BULLETIN OF THE POLISH ACADEMY OF SCIENCES
TECHNICAL SCIENCES, Vol. 72(6), 2024, Article number: e151956
DOI: 10.24425/bpasts.2024.151956

ARTIFICIAL AND COMPUTATIONAL INTELLIGENCE

Prediction and comparative analysis of emissions from gas turbines using random search optimization and different machine learning-based algorithms

Emrah ASLAN*

Dicle University, Silvan Vocational School, Diyarbakır, Turkey

Abstract. Gas turbines are widely used for power generation globally, and their greenhouse gas emissions have increasingly drawn public attention. Compliance with environmental regulations necessitates sophisticated emission measurement techniques and tools. Traditional sensors used for monitoring emission gases can provide inaccurate data due to malfunction or miscalibration. Accurate estimation of gas turbine emissions, such as particulate matter, carbon monoxide, and nitrogen oxides, is crucial for assessing the environmental impact of industrial activities and power generation. This study used five different machine learning models to predict emissions from gas turbines, including AdaBoost, XGBoost, k-nearest neighbour, and linear and random forest models. Random search optimization was used to set the regression parameters. The findings indicate that the AdaBoost regressor model provides superior prediction accuracy for emissions compared to other models, with an accuracy of 99.97% and a mean squared error of 2.17 on training data. This research offers a practical modelling approach for forecasting gas turbine emissions, contributing to the reduction of air pollution in industrial applications.

Keywords: emission; gas turbines; efficiency; machine learning; random search optimization.

1. INTRODUCTION

Gas turbines play an important role in various industries, providing a reliable power source for electricity generation and industrial processes. However, the environmental impacts of gas turbine operations, especially in terms of emissions, are receiving increasing attention [1]. Gas turbines are widely used for power generation and industrial processes and emit pollutants such as carbon monoxide (CO), nitrogen oxides (NO_x), and particulate matter during combustion. Accurate estimation of these emissions is essential for assessing environmental impacts on air quality, ecosystems, and human health. It provides valuable information on the contribution of gas turbine operations to overall air pollution [2]. Governments and environmental organizations worldwide have established strict regulations to control and limit air pollutants emitted from industrial facilities, including gas turbines. Emission estimation serves as a critical tool to ensure that industries comply with these regulations. Accurate forecasts enable organizations to proactively take measures to meet or exceed emission standards, avoid legal repercussions, and contribute to a cleaner environment [3].

By understanding the factors that influence emissions, engineers and environmentalists can develop targeted approaches to reduce pollutant levels. This may involve optimizing combustion processes, adjusting operating conditions, or adopting advanced technologies for emission reduction. Accurate emission estimation helps policymakers and health authorities assess potential health risks, enabling them to take preventive measures and reduce the impact on communities living near industrial facilities [4].

Traditionally, gas turbine emission estimation has been based on empirical models and simplified assumptions. While these simplified approaches provide a basic understanding of emissions, they are limited in their ability to capture the complex relationships between the various factors affecting gas turbine operations and pollutant release [5]. While these methods serve as initial benchmarks for regulatory compliance, they lack the precision and adaptability required for the dynamic and complex nature of modern industrial processes [6].

Despite the increasing need for sustainable energy sources, precise emissions forecasting from gas turbines is still a vital field of study and development. The integration of advanced modelling techniques, such as machine learning, offers promising opportunities to further improve the precision of emission estimation. Machine learning algorithms are characterized by their ability to process large and diverse datasets [7]. These algorithms can identify patterns and trends in data, leading to more accurate predictions. Regression models are among the machine learning algorithms that have proved to be effective for predicting gas turbine emissions. These models utilize historical emission data and a broader set of input parameters such as turbine operating conditions, fuel composition, and environmental factors. Regression models can capture complex relationships using statistical techniques, enabling more nuanced and accurate prediction of emissions [8].

Gas turbine emission estimation faces several challenges that affect the accuracy and reliability of estimates. One key challenge is the availability and quality of data. Missing or inac-

*e-mail: emrah.aslan@dicle.edu.tr

Manuscript submitted 2024-08-09, revised 2024-08-26, initially accepted for publication 2024-09-03, published in November 2024.

curate data can lead to biases in regression models and hinder their effectiveness. Furthermore, because gas turbine operations can vary based on parameters like load demand, fuel type, and ambient conditions, the dynamic nature of industrial processes creates additional complications. This variability poses a challenge in developing models that can robustly address diverse and changing operational scenarios. Another critical challenge is the interpretability of the models. Understanding how different input parameters contribute to emissions is essential for informed decision-making and targeted emission reduction strategies [9].

Future directions in gas turbine emission estimation include addressing current challenges and advancing the field through innovative approaches. One potential direction is the integration of advanced sensing technologies and real-time data streams. Another avenue for improvement is the exploration of hybrid models that combine regression techniques with artificial intelligence such as machine learning and deep learning. These hybrid models can utilize the strengths of both approaches, leveraging the interpretability of regression models and the complex pattern recognition capabilities of AI techniques. Going forward, ongoing research and collaboration between industry stakeholders, researchers, and regulatory authorities will be crucial in improving and applying regression models for gas turbine emission prediction [10, 11].

In this study, regression algorithms are used to evaluate their performance in gas turbine emission prediction. Linear, k-nearest neighbours, XGBoost, random forest, and AdaBoost regressor algorithms were used for learning. To comprehensively explore the potential of the model, an extensive hyperparameter search was carried out that affects the variables that define the regression algorithms and the optimization. Random search optimization was used for optimization. The aim is to make better predictions for measuring emissions from gas turbines using machine learning and deep learning tools.

In Section 1 of this study, general information about the problem is given. Section 2 contains literature studies for emission estimation in gas turbines. Section 3 contains the materials and methods required for the applications, Section 4 contains the studies and discussion. Section 5 presents the conclusions and future work.

2. LITERATURE REVIEW

Research on gas turbine emissions has been an important focus in the field of environmental engineering and energy studies. This chapter reviews some of the studies that have contributed to a better understanding of the environmental impact by providing valuable insights into the mechanisms affecting gas turbine emissions.

Starik *et al.* present a comparative analysis of the combustion characteristics and emissions of a gas turbine engine using various alternative fuels such as Fischer-Tropsch Synthetic Paraffinic Kerosene (FT-SPK), cryogenic methane, bioethanol, biomethanol, biobutanol, dimethyl ether, biodiesel, and the conventional aviation kerosene Jet-A. The analysis reveals that the use of alternative fuels generally increases H_2O emissions, leading to higher water vapor supersaturation and potential impacts on contrail and cirrus cloud formation. The study also reveals that there are differences in emissions of CO_2, NO_x, and N-containing species and that different alternative fuels show different effects on these emissions compared to kerosene [12]. In Lebedev *et al.* a reactor model that predicts pollutant production in a diffusion-mode combustor was constructed using methane and kerosene as fuel, and it was validated against NOx emissions from a model aviation engine combustor based on three-dimensional computational fluid dynamics (CFD) simulation. The research emphasized how heavily the applied reaction mechanism influences the expected NOx emission index and the model that was created demonstrated agreement with experimental data for combustors powered by kerosene and methane. Based on the power setting of an aviation gas turbine engine, the model has also been used to estimate emissions of different species, such as sulphur compounds, carbon oxide, and unburned hydrocarbons [13]. Taha *et al.* emphasize the importance of efficiently monitoring operational parameters and environmental variables that affect gas turbine performance to reduce maintenance costs, component defects, and manpower expenses. They emphasize that while traditional sensors can miss faults in the harsh gas turbine environment, machine learning-based monitoring systems, especially those using deep learning tools, offer a cost-effective and accurate solution to overcome these challenges and improve overall efficiency [14]. Kaya *et al.* presented new data and introduced a comparative Portable Emissions Measurement System (PEMS), focusing on the estimation of CO and NO_x emissions from gas turbines. The research aimed to improve the accuracy of emission estimates through the application of this new PEMS methodology [15]. Egware and Kwasi-Effah's research introduces a new empirical model specifically designed to estimate carbon dioxide emissions from gas turbine power plants. The research focuses on developing a model that improves the accuracy of estimating CO_2 emissions and thus contributes to understanding and managing the environmental impacts associated with gas turbine operations. The new empirical model proposed in the study provides a valuable tool for the assessment and mitigation of carbon dioxide emissions in the context of gas turbine power plants [16]. Lazzaretto and Toffolo present a prediction of both the performance and emissions of a two-shaft gas turbine based on experimental data. The model is built in MATLAB/Simulink environment. This work contributes to the understanding of gas turbine behaviour and helps to develop strategies to optimize performance and reduce emissions in two-shaft gas turbine systems [17]. Coelho *et al.* focus on estimating carbon oxide (CO) and nitrous oxide (NO_x) emissions from a gas turbine using a dataset of estimated emission monitoring systems. Innovative feature generation methods are introduced, and various regression models are evaluated after feature ranking and hyperparameter tuning. The results highlight the effectiveness of the deep forest regression (DFR) model in predicting CO and NO_x emissions and underline the impact of feature engineering and hyperparameter tuning on the overall predictive capacity of the models [18]. Faqih *et al.* introduce a semi-supervised technique for predicting the appropriate operating interval of dry-low emission (DLE) gas turbines to avoid frequent start-ups and guide efficient load planning. The hybrid model combines extreme gradient boosting and k-means algorithms using real plant data and achieves high accuracy in predicting combustion temperature, nitrous oxide, and carbon monoxide concentrations. The proposed technique defines safe operating zones for DLE gas turbines, with a typical operating range of 744.68°C–829.64°C, provides a preventive maintenance strategy to reduce tripping problems, and provides valuable information to improve control strategies in power generation areas [19]. Zhao *et al.* present a prediction model for NO_x emissions in heavy-duty gas turbine combustors using moderate and intense low oxygen dilution combustion. Using the optimum gap-filling design, the research determines the optimum combination of gas and air temperatures and mass flows, achieving a minimum NO_x emission of 24.11 mg/m^3 [20].

The motivation behind this work stems from the growing global concern about greenhouse gas emissions and their impact on climate change and public health.

This study makes a significant contribution to the literature on gas turbine emission prediction using advanced machine learning techniques, including regression algorithms such as AdaBoost, XGBoost, k-nearest neighbours, linear regression, and random forest models. For the regression algorithms to give the best results, random search optimization was used to adjust the parameters for each regression. The primary contribution of the research lies in the comprehensive evaluation of the performance of these models in predicting emissions of critical pollutants such as nitrogen oxides and carbon monoxide from gas turbines. By achieving a very high prediction accuracy with the AdaBoost regressor model (99.97% on training data and 92.04% on test data), this study not only demonstrates the potential of machine learning in environmental monitoring but also sets a new benchmark for predictive modelling in this field.

Furthermore, the practical implications of the research are important as the proposed models can serve as valuable tools for industries and policymakers aiming to reduce air pollution from gas turbine operations.

In summary, this study provides a robust framework for accurate prediction of gas turbine emissions using machine learning regression models. It highlights the potential for advanced algorithms such as AdaBoost to outperform conventional methods, thus adding valuable insights to the emission estimation literature. While the study is limited by data quality and model interpretability, it lays a strong foundation for future research aimed at overcoming these challenges and increasing the applicability of machine learning in environmental monitoring and industrial emission control.

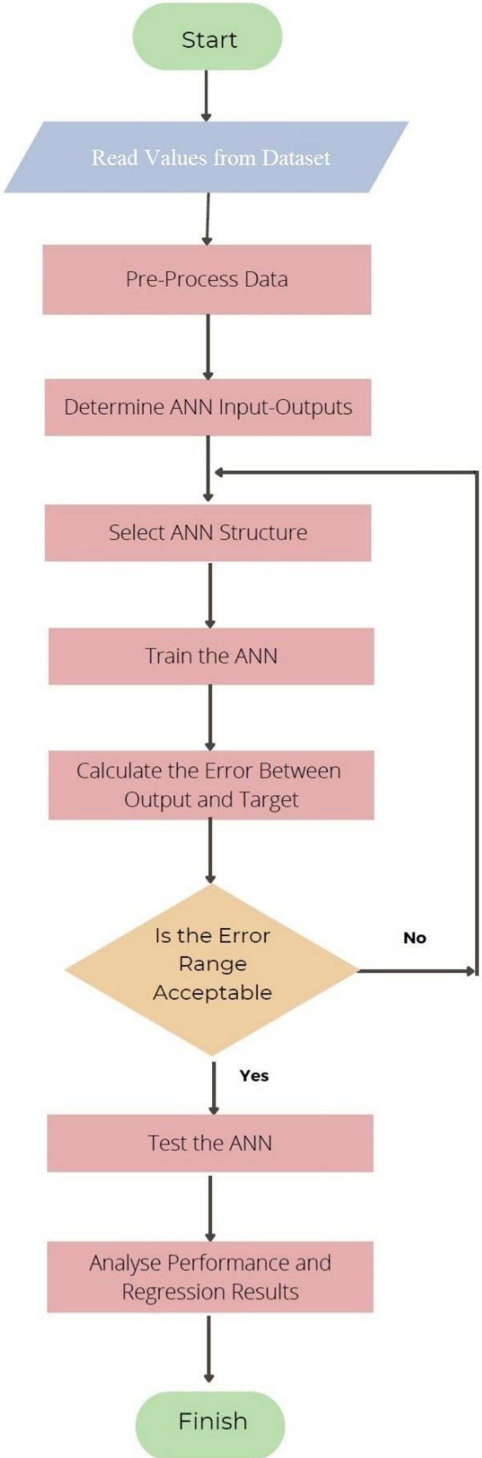

Fig. 1. Flow diagram

3. MATERIAL AND METHOD

Artificial neural networks such as linear, k-nearest neighbours, XGBoost, random forest, and AdaBoost regressor were used to detect emissions from gas turbines. Figure 1 shows a flow diagram showing the model development procedure. Initially, 80% of the DLE gas turbine data of a real plant is split into training and 20% into testing sets. Preprocessing of the training data is done, along with correlation testing, to find significant characteristics for the model input.

3.1. Dataset

Exhaust Emission Dataset (https://www.kaggle.com/datasets/muniryadi/gasturbine-co-and-nox-emission-data. Accessed January 24, 2024) contains hourly average sensor readings of eleven variables, including nine input and two target variables, collected over five years, resulting in a total of 36 733 samples [21]. Table 1 lists the nine input measures along with their names, acronyms, and basic statistics. These measurements are categorized as ambient variables (such as temperature, humidity, and pressure) and process parameters (such as turbine energy efficiency, and air filter differential pressure). The gas emissions measured by the gas analyser are NO_x and CO concentration.

Table 1
Basic statistical information about the data used in the dataset [15]

Variable	Abb	Unit	Min	Max	Mean
Ambient temperature	AT	°C	–6.23	37.10	17.71
Ambient pressure	AP	mbar	985.8	1036	1013
Ambient humidity	AH	(%)	24.08	100.2	77.87
Air filter difference pressure	AFDF	mbar	2.09	7.61	3.93
Gas turbine exhaust pressure	GTEP	mbar	17.70	40.72	25.56
Turbine inlet temperature	TIT	°C	1000.8	1100.8	1081.43
Turbine after temperature	TAT	°C	511.04	550.61	546.16
Compressor discharge pressure	CDP	mbar	9.85	15.16	12.06
Turbine energy yield	TEY	MWH	100.	179.5	13.51
Carbon monoxide	CO	mg/m^3	0.0	44.10	2.37
Nitrogen oxides	NO_x	mg/m^3	25.90	119.9	65.29

3.2. Data preprocessing

The correlation matrix result is shown in Fig. 2 to help visualize the link between the dataset parameters. The relationship is then examined using the pairwise correlation between the input and target parameters as well as the correlation of each input parameter. This correlation matrix shows the relationships between various variables measured in the gas turbine exhaust emission dataset considered in the study. The matrix contains the correlation coefficients between pairs of variables ranging from −1 to +1. When the correlation coefficient is +1, there is a positive and strong relationship between the two variables, while −1 indicates a negative and strong relationship. A correlation coefficient of 0 indicates that there is no linear relationship between the variables. In this analysis, a strong positive correlation is observed, especially between TEY and CDP (0.99). These strong correlations indicate that as the turbine energy efficiency increases, the compressor discharge pressure also increases. Conversely, a negative correlation is observed between NO_x and TAT (−0.70), indicating that higher final temperatures may be associated with lower NO_x emissions.

Overall, the correlation matrix highlights critical interdependencies between variables and provides vital insights for the development of predictive models for emission estimation. Understanding these relationships is essential for improving the accuracy of neural network models in predicting emissions and thus helping to reduce environmental impacts from gas turbine operations.

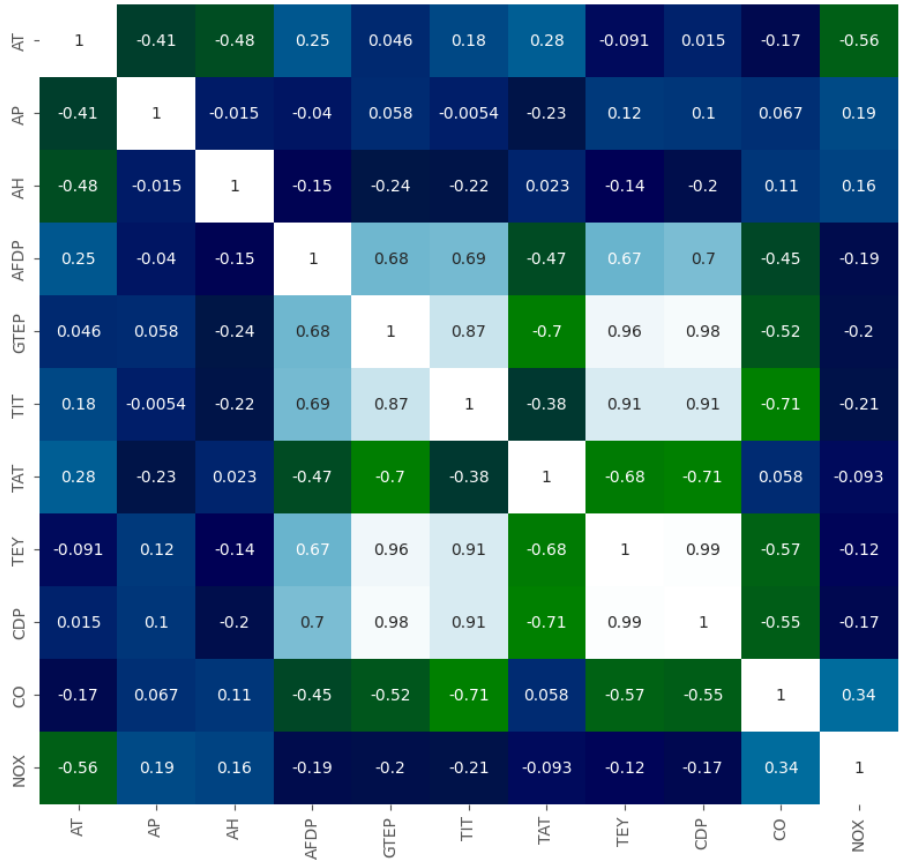

Fig. 2. Correlation matrix

3.3. Linear regression

This regression is a basic model used in statistics and machine learning [22]. This model is used to model the relationship of a dependent variable with one or more independent variables.

3.4. K-nearest neighbours

It is a basic classification and regression algorithm. This algorithm is used to predict a data point by the label or value of k-nearest neighbours around it. The working principle of k-nn is quite simple. The data point calculates the distances to all other data points according to the similarity measure. The k closest data points are selected.

3.5. XGBoost

It is a learning algorithm that has recently achieved great success in machine learning competitions and industrial applications. It is based on Gradient Boosting methods but has been improved, especially in terms of scalability, speed, and accuracy. The basic working principle of XGBoost is to combine many weak predictors into a strong predictor. The model trains these predictors sequentially, focusing on correcting the errors of previous predictions. This is a gradient descent approach to minimize the errors in the data set used to train the next predictor.

3.6. Random forest

It is an ensemble learning algorithm based on decision trees. It is essentially an ensemble method that combines multiple decision trees. Random forest trains each tree differently and then aggregates their predictions to create a more robust and balanced predictor.

3.7. AdaBoost

AdaBoost is a variation of the adaptive boosting algorithm and is an ensemble method for regression problems. AdaBoost combines weak predictors to form a strong predictor.

3.8. Random search optimization

Random search optimization is a highly effective method for tuning the hyperparameters of machine learning models. Instead of systematically evaluating every possible combination as in grid search, random search evaluates a set number of randomly selected points within the parameter space. This method greatly reduces the computational burden, especially when dealing with models that have large and intricate hyperparameter spaces, and often yields good results in a shorter amount of time. The randomness inherent in this approach also enhances the likelihood of exploring a broader region of the hyperparameter space, thereby increasing the chances of finding the global optimum.

Random search is particularly advantageous when working with high-dimensional datasets and complex models. This efficiency is particularly valuable in real-world applications, such as power system fault detection using machine learning, where the flexibility and computational efficiency of random search make it a powerful tool for model optimization. By preventing overfitting and enhancing performance metrics, this method plays a crucial role in improving the reliability and accuracy of machine learning models.

3.9. Gas turbine data acquisition

The system flow diagram and the measurement sensors for a typical dry low-emission (DLE) gas turbine are depicted in Fig. 3 [19]. Compressor components are subject to mechanical failure and are not guaranteed to burn out as components of a combustion engine. The operation of the gas turbine is dictated by the load demand, as noted in 1. The load being driven determines the power output by ensuring that the rotational speed of the mechanical turbine is maintained at a specific speed, referred to as 2. The ambient air temperature, which is 3, has a significant impact on power production since rising temperatures reduce air density, which in turn lowers mass flow through

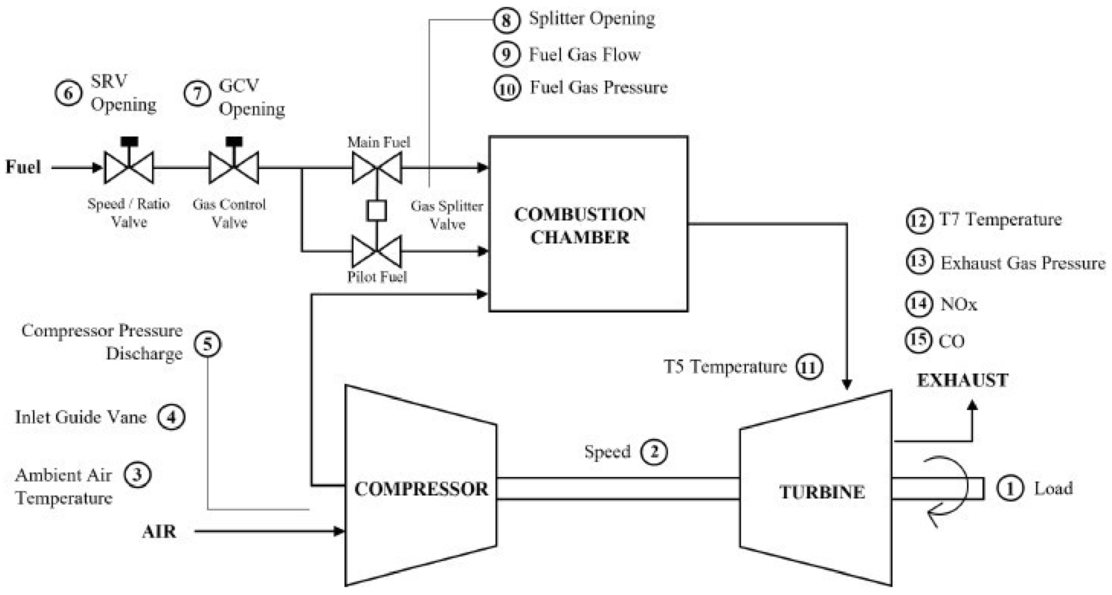

Fig. 3. Flow diagram of the DLE gas turbine

the turbine and lowers power output. As a result, keeping an eye on the outside air temperature is essential to maintaining the dependability and efficiency of a gas turbine. The compressor receives air through the inlet guide vanes (IGV), shown at 4. The air is then compressed, and the discharge pressure is measured as the compressor discharge pressure (CDP), shown at 5. The compressed air is then mixed with fuel in the combustion chamber. The fuel entering the combustion chamber passes through the stop ratio valve (SRV) indicated at 6, which maintains constant gas pressure and regulates the pressure drop. The gas control valve (GCV) at 7 controls the fuel flow necessary for the combustion process. Given that the DLE combustor requires separate sections for main fuel and pilot fuel, a splitter valve, designated at 8, manages the division of main and pilot fuel before they enter their respective chambers. At positions 9 and 10, the fuel flow and pressure are recorded. Because of the harsh circumstances and temperature differences inside the chamber, monitoring the combustion temperature is difficult. The temperature of the gas leaving the chamber has a direct bearing on the firing temperature. As direct measurement of temperature within the combustion chamber is not feasible due to sensor limitations, the temperature is instead measured at the exhaust point, labelled as T5. Consequently, in this study, T5 is considered the combustion temperature and will be employed for estimating the operating range. In the turbine exhaust section, temperature and pressure are monitored at 12 and 13, respectively. The process generates NO_x and CO emissions, which are measured at 14 and 15, respectively [19].

4. RESULTS AND DISCUSSION

In this study, the performance of regression algorithms used in gas turbine emission estimation was evaluated. Linear, k-nearest neighbours, XGBoost, random forest, and AdaBoost regressor algorithms were trained. To examine the potential of the model in detail, a comprehensive study was conducted on the variables and hyperparameters that define the regression algorithms.

The parameters employed in the training with linear regression, as utilized in the study, are listed in Table 2. Following the training process, linear regression achieved a training performance of 52.17% and a test performance of 51.37% in predicting emissions in gas turbines. The RMSE was calculated as 8.13, the MAE as 5.90, and the R^2score as 51.37%. A graph illustrating the gas turbine emissions predicted by the linear regression algorithm compared to the actual emission values is presented in Fig. 4. Additionally, the actual values from the dataset and the values predicted by the linear regression algorithm are provided in Table 7.

Table 2
Linear regression parameters

Parameter	Value
Fit_intercept	true
copy_X	true
n_job	none
positive	true
Max_iter	10
Random_state	10

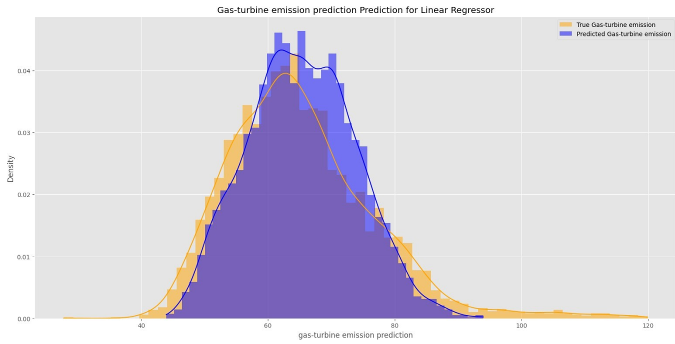

Fig. 4. Graph of actual value and predicted value with linear regression

The parameters used in training with random forest regression are given in Table 3.

Table 3
Random forest regression parameters

Parameter	Value
n_estimators	1000
criterion	log_loss
Max_depth	none
Max_iter	10
Random_state	10

As a result of the training of random forest regression, 98.05% training performance and 86.72% test performance were obtained in predicting emissions in gas turbines. RMSE is 4.24, MAE is 2.69, R^2 Score is 86.72. The graph showing the gas turbine emissions estimated by the random forest regression algorithm and the prediction of actual emission values is given in Fig. 5. Table 7 shows the values that the random forest regression algorithm predicted and the actual values in the data set.

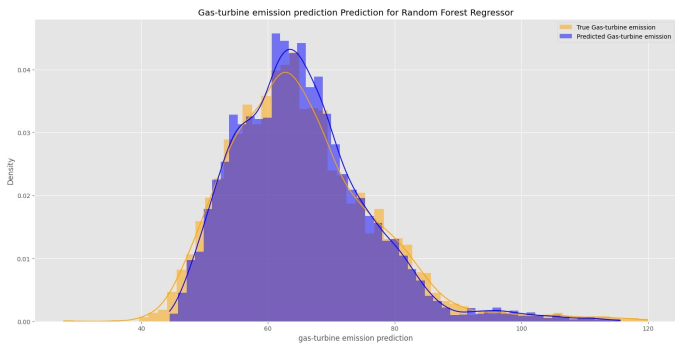

Fig. 5. Graph of actual value and predicted value with random forest regression

The parameters used for training with the k-nearest neighbour (k-nn) regression algorithm are listed in Table 4. After training, the k-nn regression achieved a training performance of 84.15% and a test performance of 80.42% in predicting emissions in gas turbines. The RMSE was 5.16, the MAE was 3.41, and the R^2score was 80.42%. A graph depicting the gas turbine emissions predicted by the k-nn regression algorithm alongside the actual emission values is shown in Fig. 6. Additionally, Table 7 provides a comparison between the actual values in the dataset and the predictions made by the k-nn regression algorithm.

Table 4
k-nn regression parameters

Parameter	Value
Probability	true
Epsilon	0.1
Degree	3
Max_iter	10
Random_state	10

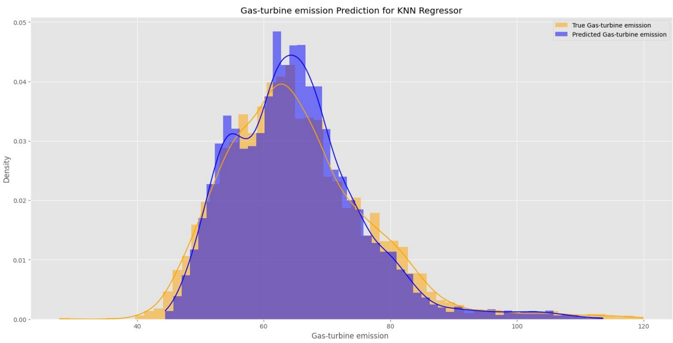

Fig. 6. Graph of actual value and predicted value with k-nn regression

The parameters used for training with the XGBoost regression algorithm are presented in Table 5. The training process with XGBoost regression resulted in a training performance of 91.81% and a test performance of 84.51% in predicting emissions in gas turbines. RMSE was 4.58, the MAE was 3.09, and the R^2score was 84.54%. Figure 7 displays a graph comparing the gas turbine emissions predicted by the XGBoost regression algorithm with the actual emission values. Additionally, Table 7 provides the actual values from the dataset alongside the predictions made by the XGBoost regression algorithm.

Table 5
XGBoost regression parameters

Parameter	Values
booster	gbtree
verbosity	1
eta	0.3
learning_rate	0.99
max_dept	6
n_estimators	50

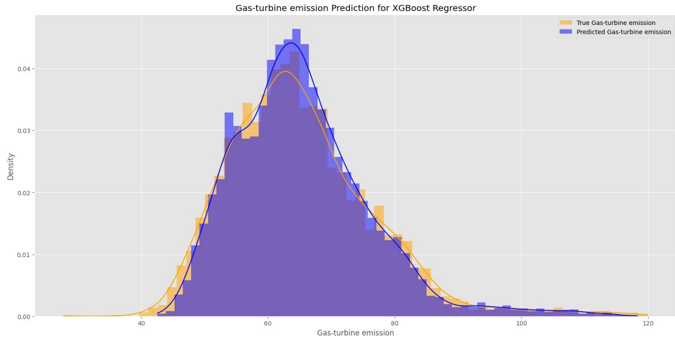

Fig. 7. Graph of actual value and predicted value with XGBoost regression

The hyperparameters of the AdaBoost regressor are presented in Table 6. AdaBoost regression training produced a test performance of 92.04% and a training performance of 99.97% for emissions prediction in gas turbines. RMSE is 3.95, MAE is 2.17 and R^2 Score is 92.04. The graph showing the gas turbine emission predicted by the AdaBoost regression algorithm and the prediction of actual emission values is given in Fig. 8. Table 7 shows the values that the AdaBoost regression method predicted and the actual values in the data set.

Table 6
AdaBoost regression parameters

Parameter	Values
Base_estimator	dtree
learning_rate	0.99
loss	linear
n_estimators	50
Max_iter	10
Random_state	10

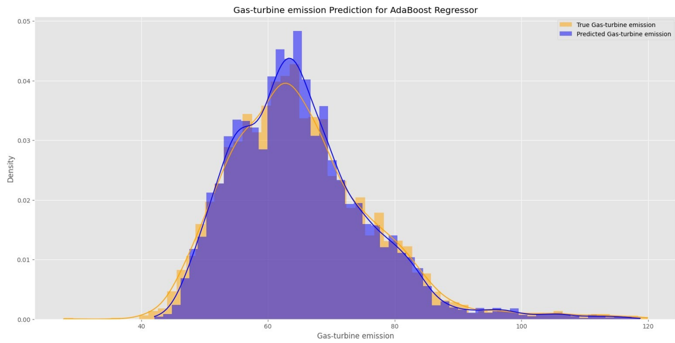

Fig. 8. Graph of actual value and predicted value with AdaBoost regression

Table 7
Actual and predicted temperature

ID	True gas turbine emission	Predicted gas turbine emission				
		Linear	RF	K-NN	XGBoost	AdaBoost
1	48.876	52.547	49.924	50.239	48.662	49.293
2	68.584	69.840	69.120	69.245	70.115	70.781
3	55.400	56.953	55.547	53.564	55.997	55.008
4	52.123	54.989	57.015	57.710	55.233	51.225
5	57.123	60.928	56.658	50.371	55.096	58.994
6	69.161	71.102	74.912	78.651	73.354	71.165
7	80.834	74.863	80.999	80.939	73.750	80.544
8	66.639	59.598	63.958	62.099	63.642	68.739
9	63.983	66.383	65.198	68.761	65.319	67.533
10	67.361	71.133	65.605	70.462	67.958	66.716

Table 7 compares the actual values of gas turbine emissions with the predicted values by five different regression algorithms. The deviations between the actual emission values and the predicted values are a critical indicator for understanding the accuracy and reliability of the model.

As seen in the table, linear and random forest algorithms generally show higher deviations in their predictions. The k-nn, XGBoost, and AdaBoost algorithms also show deviations in some cases, but generally, their predicted values are closer to the true values. In particular, the AdaBoost algorithm was more consistent than the other algorithms in many cases. For example, in ID 5, the true value was 57.123, while AdaBoost predicted 58.994, which is a lower deviation compared to the other algorithms. This shows that the prediction algorithms have different levels of performance. It is observed that the AdaBoost algorithm performs the best overall, while the linear model produces the largest deviations. These results show that the AdaBoost algorithm outperforms the other algorithms in gas turbine emission forecasting and is therefore a reliable tool for gas turbine emission forecasting.

Table 8 shows the metrics used to evaluate the scaling success and performance of the regression algorithms. Table 8 presents various metrics comparing the performance of five different regression algorithms on training and test data. These metrics evaluate the ability of the algorithms to predict gas turbine emissions and are used to determine which one performs the best. Table 8 shows that the linear regression performance is quite low compared to the other algorithms. With a training accuracy of 52.17% and a test accuracy of 51.37%, linear regression has the lowest accuracy rates compared to other models. Furthermore, with an RMS Score of 8.13 and an MAE Score of 5.90, linear regression has the largest margin of error in emission estimates. The R^2 score is also 51.37%, indicating that the model has a low capacity to explain the data. The random forest algorithm performed very well with a training accuracy of 98.05%. However, the testing accuracy drops to 86.72%, indicating that the model may be slightly overfitting. The RMS score is 4.24 and the MAE score is 2.69, indicating relatively low errors in predictions. The R^2 score is 86.72%, indicating that the model explains the data quite well. The k-nn algorithm also performed reasonably well, but with a test accuracy of 80.42%, it lags behind the random forest and AdaBoost algorithms. The RMS score of k-nn is 5.16 and the MAE score is 3.41, indicating that this model may also have significant errors in prediction in some cases. The R^2 score is 80.42%, which can reasonably explain the data.

Table 8
Comparison of regression algorithms

Regressor algorithm	RMS score	MAE score	R^2 score
Linear	8.13	5.90	51.37
Random Forest	4.24	2.69	86.72
K-NN	5.16	3.41	80.42
XGBoost	4.58	3.09	84.51
AdaBoost	3.95	2.17	92.04

The XGBoost algorithm performed very well with 91.81% training accuracy and 84.51% test accuracy. The RMS score of 4.58 and MAE score of 3.09 indicate that the predictions of this model have a relatively low margin of error. XGBoost emerges as a strong option for model performance. The AdaBoost algorithm performs the best compared to the other algorithms, offering the highest test accuracy (92.04%). The accuracy is 99.97%, indicating that the ability of this model to learn the data is very high. The RMS score of 3.95 and MAE score of 2.17 show that AdaBoost has the lowest margin of error and can produce the most accurate predictions. The R^2 score is 92.04%, indicating that the model has a very high capacity to explain the data. Table 8 clearly shows that the AdaBoost algorithm is superior to the other algorithms in predicting gas turbine emissions and performs the best in this area. Random forest and XGBoost also stand out as strong alternatives, but the overall performance of AdaBoost makes it the most suitable model for this study. However, it should be noted that each algorithm may perform differently under certain datasets and conditions.

One of the limitations of this work is the dependence on the quality and completeness of the dataset. As with most machine learning applications, the accuracy of the models is highly dependent on the quality of the input data. Missing or inaccurate data points can lead to biases, potentially affecting the reliability of the predictions. Furthermore, while the dataset used in this study is comprehensive, it may not fully capture the variability of real-world gas turbine operations across different geographic locations, fuel types, and operating conditions. This limitation may affect the generalizability of the models to other contexts. Another limitation is the complexity and interpretability, especially of well-performing machine learning models such as AdaBoost and XGBoost.

The limitations of the study can be improved by developing hybrid models that combine the strengths of traditional regression techniques with the advanced pattern recognition capabilities of machine learning and deep learning algorithms. In addition, future research could focus on expanding the dataset to include a wider range of operational scenarios, fuel compositions, and environmental conditions.

5. CONCLUSIONS

This study evaluated and compared the performance of different regression algorithms for predicting gas turbine emissions. Although gas turbines are a widely used technology in industrial power generation, their emissions are a major concern in terms of compliance with environmental regulations and protection of public health. Accurately estimating these emissions is therefore a critical requirement for sustainable energy production and environmental protection efforts. Among the five regression algorithms used in the study, the AdaBoost regressor model was the most successful in predicting gas turbine emissions, offering the highest accuracy and the lowest margin of error. The AdaBoost model outperformed the other models with a training accuracy of 99.97% and a test accuracy of 92.04%. Moreover, the low RMS and MAE scores of this model indicate that AdaBoost not only provides high accuracy but also minimizes the margin of error in predictions. Random forest and XGBoost algorithms also showed remarkable performances and achieved results close to AdaBoost. The results of this study prove that machine learning methods, especially robust models such as AdaBoost, are effective tools for emission estimation in industrial processes. By offering higher accuracy and lower margin of error compared to conventional methods, these models can make a significant contribution to assessing environmental impacts and ensuring compliance with regulatory requirements. This provides important support to efforts to reduce environmental pollution by enabling industrial facilities to more accurately predict their emissions and proactively take necessary measures. However, this study also has some limitations. For example, the size and diversity of the dataset used are important factors that can affect model performance. Future studies should test the general validity of these models by using larger and more diverse data sets and evaluate their performance under different operational conditions. Furthermore, the interpretability of the models needs to be emphasized, because in industrial applications, not only high accuracy but also comprehensibility of the model decision mechanisms is of great importance.

REFERENCES

[1] R.J. Allam *et al.*, "High efficiency and low cost of electricity generation from fossil fuels while eliminating atmospheric emissions, including carbon dioxide," *Energy Procedia*, vol. 37, pp. 1135–1149, 2013, doi: 10.1016/j.egypro.2013.05.211.

[2] T. Sebbagh, "Modeling, analysis, and techno-economic assessment of a clean power conversion system with green hydrogen production," *Bull. Pol. Acad. Sci. Tech. Sci.*, vol. 72, no. 4, p. e150115, 2024, doi: 10.24425/bpasts.2024.150115.

[3] M.C. Chiong *et al.*, "Liquid biofuels production and emissions performance in gas turbines: A review," *Energy Conv. Manag.*, vol. 173, pp. 640–658, 2018, doi: 10.1016/j.enconman.2018.0.

[4] N. Guellouh, Z. Szamosi, and Z. Siménfalvi, "Combustors with low emission levels for aero gas turbine engines," *Int. J. Eng. Manag. Sci.*, vol. 4, no. 1, pp. 503–514, 2019, doi: 10.21791/IJEMS.2019.1.62.

[5] P. Steckowicz, P. Pyrzanowski, and E. Bulut, "Development and implementation of robotized wire arc repair of gas turbine diaphram," *Bull. Pol. Acad. Sci. Tech. Sci.*, vol. 72, no. 2, p. e147920, 2024, doi: 10.24425/bpasts.2023.147920.

[6] D. Rochelle and H. Najafi, "A review of the effect of biodiesel on gas turbine emissions and performance," *Renew. Sust. Energ. Rev.*, vol. 105, pp. 129–137, 2019, doi: 10.1016/j.rser.2019.01.056.

[7] H. Mohamed, H.B. Ticha, and S. Mohamed, "Simulation of pollutant emissions from a gas-turbine combustor," *Combust. Sci. Technol.*, vol. 176, no. 5–6, pp. 819–834, 2004, doi: 10.1080/00102200490428422.

[8] D.A. Wood, "Long-term atmospheric pollutant emissions from a combined cycle gas turbine: Trend monitoring and prediction applying machine learning," *Fuel*, vol. 343, p. 127722, 2023, doi: 10.1016/j.fuel.2023.127722.

[9] Y. Liu, X. Sun, V. Sethi, D. Nalianda, Y. Li, and L. Wang, "Review of modern low emissions combustion technologies for aero gas turbine engines," *Prog. Aeosp. Sci.*, vol. 94, pp. 12–42, 2017, doi: 10.1016/j.paerosci.2017.08.001.

[10] T.A. Kuznetsova and A.A. Naborshchikov, "Optimal automatic control of NO_x emissions from combustion chambers of gas turbine aeroengine based on the Bellman method," *AIP Conf. Proc.*, vol. 2700, no. 1, p. 040013, 2023, doi: 10.1063/5.0137228.

[11] A.A. Beige and A. Mardani, "An investigation on flame structure and NO_x formation in a gas turbine model combustor using large eddy simulation," *Phys. Fluids*, vol. 35, no. 7, p. 075133, 2023, doi: 10.1063/5.0155974.

[12] A.M. Starik, A.M. Savel'ev, O.N. Favorskii, and N.S. Titova, "Analysis of emission characteristics of gas turbine engines with some alternative fuels," *Int. J. Green Energy*, vol. 15, no. 3, pp. 161–168, 2018, doi: 10.1080/15435075.2017.1324790.

[13] A.B. Lebedev, A.N. Secundov, A.M. Starik, N.S. Titova, and A.M. Schepin, "Modeling study of gas-turbine combustor emission," *Proc. Combust. Inst.*, vol. 32, no. 2, pp. 2941–2947, 2009, doi: 10.1016/j.proci.2008.05.015.

[14] S. Taha, F. Ismail, and S. Thiruchelvam, "Gas turbine performance monitoring and operation challenges: A review," *Gazi U. J. Sci.*, vol. 36, no. 1, pp. 154–171, 2023, doi: 10.35378/gujs.948875.

[15] H. Kaya, P. Tüfekci, and E. Uzun, "Predicting CO and NO_x emissions from gas turbines: Novel data and a benchmark PEMS," *Turk. J. Electr. Eng. Comput. Sci.*, vol. 27, no. 6, pp. 4783–4796, 2019, doi: 10.3906/elk-1807-87.

[16] H. Egware and C. Kwasi-Effah, "A novel empirical model for predicting the carbon dioxide emission of a gas turbine power plant," *Heliyon*, vol. 9, no. 3, p. e14645, 2023, doi: 10.1016/j.heliyon.2023.e14645.

[17] A. Lazzaretto and A. Toffolo, "Prediction of performance and emissions of a two-shaft gas turbine from experimental data," *Appl. Therm. Eng.*, vol. 28, no. 17–18, pp. 2405–2415, 2008, doi: 10.1016/j.applthermaleng.2008.01.021.

[18] L.S. Coelho, H.V. Ayala, and V.C. Mariani, "CO and NO_x emissions prediction in gas turbine using a novel modeling pipeline based on the combination of deep forest regressor and feature engineering," *Fuel*, vol. 355, p. 129366, 2024, doi: 10.1016/j.fuel.2023.129366.

[19] M. Faqih, M.B. Omar, and R. Ibrahim, "Prediction of dry-low emission gas turbine operating range from emission concentration using semi-supervised learning," *Sensors*, vol. 23, no. 8, p. 3863, 2023, doi: 10.3390/s23083863.

[20] Q. Zhao, F. Liu, A. Jiao, Q. Yang, H. Xu, and X. Liao, "Prediction model of NO_x emissions in the heavy-duty gas turbine combustor based on MILD combustion," *Energy*, vol. 282, p. 128974, 2023, doi: 10.1016/j.energy.2023.128974.

[21] M. Jadikar, "Gas-Turbine CO and NOx Emission Data," Kaggle, 2021. [Online]. Available: https://www.kaggle.com/datasets/muniryadi/gasturbine-co-and-nox-emission-data (accessed: 14.03.2024).

[22] Y. Özüpak, "Analysis and experimental verification of efficiency parameters affecting inductively coupled wireless power transfer systems," *Heliyon*, vol. 10, no. 5, p. e27420, 2024, doi: 10.1016/j.heliyon.2024.e27420.