Published by
NIGERIAN SOCIETY OF PHYSICAL SCIENCES
Available online @ https://journal.nsps.org.ng/index.php/jnsps

J. Nig. Soc. Phys. Sci. 7 (2025) 2725

Journal of the Nigerian Society of Physical Sciences

Appraising raw exhaust pollutant gases emissions from industrial generators using statistics and machine learning approaches

O. Oderinde[a,*], C. L. Mgbechidinma[b], A. O. Agbeja[c,d], A. A. Ajayi[c,e], A. O. Ogundiran[c,f], O. O. Olaide[g], O. A. Orelaja[h], C. A. Mgbechidimma[i], C. O. Ajanaku[j], K. D. Oyeyemi[k,**]

[a]*Department of Chemistry, Nile University of Nigeria, Abuja FCT 900001 Nigeria*
[b]*School of Life Sciences, The Chinese University of Hong Kong, Shatin, Hong Kong, China, & Department of Microbiology, University of Ibadan, Ibadan, Nigeria*
[c]*Department of Chemistry, Faculty of Natural and Applied Sciences, Lead City University, Ibadan, Nigeria*
[d]*Department of Sustainable Forest Management, Forestry Research Institute of Nigeria, PMB 5054, Ibadan, Nigeria*
[e]*Department of Science Laboratory Technology, School of Pure and Applied Sciences, Federal Polytechnic, PMB 50, Ilaro, Nigeria*
[f]*Department of Chemical Sciences, Faculty of Sciences, Taraba State University, PMB 1167, Jalingo, Nigeria*
[g]*Department of Chemical and Pharmaceutical Chemistry, University of South Wales, 3, Wood Road, Treforest Pontypridd, CF37 1RG, United Kingdom*
[h]*Department of Mechanical Engineering, School of Engineering, Moshood Abiola Polytechnic Abeokuta Nigeria*
[i]*Department of Computer Science, University of Ibadan, Ibadan, Nigeria*
[j]*Department of Chemistry, Landmark University, PMB 1001, Omu Aran, Nigeria*
[k]*Applied Geophysics Unit, Department of Physics, Covenant University, PMB 1023, Ota, Nigeria*

Abstract

Industrial generators, widely used for backup power generation, emit significant levels of pollutant gases such as carbon monoxide (CO), carbon dioxide (CO_2), hydrocarbons (HC), and nitrogen oxides (NOx). These emissions exacerbate air pollution and climate change, while their inhalation adversely impacts human health, leading to respiratory/cardiovascular diseases and increased mortality rates. Raw exhausts of CO, CO_2, HC, NOx, and O_2 from industrial generators were assessed using a portable analyser. Thereafter, the obtained dataset was analysed using multiple linear regression and Pearson's correlation to quantify the synergistic impact of generator characteristics, while the study equally trained 70% of the dataset using machine learning (ML) classification models. The result showed that generators' age and capacity impacted considerably on exhaust concentrations as the diesel-powered generators exhibited higher CO_2 and NOx emissions at 76.1% and 7393ppm, respectively, compared to gas-powered generators. For diesel-powered generators, there was a moderate negative correlation at -0.49142 and p-value of 0.03281 for CO and NOx. For the gas-powered generators, the correlation is statistically significant for CO and HC, while there was an inverse association between NOx and O_2. The employed ML models achieved high prediction accuracy range of 80.6–93.5 % for exhaust pollutant gases for OGEPA classification status. Based on this study, policy frameworks should be implemented up to impose stringent generator emissions standards to reduce air pollution, invest in expanding/upgrading the national electricity grid to reduce reliance, provide low-interest green loans to finance renewable energy systems, as well as access climate finance mechanisms to subsidise clean energy projects.

DOI:10.46481/jnsps.2025.2725

Keywords: Industrial emissions, Exhaust gases, Generators, Statistical modelling, Machine Learning

Article History :
Received: 02 March 2025
Received in revised form: 28 May 2025
Accepted for publication: 11 July 2025
Available online: 18 August 2025

Communicated by: B. A. Abiodun

*Corresponding author Tel. No.: +234-803-280-1872.
**Co-Corresponding author Tel. No.: +234-803-911-4029.
Email addresses: yinkaoderinde@yahoo.com (O. Oderinde), kdoyeyemi@yahoo.com (K. D. Oyeyemi)

1. Introduction

With the increasing world population, which has led to each nation's rapid rise in industrialization, the energy demand has been geometrically increasing day by day. Fossil oil ultimately serves as a primary source of fuel for this industrialization drive for the entire world's population. Fossil fuels have remained the major global source of energy, accounting for about 84% of energy demands. Even though it is widely utilised in different areas, the emissions from fossil fuel usage and combustion into the atmosphere cause major difficulties in the form of acid rain and smog, to name a few [1–3]. Exhaust emission particles from diesel combustion were classified by the International Agency for Research on Cancer as carcinogens, affecting human health in both the cardiovascular- and pulmonary systems. These occur via the triggering of numerous cell signaling pathways, causing oxidative stress, followed by inflammatory markers' release and the damage of the DNA, eventually resulting in cell death [4–6].

Air pollution promotes and exacerbates climate change while also greatly impacting human health. Children, especially, have been reported to have an elevated risk of functional and morphological consequences throughout fetal development due to air pollution. This higher risk is attributed to the accelerated rate of breathing and increased air intake per body weight [7]. Concerning outcomes include pre-term delivery, low birth weight, neurodevelopmental abnormalities, intelligent quotient (IQ) loss, pediatric malignancies, and a higher risk of adult chronic diseases. Other reported health impacts include exacerbations of respiratory disorders, impaired lung function development, and a rise in asthma occurrence. These effects are mediated by oxidative stress resulting in acute or chronic lung inflammation and injuries, endocrine disruption, as well as genetic and epigenetic pathways that occur throughout life [8–10]. Furthermore, testicular histomorphology degeneration and increased semen oxidative stress, which may ultimately affect male fertility, have been linked to exposure to fossil fuel exhaust pollution [11]. Several other published works have reported a significant association between air pollution and health due to fossil fuel usage [12–21].

Despite the huge economic potential in Africa due to its vast mineral resources and a human population of ∼1.3 billion people [22], nearly all parts of Sub-Saharan Africa (Nigeria inclusive) have the lowest per capita consumption of electricity as well as the lowest rate of electricity access, with annual per capita electricity consumption being 518 kWh, despite many countries being rich in hydropower resources, large natural gas, and coal reserves [23]. Nigeria, being the most populous black nation in the world (∼ 220 million inhabitants) [24], with vast natural resources including petroleum, natural gas, and other mineral deposits, has for decades encountered the challenges of providing stable electricity (power) to her populace, generating on-off grid performance ∼4,430.76–5,209.60 MW [25]. In fact, according to the information from the website of the Ministry of Power [26], the government body in charge of electricity generation and distribution in Nigeria, the national grid performance peaked at 5, 184.90 MW, while the off-peak generation was 4, 241.02MW as at December 22, 2024.

Due to the unstable, low electricity supply and constant shut-down of the national grid, nearly all manufacturing industries and big organisations have resorted to using generators (either diesel or gas-powered) of various capacities as their main source of power generation for continuous running of their business in order to stay afloat [27–29]. However, the environmental and health impacts of the pollutant gases exhaust emissions such as carbon monoxide CO, carbon dioxide CO_2, hydrocarbons HC, and oxides of nitrogen NOx, cannot be overlooked as these pollutants have been reported to have considerable contributions to climate change as well as delirious impact on human health [8–21]. Reports from online searches on Scopus, Web of Sciences, PubMed, and Google scholar databases revealed that there are rare reports of studies based on raw exhaust pollutants' concentration from industrial generators (diesel and gas-powered), especially in Africa.

The prediction of exhaust polluting emissions from combustion engines and industrial generators has received a lot of attention in recent years, thanks to increased environmental laws and a drive for better energy alternatives. Traditional mathematical and thermodynamic models frequently fail to adequately capture the nonlinear and dynamic nature of exhaust emissions [30–35]. As a result, machine learning (ML) methods are being used to improve prediction accuracy and model adaptability and artificial neural networks (ANNs) are amongst the most extensively used machine learning (ML) approaches in this field. Ganesan *et al.* [36] employed artificial neural networks (ANN) to forecast NOx and CO emissions from diesel generators under varied load situations, achieving a strong correlation between predicted and measured values, while Zeinalipour *et al.* [37] used multilayer perceptrons (MLPs) to model CO_2 and NOx emissions from spark ignition engines, which outperformed typical regression models. Furthermore, Random Forest (RF) and Support Vector Machine (SVM) models have been proven useful in identifying emission levels and calculating pollutant concentrations. Potts *et al.* [38] used RF to estimate PM and NOx emissions from CNG buses, highlighting its resistance to noisy input data and interpretability. Ensemble approaches, such as XGBoost, are increasingly being adopted due to their superior performance on structured emission datasets. On the afore-mentioned bases, this study presents a novel integration of statistical analysis and machine learning (ML) techniques to appraise raw exhaust pollutant gas emissions from industrial generators in Nigeria—a region where dependence on fossil-fuel-powered generators is widespread due to unreliable grid electricity. While there are very scarce research works on generator emissions especially in Africa, this research brings out novelty as it (1) focuses on raw exhaust data by offering real-time, high-fidelity insight into emission characteristics without filtration or post-combustion treatment interference, (2) addresses a data-scarce context by developing or applying predictive ML models tailored to limited and heterogeneous Nigerian industrial data, (3) combines classical statistical tools with modern machine learning to enhance prediction accuracy, as well as uncover hidden emission patterns, and support environmental risk assessment, and (4)

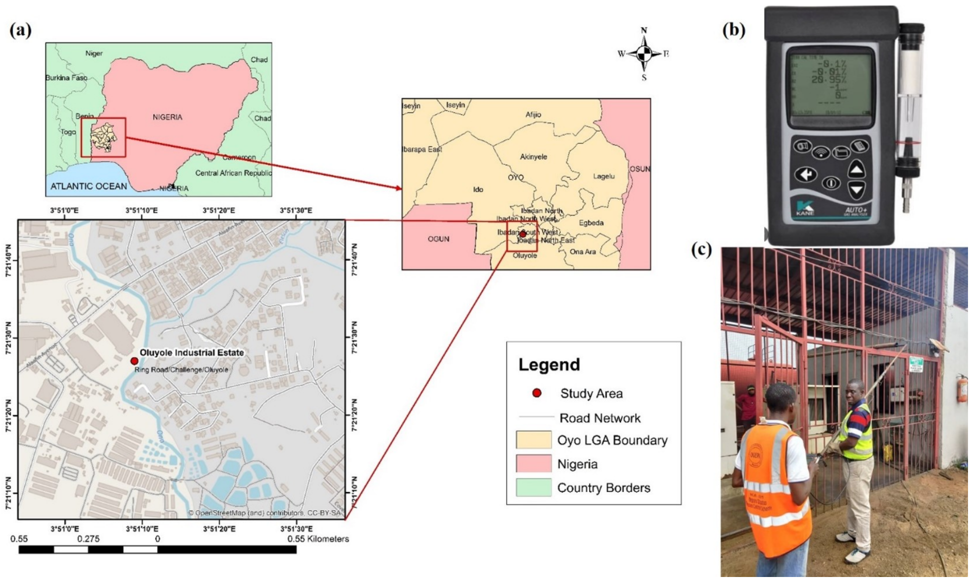

Figure 1: (a) Map showing the study area, (b) Kane gas analyzer used, (c) A scene of an industrial generator sampling.

contributes to region-specific findings that can guide policymakers, environmental agencies, and industry stakeholders in designing emission control strategies and optimising generator usage.

By merging empirical emission measurements with predictive analytics, the research provides a pioneering methodological framework that enhances environmental monitoring and regulatory planning in developing economies facing similar energy and environmental challenges. Hence, this study assesses the raw exhaust pollutant gases concentration from some selected industries using generators in a Southwest city of Ibadan in Nigeria while also using machine learning parameters to evaluate the level of accuracy and precision of the data obtained.

2. Materials and method

2.1. The sampling location

The sampling sites were some industries located within Oluyole Industrial Estate, Ibadan, Oyo State (Figure 1a).

2.2. Raw exhaust sampling and data collection measurement

Exhaust pollutant emissions from industrial generators were taken using a portable KANE 5-Gas Automotive Analyser (Model 5-2) (Figure 1b), which has been programmed to detect and determine the concentrations of CO_2, O_2, HC, CO, and NOx. Before each measurement round, the industrial generators were allowed to run while "on load" during their working hours, after which the tests were conducted. The instrument probe was inserted into the exhaust pipe of the generator end and clamped (Figure 1c). The obtained data were recorded in percentage (%) volume for CO, CO_2, and O_2, and parts per million (ppm) for NOx and HC [39, 40]. Each measurement lasted 10 min, and after each round of measurement, the analyser was calibrated to 'zero'. The testing event was done using thirty-one (31) industrial generators in June 2024, coinciding with the rainy season, while sampling events were carried out in triplicate for each generator within the sampling time.

2.3. Data analysis

2.3.1. Statistical analysis

The data obtained were subjected to statistical analysis using tools such as simple descriptive and inferential (Duncan Multiple Range Test and Principal Component Analysis) analyses Python version: 3.9.12 [MSC v.1916 64 bit (AMD64)] using seaborn as sns library, matplotlib.pyplot as plt to plot the graphical representations and determining the relationships. Also, the study utilised a multiple linear regression (MLR) approach to quantify the joint impact of industrial generator characteristics on the exhaust emission levels of CO and NOx for the generator samples tested. Equations (1) and (2) represent the MLR model generated from sampled generators for CO and NOx.

The hypothesis for this study is stated below:

Null Hypothesis: H0 = Model adequately fits the data

Alternative Hypothesis: HA= Model does not adequately fit the data

Regression Equation:

$$(\mathrm{CO}) = b_0 + b_1 G_m + b_2 E_t + b_3 A_g + b_4 C_p, \quad (1)$$

$$(\mathrm{NOx}) = b_0 + b_1 G_m + b_2 E_t + b_3 A_g + b_4 C_p, \quad (2)$$

where G_m = Generator Model; E_t = Engine fuel type; A_g = Age; C_p = Capacity; b_0 = Intercept; b_1-b_5 = Coefficients.

Furthermore, Pearson's correlation was used to assess the strength of the linear relationship between two continuous variables. The Pearson Correlation Coefficient (r) is described in Equation (3).

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}}, \quad (3)$$

where r is Pearson correlation coefficient, which ranges from -1 to +1. +1 is a perfect positive linear relationship, -1 is a perfect negative linear relationship, and 0 is a no linear relationship. n is The total number of paired observations (data points) in the dataset.

$\sum xy$ = The sum of the product of corresponding values of the two variables x and y. When $x = [x_1, x_2, \ldots, x_n]$ and $y = [y_1, y_2, \ldots, y_n]$ this is $x_1y_1 + x_2y_2 + \ldots + x_ny_n$. $\sum x$ and $\sum y$ represent the sum of all values in the first dataset (x) and the second dataset (y), respectively. $\sum x^2$ and $\sum y^2$ represent the sum of the squares of all values in x and y, respectively. For example, $x_1^2 + x_2^2 + \ldots + x_n^2$. The generator models, Caterpillar (CAT), Cummins, Heli, JMA, JMG, Marapco, Mikano, and Perkins were grouped based on their manufacturers or brand name.

2.3.2. Machine learning analysis

Exhaust emissions are governed by non-linear thermo-chemical reactions; thus, machine learning models are suitable for the predictions of these non-linear relationships. Machine learning (ML) is an artificial intelligence (AI) subset that focuses on developing algorithms and statistical models enabling computers to perform tasks without being explicitly programmed [41]. Instead of following predefined rules, ML systems learn data to make predictions through recognizing patterns. ML techniques have multivariable learning that can enable multi-input-multi-output predictions. The acquired exhaust datasets were classified into pass (1) and fail (0) according to the Ogun State Environmental Protection Agency (OGEPA) standard. Four different machine learning algorithms are employed on the datasets. They include decision trees, support vector machine (SVM), K-Nearest Neighbors (KNN), and ensemble (bagged Trees) algorithms. Though the general deep learning techniques require a large dataset, the chosen machine learning techniques (DT, SVM, KNN, and EBT) in this study captures the non-linear dependencies of the datasets, they perform well with small-to-medium datasets, and they are good with high-dimensional data. The applied ensemble (bagged Trees) model can handle complex interactions very well and is quite robust to noise within the dataset. We trained our dataset using four supervised machine learning algorithms available in MATLAB®: Decision Trees, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Ensemble Methods (Bagged Trees). These models were implemented through MATLAB®'s Regression Learner App [42]. To prevent overfitting and ensure the generalization of each model, we applied 5-fold cross-validation. This technique partitions the dataset into five equal subsets (folds), training the model in four folds and validating it on the fifth. This process is repeated five times, ensuring each fold is used once for testing. The average performance across all folds was computed and used as the final evaluation metric. The model parameters and configuration for each technique include tree depth of 10 to 100 maximum number of splits and minimum observation per leaf of 4 to 10 for the decision trees technique. The SVM used the gaussian radial basis function as the kernel function, and the regularization range of 0.1 to 10 Box constraints, and margin of tolerance regression of 0.1 to 0.5 (Epsilon). KNN algorithm used 3 to 15 optimal value and Euclidean distance and equal distance weight. The tuning for KNN was performed using cross-validation with K-fold of 5. The EBT was performed using a bootstrap aggregate with number of learning cycles of 100 trees, with minimum leaf size of 4 to 10, and learning rate of 0.1.

2.3.3. Decision trees (DT)

The tree algorithm is rooted in optimization, information theory, and probability. A decision tree is a hierarchical structure where nodes represent decision points based on a feature, edges represent outcomes of decisions (splits), and leaves represent the final decision or prediction (class or value). The splitting criterion states that at each node, the algorithm chooses the feature and threshold that best splits the data into subsets to maximize decision accuracy. The Gini impurity (G) measures how often a randomly chosen element would be incorrectly classified, as stated in Equation (4). K represents the number of classes and p_i is the proportion of samples belonging to class i in the node. Entropy (H) measures the impurity or disorder in the node, as shown in Equation (5). The lower entropy indicates pure nodes. The information gained (IG) in the classification evaluates the reduction in entropy after a split in Equation (6), where N is the total samples in the parent node, N_j is the number of samples in the child node, v is the number of split and H_j is the entropy of the subset v.

$$G = 1 - \sum_{i=1}^{K} p_i^2, \tag{4}$$

$$H = -\sum_{i=1}^{K} p_i \log_2 p_i, \tag{5}$$

$$\mathrm{IG} = H - \sum_{j=1}^{v} \frac{N_j}{N} H_j. \tag{6}$$

The classification prediction using decision trees assigns the majority class of samples in a leaf. The training time is defined as $O(n \times m \times \log n)$ where n is the number of samples and m is the number of features. $O(\text{depth})$ is the prediction time as the algorithm traverses the tree from root to leaf.

2.3.4. Support vector machine (SVM)

A support vector machine (SVM) is a powerful supervised learning algorithm that is used to classify and perform regression tasks. It relies on mathematical optimization and geometric principles to create decision boundaries in a feature space. Given a dataset in Equation (7), the goal is to find a hyperplane that separates the data points into two classes $y_i \in \{-1, +1\}$ with the largest margin. Generally, SVM guarantees the global optimization solution for linearly separable data. It is quite efficient in handling nonlinear data using kernel tricks.

$$D = \{(x_i, y_i) | x_i \in \mathbb{R}^p, y_i \in \{-1, +1\}\}_{i=1}^{n}. \tag{7}$$

Table 1: Generator brand and percentage by motorization, type, and age.

Generator Model	Engine fuel type (%)		Age in Years (%)			Capacity (%)		
	Diesel	Gas	<5	5-10	>10	<500 kV	500-1000 kVA	>1000 kVA
CATERPILLAR (CAT)	3.125	6.250	0.000	9.375	0.000	0.000	3.125	6.25
CUMMIN	3.125	0.000	0.000	0.000	3.125	3.125	0.000	0.000
HELI	6.250	0.000	0.000	6.25	0.000	3.125	3.125	0.000
JMA	3.125	0.000	0.000	0.000	3.125	3.125	0.000	0.000
JMG	3.125	3.125	0.000	6.25	0.000	3.125	0.000	3.125
MARAPCO	21.875	0.000	3.125	6.25	12.5	15.625	6.25	0.000
MIKANO	9.375	0.000	3.125	6.25	0.000	9.375	0.000	0.000
PERKINS	9.375	0.000	0.000	3.125	3.125	6.25	6.25	0.000
UNSPECIFIED	0.000	28.125	3.125	18.75	6.25	0.000	15.625	12.5

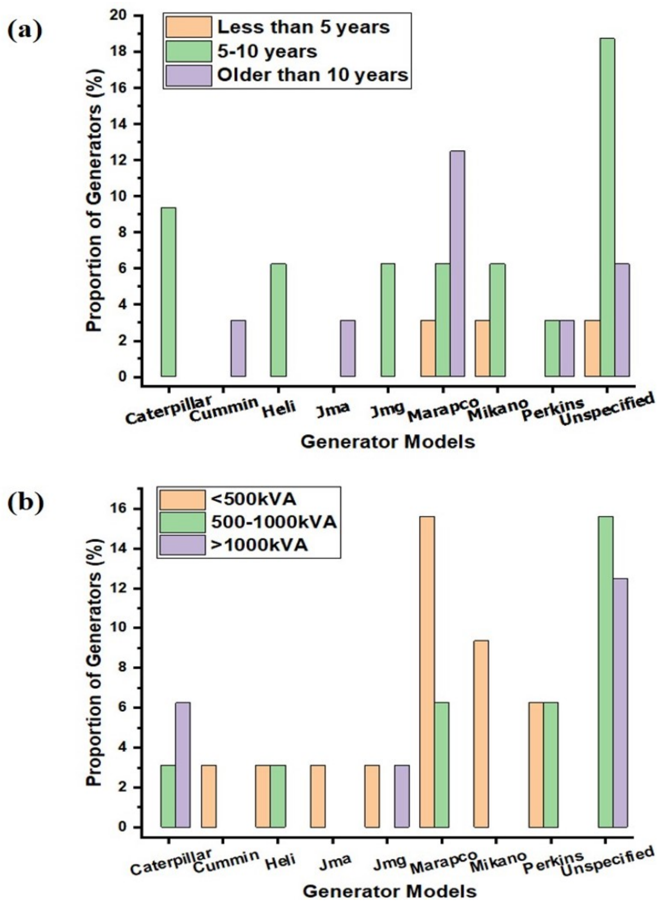

Figure 2: (a) Age of the tested generation, (b) Generator model capacity.

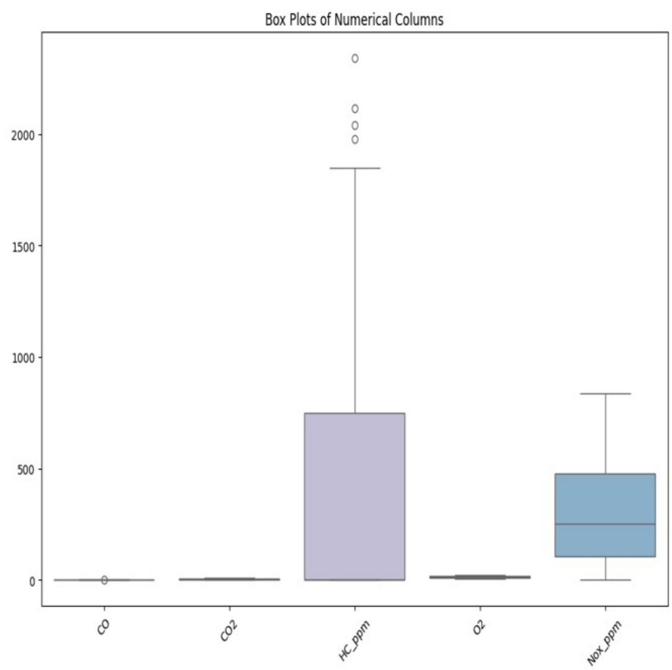

Figure 3: Box plots for numerical columns.

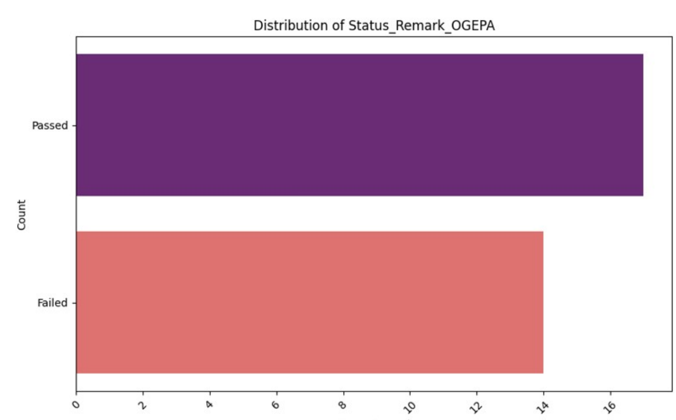

Figure 4: Distribution of status (remark) based on OGEPA exhaust emission standard.

2.3.5. K-nearest neighbor (KNN)

The K-Nearest Neighbors (KNN) algorithm is a non-parametric and instance-based earning method used for classification and regression tasks. Its foundation lies in geometry, distance metrics, and majority voting. KNN predicts the class label of a data point x_q by finding the k-nearest neighbors in the feature space. These neighbors are identified using a distance metric, and the prediction is made based on their labels. The distance metrics used to determine the nearest neighbors is the Euclidean distance in Equation (8), where x and y are two data points in a d-dimensional space, and x_i is the value of the i^{th} feature of x.

$$d(x, y) = \sqrt{\sum_{i=1}^{d} (x_i - y_i)^2}. \tag{8}$$

2.3.6. Ensemble trees (ET)

The mathematics of ensemble algorithms, specifically bagging, involves techniques that combine predictions from multiple base models to improve accuracy and robustness. Bagging leverages concepts from probability, statistics, and machine learning optimization. Bagging is generally used with decision trees - suppose we have datasets represented by Equation (9), where x_i is the feature vector, y_i the target variable, and n the data points number. The goal is to train multiple models on the subset of the data and combine their predictions.

$$D = \{(x_i, y_i)\}_{i=1}^{n}. \tag{9}$$

2.3.7. Performance indicators

The performance metrics include validation, accuracy and confusion matrix. The validation accuracy score evaluates a model's performance on new data in comparison to the training data. This score helps to choose the best model. Confusion matrix plot helps to understand how the currently selected classifier performed in each class. True positive rate (TPR) is a metric that measures how well a model can identify actual positives. It helps you understand how well a model can correctly forecast cases in the true class. The true positive rate equation is expressed in Equation (10), where TP is the true positive, FN is the false negativity, and a perfect model would have a TPR of 1.0, which means it would have 100% detection rate. It reflects the proportion of positive instances that are correctly classified by the model. False negative rate (FNR) represents the proportion of positive results that yield negative test outcomes with the test.

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}. \tag{10}$$

3. Results and discussion

3.1. Characteristics of the generator model

A summary of the generator models, indicating the percentage by engine fuel type and age, is outlined in Table 1. The findings revealed that 59.375% of the tested generators use diesel fuel, while 37.5% were equipped with gas (LPG) as fuel. The breakdown of diesel-powered generator models includes Caterpillar (3.125%), Cummin (3.125%), Heli (6.25%), JMA (3.125%), JMG (3.125%), Marapco (21.875%), Mikano (9.375%), Perkins (9.375%) and no unspecified generator models. On the other hand, a larger percentage of gas-powered generator models comprised mainly of the unspecified generator models (28.125%), Caterpillar (6.25%), and JMG (3.125%). The preference for diesel generators by most of the sampled industries indicated diesel engines have extensive usage on account of their low operating costs, energy efficiency, high durability, and reliability, as well as generally require less maintenance due to their sturdier construction and lower revolutions per minute (RPM) [43]. However, compared to diesel generators, a significant number of industries are now turning towards LPG gas due to its greenness and global acceptance, as it emits less greenhouse gases of CO and NOx, as diesel emissions are comprised of higher amounts of particulate matter (PMs) and NOx which are known to be associated with severe environmental and health cases, causing acid rains, ground-level ozone, and reduced visibility [43].

The age of the generator models tested is a crucial factor in the engine's emissions. The results, as illustrated in Figure 2a, revealed that 10% fell below 5 years, 60% of the total generator models were aged between 5 to 10 years, while 30% were above 10 years old. It was observed that the Marapco model is not only the most tested diesel generator model but is also the model with a high proportion of older than 10 years in age. The capacity of the generator models, distinguishing between a low capacity < 500 kVA, middle capacity 500-1000kVA, and high capacity, >1000 kVA was illustrated in Figure 2b. The findings indicated that 43.754% of the generator models have less than 500 kVA capacity while the least 21.875% are classified to have the capacity above 1000 kVA.

3.2. Distribution and model based exhaust concentration

From Figure 3, the result reveals exhaust concentrations in comparison with the distribution. The distribution of CO emissions is somewhat narrow, with a median of 0.02% and an interquartile range of 0.04%. One outlier, at 0.18%, denotes a solitary occurrence of exceptionally high CO emissions. Compared to CO, CO_2 emissions are more widely distributed, with a median of 5% and an interquartile range of 5.2%. Interestingly, CO_2 emissions do not exhibit any anomalies. Two outliers, with HC emissions of 2339ppm and 2117ppm, respectively, revealed very high HC emissions, while the distribution of O_2 emissions indicates moderacy, with an interquartile range of 7.99% and a median of 13.31%, as the emissions of O_2 are not anomalous. Furthermore, NOx emissions revealed a median of 254ppm and an interquartile range of 377ppm, indicating a moderate dispersion, of which the NOx emission itself does not exhibit any anomalies. With the result, it could be inferred that while HC emissions have noticeable outliers indicating higher exhaust concentration, CO and CO_2 emissions are generally well-contained, while there are no outliers in the more modest distributions for O_2 and NOx emissions.

From Figure 3, the results indicate that the CO emissions are tightly clustered around the median value of 0.02%, indicating that CO emissions are generally low and not very spread out. But where one outlier for CO at 0.18% was observed, this denotes a solitary occurrence (one data point) where the CO level is much higher than the rest (an outlier). For the CO_2, the emissions are more widely distributed, which be explained that the CO_2 values vary more. The median is 5%, indicating that the data is more spread out. Despite the wider range of the CO_2 values, there are no extreme outliers (unusual values) in the data. Also, the HC emissions data contains two outliers—values much higher than the rest (2339ppm and 2117ppm), indicating significant spikes in HC exhaust emissions, which could be due to engine issues, incomplete combustion, just to mention a few. Additionally, the O_2 emission values were revealed to have a moderate spread, centering around a median of 13.31%, with 50% of values within a 7.99% range. Since there

Table 2: Ogun State Environmental Protection Agency (OGEPA) standard limit on industrial generators exhaust emissions.

Industrial Generators	CO (%)	CO_2 (%)	HC (ppm)	NOx (ppm)
Diesel-powered	0.1	10.0	–	600
Gas-powered	1.0	10.0	400	300

*Note: This standard was obtained from the Bye-Laws Document at OGEPA [44].

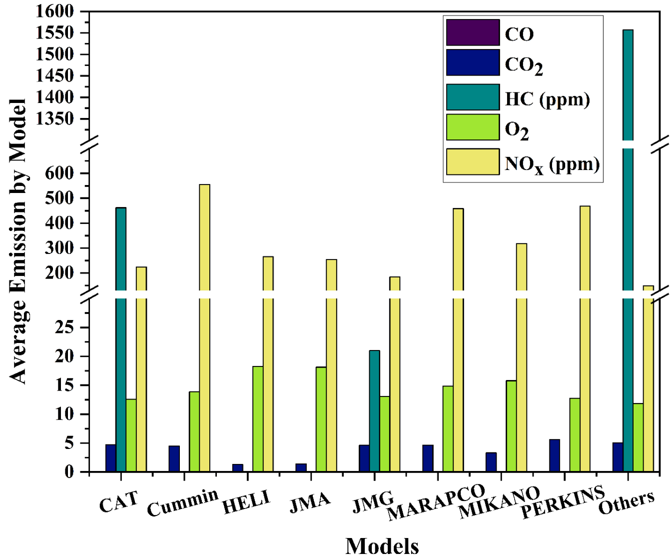

Figure 5: Bar plot of average emissions by model.

are no outliers (unusual oxygen levels), the data is therefore considered normal.

Furthermore, NOx emissions revealed a median of 254ppm and an interquartile range of 377ppm, indicating a moderate dispersion, of which the NOx emission itself does not exhibit any anomalies, suggesting a relatively consistent pattern. For the HC exhibiting noticeable outliers (unusually high values), this could be somewhat due to the abnormal working condition or exhaust gas recirculation (EGR) and other internal combustion working mechanisms of the generators, while the CO and CO_2 values revealed that the emissions are controlled and consistent, with CO having a small outlier, while the O_2 and NOx values obtained indicate that distributions are moderate, indicating stable performance.

The status/remark as compared with the Ogun State Environmental Protection Agency, OGEPA exhaust emission standards limit, as shown in Figure 4 and Table 2 revealed that a total of thirteen (13) of the tested generators "failed" while eighteen (18) "passed". The status/remark of Pass or Failed was related to OGEPA standards, as the Agency is the only State Agency in the whole of Nigeria to have successfully carried out large-scale exhaust emissions tests on automobiles, domiciling such in the State's Bye Laws, hence the emissions standard used in this study. Of the "Failed" test remarks, diesel-powered generators have three (3), and gas-powered generators have ten (10), while from the "Passed" distribution, diesel-powered generators were sixteen (16) while gas has two (2). This explains the higher pass count in which diesel generators performed.

Utilising the model distribution as seen in Figure 5, only eight (8) models were observed and reported. The HC of CAT (from Caterpillar) model possessed the highest emission concentration with 1386ppm, while its lowest emission concentration comes from CO at 0.26%. For the CUMMIN model, its NOx emissions have the highest concentration at 555ppm. For the HELI model, the HC has the lowest emission concentration with 0ppm as its NOx has the highest concentration of 532ppm, while for JMA model, its highest emission concentration was observed in O_2 at 26.16%, as its lowest emission concentration was observed in HC at 0ppm. In addition, the JMG model has its CO possessing the lowest concentration at 0.02%, as its NOx was reported to have the highest concentration at 369ppm. Moreover, the NOx of the MARAPCO model was observed to have the highest emission concentration of 727ppm while HC (ppm) has zero (0ppm) emissions concentration, while for the MIKANO, the NOx has the highest emission of NOx (ppm) of 479ppm while HC has the lowest exhaust concentration of 0ppm. Furthermore, the NOx of PERKINS was observed to have its highest exhaust concentration at 1405ppm, while the exhaust concentration of the HC recorded the highest at 0 ppm.

For Figure 5, it could be inferred from the results that the CAT model exhibited the highest HC emissions at 1386ppm while its lowest CO emission value was 0.26%. The high HC exhaust emissions is attributable to incomplete fuel combustion resulting from ignition misfire or misfire due to excessively lean or rich air/fuel mixtures. Other causes are excessive EGR dilution, restricted or plugged fuel injectors, exhaust leakage past exhaust valves, incorrect spark timing, insufficient cylinder compression, faulty signal to the electronic control module (ECM), just to mention a few. The high NOx exhaust values observed for CUMMIN (555ppm), HELI (532ppm), JMG (369ppm), MARAPCO (727ppm), MIKANO (479ppm) and PERKINS (1405ppm) are ascribed to higher combustion temperatures or engine load at the time of the exhaust testing. This high combustion temperature is usually caused by cooling system problems, leaky intake manifold gasket, improper oxygen sensor and spark advance systems, inefficient EGR system operation, just to mention a few.

3.3. *Effect of the generator's year of manufacture on exhaust gas' concentration*

The year of manufacture (age) of the generator is another key factor, playing a decisive role in the determination of the exhaust gas' concentration. Figure 6a illustrates the relationship between the year of manufacture of the tested generator and the exhaust concentrations of CO_2 and NOx. The generators proportion in the age groups of ≥5 years remarkably contributes to increased exhaust concentrations in both engine

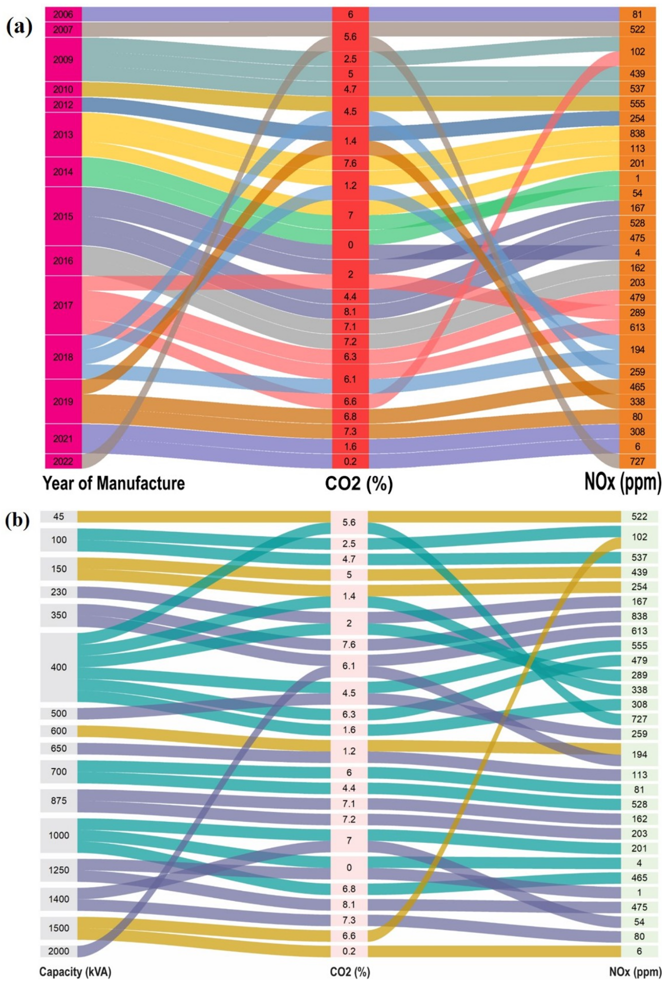

Figure 6: Relationship between (a) year of manufacture and exhaust gas' concentration, (b) generator capacity and exhaust gas' concentration.

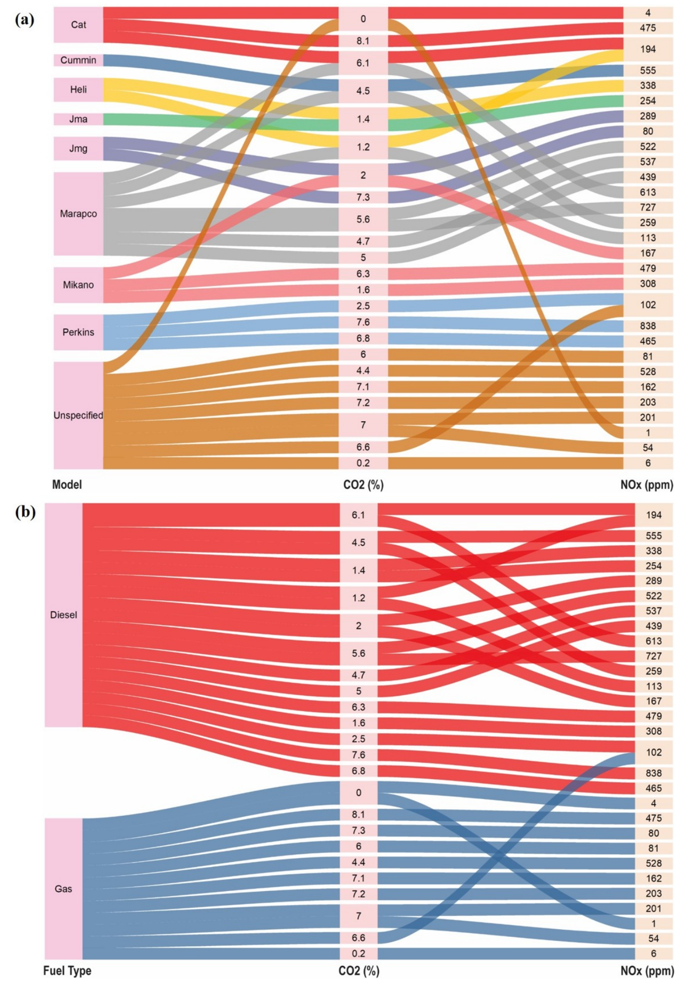

Figure 7: Relationship between (a) the generator model and exhaust gas' concentration, (b) the engine fuel type and exhaust gas' concentration.

types. Noteworthy, in the case of diesel-powered engines, a considerable rise in CO_2 and NOx concentrations was observed due to a higher proportion of Marapco and Perkins models aged more than 10 years, while in gas-powered engines, the exhaust concentrations are predominantly influenced by the older Caterpillar and unspecified models.

3.4. Effect of generator capacity on exhaust gas' concentration

The capacity (KVA) of the generator models is another important factor in the determination of pollutant exhaust concentration, knowing fully well that the higher the capacity, the more fuel consumed. Figure 6b illustrates the relationship between the generator capacity and polluting exhaust concentrations of CO_2 and NOx for both engine fuel types. The highest and lowest CO_2 concentration (8.1% and 0% respectively) emanated from the generator with a capacity of 1250 (¿1000kVA). Also, the corresponding NOx were relatively high and low (475ppm and 1ppm). This proposes a possible relationship between the CO_2 and NOx concentration in the exhaust emission from the industrial generator models tested.

3.5. Effect of generator model and fuel type on exhaust gas' concentration

Mapping the tested generator model revealed a discontinuity in the level of emitted CO_2 and NOx (Figure 7a). However, Caterpillar exhibited the highest CO_2 emission (8.1%), while Perkins revealed the highest NOx emission (838ppm) (Figure 7). As regards the fuel type on exhaust concentration, the total percentage of CO_2 emissions was 76.1%, and the total percentage of NOx 7393ppm was observed for diesel, while the total percentage of CO_2 emissions was 60.9% and the total percentage of NOx stood at 1897ppm for gas fuel (Figure 7b).

3.6. The correlation analysis for the exhaust gases

According to the high p-value (0.65499) and weak negative correlation (-0.10964) in diesel-powered generator's exhausts, as shown in Figure 8a, there is no significant linear association between CO% and CO_2%. This suggests that, given the settings under investigation, variations in CO levels do not forecast variations in CO_2 levels. In a similar vein, the high p-value (0.89239) and extremely weak negative correlation (-0.03329)

(a)

Descriptive Statistics

	N	Mean	SD	Sum	Min	Max
"CO (%)"	19	0.02737	0.04344	0.52	0	0.18
"CO2 (%)"	19	4.00526	2.20138	76.1	1.2	7.6
"HC (ppm)"	19	0	0	0	0	0
"O2 (%)"	19	15.07947	3.02073	286.51	10.42	19.33
"NOx (ppm)"	19	389.10526	209.19313	7393	102	838

Pearson Correlations

		CO (%)	CO2 (%)	HC (ppm)	O2 (%)	NOx (ppm)
"CO (%)"	Pearson Corr.	1	-0.10964	--	-0.03329	-0.49142*
	p-value	--	0.65499	--	0.89239	0.03261
"CO2 (%)"	Pearson Corr.	-0.10964	1	--	-0.94567*	0.73512*
	p-value	0.65499	--	--	1.01615E-9	3.36053E-4
"HC (ppm)"	Pearson Corr.	--	--	--	--	--
	p-value	--	--	--	--	--
"O2 (%)"	Pearson Corr.	-0.03329	-0.94567*	--	1	-0.70955*
	p-value	0.89239	1.01615E-9	--	--	6.67708E-4
"NOx (ppm)"	Pearson Corr.	-0.49142*	0.73512*	--	-0.70955*	1
	p-value	0.03261	3.36053E-4	--	6.67708E-4	--

2-tailed test of significance is used
*: Correlation is significant at the 0.05 level

(b)

Descriptive Statistics

	N	Mean	SD	Sum	Min	Max
"CO (%)"	12	0.03	0.02045	0.36	0	0.06
"CO2 (%)"	12	5.075	3.14848	60.9	0	8.1
"HC (ppm)"	12	1286.91667	818.79706	15443	42	2339
"O2 (%)"	12	11.83417	5.62555	142.01	6.14	20.92
"NOx (ppm)"	12	158.08333	175.47673	1897	1	528

Pearson Correlations

		CO (%)	CO2 (%)	HC (ppm)	O2 (%)	NOx (ppm)
"CO (%)"	Pearson Corr.	1	0.57749*	0.55988	-0.57885*	0.42789
	p-value	--	0.04926	0.05835	0.04861	0.16525
"CO2 (%)"	Pearson Corr.	0.57749*	1	0.383	-0.99726*	0.44174
	p-value	0.04926	--	0.21912	1.21988E-12	0.15051
"HC (ppm)"	Pearson Corr.	0.55988	0.383	1	-0.40181	0.29722
	p-value	0.05835	0.21912	--	0.19541	0.34814
"O2 (%)"	Pearson Corr.	-0.57885*	-0.99726*	-0.40181	1	-0.4905
	p-value	0.04861	1.21988E-12	0.19541	--	0.10544
"NOx (ppm)"	Pearson Corr.	0.42789	0.44174	0.29722	-0.4905	1
	p-value	0.16525	0.15051	0.34814	0.10544	--

2-tailed test of significance is used
*: Correlation is significant at the 0.05 level

Figure 8: The descriptive and Pearson's correlation of the exhaust emission parameters (a) using diesel, (b) using gas.

suggest that there is no meaningful link between CO% and O_2%, implying that there is no direct correlation between CO emissions and O_2 levels. Between CO and NOx, there is a statistically significant inverse link, as indicated by the moderately negative correlation (-0.49142) and p-value of 0.03281. This may be attributable to incomplete combustion circumstances resulting in higher CO emissions, hence limiting the production of NOx [45]. Q. Malé and coworkers [46] in their study on plasma-assisted combustion in sequential combustors, also corroborated that incomplete combustion can result in higher CO emissions, which may influence NOx production, stating that increased CO levels may affect the pathway for NOx formation. A very low p-value (1.01615e-9) coupled with a substantial negative correlation (-0.94567) points to a highly significant unfavorable association. This establishes a fact because O_2 is used during complete combustion with the fuel, produc-

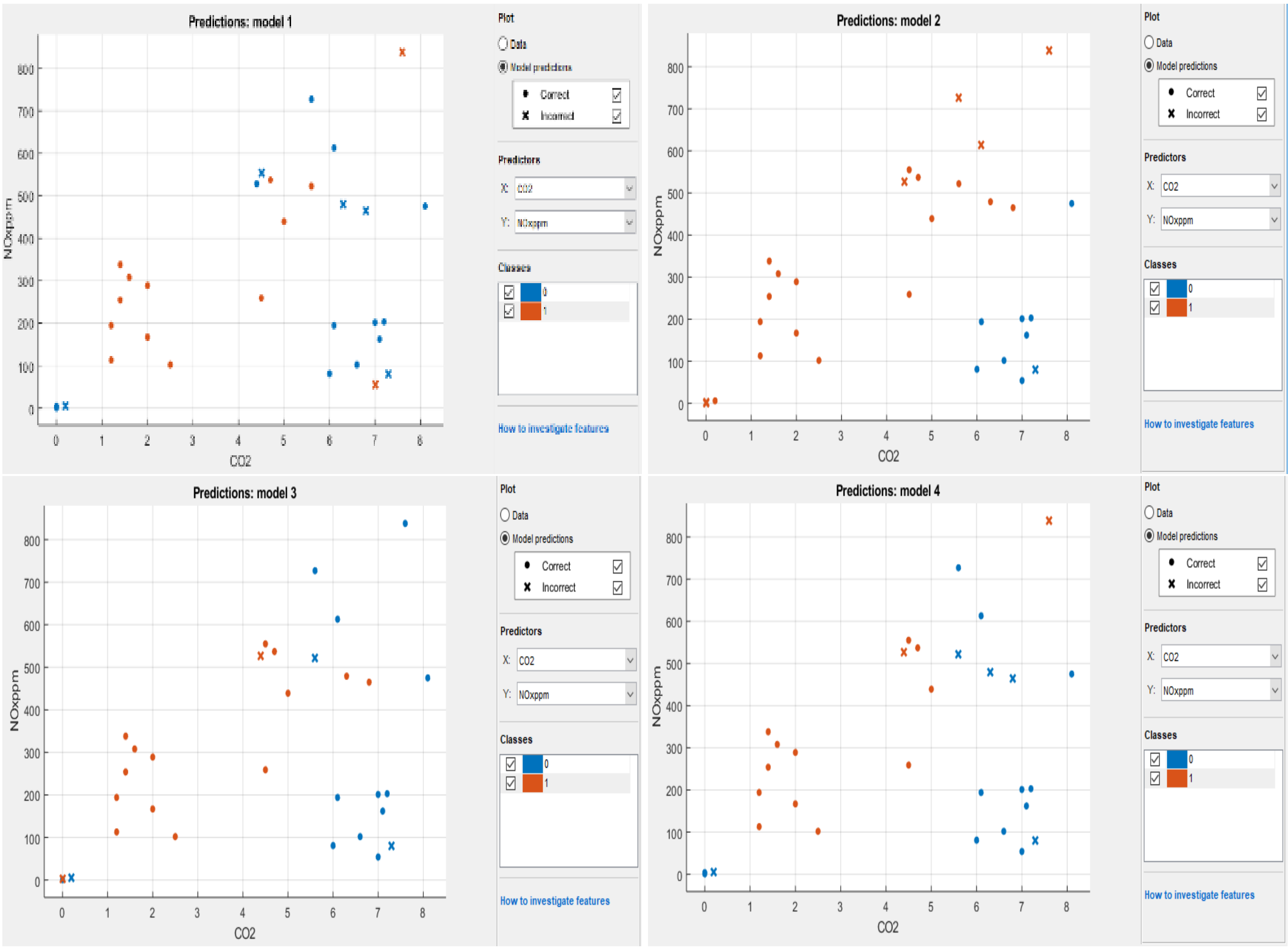

Figure 9: Scatter plots for all the predicting models (DT (model1), SVM (model 2), KNN (model 3), EBT (model 4)) of gases CO_2 and NOx, where the dot represents correct prediction, and cross represents incorrect prediction.

ing CO_2. The results indicated a substantial positive correlation (0.73512) between higher CO_2 and higher NOx levels, with a p-value of 3.36053e-4, as previously reported by Kozlov and Titova [47].

Another higher possibility of this can occur in situations when there is complete combustion at very high temperatures, causing high NOx production. The substantial p-value (1.01615e-9) and robust negative correlation (-0.94567) for O_2 and CO_2 supported the anticipated adverse association resulting from combustion. The results indicate a significant negative correlation (-0.70955) between greater O_2 levels and lower NOx emissions, with a p-value of 6.67708e-4. There is a significant p-value of 0.03261 and a considerable negative correlation of -0.49142 between greater CO and lower NOx levels. The combustion theory provides a framework for interpreting the correlation between NOx and CO2, as NOx primarily forms through the thermal NOx mechanism, requiring high temperatures to dissociate N2 and O2, forming NO. This is prevalent in complete combustion conditions, explaining the positive CO2-NOx correlation, in this study for diesel-powered exhaust emissions. Herein, a substantial p-value (3.36053e-4) and a strong postive association (0.73512) indicate that circumstances that support increased CO_2 generation also support the development of NOx. Similarly, Shahariar *et al.* reported a strong linear correlation between NOx and CO_2 emissions in urban driving scenarios of diesel vehicles, with a Pearson correlation coefficient of 0.82 at a p-value of 0.001, indicating that higher fuel consumption (reflected in CO_2 emissions) generally aligns with increased NOx emissions [48]. However, the analysis highlights that driving style and route features significantly modulate this relationship, as aggressive driving behaviors (such as rapid acceleration and hard braking) were linked to spikes in NOx emissions, even when CO_2 levels (fuel consumption) remained comparable across drivers. This suggests that NOx emissions are more sensitive to transient driving dynamics (e.g., turbocharger lag during sudden acceleration) than CO_2, which correlates more strongly with trip duration and steady-state driving. Nevertheless, a recent study has shown that carbon reduction is a significant driver of NOx depletion in industrial sectors [49]. Urban infrastructure features like motorway entrances, traffic signals, and intersections can be emerging emission hotspots, where abrupt speed changes cause sharp increases in both pol-

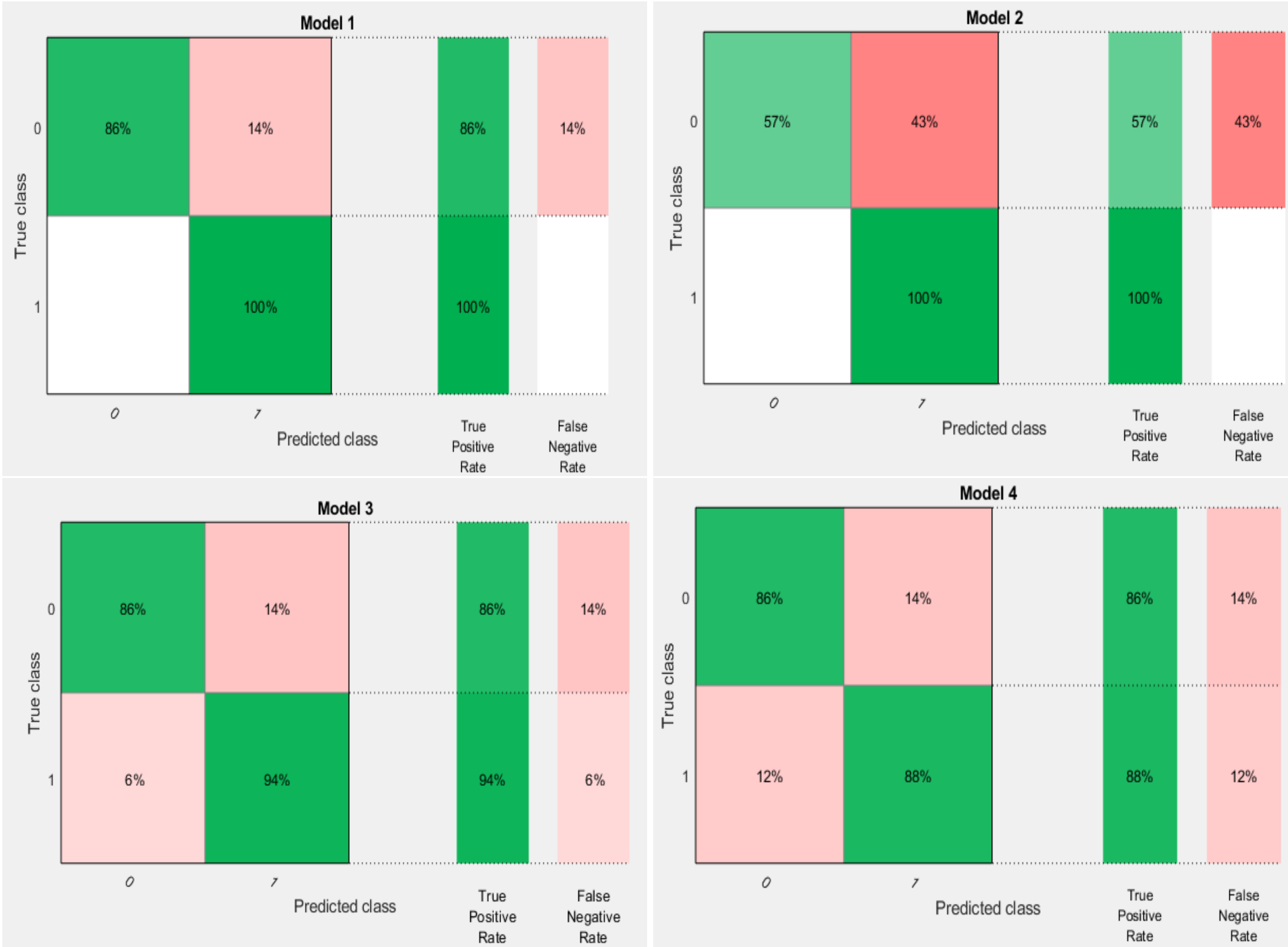

Figure 10: Confusion matrix for model 1 (DT), model 2 (SVM), model 3 (KNN), and model 4 (EBT).

lutants. Despite this overlap, the study by Shahariar *et al.* underscores that NOx emissions exhibit greater variability due to driver behavior, while CO_2 is more consistently tied to energy consumption patterns [48]. Therefore, there is a need for targeted interventions addressing driving habits and urban planning to mitigate both greenhouse gases and local air pollutants. The conclusion that lower O_2 levels correlate with increased NOx emissions is supported by the strong negative correlation (-0.70955) and significant p-value (6.67708e-4). There is no discernible relationship between HC and CO, CO_2, O_2, or NOx. This suggests that, in contrast to the other pollutants examined, the mechanisms influencing HC emissions are distinct or more intricate.

With a p-value of 0.04926 and a moderate positive correlation (0.57749) for the exhausts from gas-powered generators, as shown in Figure 8b, the correlation is statistically significant. This implies that maybe as a result of changes in combustion efficiency, CO_2 emissions tend to increase along with CO emissions. Near the significance threshold, the moderately positive correlation (0.55988) with a p-value of 0.05835 points to a possible link between higher CO levels and higher HC emissions, most likely in the context of incomplete combustion. There is a substantial adverse link, as indicated by the moderately negative correlation (-0.57885) and p-value of 0.04861. This shows that incomplete combustion, where less oxygen is used, is linked to lower O_2 levels and higher CO emissions. Also, it is clear from the slight positive correlation (0.42789) and p-value of 0.16525 that there is no significant association between CO and NOx emissions under the conditions of the study. The lack of a significant association is suggested by the modest positive correlation (0.383) and high p-value (0.21912), which show that CO_2 and HC emissions are not highly correlated. As would be predicted given that O_2 is consumed to produce CO_2 during combustion, the extremely strong negative correlation (-0.99726) with a highly significant p-value (1.21988e-12) shows an almost perfect inverse association.

There appears to be no significant association between CO_2 and NOx emissions, as indicated by the modest positive correlation (0.44174) and p-value of 0.15051. Unlike the diesel-powered exhaust emission, the non-significant correlation between NOx and CO_2 in the gas-powered exhaust emission may reveal how the incorporation of modern technologies tends to decouple the relationship between NOx and CO_2. Interestingly, factors such as their distinct formation mechanisms, independent control technologies, and varying responses to driving conditions can contribute to the decoupling. NOx is temperature-dependent and managed by catalytic systems, while CO_2 is fuel consumption-driven and less affected by these controls. Higher CO levels are linked to decreased O_2 levels, according to the significant p-value (0.04861) and moderate negative correlation (-0.57885). The predicted unfavorable link is confirmed by the extremely significant p-value (1.21988e-12) and very substantial negative correlation (-0.99726). Higher O_2 levels may be linked to decreased NOx emissions, according to a tendency shown by the somewhat negative correlation (-0.4905) and non-significant p-value (0.105544), even though this association is not statistically significant. Between NOx and O_2, there may be an inverse association, but it is not statistically significant, as shown by the moderately negative correlation (-0.4905) with a non-significant p-value (0.105544). Between HC and O_2, the modest negative correlation (-0.40181) and non-significant p-value (0.19541) indicate no significance. The high p-value (0.34814) and modest positive correlation (0.29722) between HC and NOx suggest that there is no meaningful association. Therefore, a positive r indicates that as one variable increases, the other tends to increase. A negative r indicates that as one variable increases, the other tends to decrease. The closer $|r|$ is to 1, the stronger the linear relationship between the variables.

3.7. Machine learning analysis of the exhaust gases

The Decision Tree, the support vector machine, the KNN, and the Tree Ensemble were trained on the MATLAB using 70% of the dataset. The scattered plots for the evaluation of the predictive models (1-4) using the remaining 30% of the data are presented in Figure 9. The prediction accuracies for the classification status of the exhaust gases are 93.5 %, 80.6 %, 90.3 %, and 87.1 % for the decision tree learner, support vector machine learner, KNN learner, and tree ensemble learner, respectively (Table 3). In terms of the prediction speed and model training time, the support vector machine learner has the greatest prediction speed of 850 observations/second, while the tree ensemble learner has the lowest prediction speed of 260 observations/second. Also, the decision tree learner has the longest training time of 4.57 s compared to that of the support vector machine learner of 1.31 s.

The confusion matrices for all the models are presented in Figure 10. Both DT and SVM learners have the highest true positive rate prediction of 100 % for the classification status of pass (1), while the EBT has the lowest true positive rate prediction of 88 %. Generally, all the machine learning classifications have an excellent true positive rate prediction for pass (1) and fail (0) classification status, except SVM, which performs poorly with a prediction accuracy of 57 % for fail (0) classification status. These results are similar to that of the false negative rate, where SVM equally performed poorly in the prediction of fail (0) classification status.

The decision tree achieved high accuracy and perfect classification of the pass (1) class. However, it exhibited the longest training time (4.57 s), likely due to recursive binary splitting and deep tree structures. Decision trees perform well when the data exhibits clear logical partitions, which may indicate that the features in this dataset align well with class boundaries. Even though the SVM has the fastest prediction speed (850 obs/s) and short training time (1.31 s), it suffered from low accuracy in predicting the fail (0) class (57 %). This is likely due to imbalance class distributions where SVMs tend to favour the majority class (pass (1)), especially when the kernel parameters are not well-tuned. K-Nearest Neighbors (KNN) produced a balanced performance, combining high overall accuracy (90.3%) and moderately fast predictions. The strength of this algorithm lies in capturing local data structures; its effectiveness suggests that the dataset contains meaningful clusters corresponding to the classification labels. However, KNN can suffer in high-dimensional or noisy data, and performance can degrade with large dataset due to the computational cost of distance calculations. The employed Tree Ensemble (Bagged Trees) provided a robust compromise between bias and variance, with accuracy of 87.1%, though with the slowest prediction speed (260 obs/s). The ensemble techniques are particularly good at generalizing especially on complex or non-linear datasets. The reduced true positive rate (88%) for the pass class could be due to model averaging, which smoothens decision boundaries, slightly compromising peak performance for increased robustness.

In environmental monitoring, especially when tracking exhaust gas emissions, The false negative rate is high for SVM technique. The failure of SVM to detect a pollutant breach will have serious implications for environmental monitoring. The false negatives are costly for environmental monitoring, as they may lead to regulatory violations and environmental harm. Therefore, Decision trees and Tree Ensemble learners are preferable for environmental monitoring, as they have higher sensitivity to the fail (0) class, despite slightly higher computational costs.

4. Conclusion

Most of the sampled industrial generators utilise a higher number of diesel generators than gas-powered ones, possibly due to their energy efficiency, high durability, and reliability. The diesel-powered generators generally require less maintenance due to their sturdier construction and lower revolutions per minute (RPM). Also, the age and capacity (KVA) of the generators have been observed to have considerably impacted high pollutant exhaust concentrations in both the diesel- and gas-powered engines; regardless of generator models, there was a considerable rise in CO_2 and NOx concentrations from generators of more than 10 years and capacity of more than 1000KVA. Furthermore, the total percentage of CO_2 and NOx emissions reported for diesel were 76.1% and 7393ppm, while for gas-powered, they were 60.9% and 1897ppm. With respect to the concentration of pollutants, it was observed that there is a positive correlation between CO_2 and NOx for diesel-powered generators, while for gas, there is a positive correlation between CO and CO_2, the same for CO and HC. Interestingly, a negative correlation was observed for CO_2 and O_2 in both the diesel- and gas-powered generators. The study equally trained four machine learning classification models on MATLAB using

Table 3: Performance indicators of machine learning algorithms.

	Accuracy	Training Time (sec)	Prediction Speed (obs/sec)	True Positive (%)		False Negative (%)	
				1 (Pass)	0 (Fail)	1 (Pass)	0 (Fail)
DT	93.5	4.57	410	100	86	0	14
SVM	80.6	1.87	850	100	57	0	43
KNN	90.3	1.31	760	94	86	6	14
EBT	87.1	3.15	260	88	86	12	14

70% of the dataset. The models achieved prediction accuracies of 93.5%, 80.6%, 90.3%, and 87.1% for exhaust gas OGEPA classification status. The support vector machine learner had the highest prediction speed of 850 observations/second, while the tree ensemble learner had the lowest. The confusion matrices showed that DT and KNN learners possessed the highest true positive rate prediction of 100% for pass classification status, respectively. The SVM suffered from low accuracy in predicting the fail (0) class (57 %). The inability of the Support Vector Machine (SVM) model to accurately detect pollutant threshold breaches can have significant consequences for environmental monitoring. False negative predictions are particularly critical, as they may result in undetected regulatory violations and potential environmental damage. In contrast, Decision Tree and Tree Ensemble models demonstrated greater sensitivity in identifying the ”fail” (0) class. Although these models may incur slightly higher computational costs, their enhanced reliability makes them more suitable for applications in environmental compliance and risk mitigation. From the results, it could be observed that gas-powered (LPG) generators are better and offer several advantages over diesel-powered generators as the LPG generators produce significantly fewer greenhouse gases and pollutants like NOx, when compared to diesel, making them more environmentally friendly.

In an attempt to ameliorate the usage of fossil-fuel generators in developing countries including Nigeria, strategic frameworks that can guide multi-pronged policy formulation policy are crucial. A strategy to establish phasing out fossil-fuel generators in favour of cleaner alternatives should be adopted, while stringent emissions standards should be set to reduce air pollution. Also, incentives and low-interest loans should be offered to citizens to adopt renewable energy systems, while the expansion and upgrading the national electricity grid should be implemented to reduce reliance on fossil-fuel generators, amongst other policies. Furthermore, future studies can broaden the dataset to include a broader range of generator types and regional variances in order to increase model generalisability, while machine learning (ML) models and Internet of Things (IoT) could be integrated for real-time emission diagnosis and monitoring. Looking at the policy level, our study contributes to the development of data-driven air quality legislation, generator usage regulations as well as overall emissions monitoring in Nigeria and other developing nations. Finally, we recognise that the sample size of 31 generators may be insufficient for drawing broad conclusions about industrial generator emissions over the entire region, as lower sample size can reduce statistical power and fail to capture the full variability found in the larger population of industrial generators. This limitation may affect the generalisability of the results, especially regarding emission trends and model-specific performance, nevertheless, the dataset still offers meaningful patterns and highlights key emission concerns which can inform future, larger-scale studies.

Data availability

The datasets used and/or analysed in the current study are available from the corresponding author upon reasonable request.

Acknowledgment

The authors appreciate the Editor and anonymous reviewers for their efforts in improving the quality of this manuscript.

References

[1] D. Akal, S. Öztuna & M. K. Büyükakın, “A review of hydrogen usage in internal combustion engines (gasoline-Lpg-diesel) from combustion performance aspect”, International Journal of Hydrogen Energy **45** (2020) 35257. https://doi.org/10.1016/j.ijhydene.2020.02.001.

[2] C. S. Sriharikota, K. Karuppasamy, V. Nagarajan, R. Sathyamurthy, B. Ramani, V. Muthu & S. Karuppiah, “Experimental investigation of the emission and performance characteristics of a DI diesel engine fueled with the vachellia nilotica seed oil methyl ester and diesel blends”, ACS Omega **6** (2021) 14068. https://doi.org/10.1021/acsomega.1c00437.

[3] I. Abrar, T. Arora & R. Khandelwal, “Bioalcohols as an alternative fuel for transportation: Cradle to grave analysis”, Fuel Processing Technology **242** (2023) 107646. https://doi.org/10.1016/j.fuproc.2022.107646.

[4] A. Zerboni, T. Rossi, R. Bengalli, T. Catelani, C. Rizzi, M. Priola, S. Casadei & P. Mantecca, “Diesel exhaust particulate emissions and in vitro toxicity from Euro 3 and Euro 6 vehicles”, Environmental Pollution **297** (2022) 118767. https://doi.org/10.1016/j.envpol.2021.118767.

[5] D. E. Schraufnagel, “The health effects of ultrafine particles”, Experimental & Molecular Medicine **52** (2020) 311. https://doi.org/10.1038/s12276-020-0403-3.

[6] S. Steiner, C. Bisig, A. Petri-Fink & B. Rothen-Rutishauser, “Diesel exhaust: current knowledge of adverse effects and underlying cellular mechanisms”, Archives of Toxicology **90** (2016) 1541. https://doi.org/10.1007/s00204-016-1736-5.

[7] W. A. Khan, F. Sharif, M. F. Khokhar, L. Shahzad, N. Ehsan & M. Jahanzaib, “Monitoring of ambient air quality patterns and assessment of air pollutants’ correlation and effects on ambient air quality of Lahore, Pakistan”, Atmosphere **14** (2023) 1257. https://doi.org/10.3390/atmos14081257.

[8] H. L. Brumberg & C. J. Karr, “Ambient air pollution: Health hazards to children”, Pediatrics **147** (2021) e2021051484. https://doi.org/10.1542/peds.2021-051484.

[9] J. Moya, L. Phillips, L. Schuda, P. Wood, A. Diaz, R. Lee, R. Clickner, R. J. Birch, N. Adjei, P. Blood, K. Chapman, R. de Castro & K. Mahaffey, "EPA's exposure factors handbook (EFH)" (USEPA, 2011). [Online]. https://www.google.com/url?sa=t&source=web&rct=j&opi=89978449&url=https://rais.ornl.gov/documents/EFH_2011.pdf.

[10] M. Guarnieri & J. R. Balmes, "Outdoor air pollution and asthma", The Lancet **383** (2014) 1581. https://doi.org/10.1016/S0140-6736(14)60617-6.

[11] C. F. Oguejiofor, U. U. Eze, I. G. Eke, A. A. Eze, O. B. Onyejekwe & B. M. Anene, "Adverse effects of exposure to petrol-generator exhaust fumes on the reproductive hormones, testis and spermatozoa in male dogs", Reproductive Toxicology **123** (2024) 108516. https://doi.org/10.1016/j.reprotox.2023.108516.

[12] J. Valente, C. Pimentel, R. Tavares, J. Ferreira, C. Borrego, P. Carreiro-Martins, I. Caires, N. Neuparth & M. Lopes, "Individual exposure to air pollutants in a Portuguese urban industrialized area", Journal of Toxicology and Environmental Health - Part A: Current Issues **77** (2014) 888. https://doi.org/10.1080/15287394.2014.910159.

[13] P. Rossner, T. Cervena & M. Vojtisek-Lom, "In vitro exposure to complete engine emissions – a mini-review", Toxicology **462** (2021) 152953. https://doi.org/10.1016/j.tox.2021.152953.

[14] A. Iqubal, M. Ahmed, S. Ahmad, C. R. Sahoo, M. K. Iqubal & S. E. Haque, "Environmental neurotoxic pollutants: review", Environmental Science and Pollution Research **27** (2020) 41175. https://doi.org/10.1007/s11356-020-10539-z.

[15] L. Müller, J. Usemann, M. P. Alves & P. Latzin, "Diesel exposure increases susceptibility of primary human nasal epithelial cells to rhinovirus infection", Physiological Reports **9** (2021) e14994. https://doi.org/10.14814/phy2.14994.

[16] G. Y. Kim, I. Jung, M. Park, K. Park, S. H. Lee & W. H. Kim, "Diesel exhaust particles induce human umbilical vein endothelial cells apoptosis by accumulation of autophagosomes and caspase-8 activation", Scientific Reports **12** (2022) 16492. https://doi.org/10.1038/s41598-022-21044-3.

[17] X. Wu, C. Ciminieri, I. S. T. Bos, M. E. Woest, A. D'Ambrosi, R. Wardenaar, D. C. J. Spierings, M. Königshoff, M. Schmidt, L. E. M. Kistemaker & R. Gosens, "Diesel exhaust particles distort lung epithelial progenitors and their fibroblast niche", Environmental Pollution **305** (2022) 119292. https://doi.org/10.1016/j.envpol.2022.119292.

[18] Y. Sun, F. Bochmann, A. Nold & M. Mattenklott, "Diesel exhaust exposure and the risk of lung cancer-a review of the epidemiological evidence", International Journal of Environmental Research and Public Health **11** (2014) 1312. https://doi.org/10.3390/ijerph110201312.

[19] E. Long, C. Schwartz & C. Carlsten, "Controlled human exposure to diesel exhaust: a method for understanding health effects of traffic-related air pollution", Particle and Fibre Toxicology **19** (2022) 15. https://doi.org/10.1186/s12989-022-00454-1.

[20] L. P. Wong, H. Alias, N. Aghamohammadi, A. Ghadimi & N. M. N. Sulaiman, "Control measures and health effects of air pollution: A survey among public transportation commuters in Malaysia", Sustainability (Switzerland) **9** (2017) 1616. https://doi.org/10.3390/su9091616.

[21] P. E. Rosenfeld, K. R. Spaeth, R. Hallman, R. Bressler & G. C. Smith, "Cancer risk and diesel exhaust exposure among railroad workers", Water, Air, and Soil Pollution **233** (2022) 171. https://doi.org/10.1007/s11270-022-05651-4.

[22] I. O. Abramova, "The population of Africa under the conditions of transformation of the world order", Herald of the Russian Academy of Sciences **92** (2022) S1306. https://doi.org/10.1134/S0018151X21040131.

[23] O. Rosnes & H. Vennemo, "The cost of providing electricity to Africa", Energy Economics **34** (2012) 1318. https://doi.org/10.1016/j.eneco.2012.06.008.

[24] R. O. Adesola, E. Opuni, I. Idris, O. J. Okesanya, O. Igwe, M. D. Abdulazeez & D. E. Lucero-Prisno, "Navigating Nigeria's health landscape: population growth and its health implications", Environmental Health Insights **18** (2024). https://doi.org/10.1177/11786302241250211.

[25] "Nigeria - country commercial guide- electricity", Power Systems and Renewable Energy. International Trade Administration, (2023). [Online]. https://www.trade.gov/country-commercial-guides/electricity-power-systems-and-renewable-energy.

[26] Ministry of Power. [Online]. Available at https://power.gov.ng/.

[27] M. Izuaka, "Nigeria suffers power outage as grid collapses", Premium Times, (2024). [Online]. https://www.premiumtimesng.com/news/top-news/760693-nigeria-suffers-power-outage-as-grid-collapses.html.

[28] "National power grid collapses again, 12th time this year alone", This Day, (2024). [Online]. https://www.thisdaylive.com/index.php/2024/12/12/national-power-grid-collapses-again-12th-time-this-year-alone/#google_vignette.

[29] D. Olawin, D. Aina, C. Azubuike, A. Amshi & R. Ede, "Nationwide blackout as grid suffers 10th collapse in 2024", Punch Newspaper, (2024). [Online]. https://punchng.com/nationwide-blackout-as-grid-suffers-10th-collapse-in-2024/.

[30] Shivakumar, P. Srinivasa Pai & B. R. Shrinivasa Rao, "Artificial neural network based prediction of performance and emission characteristics of a variable compression ratio CI engine using WCO as a biodiesel at different injection timings", Applied Energy **88** (2011) 2344. https://doi.org/10.1016/j.apenergy.2010.12.030.

[31] J. Liu, Q. Huang, C. Ulishney & C. E. Dumitrescu, "Machine learning assisted prediction of exhaust gas temperature of a heavy-duty natural gas spark ignition engine", Applied Energy **300** (2021) 117413. https://doi.org/10.1016/j.apenergy.2021.117413.

[32] B. Ochieng, F. Onyango, P. Kuria, M. Wanjiru, B. Maake & M. Awuor, "AI-driven carbon emissions tracking and mitigation model", 2024 IST-Africa Conf. (IEEE, 2024), pp. 1. https://doi.org/10.23919/IST-Africa63983.2024.10569242.

[33] T. F. Yusaf, D. R. Buttsworth, K. H. Saleh & B. F. Yousif, "CNG-diesel engine performance and exhaust emission analysis with the aid of artificial neural network", Applied Energy **87** (2010) 1661. https://doi.org/10.1016/j.apenergy.2009.10.009.

[34] T. Van Hung, H. H. Alkhamis, A. F. Alrefaei, Y. Sohret & K. Brindhadevi, "Prediction of emission characteristics of a diesel engine using experimental and artificial neural networks", Applied Nanoscience **13** (2023) 433. https://doi.org/10.1007/s13204-021-01781-z.

[35] M. B. Sanjeevannavar, N. R. Banapurmath, V. D. Kumar, A. M. Sajjan, I. A. Badruddin, C. Vadlamudi, S. Krishnappa, S. Kamangar, R. U. Baig & T. M. Y. Khan, "Machine learning prediction and optimization of performance and emissions characteristics of IC Engine", **15** (2023) 13825. https://doi.org/10.3390/su151813825.

[36] P. Ganesan, S. Rajakarunakaran, M. Thirugnanasambandam & D. Devaraj, "Artificial neural network model to predict the diesel electric generator performance and exhaust emissions", Energy **83** (2015) 115. https://doi.org/10.1016/j.energy.2015.02.094.

[37] K. Zeinalipour, L. Barriere, D. Ghelardi & M. Gori, "Application of machine learning models for carbon monoxide and nitrogen oxides emission prediction in gas turbines", arXiv, (2025). [Online]. https://doi.org/https://doi.org/10.48550/arXiv.2501.17865.

[38] R. Potts, R. Hackney & G. Leontidis, "Tabular machine learning methods for predicting gas turbine emissions", Machine Learning and Knowledge Extraction **5** (2023) 1055. https://doi.org/10.3390/make5030055.

[39] S. Moonsammy, T. D. T. T. Oyedotun, O. Oderinde, M. Durojaiye & A. Durojaye, "Exhaust determination and air-to-fuel ratio performance of end-of-life vehicles in a developing African country: a case study of Nigeria", Transportation Research Part D: Transport and Environment **91** (2021) 102705. https://doi.org/10.1016/j.trd.2021.102705.

[40] P. K. Alimo, G. Lartey-Young, S. Agyeman, T. Y. Akintunde, E. Kyere-Gyeabour, F. Krampah, A. Awomuti, O. Oderinde, A. O. Agbeja & O. G. Afolabi, "Spatial distribution and policy implications of the exhaust emissions of two-stroke motorcycle taxis: a case study of southwestern state in Nigeria", Journal of the Nigerian Society of Physical Sciences **6** (2024) 1898. https://doi.org/10.46481/jnsps.2024.1898.

[41] T. M. Mitchell, "Does machine learning really work?", AI Magazine **18** (1997) 5. https://doi.org/10.1609/aimag.v18i3.1303.

[42] "MATLAB and Statistics Toolbox Version: 24.1.7 (R2024a)", The MathWorks, Inc., Natick, Massachusetts, United States, (2024). [Online]. https://www.mathworks.com.

[43] İ. A. Reşitoğlu, K. Altinişik & A. Keskin, "The pollutant emissions from diesel-engine vehicles and exhaust aftertreatment systems", Clean Technologies and Environmental Policy **17** (2015) 15. https://doi.org/10.1007/s10098-014-0793-9.

[44] Ogun State Environmental Protection Agency (OGEPA). [Online]. Available ar https://ogepa.og.gov.ng/.

[45] P. Brijesh & S. Sreedhara, "Exhaust emissions and its control methods in compression ignition engines: a review", International Journal of Automotive Technology **14** (2013) 195. https://doi.org/10.1007/s12239-013-0022-2.

[46] Q. Malé, N. Barléon, S. Shcherbanev, B. Dharmaputra & N. Noiray, "Numerical study of nitrogen oxides chemistry during plasma assisted combustion in a sequential combustor", Combustion and Flame **260** (2024) 113206. https://doi.org/10.1016/j.combustflame.2023.113206.

[47] V. E. Kozlov & N. S. Titova, "Simulation of the emission of nitrogen and carbon oxides during the turbulent combustion of a partially mixed methane–air mixture", High Temperature **60** (2022) S67. https://doi.org/10.1134/S0018151X21040131.

[48] G. M. Hasan Shahariar, M. Sajjad, K. A. Suara, M. I. Jahirul, T. Chu-Van, Z. Ristovski, R. J. Brown & T. A. Bodisco, "On-road CO_2 and NO_x emissions of a diesel vehicle in urban traffic", Transportation Research Part D: Transport and Environment **107** (2022) 103326. https://doi.org/10.1016/j.trd.2022.103326.

[49] X. Qin, P. Xie & C. Liao, "Study on the synergistic effect of NO_x and CO_2 emission reduction in the industrial sector of Guangzhou", Frontiers in Environmental Science **13** (2025) 1497121. https://doi.org/10.3389/fenvs.2025.1497121.