7118
	NO	0.9763	16.6484	
	CO ₂	0.9706	0.2275	
Model #6 128 × 128 × 128	O ₂	0.9779	0.2194	4.5672
	NO	0.9871	13.2993	
	CO ₂	0.9816	0.1829	

All models achieved high coefficients of determination ($R^2 \geq 0.92$) across the three estimated gas species (O₂, NO, CO₂), demonstrating the general suitability of the proposed CNN-based approach for gas concentration regression. Across all models and resolutions, CO₂ yielded the lowest mean absolute error (MAE range: 0.18–0.32), while NO exhibited the highest absolute MAE values (13.1–26.0), which is attributable to its substantially wider concentration range. Nonetheless, the R^2 values for NO remained high (>0.97) for most configurations. Among all architectures, Model #4 achieved the most consistent and accurate predictions, attaining the lowest global MAE of 4.52 at an input resolution of $128 \times 96 \times 96$, with R^2 values of 0.975, 0.987, and 0.980 for O₂, NO, and CO₂, respectively. Model #6 performed comparably at the highest resolution ($128 \times 128 \times 128$, global MAE = 4.57), whereas Models #1 and #3 exhibited a clear sensitivity to increasing spatial resolution, with global MAE degrading to 8.34 and 8.91 at $128 \times 128 \times 128$ and $128 \times 96 \times 96$, respectively. Overall, the $128 \times 96 \times 96$ resolution offered the best trade-off between spatial detail and model accuracy for most architectures, with further increases in input dimensionality providing no consistent improvement unless accompanied by a proportionally larger model capacity. These findings suggest that the models demonstrate a high degree of accuracy; nevertheless, given the restricted nature of the dataset and the lack of an external dataset for validation, the high R^2 and MAE values reported here should be interpreted with caution.

The real-versus-predicted scatter plots in Figure 6 indicate that the optimal model yields accurate and well-calibrated estimates of O₂, NO, and CO₂ concentrations. For all three gases, the data points are closely clustered around the 1:1 reference line across their respective operating ranges, with only modest and largely symmetric dispersion. For O₂ and CO₂, the scatter is especially narrow, with only a slight increase in spread at the lowest concentration levels. These observations are consistent with the low MAE values reported in Table 1 and support the conclusion that the network captures global trends and fine-scale variability in gas concentrations.

Another qualitative assessment provided the relative error distribution per gas in the form of boxplots, as shown in Figure 7.

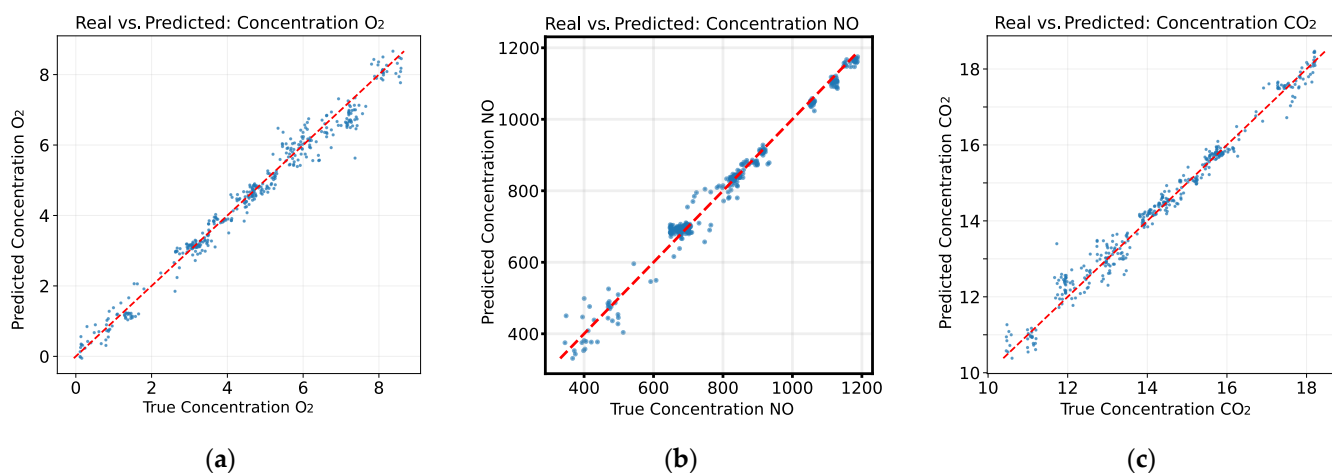


Figure 6. The real-versus-predicted scatter plots obtained for Model #4, $128 \times 96 \times 96$: (a) O_2 , (b) NO, and (c) CO_2 . The predicted concentrations were subsequently converted back to their original physical units (de-standardized).

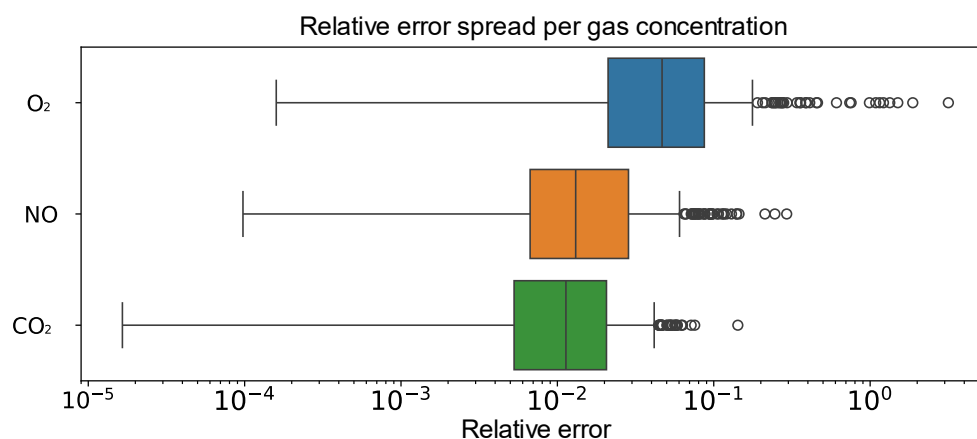


Figure 7. Boxplots of the relative error for Model #4, $128 \times 96 \times 96$.

For the best model (Model No. 4, $128 \times 96 \times 96$), the quantitative error metrics confirm that the model achieved high forecasting precision for all three gases. The MAE values are 0.2258% by volume (O_2), 13.1520 mg/m³ (NO), and 0.2169% vol. (CO_2), with corresponding RMSE values of approximately 0.3006, 19.2341, and 0.2899, indicating that absolute deviations remain small across the entire operating range. Average relative errors of approximately 11.2% for O_2 and below 2% for NO and CO_2 , as well as MAE = 4.5220, RMSE = 11.1591, and MRE = 4.1% (mean relative error), demonstrate that this model captures dominant trends and fine changes in gas concentrations with limited deviation.

The relative error analysis confirms that the optimal model provides accurate predictions for all three gas species. The relative error boxplots reveal that most samples for O_2 , NO, and CO_2 lie within a narrow band around zero, in agreement with the real-versus-predicted scatter plots. CO_2 shows the tightest error distribution and only a few outliers, consistent with its lowest absolute MAE. O_2 and NO display slightly broader spreads, with a limited number of larger relative errors corresponding to isolated points that differ from the 1:1 line in the scatter diagrams. However, these cases are rare and do not alter the overall error structure.

It should be emphasized that the presented results were obtained from a single combustion experiment of limited scope. Therefore, they primarily reflect interpolation within a single study rather than generalization over different measurement series. Furthermore, the training and test datasets were generated from temporally adjacent intervals within the

same measurement sessions, which may have further skewed the results towards overly optimistic values; this should be kept in mind when interpreting the results.

4. Discussion

The influence of architecture type on prediction accuracy was evaluated by comparing three backbone–branch combinations across two backbone families. The choice of backbone constituted the most significant architectural factor: the 3D ResNet50-based models (Models #4–#6) consistently outperformed their 3D CNN counterparts (Models #1–#3) across all input resolutions and gas species, with the plain 3D ResNet50 backbone (Model #4) reducing the global MAE from 5.69 to 4.96 relative to the plain 3D CNN (Model #1) at $128 \times 64 \times 64$, and from 8.34 to 4.92 at $128 \times 128 \times 128$. This improvement can be attributed to the residual connections inherent to the ResNet50 architecture, which enable more effective gradient propagation, enable the extraction of richer spatio-temporal features, and confer greater tolerance to increasing spatial resolution.

A higher spatial resolution of the input data did not automatically translate to better forecasting performance across all architectures. In the shallow 3D CNN models (#1–3), enlarging the input size from 64×64 to 128×128 raised the number of input pixels per sequence without a proportional increase in model capacity or training data; as a result, these networks became more prone to fitting minor, low-information details of the flame image and, consequently, to overfitting. In contrast, the 3D ResNet-50 models (#4–#6), equipped with residual connections and a squeeze and excitation attention mechanism, maintained or slightly improved accuracy at higher resolutions. This indicates that sufficient depth and regularization are required to exploit the additional spatial information effectively. Overall, for the available dataset of 1890 sequences, moderate resolutions on the order of $128 \times 96 \times 96$ offered the best trade-off between spatial detail, model complexity, and generalization, and further gains from higher resolutions are likely to require either larger training datasets or stronger capacity control.

The incorporation of auxiliary branches produced architecture-dependent effects. The brightness statistics branch improved performance when combined with the plain 3D CNN backbone (Model #2 vs. Model #1, global MAE: 5.35 vs. 8.34 at $128 \times 128 \times 128$), most likely because the compact global intensity descriptors compensate for the limited representational capacity of the shallower backbone. However, no consistent benefit was observed when the same branch was paired with the ResNet50 backbone (Model #5 vs. Model #4), which suggests that the deeper architecture already captures equivalent information within its spatial feature maps. The STFT branch yielded the most nuanced behavior. When combined with the plain 3D CNN backbone (Model #3), it caused a marked performance decline at intermediate resolution (global MAE = 8.91 at $128 \times 96 \times 96$, $R^2 = 0.928$ for NO), indicating that the plain backbone lacks the capacity to jointly optimize spatial and frequency representations as the input dimensionality grows. In contrast, pairing the STFT branch with the ResNet50 backbone (Model #6) yielded the best-performing multi-branch configuration, achieving a global MAE of 4.57 at $128 \times 128 \times 128$ and an R^2 of 0.987 for NO, which closely approaches the best overall result.

Taken together, these results indicate that, although the time–frequency features derived from the STFT are informative for gas concentration estimation, their effective integration requires a backbone with sufficient depth and residual connectivity to resolve the competing gradient signals introduced by the multi-branch design.

Several limitations should be acknowledged when interpreting these results. Although the collected dataset spans multiple operating regimes (high- O_2 /low- CO_2 , low- O_2 /high- CO_2 , varying NO) with wide dynamic ranges for all three gases, it originated from a single 0.5 MW test rig operated with a single fuel mixture containing 10% biomass. The ability

of the proposed models to generalize to other fuels and blend ratios remains unclear. In addition, the flame-image sequences are strongly constrained by the spatial orientation of the borescope relative to the burner, which may further limit generalization.

Each configuration was also trained only once, and thus, the reported MAE and R^2 values do not include variability across different random initializations and data shuffles. A more statistically robust assessment would require retraining the key models with 3–5 random seeds per configuration and reporting the mean and standard deviation of the metrics, which is planned for future work.

Furthermore, the present study was restricted to spatio-temporal modeling utilizing 3D CNN and 3D ResNet-50 architectures, with no explicit comparison with recurrent or attention-based video models such as ConvLSTMs or vision transformers. While the findings indicate that 3D convolutions provide promising multi-gas concentration estimates from flame sequences, systematic benchmarking against ConvLSTM, temporal attention mechanisms, and Transformer-style backbones would help to clarify the relative strengths and efficiency of these alternatives. Such comparisons will be pursued in future work, once larger datasets spanning multiple biomass fractions are available. This will allow the higher capacity of attention-based models to be exploited without overfitting.

Finally, the high accuracy reported in this study must be interpreted within the context of the limited experimental diversity. All models were trained and tested on data from a single co-firing test with a single fuel mixture, with the train–test split drawn from the same set of recording sessions. Therefore, they may have partially exploited experiment-specific flame patterns and operating regimes. As such, the present work should be seen as a feasibility study that demonstrates internal consistency rather than a full external generalization. A more rigorous robustness assessment will require external validation or cross-condition testing on datasets acquired using different biomass fractions, loads, and operating runs, ideally on independent combustion units.

5. Conclusions

This study analyzed a deep learning-based approach in the spatiotemporal domain for the optical, “soft” detection of exhaust emissions during coal and biomass co-combustion. Three shallow 3D CNNs and three 3D ResNet-50 models with a “squeeze and excitation” attention mechanism were trained to estimate the concentrations of O_2 , CO_2 , and NO directly from high-speed RGB flame image sequences. All models achieved high accuracy ($R^2 \geq 0.92$ for each gas). The best results were obtained by model #4, a single-data-stream 3D ResNet-50 with dimensions of $128 \times 96 \times 96$ ($R^2 = 0.975, 0.987$, and 0.980 ; MAE = 0.23% by volume, 13.15 mg/m^3 , and 0.19% by volume for O_2 , NO, and CO_2 , respectively). The following conclusions can be drawn:

- The residual ResNet-50 backbone was the dominant factor and exhibited significantly greater tolerance for higher resolutions than shallow CNNs.
- A moderate resolution of $128 \times 96 \times 96$ provided the best compromise between accuracy and complexity.
- The auxiliary brightness and STFT branches were helpful only when combined with a sufficiently deep backbone.

These results provide a promising foundation for soft optical sensors in combustion diagnostics. However, these results are based on a single experiment with a single biomass fraction and borescope orientation. The inference delay was not quantified, so the reported accuracy should be considered evidence of internal consistency rather than full generalizability. Future work will validate this approach using independent datasets covering various fuel types and loads. It will also assess variability across different random initial-

izations, conduct comparative tests with ConvLSTM- and Transformer-based video models, and evaluate real-time performance for practical implementation.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study were collected during previous technical research and are available from the corresponding author upon reasonable request; however, due to the very large volume of raw measurements, only curated subsets can be shared.

Conflicts of Interest: The author declares no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

3D CNN	Three-dimensional convolutional neural network
STFT	Short-time Fourier transform
DBN	Deep belief network
PCA	Principal component analysis
PLS	Partial Least Squares
SVR	Support vector regression
WGAN	Wasserstein generative adversarial network
GAP	Global average pooling
SE	Squeeze-and-excitation
MLP	Multi-layer perceptron
MRE	Mean relative error
MSE	Mean squared error
RMSE	Root mean squared error
MAE	Mean absolute error

References

1. Roni, M.S.; Chowdhury, S.; Mamun, S.; Marufuzzaman, M.; Lein, W.; Johnson, S. Biomass Co-Firing Technology with Policies, Challenges, and Opportunities: A Global Review. *Renew. Sustain. Energy Rev.* **2017**, *78*, 1089–1101. [\[CrossRef\]](#)
2. Al-Mansour, F.; Zuwala, J. An Evaluation of Biomass Co-Firing in Europe. *Biomass Bioenergy* **2010**, *34*, 620–629. [\[CrossRef\]](#)
3. Zhai, M.; Tian, X.; Liu, Z.; Zhao, Y.; Deng, Y.; Yang, W. Advancing Just Transition: The Role of Biomass Co-Firing in Emission Reductions and Employment for Coal Regions. *Sustain. Energy Technol. Assess.* **2025**, *75*, 104246. [\[CrossRef\]](#)
4. Olsztyńska, I. Biomass in the Fuel Mix of the Polish Energy and Heating Sector. *Polityka Energetyczna Energy Policy J.* **2019**, *22*, 99–118. [\[CrossRef\]](#)
5. Baxter, L. Biomass-Coal Cofiring: An Overview of Technical Issues. In *Solid Biofuels for Energy*; Grammelis, P., Ed.; Green Energy and Technology; Springer: London, UK, 2011; pp. 43–73. ISBN 978-1-84996-392-3.
6. Jappe, F. Utilizing Biomass and Waste for Power Production—A Decade of Contributing to the Understanding, Interpretation and Analysis of Deposits and Corrosion Products. *Fuel* **2005**, *84*, 1277–1294. [\[CrossRef\]](#)
7. Damstedt, B.; Pederson, J.M.; Hansen, D.; Knighton, T.; Jones, J.; Christensen, C.; Baxter, L.; Tree, D. Biomass Cofiring Impacts on Flame Structure and Emissions. *Proc. Combust. Inst.* **2007**, *31*, 2813–2820. [\[CrossRef\]](#)
8. Dai, J.; Sokhansanj, S.; Grace, J.R.; Bi, X.; Lim, C.J.; Melin, S. Overview and Some Issues Related to Co-firing Biomass and Coal. *Can. J. Chem. Eng.* **2008**, *86*, 367–386. [\[CrossRef\]](#)
9. Spliethoff, H.; Hein, K.R.G. Effect of Co-Combustion of Biomass on Emissions in Pulverized Fuel Furnaces. *Fuel Process. Technol.* **1998**, *54*, 189–205. [\[CrossRef\]](#)
10. Pronobis, M. The Influence of Biomass Co-Combustion on Boiler Fouling and Efficiency. *Fuel* **2006**, *85*, 474–480. [\[CrossRef\]](#)
11. Lu, G.; Yan, Y.; Cornwell, S.; Whitehouse, M.; Riley, G. Impact of Co-Firing Coal and Biomass on Flame Characteristics and Stability. *Fuel* **2008**, *87*, 1133–1140. [\[CrossRef\]](#)
12. Nunes, L.J.R.; Matias, J.C.O.; Catalão, J.P.S. Biomass Combustion Systems: A Review on the Physical and Chemical Properties of the Ashes. *Renew. Sustain. Energy Rev.* **2016**, *53*, 235–242. [\[CrossRef\]](#)
13. Akhmetshin, M.R.; Nyashina, G.S.; Strizhak, P.A. Comparative Analysis of Factors Affecting Differences in the Concentrations of Gaseous Anthropogenic Emissions from Coal and Slurry Fuel Combustion. *Fuel* **2020**, *270*, 117581. [\[CrossRef\]](#)

14. Jia, L.; Naber, J.D.; Blough, J.R. Review of Sensing Methodologies for Estimation of Combustion Metrics. *J. Combust.* **2016**, *2016*, 8593523. [\[CrossRef\]](#)
15. Guth, U.; Shuk, P.; McGuire, C. Gas Sensing Technologies in Combustion: A Comprehensive Review. *J. Electrochem. Sci. Eng.* **2020**, *10*, 103–110. [\[CrossRef\]](#)
16. Fonollosa, J.; Solórzano, A.; Marco, S. Chemical Sensor Systems and Associated Algorithms for Fire Detection: A Review. *Sensors* **2018**, *18*, 553. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Lee, S.-M.; Choi, S.-W.; Lee, E.-B. Prediction Modeling of Flue Gas Control for Combustion Efficiency Optimization for Steel Mill Power Plant Boilers Based on Partial Least Squares Regression (PLSR). *Energies* **2023**, *16*, 6907. [\[CrossRef\]](#)
18. Copik, P.; Szlek, A.; Ditaranto, M. Simplified Mathematical Model of Oxy-Fuel Combustion of Municipal Solid Waste in the Grate Furnace: Effect of Different Flue Gas Recirculation Rates and Comparison with Conventional Mode. *Arch. Thermodyn.* **2024**, *45*, 13–25. [\[CrossRef\]](#)
19. Tumuluru, J.S.; Wright, C.T.; Boardman, R.D.; Yancey, N.A.; Sokhansanj, S. A Review on Biomass Classification and Composition, Co-Firing Issues and Pretreatment Methods. In Proceedings of the 2011 ASME International Conference on Fuel Processing Technology, Louisville, KY, USA, 7–10 August 2011.
20. Basu, P.; Butler, J.; Leon, M.A. Biomass Co-Firing Options on the Emission Reduction and Electricity Generation Costs in Coal-Fired Power Plants. *Renew. Energy* **2011**, *36*, 282–288. [\[CrossRef\]](#)
21. Lovcha, Y.; Perez-Laborda, A.; Sikora, I. The Determinants of CO₂ Prices in the EU Emission Trading System. *Appl. Energy* **2022**, *305*, 117903. [\[CrossRef\]](#)
22. Hannun, R.M.; Abdul Razzaq, A.H. Air Pollution Resulted from Coal, Oil and Gas Firing in Thermal Power Plants and Treatment: A Review. *IOP Conf. Ser. Earth Environ. Sci.* **2022**, *1002*, 012008. [\[CrossRef\]](#)
23. Zhang, J.; Shen, J.; Fang, J.; Chen, L.; Wang, C.; Che, D. Study on NO Formation and Ash Characteristics during Co-Combustion of Semi-Coke and Biomass under O₂/CO₂ Conditions. *Fuel* **2025**, *390*, 134717. [\[CrossRef\]](#)
24. Nussbaumer, T. Combustion and Co-Combustion of Biomass: Fundamentals, Technologies, and Primary Measures for Emission Reduction. *Energy Fuels* **2003**, *17*, 1510–1521. [\[CrossRef\]](#)
25. Trindade, T.; Ferreira, A.; Fernandes, E. Characterization of Combustion Chemiluminescence: An Image Processing Approach. *Procedia Technol.* **2014**, *17*, 194–201. [\[CrossRef\]](#)
26. Sujatha, K.; Venmathi, M.; Pappa, N. Flame Monitoring in Power Station Boilers Using Image Processing. *IJIVP* **2012**, *2*, 427–434. [\[CrossRef\]](#)
27. Hernández, R.; Ballester, J. Flame Imaging as a Diagnostic Tool for Industrial Combustion. *Combust. Flame* **2008**, *155*, 509–528. [\[CrossRef\]](#)
28. Liu, Y.; Fan, Y.; Chen, J. Flame Images for Oxygen Content Prediction of Combustion Systems Using DBN. *Energy Fuels* **2017**, *31*, 8776–8783. [\[CrossRef\]](#)
29. Ganpati, K.; Bhusnur, S. Utilizing Advanced Machine Learning Techniques for Accurate Prediction of Oxygen Quantity in Gas-Fired Boiler Combustion to Enhance Environmental Pollution Control. *JART* **2024**, *22*, 873–885. [\[CrossRef\]](#)
30. Gao, S.; Liu, A.; Jiang, Y.; Liu, Y. Online Estimation of Combustion Oxygen Content with an Image-Augmented Soft Sensor Using Imbalanced Flame Images. *ACS Omega* **2023**, *8*, 40657–40664. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Qin, L.; Lu, G.; Hossain, M.M.; Morris, A.; Yan, Y. A Flame Imaging-Based Online Deep Learning Model for Predicting NO_x Emissions From an Oxy-Biomass Combustion Process. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 2501811. [\[CrossRef\]](#)
32. Cugley, J. Advanced Flame Monitoring and Emission Prediction Through Digital Imaging and Spectrometry. Ph.D. Thesis, University of Kent, Canterbury, UK, 2019.
33. Shahid, M.; Chien, I.-F.; Sarapugdi, W.; Miao, L.; Hua, K.-L. Deep Spatial-Temporal Networks for Flame Detection. *Multimed. Tools Appl.* **2021**, *80*, 35297–35318. [\[CrossRef\]](#)
34. Dimitropoulos, K.; Barmpoutis, P.; Grammalidis, N. Spatio-Temporal Flame Modeling and Dynamic Texture Analysis for Automatic Video-Based Fire Detection. *IEEE Trans. Circuits Syst. Video Technol.* **2015**, *25*, 339–351. [\[CrossRef\]](#)
35. Sanz, A.; Ballester, J.; Hernández, R.; Cerecedo, L.M. Advanced Monitoring of Industrial Burners Based on Fluctuating Flame Signals. *Fuel* **2008**, *87*, 1063–1075. [\[CrossRef\]](#)
36. Ronquillo-Lomeli, G.; Romero, C.E.; Yao, Z.; Si, F.; Coria-Silva, R.; Hernandez-Rosales, F.; Sanchez-Gaytan, J.L.; Trejo-Morales, A. On-Line Flame Signal Time Series Analysis for Oil-Fired Burner Optimization. *Fuel* **2015**, *158*, 416–423. [\[CrossRef\]](#)
37. Kotyra, A.; Gromaszek, K. Determination of Coal and Biomass Co-Combustion Process States Using Convolutional Neural Networks. *Energies* **2025**, *18*, 5219. [\[CrossRef\]](#)
38. De Giorgi, M.G.; Sciolti, A.; Pescini, E.; Ficarella, A. Frequency Analysis and Predictive Identification of Flame Stability by Image Processing. In *Proceedings of the Volume 2: Economic, Environmental, and Policy Aspects of Alternate Energy; Fuels and Infrastructure, Biofuels and Energy Storage; High Performance Buildings; Solar Buildings, Including Solar Climate Control/Heating/Cooling; Sustainable Cities and Communities, Including Transportation; Thermofluid Analysis of Energy Systems, Including Exergy and Thermoeconomics*; American Society of Mechanical Engineers: Boston, MA, USA, 2014; p. V002T04A014.

39. Saleem, N.; Elmannai, H.; Bourouis, S.; Trigui, A. Squeeze-and-Excitation 3D Convolutional Attention Recurrent Network for End-to-End Speech Emotion Recognition. *Appl. Soft Comput.* **2024**, *161*, 111735. [[CrossRef](#)]
40. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: Las Vegas, NV, USA, 2016; pp. 770–778.
41. He, K.; Zhang, X.; Ren, S.; Sun, J. Identity Mappings in Deep Residual Networks. In *Computer Vision—ECCV 2016*; Leibe, B., Matas, J., Sebe, N., Welling, M., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2016; Volume 9908, pp. 630–645. ISBN 978-3-319-46492-3.
42. Ansari, G.; Singh, R.; Kumar, S.; Soliman, N.F. Fast and Robust Mixed Gas Identification and Recognition Using Tree-Based Machine Learning and Sensor Array Response. *Sci. Rep.* **2025**, *15*, 34472. [[CrossRef](#)] [[PubMed](#)]
43. Zeng, X.; Shahzeb, M.; Cheng, X.; Shen, Q.; Xiao, H.; Xia, C.; Xia, Y.; Huang, Y.; Xu, J.; Wang, Z. An Enhanced Gas Sensor Data Classification Method Using Principal Component Analysis and Synthetic Minority Over-Sampling Technique Algorithms. *Micromachines* **2024**, *15*, 1501. [[CrossRef](#)] [[PubMed](#)]

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.