



# Data-Driven Modeling for the Prediction of Stack Gas Concentration in a Coal-Fired Power Plant in Türkiye

Mandana Mohammadi · Didem Saloglu ·  
Halil Dertli · Mitra Mohammadi ·  
Mansour Ghaffari-Moghaddam

Received: 9 January 2024 / Accepted: 19 April 2024  
© The Author(s) 2024

**Abstract** In this research, deep learning and machine learning methods were employed to forecast the levels of stack gas concentrations in a coal-fired power plant situated in Türkiye. Real-time data collected from continuous emission monitoring systems (CEMS) serves as the basis for the predictions. The dataset includes measurements of carbon monoxide (CO), sulfur dioxide (SO<sub>2</sub>), nitrogen oxides (NO<sub>x</sub>), oxygen (O<sub>2</sub>), and dust levels, along with temperatures recorded. For this analysis, deep learning methods such as multi-layer perceptron network (MLP) and long short-term memory (LSTM) models were used, while machine

learning techniques included light gradient boosted machine (LightGBM) and stochastic gradient descent (SGD) models were applied. The accuracy of the models was determined by analysing their performance using mean absolute error (MAE), root means square error (RMSE), and R-squared values. Based on the results, LightGBM achieved the highest R-squared (0.85) for O<sub>2</sub> predictions, highlighting its variance-capturing ability. LSTM excelled in NO<sub>x</sub> (R-squared 0.87) and SO<sub>2</sub> (R-squared 0.85) prediction, while showing the top R-squared (0.67) for CO. Both LSTM and LGBM achieved R-squared values of 0.78 for dust levels, indicating strong variance explanation. Conclusively, our findings highlight LSTM as the most effective approach for stack gas concentration forecasting, closely followed by the good performance of LightGBM. The importance of these results lies in their potential to effectively manage emissions in coal-fired power plants, thereby improving both environmental and operational aspects.

---

M. Mohammadi  
Research and Development Unit, AirpaajCo., Mashhad,  
Iran

D. Saloglu   
Disaster and Emergency Management Department,  
Disaster Management Institute, Istanbul Technical  
University, 34730 Istanbul, Turkey  
e-mail: saloglu@itu.edu.tr

H. Dertli  
Chemical Engineering Department, Chemical  
and Metallurgical Engineering Faculty, Istanbul Technical  
University, 34730 Istanbul, Turkey

M. Mohammadi  
Department of Environmental Science, Kheradgarayan  
Motahar Institute of Higher Education, Mashhad, Iran

M. Ghaffari-Moghaddam  
Department of Chemistry, Faculty of Science, University  
of Zabol, Zabol, Iran

**Keywords** Coal-fired power plant · Emissions prediction · Environmental monitoring · Deep learning · Machine learning

## 1 Introduction

Energy consumption is one of the most important indicators of the development levels of countries and has recently increased due to industrialization,

technological developments, population growth, and urbanization (Dai et al., 2022; Fontes et al., 2014; Khan & Khan, 2023; Li et al., 2019). Significant environmental problems, such as the greenhouse effect, occur with the increase in energy consumption (Cui et al., 2023; Dong et al. 2022a, b; Hu et al., 2024; Khan & Khan, 2023). In particular, coal-fired thermal power plants both play an active role in energy production and cause environmental problems due to the use of fossil fuels (Hu et al., 2024; Kasman & Duman, 2015; Munawar, 2018; Osobajo et al., 2020; Song et al. 2021a, b). It is crucial to address the environmental issues arising from the coal-fired power sector to promote sustainable development and mitigate air pollution and global climate change (Asif et al., 2022; Hu & Shi, 2021; Song et al. 2021a, b). This requires a concerted effort to increase the efficiency of energy production and embrace innovation in the energy sector (Song et al. 2021a, b). Recognizing the urgency of these issues, Türkiye joined the Climate Change Framework Convention in 2004 and endorsed the Kyoto Protocol in 2009 (Kerem, 2022; Yagmur et al., 2023). Furthermore, at the 2015 Conference of the Parties-UNFCCC, Türkiye committed to reducing greenhouse gas emissions by 21% of projected baseline emissions between 2021 and 2030 (Birpınar & Tuğaç, 2022).

Considering Türkiye's heavy dependence on lignite for electricity generation, as its lignite resources reach 17.3 billion tonnes, it's important to acknowledge the challenges posed by its characteristics. The calorific value of Türkiye's lignite is notably low, with 90% of resources having a calorific value below 3,000 kcal/kg, and 71% below 1,500 kcal/kg. Moreover, more than half of Türkiye's lignite has a moisture content of more than 20% (Atukalp & Kesimal, 2023). Despite, Türkiye's involvement in the global climate change and carbon emissions agreements, it generally provides its energy needs from coal-fired thermal power plants. Therefore, Türkiye follows current innovations in thermal power plant processes and complies with legal restrictions and agreements (Vardar et al., 2022). Although their environmental impacts are well known, coal-fired power plants are preferred for many reasons, such as the ability to use domestic and/or cheap resources, the small number of large energy production capacities compared to renewable energy plants, and the attractiveness of uninterrupted production periods. In addition, coal

has advantages such as being geographically widespread, easy and low-cost production compared to other fossil fuels, being an energy source that can be transformed into energy with simpler technology. However, it is important to recognize the environmental consequences associated with coal combustion, particularly its high carbon emission characteristics, and its potential contribution to climate change issues on a global scale (Atukalp & Kesimal, 2023).

The coal-based energy industry in Türkiye is at the center of air pollution and quality concerns and emits various pollutants such as particulate matter, ozone (O<sub>3</sub>), nitrogen oxides (NO<sub>x</sub>), carbon monoxide (CO), carbon dioxide (CO<sub>2</sub>), sulphur dioxide (SO<sub>2</sub>), methane (CH<sub>4</sub>), volatile organic compounds (VOCs) and over 80 hazardous air pollutants such as lead, arsenic, and benzene (Filonchik & Peterson, 2023; Meikandaan et al., 2019; Moharreri et al., 2020; Xie et al., 2021; Xu et al., 2022; Zhou et al., 2023). Among these emissions, the main components of the stack gas from a coal fired power plant are CO<sub>2</sub> and CO. Direct reduction and indirect reduction reactions leading to the conversion of coal to CO and CO to CO<sub>2</sub> are active. The O<sub>2</sub> concentration and the structure of the coal are considered to be the main factors influencing the degree of reduction reaction related to CO<sub>2</sub> emissions (Cui et al., 2023; Hu et al., 2024). According to Tunckaya and Koklukaya (2015), strict regulations require continuous monitoring and control of these emissions to mitigate environmental damage, including acid rain and climate change. Accurate forecasting of stack gas concentrations is essential for optimizing power plant operations and ensuring compliance with emission standards (Gaffney & Marley, 2009; Hu & Shi, 2021). As energy demand increases due to economic growth, the ability to predict emissions using available input data for operational adjustment becomes critical. Mathematical models and machine learning have been considered for estimating greenhouse gas emissions from coal-fired power plants (Adams et al., 2020; Sharma et al., 2023; Vo et al., 2019). Mathematical models, while popular (Tunckaya & Koklukaya, 2015; Vo et al., 2019; Wu et al., 2023), struggle to capture the complex behaviour of coal-fired plants due to simplifying assumptions. Additionally, they can be computationally expensive. Data-driven models provide a promising alternative, as shown in various power generation applications. In recent years, there has been a

tremendous development in data-driven modelling applications (Krzywanski & Nowak, 2016; Li et al., 2017; Luo et al., 2024; Lv et al., 2017; Shang & Luo, 2021; Shi et al., 2019; Tunckaya & Koklukaya, 2015; Tuttle et al., 2019; Xu et al., 2023).

Data-driven modelling of air pollution and air quality problems can be done by recording stack gas in a coal-fired thermal power plant. Due to the temporal nature of emitted pollutants, coal-fired power plants use data-driven time series analysis (Laubscher, 2019). Data-driven models are often developed and implemented using artificial intelligence (AI) techniques. Due to the capacity of AI to detect complex temporal patterns and non-linear correlations in data, these models are more accurate in predicting the stack-gas emissions of coal-fired power plants than conventional techniques (Alnaim et al., 2022; Laubscher, 2019). Previous studies, such as those by Krzywanski & Nowak, 2016, Vujić et al., 2019, Tang et al., 2022, Chikobvu & Mamba, 2023, Movahed et al., 2023, and Josimović et al., 2023 have explored AI-driven models for forecasting specific pollutants such as CO, NO<sub>x</sub>, NO<sub>2</sub>, PM<sub>2.5</sub>, and SO<sub>2</sub>. These researches have primarily focused on individual pollutants, resulting in a limited number of studies that have comprehensively addressed the forecasting of all pollutants emitted from coal-fired power plants. This highlights a significant gap in the existing research and emphasizes the need for a more integrated approach to emissions forecasting. However, our study goes a step further and expands the current knowledge by considering a broader variety of pollutants into account. Hence, it is essential to use a methodology that employs diverse AI tools to forecast integrated emissions of O<sub>2</sub>, SO<sub>2</sub>, CO, NO<sub>x</sub>, and dust.

For this purpose, a combination of multilayer perceptron (MLP), short-term memory (LSTM) networks, light gradient augmented machine (Light-GBM) and random gradient (SGD) regression models were applied to model the stack gas amount of a coal-fired power plant using real-time measurements from continuous emission monitoring systems (CEMS). Compared to previous studies that might have only employed one approach, this study provides valuable insights into the suitability of these different algorithms for predicting a wider range of pollutants. In addition, the practical innovation contributions of this research can be summarized as follows: Firstly, the use of machine learning and deep learning methods

provides a comprehensive assessment of the efficiency levels of different decision units. Secondly, the comprehensive model shows excellent performance in stack gas emission prediction, the accuracy of the prediction results is fully confirmed, thus providing an effective tool to control and reduce stack gas emissions (Luo et al., 2024). It can be safely stated that, it is the first holistic study for a coal-fired power plant in Türkiye to apply machine learning and deep learning models based on CEMS values for one year of data.

## 2 Methods

### 2.1 The Reactor and Coal Combustion Mechanism in the Reactor of Coal-Fired Plant

The stack (exhaust) gas emission values of a real coal-fired thermal power plant in Türkiye were the source of the data used in this paper, conducted between January and December 2017. In this plant, Columbia lignite coal with low sulfur content was burned instead of lignite of Turkish origin. In general, excess oxygen (excess air) is needed to begin the combustion of the coal fed to the reactor. As a result of coal combustion, the carbon is completely converted to CO<sub>2</sub>. In addition, sulfur and nitrogen in the coal are oxidized. The chemical reactions in the reactor of the stack gas formed in the coal-fired thermal power plant in Eqs. 1-7 are shown.



In addition to carbon, other elements can undergo oxidation. Among them, the most important air pollutant is sulphur oxidation;



Nitrogen is less of a component in coal. Nitrogen oxides in the stack gas are usually unstable compounds formed by heat reactions above the molecular nitrogen in the air. Eq. 6 and 7 can occur after H<sub>2</sub>O formation.

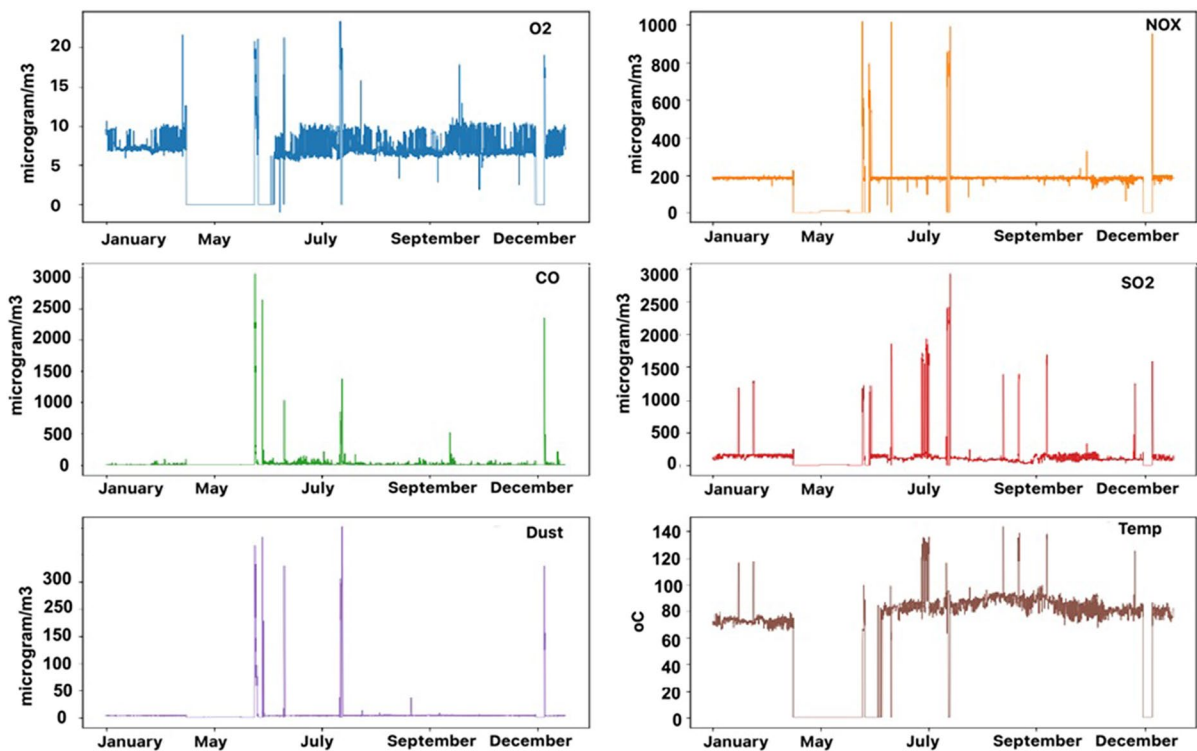


As a result of coal combustion in thermal power plants, the reactor's internal temperature is typically between 880–1150°C. At this temperature, the primary components of coal, C, H, as well as the reactions described above, oxidize into stack gases. Thus, gases such as SO<sub>2</sub> and nitrogen oxides form in stack gases based on the amount of S and N in the coal and the ratio of CO<sub>2</sub>, CO, and H<sub>2</sub>. Although SO<sub>2</sub> is dangerous among these gases, coal combustion is caused by the sulphury components in coal. However, the nitrogen oxides are produced by high-temperature reactions that participate in the combustion event and contain 78% nitrogen. Although nitrogen gas is a difficult reacting gas, if the reactor temperature increases above 950°C, it can react with oxygen with the effect of the temperature and convert to harmful

gases such as NO<sub>2</sub> and N<sub>2</sub>O. Thus, in the stack gas, nitrogen oxides are formed, as well as SO<sub>2</sub>.

## 2.2 Data Collection and Preprocessing of CEMS Data

CEMS provides basic information such as location and sector, as well as hourly stack-level data such as emission rates of CO, SO<sub>2</sub>, NO<sub>x</sub>, O<sub>2</sub> and dust, which are valuable for fine temporal emission estimates. The CEMS dataset used here covers a coal-fired power plant in Zonguldak, Türkiye, during 2017. It is worth noting that the CEMS dataset contains outliers and missing values, although the Industrial Air Pollution Control Regulation has been issued to guide the installation, operation and management of CEMS. Fig. 1 illustrates the stack gas levels, dust, and temperature during the hourly data collection of the combustion process using CEMS of plant. It is worth emphasizing that this study's findings were derived solely from hourly data because there are more of them than daily data, which aligns with one of the



**Fig. 1** Hourly levels of emitted gases, dust, and temperature during the combustion process

**Table 1** Variables description

Variables	Units	Range
NO <sub>x</sub>	µg/m <sup>3</sup>	0 and 1019.47
CO	µg/m <sup>3</sup>	-6.32 and 3053.05
SO <sub>2</sub>	µg/m <sup>3</sup>	0 and 2922.90
Dust	µg/m <sup>3</sup>	0 and 351.53
O <sub>2</sub>	µg/m <sup>3</sup>	-1 and 23.30
Temperature	°C	0 and 143.48

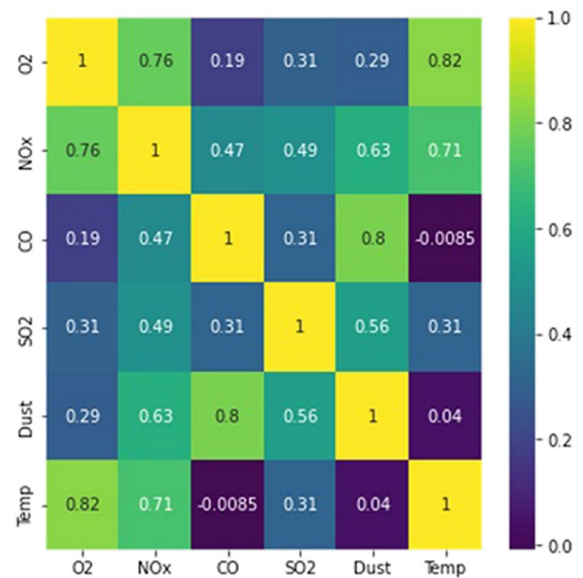
requirements of AI. For further details on the variables and their measuring units, refer to Table 1.

In the current study, there were 8,496 records in the CEMS data set; 1,699 were used for model testing, while 6,797 were used for training (8:2 ratio for the train and test datasets). Hourly measurements of the six selected features were used as inputs to the models over a twenty-four-hour period. Importantly, the training and testing datasets had non-overlapping durations. The test dataset served as the basis for model validation and evaluation, while the training dataset was used for all training purposes. To facilitate time-series forecasting, the data was restructured accordingly.

### 2.3 Interrelationships Among Emission Gases and Other Variables

Analyzing the connection between variables is crucial. So, the magnitude of this phenomenon is illustrated through the heat map of the correlation analysis in Fig. 2. The factors with a stronger correlation with O<sub>2</sub> concentrations are displayed in dark yellow. In contrast, those with a weaker correlation are displayed in dark blue. As examples of the relation between variables, specifically concerning O<sub>2</sub>, the coefficients of correlation between temperature and nitrogen oxides, and O<sub>2</sub> concentrations, are 0.82 and 0.76, respectively, indicating a robust link between these two variables. On the other hand, the coefficients of 0.31 for SO<sub>2</sub>, 0.29 for dust, and 0.19 for CO suggest a weaker association with O<sub>2</sub>. These correlation coefficients provide insights into the varying strengths of the relationships among the variables under study.

The heatmap derived from the datasets (Hael, 2023) of a real thermal power plant examined in this

**Fig. 2** Heatmap of correlation between stuck gases and temperature

article reveals that the relationship between O<sub>2</sub> and nitrogen oxides is stronger than that of O<sub>2</sub> and other gases. This means that excess air is supplied to meet the O<sub>2</sub> requirement for the combustion process from the air, and the excess air supplied includes nitrogen along with oxygen. Therefore, the relationship between O<sub>2</sub> and N<sub>2</sub> was high in the heatmap.

### 2.4 Data Engineering and Forecasting Pipeline

Several essential data engineering techniques were used to guarantee the quality and suitability of the data for forecasting stack gas levels. To ensure data accuracy and consistency, unexpected categorical annotations were removed from numeric values. These anomalies may have resulted from data input errors or inconsistencies in data collection; their elimination contributed to the development of a reliable dataset for the prediction of stack gas level. Additionally, the column names were changed to enhance data readability and understanding. A careful method was used to deal with the missing values. The average of the previous and subsequent observations at the time of the missing value was utilized to impute the missing data effectively.

To tackle outliers, a robust outlier detection method Z-score to carefully identify and flag potential outliers was employed. A cautious approach

was taken with the outliers found. Instead of complete removal, robust imputation techniques, replacing outliers with the median, a measure that is less sensitive to extreme values after removing outliers were applied. This ensured a reduction in the impact of outliers on the predictions while maintaining the integrity of the dataset (Maltare & Vahora, 2023).

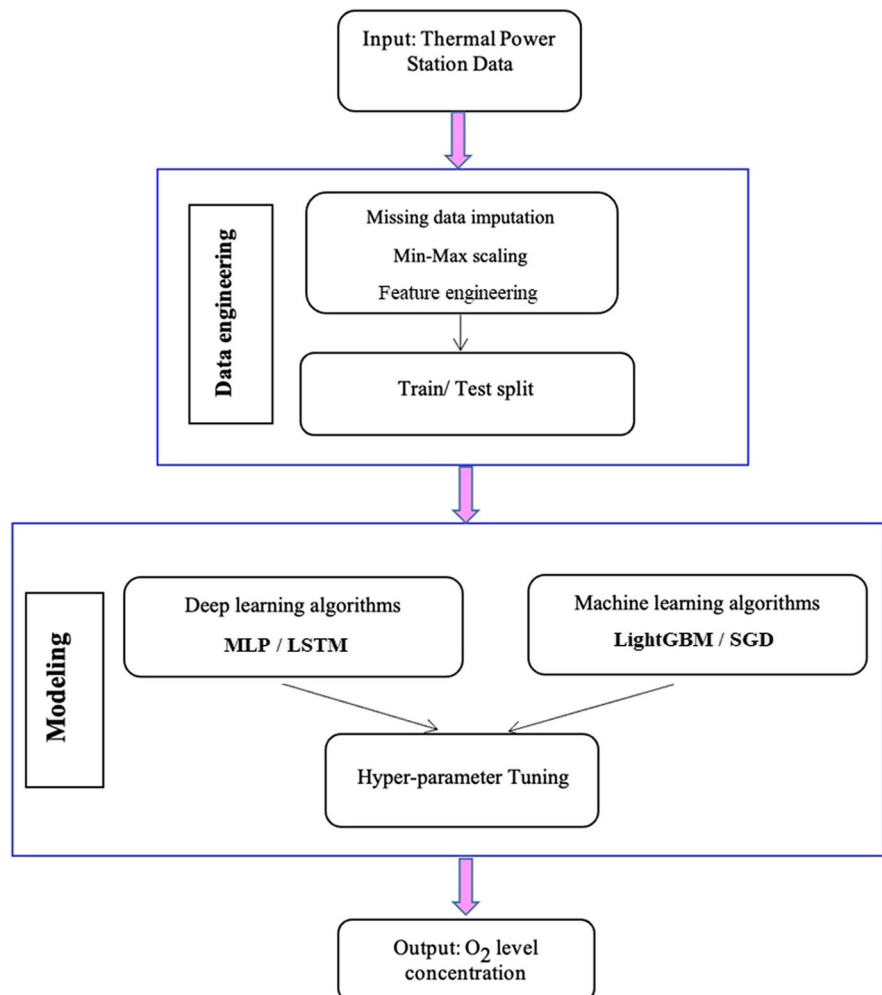
Moreover, to mitigate the impact of numerical variations on forecast accuracy, the data were transformed using the min-max function, scaling it to the range [0, 1] (Eq. 8).

$$y' = \frac{y - \min(y)}{\max(y) - \min(y)} \quad (8)$$

The prediction results were acquired after model training by using the reverse transformation to

rescale. The inverse transformation delivered predicted stack gas levels at their original scale, allowing for meaningful and interpretable projections in thermal power plants. The integrated pipeline for stack gas level forecasting is illustrated in Fig. 3. The pipeline commences with data collection from coal-fired thermal power stations, followed by data preprocessing and feature engineering to ensure data quality. Afterward, the data is partitioned and scaled before feeding into various suggested models, such as MLP, LSTM, LightGBM, and SGD models. Comprehensive model training allows for a thorough comparison and evaluation, ultimately identifying the optimal approach for stack gas level forecasting.

**Fig. 3** Data processing and model selection pipeline for stack gas level foreca





### 3 Modelling

Machine learning approaches offer flexible tools for data analysis and prediction, allowing us to derive significant insights from the massive stack of gas emission data collected over time. These methods aim to improve stack gas level forecasts for better environmental management and energy conservation (Kumar & Pande, 2023). Neural networks, also known as artificial neural networks (ANNs), are the foundation of deep learning algorithms and constitute a subfield of machine learning. Deep learning as a subset of artificial neural networks, can tackle a wide range of issues, including time-series analysis. This method allows the model to understand complex relationships by processing data through numerous layers of interconnected neurons. Model depth boosts its ability to capture complicated patterns and features (Zhu et al., 2023). To assess the performance and generalizability of the implemented model, a time series cross-validation approach was implemented. Specifically, a sliding-window validation strategy, where the dataset has been divided into overlapping training and testing sets were employed, respectively. The model was trained on a subset up to a certain time, tested in the next period, and evaluated iteratively. This dynamic approach preserves the temporal dependencies that are necessary to evaluate the predictive performance of time series data.

In this paper, LSTM and MLP models were selected as deep learning methods; LightGBM and SGD models were selected as machine learning methods., and the models were applied to real data collected from stack gases.

#### 3.1 Multi-Layer Perceptron (MLP) Neural Network Model

The multi-layer perceptron (MLP) is a type of ANN within the feedforward neural network family, serving as a fundamental form of deep learning techniques (Ehteram et al., 2023). MLP involves numerous layers of interconnected perceptrons or artificial neurons. The network consists of an input layer, one or more hidden layers, and an output layer. The data is first received by the input layer, then processed and learned from by the hidden layers, and the output layer produces the final results or predictions. Each perceptron in the hidden and output layers is directly

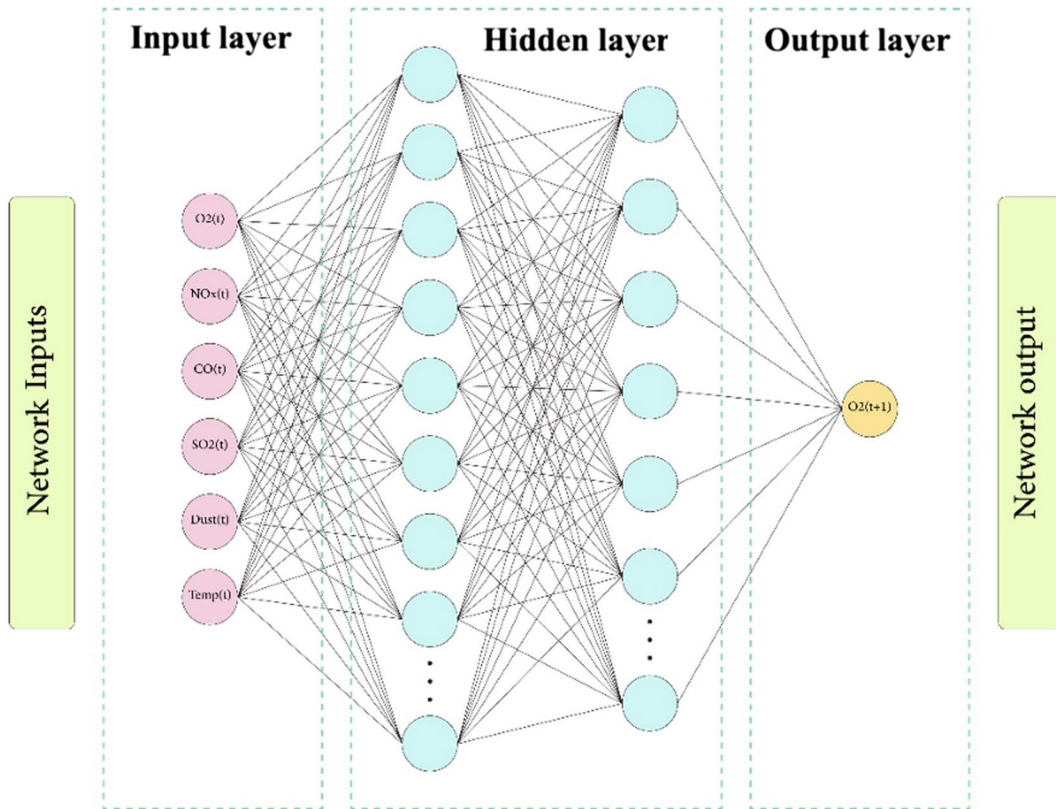
coupled to the perceptron in the preceding and next layers. This connectivity enables the network to process and learn complex patterns from input data (Rumelhart et al., 1986). Each perceptron takes a set of inputs and produces a single output. A perceptron's inputs are multiplied by weights and the sum of the products is processed through an activation function. The following equation (Eq. 9) can represent how each perceptron functions mathematically:

$$y_{xw} = f\left(\sum_{i=1}^m w_i x_i + b\right) \quad (9)$$

The output of the perceptron is represented as  $y_{xw}$ , where  $y$  denotes the output,  $x$  is the input data vector of size  $m$ ,  $w_i$  denotes the weights, and  $b$  represents the bias term (Eq. 9). The activation function  $f$  introduces non-linearity to the output by applying it to the sum of input features and corresponding weights. The main objective is to check if the activation function ( $f$ ) generates a non-zero value by calculating the weighted sum of input features, which are learned through supervised learning. An MLP typically consists of three or more fully connected layers. Fig. 4 depicts a schematic illustration of an ordinary MLP design. The minimal number of layers required is an input, hidden, and output layer. Specifically, back-propagation is used to train the MLP network. It is important to note that the MLP method is a type of deep learning.

#### 3.2 Long Short-Term Memory (LSTM) Model

A particular type of recurrent neural network (RNN) called Long Short-Term Memory (LSTM) is made by Hochreiter and Schmidhuber (1997) to deal with time series and sequential data. LSTM networks are extremely efficient for tasks requiring sequences and temporal dependencies due to their unique ability to store and utilize information from previous time steps, in contrast to standard feedforward neural networks that process individual input points separately. LSTM manages information through units called cells. Each cell has three main components: a forget gate ( $f_t$ ), an input gate ( $i_t$ ) and an output gate ( $o_t$ ). The forget gate decides what information from the previous cell state ( $C_{t-1}$ ) is allowed to be forgotten. The input gate, decides how much of the current input is allowed to enter the cell. The output gate decides how



**Fig. 4** Structure of an MLP design

much of the current state ( $c_t$ ) is allowed to be output to the next hidden state ( $h_t$ ). The candidate cell state ( $c'_t$ ) represents new information that can be incorporated into the cell state. The LSTM architecture is shown in Fig. 5.

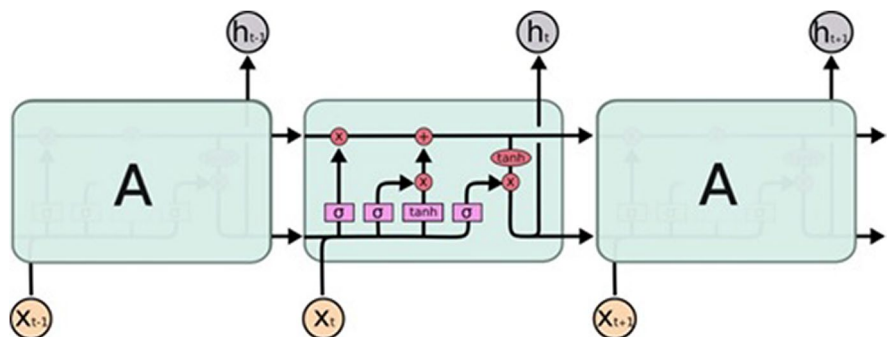
The following are the equations (Eq. 10-15) that accomplish these operations:

$$f_t = \sigma(W_f * [h_{t-1}, x_t] + b_f) \tag{10}$$

$$i_t = \sigma(W_i * [h_{t-1}, x_t] + b_i) \tag{11}$$

$$C'_t = \tanh(W_c * [h_{t-1}, x_t] + b_c) \tag{12}$$

**Fig. 5** Structure of a LSTM design





$$C_t = f_t * C_{t-1} + i_t * C'_t \quad (13)$$

$$O_t = \sigma(W_o * [h_{t-1}, x_t] + b_o) \quad (14)$$

$$h_t = O_t * \tanh(C_t) \quad (15)$$

$W$  represents the weight matrix,  $b$  represents the bias vector, and  $\sigma$  denotes the sigmoid activation function. The tanh activation function is utilized in LSTM models and other ANNs. It processes and captures complex patterns in input data by incorporating non-linearity to the LSTM model.

### 3.3 Light Gradient Boosted Machine (LightGBM) Model

The LightGBM algorithm was developed by Ke et al. (2017) as a highly efficient and scalable gradient-boosting method for handling large-scale data. LightGBM employs tree-based learning techniques which is widely used for regression tasks. The main goal of this approach is to iteratively minimize an objective function by incorporating weak learners (decision trees) into the model. The objective function consists of a loss function that measures the disparity between the actual and predicted target values. LightGBM's objective is to find the optimal parameters ( $\omega$ ) that minimize the objective function ( $Q(\omega)$ ), which is expressed in a specific form (Eq. 16):

$$Q(\omega) = \sum (L(y_i, \hat{y}_i)) + \Omega(\omega) \quad (16)$$

In this equation,  $i \in [1, \dots, n]$  where  $n$  represents the total number of data samples. In the aforementioned equation: The variable  $y_i$  denotes the actual target value of the  $i$ -th data point. The variable  $\hat{y}_i$  is the predicted target value of the  $i$ -th data point based on the current set of parameters. The loss function, denoted as  $L(y_i, \hat{y}_i)$  measures the difference between the actual target value,  $y_i$ , and the predicted target value,  $\hat{y}_i$ . The regularization term  $\Omega(\omega)$  serves as a penalty for complex models to prevent overfitting.

### 3.4 Stochastic Gradient Descent (SGD) Model

The Stochastic Gradient Descent Regressor utilizes a stochastic gradient descent algorithm, a gradient-based

optimization method commonly used to minimize the cost function associated with linear regression. Due to the usage of a randomly selected subset of the training data, known as a mini-batch, in updating the model's parameters, this approach is particularly appropriate for handling large datasets. This makes SGD computationally efficient and able to solve large-scale regression problems. Gradient descent starts with an initial guess for the parameters and iteratively updates them in the direction of the negative gradient of the cost function, which indicates the direction of the steepest descent (Tian et al., 2023). The gradient is the partial derivative of the cost function with respect to each parameter. SGD is an iterative technique used to find the optimal set of parameters ( $\omega$ ) that minimizes an objective function ( $Q(\omega)$ ) of the form of (Eqs. (17), (18)),

$$Q(\omega) = \frac{1}{n} \sum_{i=1}^n Q_i(\omega). \quad (17)$$

The goal is to estimate the parameters that result in the lowest overall loss. Typically, each summand function  $Q_i$  corresponds to the  $i^{\text{th}}$  observation in the training data set. The parameters  $\omega$  of the objective  $Q(\omega)$  have been updated as,

$$\omega : \omega - \eta \nabla Q(\omega) = \omega - \frac{\eta}{n} \sum_{i=1}^n \nabla Q_i(\omega) \quad (18)$$

In which  $\eta$  represents a step size (sometimes called the learning rate in machine learning). At each iteration, SGD computes the gradients for each observation and subsequently updates the parameters accordingly. The algorithm repeats this process for multiple epochs or until the parameters converge to an optimal solution. The learning rate ( $\eta$ ) is a hyperparameter that must be thoroughly tuned for optimal model training performance.

### 3.5 Model Settings

The performance and convergence of the models are significantly influenced by their relevant hyperparameters. Proper tuning of these parameters plays a crucial role to optimize their performance and achieve accurate stack gas level forecasting results. The main hyperparameters in the MLP model involve the learning rate, the number of hidden layers, the number of

neurons in the hidden layers, and the choice of the solver. Important hyperparameter in the context of LSTM include the number of LSTM units, dropout rate, and activation function. LightGBM's performance depends on the boosting type, learning rate, number of estimators (or trees), and the maximum number of leaves per tree. The learning rate, regularization intensity (alpha), number of iterations (epsilon), and the choice of the loss function are important hyperparameters that need to be tuned for the SGD regressor (Badriyah et al., 2020; Lee et al., 2020).

The model's forecasting ability depends on the hyperparameters. If the default hyperparameter values are used, this combination may not be ideal for the forecast. Hence, the grid search method was implemented to establish the network's architecture, adjust its hyperparameters, and fine-tune them (Liashchynskiy & Liashchynskiy, 2019). Before adjusting the parameters, the optimization range of hyperparameters was set first. After determining the ideal parameter combination, the prediction was

executed. Scikit-learn Python library (Miranda et al., 2023), which was a robust, up-to-date, and freely available machine-learning toolkit was used in the present paper. The optimum setting parameter combinations for the implemented method as determined by the tuning algorithm are displayed in Table 2.

### 3.6 Merits of the Multi-Model Approach

By utilizing various models, like LSTM, MLP, LightGBM, and SGD, it can be evaluated how well they perform and determine which model best captures the underlying trends and patterns within the using data. The reliability of findings can be strengthened by this comparative methodology. Besides, various models possess unique strengths. While LSTMs are best suited for sequential data, such as time series (Dou et al., 2023), MLPs are excellent at learning complex, non-linear relationships (Naskath et al., 2023). Light GBM performs well in handling large datasets and can provide high predictive accuracy.

**Table 2** Optimum setting parameters of MLP, LSTM, LightGBM, and SGD models

	Hyperparameter	Search range	O <sub>2</sub>	NOx	CO	SO <sub>2</sub>	Dust
MLP	Activation function	Relu, Sigmoid, Tanh, Leakyrelu	Relu	Relu	Relu	Tanh	Relu
	Regularization parameter	0.0001-1.0	0.001	0.0001	0.01	0.01	0.0001
	Number of neurons	50-200	200,100	100,50	50	150,50	100,75
	Number of hidden layers	1-3	2	2	1	2	2
	Optimization solver	Adam, SGD, RMSProp, L-BFGS	ALBFGS	ALBFGS	Adam	L-BFGS	L-BFGS
LSTM	Number of neurons	128-384	128	128	128	384	384
	Number of hidden layers	1-4	4	4	4	4	4
	Dropout rate	0.2-0.4	0.2	0.4	0.3	0.3	0.2
	Number of epochs	50-100	50	100	50	74	100
LightGBM	Boosting type	GBDT, DART, GOSS	GBDT	GOSS	GBDT	GOSS	GOSS
	Learning rate	0.01-1.0	0.01	0.05	0.075	0.2	0.15
	Number of trees	100-300	300	150	150	150	200
	Number of leaves	10-40	31	40	25	40	30
SGD	Alpha	0.0001-0.01	0.001	0.001	0.001	0.001	0.0001
	Epsilon	0.001-1.0	0.1	0.05	0.01	0.05	0.1
	Loss function	Squared error, Epsilon_insensitive, Huber	Squared error	Epsilon_insensitive	Epsilon_insensitive	Epsilon_insensitive	Huber

Compared to more complex models like MLP and LSTMs, SGD regression is simpler. Utilizing a diverse set of models leverages these strengths, which may result in more precise forecasts. In addition, using several models increases the overall strength of a study. If one model performs poorly due to unforeseen factors, the results of other models can still provide valuable insight. This reduces the reliance on a model that might be inaccurate. Training several models, particularly complex models such as LSTMs and MLP can be computationally intensive (Song et al. 2021a, b).

As it is known that LSTMs are a type of Recurrent Neural Network (RNN) that was developed to deal with time series and sequential data. They are especially beneficial for time series forecasting tasks because they can remember information for a long period and capture long-term dependencies. LSTMs are effective for capturing dynamics and trends over time and provide a more accurate representation of emission behaviour. In addition, MLPs are strong neural network models that can represent complex nonlinear relationships in data. They are good at learning complex patterns and are suitable for tasks with multiple input features, which makes them suitable for tasks requiring the prediction of numerous pollutants. Given the diverse set of pollutants and the complexity of their interactions, MLPs can effectively model and capture the underlying patterns in emission data because of their capacity to discover non-linear relationships through hidden layers. LightGBM is a gradient-boosting framework that performs well in handling large datasets and can provide high predictive accuracy. It is selected to complement deep learning models and provides an alternative approach that can capture complex relationships in data. SGD regression is a less complicated model than more intricate ones like MLP and LSTMs.

### 3.7 Model Evaluation

The performance of the models in the present paper was assessed using the mean absolute error (MAE), root mean square error (RMSE), and R squared indicators defined by Eq. 19-21, respectively.

The average of the absolute error values is measured by the MAE score, which is calculated without taking into account the direction of the errors.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}| \quad (19)$$

The square root of the MSE value yields the RMSE value. RMSE is commonly used as a benchmark metric to evaluate model performance since it is a trustworthy indicator of anticipated accuracy.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y})^2} \quad (20)$$

Coefficient of determination, or R-squared, measures how much of a change in one variable can be predicted by changing another(s).

$$R^2 = 1 - \frac{(y_i - \hat{y})^2}{(y_i - \bar{y})^2} \quad (21)$$

Here, the observed value is represented by  $y_i$ , the prediction value by  $\hat{y}$ , the mean value of  $y$  by  $\bar{y}$  and  $n$  is the length of the time series.

The R-squared value represents the proportion of the variance in the dependent variable that is explained by the independent variables within the models. The R-squared values along with other metrics RMSE and MAE obtained offer a valuable metric for comparing the performance of the different models. The model's strength can be gauged by its small MAE and RMSE values and its high  $R^2$  value.

## 4 Results and Discussion

This section delves into the key findings of the study, encompassing the performance of various machine learning and deep learning models in forecasting stack gas concentrations. Different models were utilized to analyze the actual data obtained from the stack gases across one-year period. MLP, LSTM, LightGBM, and SGD regressors were evaluated for stack gas level forecasting. In three key steps, we will examine the results.

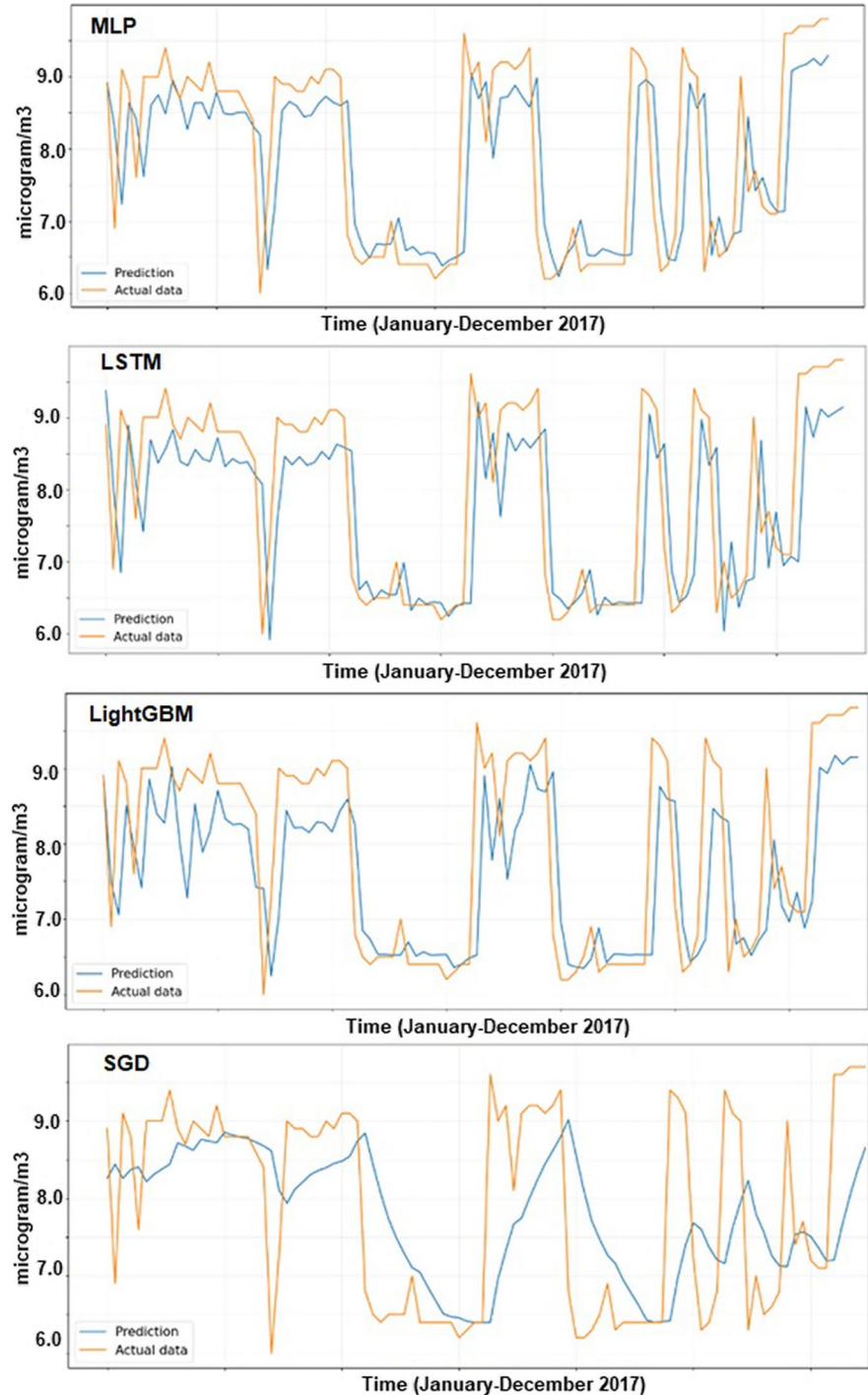
### 4.1 Visualization of Model Performance

The visual representations of the actual data alongside the predicted produced by each of the four models employed: MLP, LSTM, LightGBM, and SGD

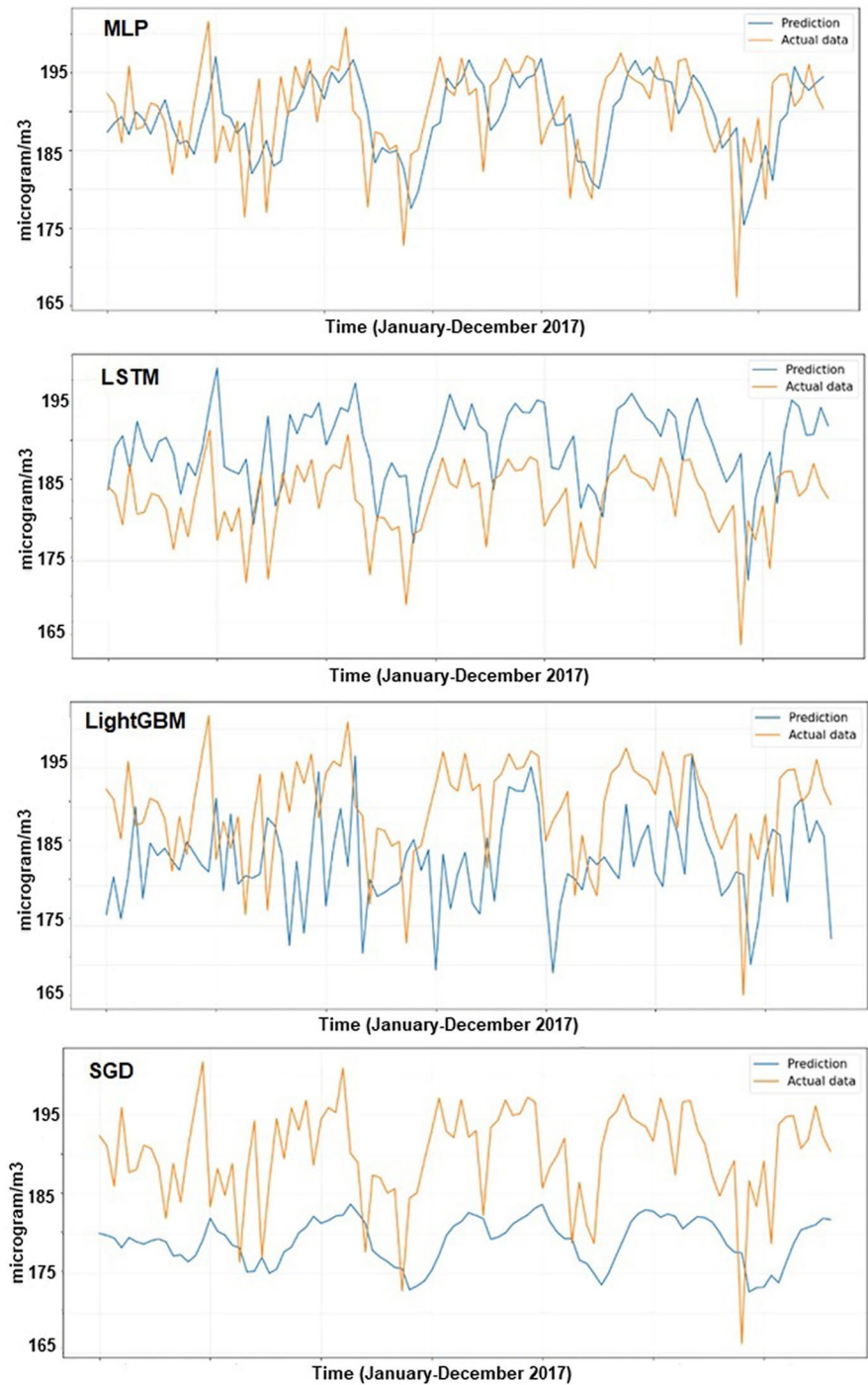
regressors are provided in this subsection. These visualizations provide a qualitative assessment and also initial insight into each model's ability to capture underlying trends and patterns in pollution data.

Figs. 6, 7, 8, 9, and 10 depict the time series plots illustrating the actual data (represented by a solid orange line) and the predicted data (represented by a solid blue line) from four distinct regression models:

**Fig. 6** Time series plot of  $O_2$  concentration level



**Fig. 7** Time series plot of NO<sub>x</sub> concentration level

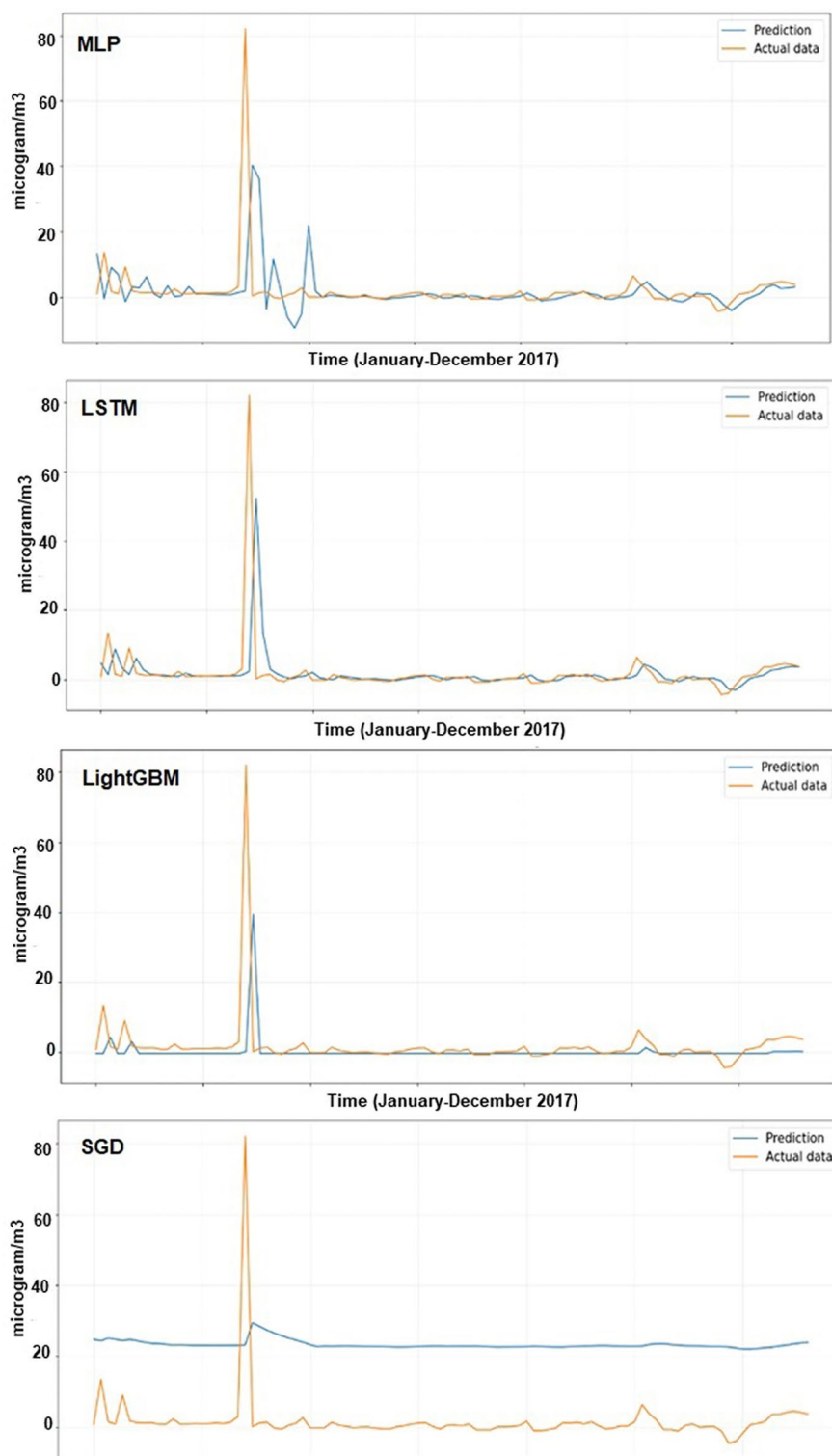


MLP, LSTM, LightGBM, and SGD regressors. The timestamp is represented by the x-axis, while the gas concentration level is represented by the y-axis. Fig. 6 reveals that the actual O<sub>2</sub> concentration levels exhibited various cyclical patterns and temporal

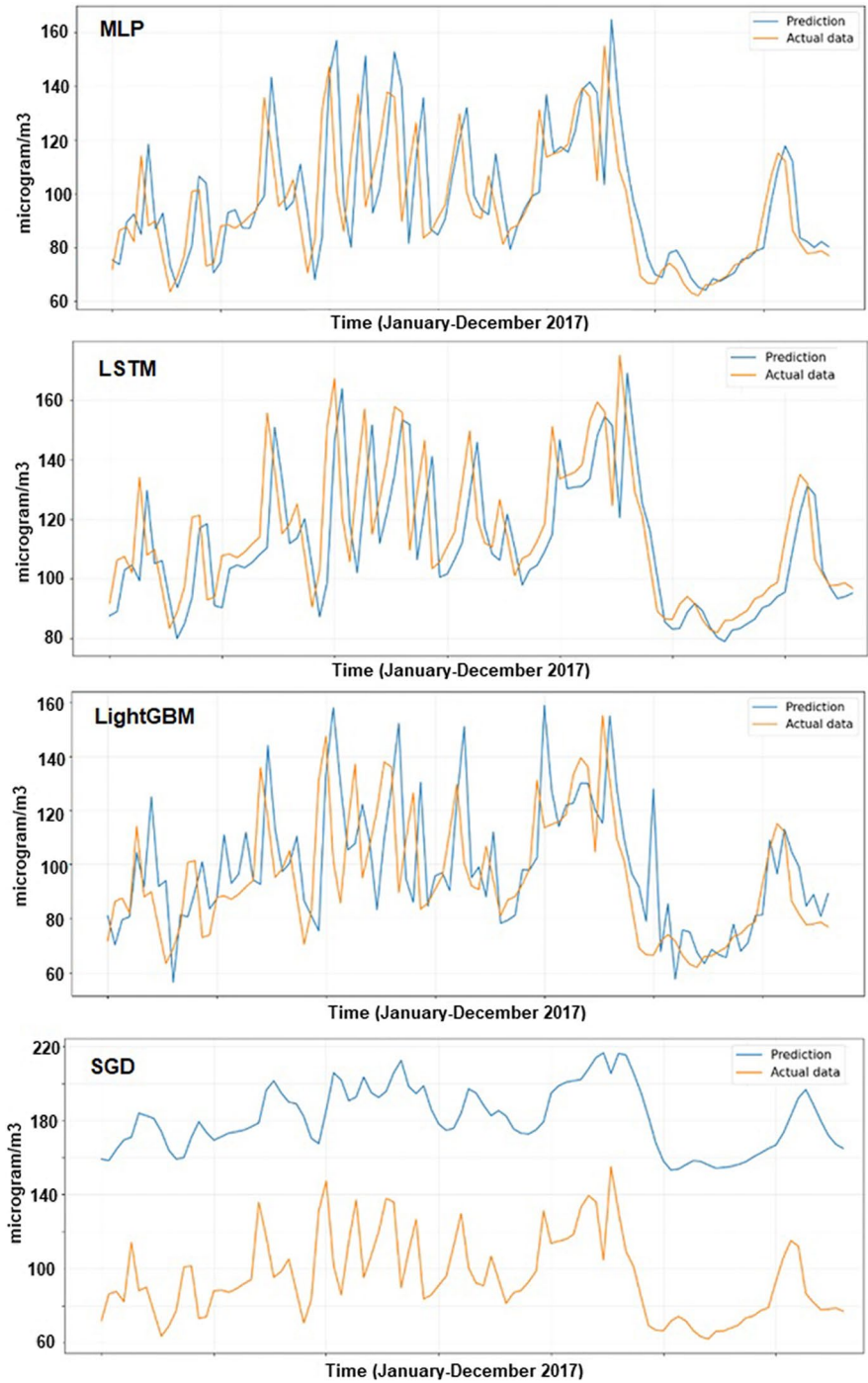
changes. MLP, LSTM, and LightGBM models attempted to define these patterns and predicted future O<sub>2</sub> concentration levels. The forecast fitted the actual values quite well with the exception of the SGD model.



**Fig. 8** Time series plot of CO concentration level



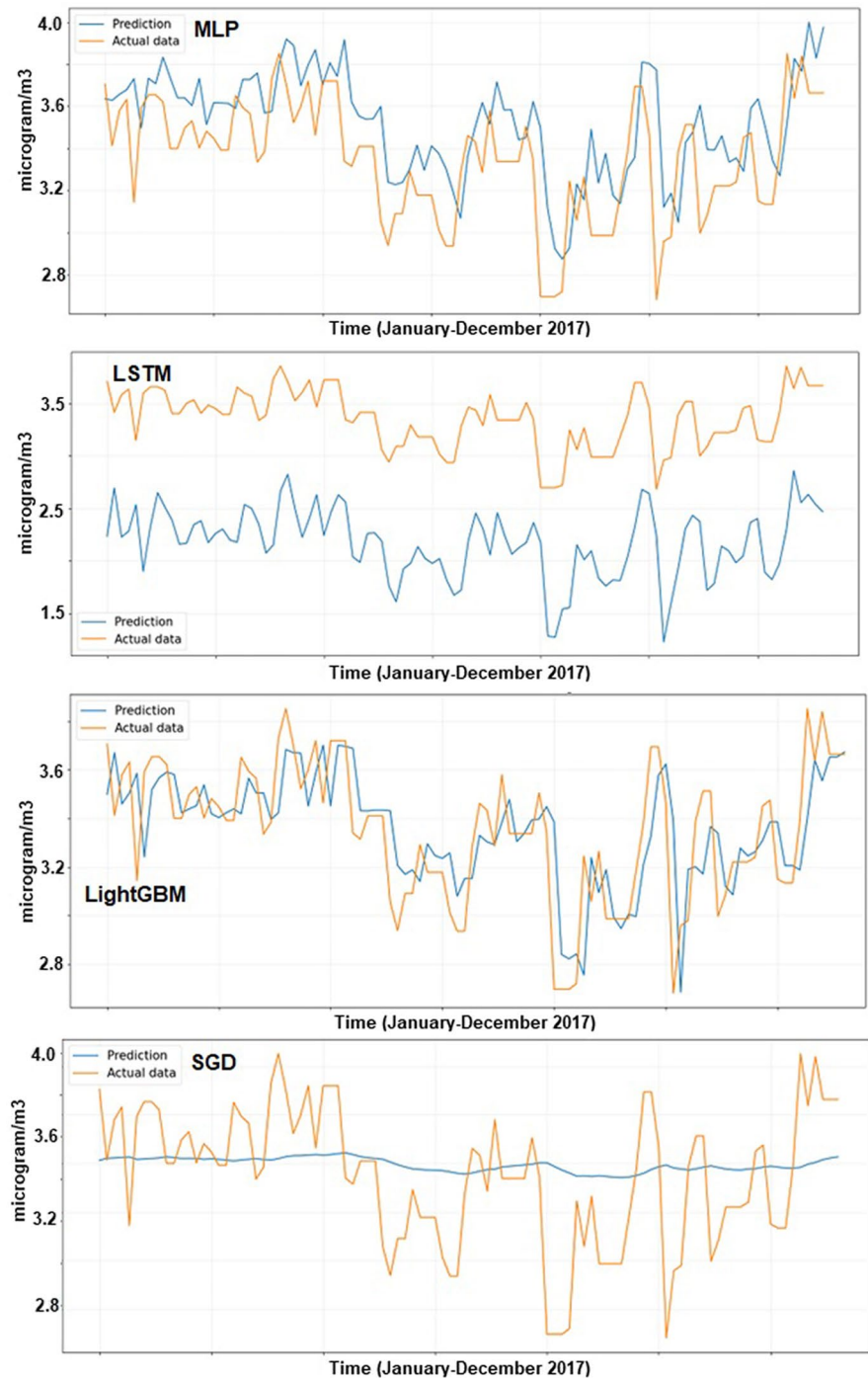
**Fig. 9** Time series plot of SO<sub>2</sub> concentration level



The time series plot of NO<sub>x</sub> concentration level is shown in Fig. 7. The MLP and LSTM models exhibited a good capacity for pattern learning and capture in the concentration data. LSTM was an effectively model both long-term trends and short-term

fluctuations, indicating its proficiency in time series analysis. The LSTM model accurately reflected concentration levels. The models' ability to capture patterns, including rapid concentration fluctuations,

**Fig. 10** Time series plot of dust concentration level



showed their robustness and applicability for accurate forecasts in dynamic environments.

In Fig. 8, the LSTM model captured the data's general trend and patterns well, making it a potential choice for accurate forecasting of the CO

concentration data. It accurately predicted short-term trends by identifying and modeling data peaks and valleys. Additional variations did not prevent the actual data from being seen in the predictions made by the MLP model. In other words, the MLP model

overfitted to noise and random variations in the training data, resulted in higher oscillations or unnecessary fluctuations in its predictions.

Fig. 9 shows the time series of SO<sub>2</sub> in detail. The LSTM model performed exceptionally well at identifying both short-term and long-term trends in the data. It provided precise short-term predictions by successfully identifying and modelled the peaks and troughs contained in the actual data. It also effectively captured the time series' overall upward tendency, making it an appropriate choice for accurate forecasting tasks (Krzywanski & Nowak, 2016). As the best alternative, the MLP model made a respectable prediction of capturing the overall trend in the data.

According to Fig. 9, the performance of the LightGBM model was reasonable, as it was able to capture the overall long-term trend. However, it encountered difficulties in effectively modeling the short-term variations.

Fig. 10 illustrates the predicted time series plot of dust. The LSTM model well captured the trend and pattern of the data, despite a little discrepancy between the predicted and real values. It indicated the model's efficiency in time series analysis by showing that it can discover and describe the underlying structure of the data.

#### 4.2 Quantitative Assessment Using Performance Metrics

The performance of these models is evaluated using the established mathematical criteria to conduct a quantitative evaluation. Three important metrics will be considered: mean absolute error (MAE), root mean square error (RMSE), and R-squared (R<sup>2</sup>). By examining these criteria, it is possible to compare models objectively and identify the model that demonstrates the most accurate and consistent forecasting ability for the chosen pollutants. The results of the LSTM,

MLP, LightGBM, and SGD models' assessment for the forecasting of stack gas levels from a thermal power plant in Türkiye are displayed in Table 3. The evaluation was based on commonly used accuracy measures, namely MAE, RMSE, and R<sup>2</sup>. Although the coefficient of determination (R<sup>2</sup>) was crucial for explaining overall variance, MAE and RMSE should also be taken into account when evaluating how well models predicted specific data points. MAE and RMSE are useful metrics for assessing the point-by-point performance of models. The model's strength can be defined by its small MAE and RMSE values and its high R<sup>2</sup> value (Chikobvu & Mamba, 2023).

According to the modelling results for all stack gases, the LSTM network offers tremendous potential as a method for simulating thermal power plants, particularly ones with large time delays, due to its adaptability and efficacy (Liu et al., 2020). The LSTM model had a remarkable ability to explain the overall trends and patterns for all gases, as it consistently achieved the highest R<sup>2</sup> values across the examined models. This indicated its superior capacity to explain data variance and model gas concentrations over time (Vujić et al., 2019). The results (Table 3) show that LSTM was the best-fit model for predicting stack gas levels, with an MAE of 0.513, 12.45, 5.77, 12.89, and 1.69; RMSE of 1.021, 23.26, 51.67, 51.95, and 5.95; R<sup>2</sup> of 0.84, 0.87, 0.67, 0.85, and 0.78 for O<sub>2</sub>, NO<sub>x</sub>, CO, SO<sub>2</sub>, and dust, respectively. When the LSTM model was applied to CO data, MAE and RMSE values were lower than those of the other stack gases. As the MAE and RMSE values decreased, the prediction success of the model increased (Josimović et al., 2023; Yuan et al., 2021).

The results show that MLP was one of the most accurate models for predicting O<sub>2</sub>, NO<sub>x</sub>, and SO<sub>2</sub> levels, with an MAE of 0.506, 7.47, and 13.32, RMSE of 1.031, 26.13, and 53.62, and R<sup>2</sup> of 0.84, 0.84, and 0.84, respectively. In addition, a strong correlation

**Table 3** Model assessment results for the forecasting of stack gas levels using MLP, LSTM, LightGBM, and SGD models

	MLP			LSTM			LightGBM			SGD		
	MAE	RMSE	R <sup>2</sup>	MAE	RMSE	R <sup>2</sup>	MAE	RMSE	R <sup>2</sup>	MAE	RMSE	R <sup>2</sup>
O <sub>2</sub>	0.506	1.031	0.84	0.513	1.00	0.84	0.529	1.015	0.85	0.82	1.394	0.71
NO <sub>x</sub>	7.47	26.13	0.84	12.45	23.26	0.87	10.33	26.40	0.84	15.69	32.76	0.75
CO	10.49	59.06	0.54	5.77	51.67	0.67	9.21	66.10	0.42	24.76	79.97	0.16
SO <sub>2</sub>	13.32	53.62	0.84	12.89	51.95	0.85	22.69	62.06	0.78	33.24	91.55	0.53
Dust	0.84	7.63	0.63	1.69	5.95	0.78	0.66	5.85	0.78	1.41	11.78	0.13

was considered to exist when the  $R^2$  value was greater than 0.8, and a low correlation was considered to exist when the  $R^2$  value was less than 0.5. In this case, it can be said that CO had a low correlation when the MLP model was applied. Similarly, MLP was not a very good fit for dust. The  $R^2$  values in the CO and dust data to which the MLP model was applied are far from 1.0, which means that the values predicted by the model with these data and the actual values were quite far from each other (Movahed et al., 2023; Tang et al., 2022). As it can be seen from the  $R^2$  values, the predicted and the actual data of  $O_2$ , NOx, and  $SO_2$  levels were closer to each other when using MLP model.

Comparing the LSTM and MLP models of  $O_2$ , the lowest RMSE value was obtained in the LSTM, while the lowest MAE value was obtained in the MLP model. In this case, the decision is up to the user. The MLP model should be used if the correlation coefficient between them was more important than the error value, and the LSTM model should be used when it was desired to obtain results with a lower error rate.

Using data for  $O_2$ , NOx,  $SO_2$ , and dust, the LightGBM model was used to compare the actual and predicted values. The MAE and RSME values for  $O_2$  and dust were low enough, and the R-square values were good, so it can be said that  $O_2$  and dust data were suitable for LightGBM model compared to other gases (Table 3). CO data has the worst results among the evaluation criteria in the LightGBM model. The results show that LightGBM was the worst fit model for predicting CO level, with an MAE of 9.21, RMSE of 66.10, and an R-square of 0.42.

Consequently, it can be safely stated that the models showing higher R-squared values and lower RMSE and MAE demonstrate a stronger ability to explain the variations in the stack gas concentrations of  $SO_2$ , CO,  $O_2$ , NOx, and dust. While both LSTM and in most cases MLP and the Light GBM models show high R-squared values, indicating statistically significant relationships between pollutant concentration and the other emitted pollutants, dust, and temperature, we acknowledge the need to investigate the practical implications of these results in more detail.

For SGD model, the relationship between actual and predicted values was modelled when using stack gas data (Table 3). Especially the R-square values of dust, CO, and  $SO_2$  were extremely low. Likewise, the MAE and RMSE values for  $SO_2$  were also extremely

high. The SGD model performed quite poorly when it comes to CO and dust gases, in particular. In compared to the other models, it has much lower  $R^2$  values of 0.16 and 0.13 for CO and dust, respectively. Therefore, it can be said that the SGD model is not suitable for all stack gases.

According to Table 3, there are two models with the highest  $R^2$  values: MLP with  $R^2$  values between 0.84-0.63 and LSTM with  $R^2$  values between 0.84-0.78. In these models, RMSE values of  $O_2$  were 1.031 and 1.021, and the RMSE values of NOx were 26.13 and 23.26 for MLP and LSTM, respectively. MLP is the second-best model. For NOx, the highest R-square and lowest RMSE values are obtained in the LSTM, while the lowest MAE value is obtained in the MLP model. In this case, the decision is left to the user. The MLP algorithm should be used if the correlation coefficient between them is more important than the error value in the established model. In contrast, the LSTM algorithm should be used when desired to obtain results with a lower error rate. LightGBM, with  $R^2$  values in the range of 0.85-0.30 and RMSE values in the range of 1.015 to 72.90 for all gas types, was the method with lower performance than LSTM and MLP models. The SGD method, with  $R^2$  values in the range of 0.71-0.13 and RMSE values of 1.394-107.25, had similar performance values to LightGBM.

To summarize all the results of the modelling studies as a whole: the LSTM model had a remarkable ability to explain the overall trends and patterns for all gases as it consistently achieved the highest  $R^2$  values of all the models examined. When the LSTM model was applied to CO data, the MAE and RMSE values were lower than for the other stack gases. Also, the results show that MLP was one of the most accurate models for predicting  $O_2$ , NOx and  $SO_2$  levels, while CO and dust had a low correlation when the MLP model was applied. The predicted and actual data for  $O_2$ , NOx and  $SO_2$  were closer when the MLP model was used. Using data for  $O_2$ , NOx,  $SO_2$ , and dust, the LightGBM model was used to compare actual and predicted values. It can be said that the  $O_2$  and dust data were suitable for the LightGBM model compared to other gases. For the SGD model, the relationship between actual and predicted values was modelled using stack gas data. In particular, the R-squared values of dust, CO and  $SO_2$  were extremely low. Therefore, it can be said that the SGD model is



not suitable for all stack gases. It can be safely stated that the models showing higher R-squared values and lower RMSE and MAE demonstrate a stronger ability to explain the variations in the stack gas concentrations of SO<sub>2</sub>, CO, O<sub>2</sub>, NO<sub>x</sub>, and dust. While both LSTM and in most cases MLP and the Light GBM models show high R-squared values, indicating statistically significant relationships between pollutant concentration and the other emitted pollutants, dust, and temperature, practical implications of these results should be investigated in more detail. In addition, according to the models' R-squared, MAE, and RMSE values according to the obtained prediction results and actual values, LSTM and MLP values were found to be better models than the others (Liu et al., 2020). A large number of raw data sets and many test data sets separated from these data sets can be considered reasons why the results of the LSTM and MLP models are more significant than the others. As seen in the models we have established, the large amount of data used for the models has enabled them to make more successful predictions. The results we obtained in the LSTM and MLP models were also quite successful when translated into real values. According to these results, it can be said that a new model has been created using artificial neural networks to determine the levels of stack gases.

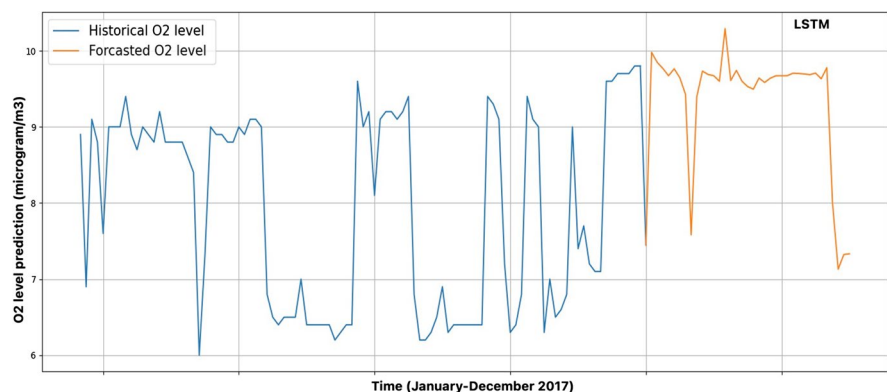
### 4.3 Demonstrating Forecasting Potential

Finally, using the insights gained from the visual and quantitative evaluations, the best performing model based on the selected criteria has been chosen. This

selected model is then used to predict pollution levels for a particular pollutant in the next 36 hours. This practical demonstration highlights the model's ability to predict future trends and allows us to evaluate its potential utility in real-world applications. Considering the results obtained in the study, the LSTM method gave the best results. Therefore, a study was conducted to estimate the oxygen content in future periods. Various approaches were explored within the LSTM model by adjusting parameters such as epochs, dropouts, and units. The best result was found in the experiment with 64 hidden units, 0.2 dropout rates, and 40 epochs. These parameters used in the model resulted in data that was closest to the real value.

As shown in Fig. 11, the LSTM model was employed to predict O<sub>2</sub> concentrations for the next 36 hours following the research period. The model utilized historical gas concentration data for training, and a sliding window approach was employed for making predictions. The research period spanned one year, and the predictions exclusively focused on O<sub>2</sub> gas. "Historical O<sub>2</sub> level," the blue line, shows past O<sub>2</sub> emission levels. This line displays actual emissions up to the point where the future prediction begins. The orange line, labeled "Forecasted future O<sub>2</sub> level," on the other hand, goes beyond the historical data range and shows the model's predictions for O<sub>2</sub> emission levels in the future. These projected values are the result of the trained LSTM model's forecasting capabilities. Close alignment between the blue (historical) and orange (predicted) lines suggests the model's projections closely match the actual past emissions. The plot illustrates that the predicted values are generally close to the actual values, but there is some deviation, particularly for the longer-term predictions.

**Fig. 11** LSTM prediction of O<sub>2</sub> level for the 36 hours



## 5 Limitations of the Study

It is important to acknowledge that our study has some limitations. Training several models, particularly complex models such as LSTMs and MLP can be computationally intensive. We addressed this issue in our study by implementing various strategies to mitigate computational costs effectively. First, to achieve a balance between computational efficiency and predictive performance, we carefully chose model designs and parameters. Moreover, we leveraged cloud computing resources, which allowed us to handle computational demands with flexibility and scalability. Overall, by employing these measures, we effectively managed the computational cost associated with training multiple models, ensuring efficient utilization of computational resources while maintaining the quality of our results. In addition, depending on the models chosen, interpreting their predictions may be other challenging. While interpretability is not critical to our current program, which focuses on predictive accuracy, we acknowledge its importance in some contexts. In future work exploring interpretable models, we might consider strategies like employing feature importance using SHAP (SHapley Additive exPlanations) value analysis. In future work, we may explore techniques to address interpretability and potential bias in model selection.

## 6 Conclusion

In this paper, air pollutant levels were modeled with artificial neural networks using actual emission data from a coal-fired thermal power plant in Türkiye that is in operation. The levels of oxygen (O<sub>2</sub>), nitrogen oxides (NO<sub>x</sub>), carbon monoxide (CO), sulfur dioxide (SO<sub>2</sub>), and dust are modeled using MLP, LSTM, LightGBM and SGD methods representing deep learning and machine learning algorithms. The dataset obtained from the thermal power plant contained a total of 8496 records, of which 1699 were activated for modelling. The efficiency of applied models for forecasting the stack gas emission from a coal-thermal power plant was evaluated by RMSE, MAE, and R<sup>2</sup>. In conclusion, this paper demonstrated that the LSTM model performed well in explaining

the overall trends and patterns for all stack gases due to its highest R<sup>2</sup> values across the other models. The emission forecasts, particularly with LSTM, had good performance results and that this model can be utilized effectively for air pollution forecasting. This proved its excellent capacity to accommodate for data variances and simulated the temporal development of gas concentrations. In general, forecasting the emission of air pollutants from a thermal power plant is essential to developing efficient emission-reduction systems.

Consequently, it can be safely stated that for real-world use, high-accuracy predictive models may be applied to improve plant operations and reduce emissions. This requires modifying emission control systems, fuel mixes, and combustion techniques in accordance with variables affecting stack gas concentrations. Precise forecasting models facilitate compliance with emission standards and environmental regulations while also empowering the facility to proactively maintain acceptable limits.

**Availability of Data and Materials** Authors confirm that all data and materials support their published claims and comply with standards. Data will be made available on request.

**Authors' Contribution** Mandana Mohammadi: conceptualization, methodology, modelling, writing- original draft, software. Didem Saloglu: conceptualization, methodology, writing- original draft, software. Halil Dertli: writing- original draft, editing. Mitra Mohammadi: writing- original draft, editing. Mansour Ghaffari-Modgahham: writing- original draft, editing.

**Funding** Open access funding provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK). No funds, grants, or other support was received.

### Declarations

**Ethical Approval** There is no involvement of human/animal participants in the reported work.

**Consent to Participate** All the authors gave their consent to participate.

**Consent to Publish** All the authors agreed to the publication and current article does not contain data from any individual person. The authors have not posted the manuscript at a preprint server prior to submission.

**Competing Interests** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Adams, D., Oh, D. H., Kim, D. W., Lee, C. H., & Oh, M. (2020). Prediction of SO<sub>x</sub>–NO<sub>x</sub> emission from a coal-fired CFB power plant with machine learning: Plant data learned by deep neural network and least square support vector machine. *Journal of Cleaner Production*, 270, 122310.
- Alnaim, A., Sun, Z., & Tong, D. (2022). Evaluating machine learning and remote sensing in monitoring NO<sub>2</sub> emission of power plants. *Remote Sensing*, 14(3), 729. <https://doi.org/10.3390/rs14030729>
- Asif, Z., Chen, Z., Wang, H., & Zhu, Y. (2022). Update on air pollution control strategies for coal-fired power plants. *Clean technologies and environmental policy*, 24(8), 2329–2347. <https://doi.org/10.1007/s10098-022-02328-8>
- Atukalp, M. E., & Kesimal, A. (2023). Efficiency Change in Coal Mining in Türkiye. *Gazi Journal of Engineering Sciences (GJES)*, 9(1). <https://doi.org/10.30855/gmbd.0705049>
- Badriyah, T., Sakinah, N., Syarif, I., & Syarif, D. R. (2020). Machine learning algorithm for stroke disease classification. In *In 2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)* (pp. 1–5). IEEE.
- Birpınar, M. E., & Tuğaç, Ç. (2022). Climate Security and Migration. *Insight Turkey*, 24(1), 105–134. <https://www.jstor.org/stable/e48509067>
- Chikobvu, D., & Mamba, M. (2023). Modelling emissions from Eskom's coal fired power stations using Generalised Linear Models. *Journal of Energy in Southern Africa*, 34(1), 1–14. <https://doi.org/10.17159/2413-3051/2022/v33i4a13819>
- Cui, Y., Su, W., Xing, Y., Hao, L., Sun, Y., & Cai, Y. (2023). Experimental and simulation evaluation of CO<sub>2</sub>/CO separation under different component ratios in blast furnace gas on zeolites. *Chemical Engineering Journal*, 472, 144579. <https://doi.org/10.1016/j.cej.2023.144579>
- Dai, L., Jia, R., & Wang, X. (2022). Relationship between economic growth and energy consumption from the perspective of sustainable development. *Journal of Environmental and Public Health*, 2022. <https://doi.org/10.1155/2022/6884273>
- Dong, J., Dou, Y., Jiang, Q., & Zhao, J. (2022a). Can financial inclusion facilitate carbon neutrality in China? The role of energy efficiency. *Energy*, 251, 123922. <https://doi.org/10.1016/j.energy.2022.123922>
- Dong, F., Li, Y., Qin, C., Zhang, X., Chen, Y., Zhao, X., & Wang, C. (2022b). Information infrastructure and greenhouse gas emission performance in urban China: A difference-in-differences analysis. *Journal of Environmental Management*, 316, 115252. <https://doi.org/10.1016/j.jenvman.2022.115252>
- Dou, G., Zhao, K., Guo, M. E. I., & Mou, J. U. N. (2023). Memristor-based LSTM network for text classification. *Fractals*, 31(06), 2340040. <https://doi.org/10.1142/S0218348X23400406>
- Ehteram, M., Ahmed, A. N., Khozani, Z. S., & El-Shafie, A. (2023). Graph convolutional network–Long short term memory neural network–multi layer perceptron–Gaussian process regression model: A new deep learning model for predicting ozone concentration. *Atmospheric Pollution Research*, 14(6), 101766. <https://doi.org/10.1016/j.apr.2023.101766>
- Filonchik, M., & Peterson, M. P. (2023). An integrated analysis of air pollution from US coal-fired power plants. *Geoscience Frontiers*, 14(2), 101498. <https://doi.org/10.1016/j.gsf.2022.101498>
- Fontes, T., Silva, L. M., Silva, M. P., Barros, N., & Carvalho, A. C. (2014). Can artificial neural networks be used to predict the origin of ozone episodes? *Science of the Total Environment*, 488, 197–207. <https://doi.org/10.1016/j.scitotenv.2014.04.077>
- Gaffney, J. S., & Marley, N. A. (2009). The impacts of combustion emissions on air quality and climate—From coal to biofuels and beyond. *Atmospheric Environment*, 43(1), 23–36. <https://doi.org/10.1016/j.atmosenv.2008.09.016>
- Hael, M. A. (2023). Unveiling air pollution patterns in Yemen: a spatial–temporal functional data analysis. *Environmental Science and Pollution Research*, 30(17), 50067–50095. <https://doi.org/10.1007/s11356-023-25790-3>
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Hu, Y., & Shi, Y. (2021). Estimating CO<sub>2</sub> emissions from large scale coal fired power plants using OCO-2 observations and emission inventories. *Atmosphere*, 12(7), 811. <https://doi.org/10.3390/atmos12070811>
- Hu, J., Zou, Y., & Zhao, Y. (2024). Robust operation of hydrogen-fueled power-to-gas system within feasible operating zone considering carbon-dioxide recycling process. *International Journal of Hydrogen Energy*, 58, 1429–1442. <https://doi.org/10.1016/j.ijhydene.2024.01.337>
- Josimović, B., Todorović, D., Jovović, A., & Manić, B. (2023). Air pollution modeling to support strategic environmental assessment: case study—National Emission Reduction Plan for coal-fired thermal power plants in Serbia. *Environment, Development and Sustainability*, 1–17. <https://doi.org/10.1007/s10668-023-03186-0>
- Kasman, A., & Duman, Y. S. (2015). CO<sub>2</sub> emissions, economic growth, energy consumption, trade and urbanization in new EU member and candidate countries: a panel data analysis. *Economic Modelling*, 44, 97–103. <https://doi.org/10.1155/2020/8842770>

- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., et al. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 30.
- Kerem, A. (2022). Investigation of carbon footprint effect of renewable power plants regarding energy production: A case study of a city in Turkey. *Journal of the Air & Waste Management Association*, 72(3), 294–307. <https://doi.org/10.1080/10962247.2022.2028690>
- Khan, H., & Khan, I. (2023). The effect of technological innovations, urbanization and economic growth on environmental quality: Does governance matter? *Frontiers in Environmental Science*. <https://doi.org/10.3389/fenvs.2023.1239288>
- Krzywanski, J., & Nowak, W. (2016). Artificial intelligence treatment of SO<sub>2</sub> emissions from CFBC in air and oxygen-enriched conditions. *Journal of Energy Engineering*, 142(1), 04015017. [https://doi.org/10.1061/\(ASCE\)EY.1943-7897.0000280](https://doi.org/10.1061/(ASCE)EY.1943-7897.0000280)
- Kumar, K., & Pande, B. P. (2023). Air pollution prediction with machine learning: a case study of Indian cities. *International Journal of Environmental Science and Technology*, 20(5), 5333–5348. <https://doi.org/10.1007/s13762-022-04241-5>
- Laubscher, R. (2019). Time-series forecasting of coal-fired power plant reheater metal temperatures using encoder-decoder recurrent neural networks. *Energy*, 189, 116187. <https://doi.org/10.1016/j.energy.2019.116187>
- Lee, M., Lin, L., Chen, C. Y., Tsao, Y., Yao, T. H., Fei, M. H., & Fang, S. H. (2020). Forecasting air quality in Taiwan by using machine learning. *Scientific Reports*, 10(1), 4153. <https://doi.org/10.1038/s41598-020-61151-7>
- Li, G. Q., Qi, X. B., Chan, K. C., & Chen, B. (2017). Deep bidirectional learning machine for predicting NO<sub>x</sub> emissions and boiler efficiency from a coal-fired boiler. *Energy & Fuels*, 31(10), 11471–11480. <https://doi.org/10.1021/acs.energyfuels.7b01415>
- Li, M., Li, L., & Strielkowski, W. (2019). The impact of urbanization and industrialization on energy security: A case study of China. *Energies*, 12(11), 2194. <https://doi.org/10.3390/en12112194>
- Liashchynskiy, P., & Liashchynskiy, P. (2019). Grid search, random search, genetic algorithm: a big comparison for NAS. <https://doi.org/10.48550/arXiv.1912.06059>
- Liu, X., Zhang, H., Niu, Y., Zeng, D., Liu, J., Kong, X., & Lee, K. Y. (2020). Modelling of an ultra-supercritical boiler-turbine system with stacked denoising auto-encoder and long short-term memory network. *Information Sciences*, 525, 134–152. <https://doi.org/10.1016/j.ins.2020.03.019>
- Luo, J., Zhuo, W., Liu, S., & Xu, B. (2024). The optimization of carbon emission prediction in low carbon economy under big data. *IEEE Access* [https://ui.adsabs.harvard.edu/link\\_gateway/2024IEEA..1214690L/doi:10.1109/ACCESS.2024.3351468](https://ui.adsabs.harvard.edu/link_gateway/2024IEEA..1214690L/doi:10.1109/ACCESS.2024.3351468)
- Lv, Y., Hong, F., Yang, T., Fang, F., & Liu, J. (2017). A dynamic model for the bed temperature prediction of circulating fluidized bed boilers based on least squares support vector machine with real operational data. *Energy*, 124, 284–294. <https://doi.org/10.1016/j.energy.2017.02.031>
- Maltare, N. N., & Vahora, S. (2023). Air Quality Index prediction using machine learning for Ahmedabad city. *Digital Chemical Engineering*, 7, 100093. <https://doi.org/10.1016/j.dche.2023.100093>
- Meikandaan, T. P., Hemapriya, M., & Vidhya, V. (2019). NO<sub>x</sub> Emission Analysis in Flue Gas of Coal Fired Thermal Power Station and it's Control Measure. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 8(12S). <https://doi.org/10.35940/ijitee.K1324.10812S19>
- Miranda, F. M., Köhnecke, N., & Renard, B. Y. (2023). Hiclass: a python library for local hierarchical classification compatible with scikit-learn. *Journal of Machine Learning Research*, 24(29), 1–17.
- Moharreri, M. A., Arkian, F., Lari, K., & Salehi, G. R. (2020). PM10 and CO dispersion modeling of emissions from four thermal power plants in Mashhad, Iran. *Scientia Iranica*, 27(5), 2433–2442. <https://doi.org/10.24200/sci.2019.52259.2625>
- Movahed, P., Reza zadeh, A. A., Avami, A., Soleymani Baghshah, M., & Mashayekhi, M. (2023). Modeling and optimization of NO emission for a steam power plant by data-driven methods. *Environmental Progress & Sustainable Energy*, 42(3), e14059. <https://doi.org/10.1002/ep.14059>
- Munawer, M. E. (2018). Human health and environmental impacts of coal combustion and post-combustion wastes. *Journal of Sustainable Mining*, 17(2), 87–96. <https://doi.org/10.1016/j.jsm.2017.12.007>
- Naskath, J., Sivakamasundari, G., & Begum, A. A. S. (2023). A study on different deep learning algorithms used in deep neural nets: MLP SOM and DBN. *Wireless Personal Communications*, 128(4), 2913–2936. <https://doi.org/10.1007/s11277-022-10079-4>
- Osobajo, O. A., Otitoju, A., Otitoju, M. A., & Oke, A. (2020). The impact of energy consumption and economic growth on carbon dioxide emissions. *Sustainability*, 12(19), 7965. <https://doi.org/10.3390/su12197965>
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533–536.
- Sharma, H., Marinovici, L., Adetola, V., & Schaeff, H. T. (2023). Data-driven modeling of power generation for a coal power plant under cycling. *Energy and AI*, 11, 100214. <https://doi.org/10.1016/j.egyai.2022.100214>
- Shang, M., & Luo, J. (2021). The tapio decoupling principle and key strategies for changing factors of Chinese urban carbon footprint based on cloud computing. *International Journal of Environmental Research and Public Health*, 18(4), 2101. <https://doi.org/10.3390/ijerph18042101>
- Shi, Y., Zhong, W., Chen, X., Yu, A. B., & Li, J. (2019). Combustion optimization of ultra supercritical boiler based on artificial intelligence. *Energy*, 170, 804–817. <https://doi.org/10.1016/j.energy.2018.12.172>
- Song, F., Mehedi, H., Liang, C., Meng, J., Chen, Z., & Shi, F. (2021a). Review of transition paths for coal-fired power plants. *Global Energy Interconnection*, 4(4), 354–370. <https://doi.org/10.13140/RG.2.2.15151.30884>
- Song, J., Liu, G., Jiang, J., Zhang, P., & Liang, Y. (2021b). Prediction of protein-ATP binding residues based on ensemble of deep convolutional neural networks and lightGBM

- algorithm. *International Journal of Molecular Sciences*, 22(2), 939. <https://doi.org/10.3390/ijms22020939>
- Tang, Z., Wang, S., Chai, X., Cao, S., Ouyang, T., & Li, Y. (2022). Auto-encoder-extreme learning machine model for boiler NO<sub>x</sub> emission concentration prediction. *Energy*, 256, 124552. <https://doi.org/10.1016/j.energy.2022.124552>
- Tian, Y., Zhang, Y., & Zhang, H. (2023). Recent advances in stochastic gradient descent in deep learning. *Mathematics*, 11(3), 682. <https://doi.org/10.3390/math11030682>
- Tunckaya, Y., & Koklukaya, E. (2015). Comparative analysis and prediction study for effluent gas emissions in a coal-fired thermal power plant using artificial intelligence and statistical tools. *Journal of the Energy Institute*, 88(2), 118–125. <https://doi.org/10.1016/j.joei.2014.07.003>
- Tuttle, J. F., Vesel, R., Alagarsamy, S., Blackburn, L. D., & Powell, K. (2019). Sustainable NO<sub>x</sub> emission reduction at a coal-fired power station through the use of online neural network modelling and particle swarm optimization. *Control Engineering Practice*, 93, 104167. <https://doi.org/10.1016/j.conengprac.2019.104167>
- Vardar, S., Demirel, B., & Onay, T. T. (2022). Impacts of coal-fired power plants for energy generation on environment and future implications of energy policy for Turkey. *Environmental Science and Pollution Research*, 29(27), 40302–40318. <https://doi.org/10.1007/s11356-022-19786-8>
- Vo, N. D., Oh, D. H., Hong, S. H., Oh, M., & Lee, C. H. (2019). Combined approach using mathematical modeling and artificial neural network for chemical industries: Steam methane reformer. *Applied Energy*, 255, 113809. <https://doi.org/10.1016/j.apenergy.2019.113809>
- Vujić, B. B., Marčeta, U. V., Popescu, F., & Tot, B. I. (2019). Air quality monitoring and modeling near coal fired power plant. *Thermal Science*, 23(6 Part B), 4055–4065. <https://doi.org/10.2298/TSCI190611385V>
- Wu, Z., Zhang, Y., & Dong, Z. (2023). Prediction of NO<sub>x</sub> emission concentration from coal-fired power plant based on joint knowledge and data driven. *Energy*, 271, 127044. <https://doi.org/10.1016/j.energy.2023.127044>
- Xie, J., Niu, X. D., He, K. Q., Shi, M. D., Yu, S. J., Yuan, C. G., & Liu, J. F. (2021). Arsenic and selenium distribution and speciation in coal and coal combustion by-products from coal-fired power plants. *Fuel*, 292, 120228. [https://doi.org/10.1016/S1872-5813\(21\)60193-7](https://doi.org/10.1016/S1872-5813(21)60193-7)
- Xu, J., Zhu, F., Ge, X., Li, H., Zhao, X., Tian, W., et al. (2022). Research progress on volatile organic compounds emissions from coal-fired power plants. *Current Pollution Reports*, 8(3), 303–314. <https://doi.org/10.1007/s40726-022-00225-8>
- Xu, T., Li, T., Long, J., Zhao, L., & Du, W. (2023). Data-driven multi-period modeling and optimization for the industrial steam system of large-scale refineries. *Chemical Engineering Science*, 119112. <https://doi.org/10.1016/j.ces.2023.119112>
- Yagmur, A., Kayakus, M., & Terzioglu, M. (2023). Predicting renewable energy production by machine learning methods: The case of Turkey. *Environmental Progress & Sustainable Energy*, e14077. <https://doi.org/10.1002/ep.14077>
- Yuan, Z., Meng, L., Gu, X., Bai, Y., Cui, H., & Jiang, C. (2021). Prediction of NO<sub>x</sub> emissions for coal-fired power plants with stacked-generalization ensemble method. *Fuel*, 289, 119748. <https://doi.org/10.1016/j.fuel.2020.119748>
- Zhou, S., He, H., Zhang, L., Zhao, W., & Wang, F. (2023). A data-driven method to monitor carbon dioxide emissions of coal-fired power plants. *Energies*, 16(4), 1646. <https://doi.org/10.3390/en16041646>
- Zhu, L., Husny, Z. J. B. M., Samsudin, N. A., Xu, H., & Han, C. (2023). Deep learning method for minimizing water pollution and air pollution in urban environment. *Urban Climate*, 49, 101486. <https://doi.org/10.1016/j.uclim.2023.101486>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.