





Hybrid Particle Swarm Optimization and Feedforward Neural Network Model for Enhanced Prediction of Gas Turbine Emissions

Ahmed N. Awad¹, Tanya Shakir Jarad^{2*}

¹ Department of Fuel and Energy Techniques Engineering, Al-Huda University College, Anbar, Ramadi 31001, Iraq

² Department of Computer Science, College of Computer Science and Information Technology, University of Anbar, Anbar, Ramadi 31001, Iraq

Corresponding Author Email: tanyashaker6@gmail.com

Copyright: ©2024 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ijepm.090204>

ABSTRACT

Received: 10 February 2024

Revised: 18 March 2024

Accepted: 7 May 2024

Available online: 30 June 2024

Keywords:

gas turbine emissions prediction, FNN-based PSO approach, K-Nearest Neighbor (KNN) algorithm, prediction accuracy measurements

Gas emissions, particularly carbon monoxide (CO) and nitrogen oxide (NO_x), pose significant operational and environmental challenges in gas-fired power plants, especially under low ambient temperatures that reduce turbine efficiency and power output. This study introduces a hybrid model that combines Particle Swarm Optimization (PSO) with a Feedforward Neural Network (FNN) to enhance the prediction accuracy of CO and NO_x emissions. The PSO method optimizes the FNN weights, improving prediction capabilities. A unique feature of the PSO is its integration of a K-Nearest Neighbor (KNN) algorithm in its random number selection strategy, aiming to minimize prediction errors. Constructed, trained, and validated using publicly accessible datasets, the model demonstrated significant improvements in prediction accuracy, evidenced by low values of Mean Square Error (MSE), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE). The model's efficacy was further validated through sensitivity analysis of hyperparameters and comparisons with conventional models like multiple linear regression and standalone neural networks. These tests confirmed the superior predictive accuracy and reliability of our hybrid model, suggesting its potential as a valuable tool for optimizing operational efficiency and environmental compliance in power plants.

1. INTRODUCTION

Gas turbines, recognized for their versatility and efficiency, are pivotal in diverse sectors including oil and gas exploration, aviation, and power generation. These systems operate by compressing air and fuel, igniting the mixture to produce a high-temperature airflow that drives a turbine, which in turn generates electricity via a shaft generator. Despite their efficacy, the combustion process in gas turbines inevitably leads to the emission of pollutants such as carbon monoxide (CO) and nitrogen oxides (NO_x) due to incomplete combustion reactions. These emissions contribute to environmental and health issues, including ozone depletion, alterations in atmospheric chemistry [1], acid rain [2], and oxygen depletion, which could otherwise react with gaseous fuels [3].

The impact of carbon monoxide is twofold: directly, by affecting respiratory health—manifesting as headaches, fatigue, dizziness, weakness, and mental disorientation [4]; and indirectly, by contributing to broader global challenges such as acid rain, climate change, and global warming [5, 6]. Mitigating these emissions is thus crucial not only for environmental preservation but also for public health. Given these concerns, developing precise and efficient models for predicting gas emissions from turbines is vital. Such models

not only facilitate the creation of cleaner energy production technologies but also enhance operational efficiency and environmental compliance. This paper proposes a machine learning-based approach to predict CO and NO_x emissions, integrating ambient weather conditions, emission outputs, and gas turbine operational parameters. By employing advanced algorithms such as Neighbor Component Analysis (NCA), the model assesses the correlation between process parameters and emission levels, thereby identifying critical factors that influence emissions. This methodology is supported by conventional statistical techniques to validate the performance of the proposed machine learning models, ensuring robust and reliable predictions.

2. LITERATURE SURVEY IN GAS TURBINE EMISSION PREDICTIONS

Monitoring industrial plant emissions, such as flue gas emissions from turbines, especially CO and NO_x emissions, are of increasing environmental and regulatory concern, and there is an increasing demand for monitoring them by estimation and measurement precision. While direct measurement is an effective approach, such measurements often involve high maintenance, equipment and operating

costs. As a result, there has been much interest in estimating or predicting these emissions using other, simpler-to-measure variable, such as temperature and pressure in various parts of the turbine.

Many emission prediction approaches have been developed by many researchers, including but not limited to Zhao et al. [7], who developed a numerical model for predicting CO and NO_x emissions using computational fluid dynamics (CFD). Numerical simulations were used to determine the NO_x formation mechanism and CO behavior concerning changes in inlet air temperature and evaluating the model's ability to forecast the isothermal flow field inside the combustor by contrasting the numerical outcomes with the atmospheric observations test rig for the primary and dilution zones' velocities. Said et al. [8] implemented a statistical method to optimize the diesel engine NO_x and CO emissions prediction using a Box-Behnken design based on response surface methods.

On the other hand, data-driven approaches are embraced because they can predict outcomes by learning from data. This simplifies calculations and reduces the possibility that the physics laws will be missing during model development. Data-driven approaches commonly use three approaches: supervised, unsupervised, and semi-supervised. In emissions prediction, supervised learning is mainly used to predict the amount of emissions. Gordon et al. [9] and Faqih et al. [10] implemented a Support Vector Machine (SVM) regressor for NO_x emission prediction for diesel engines and gas turbines, respectively. Tuttle et al. [11] used SVM and a Neural Network (NN) to categorize and predict fuel emissions. Coelho et al. [12] also used many methods to estimate carbon oxides (CO) and nitrogen oxides (NO_x) emissions from a gas turbine using the predictive emission monitoring systems dataset. First, four methods were developed for feature generation: Principal Component Analysis, t-Distributed Stochastic Neighbor Embedding, Uniform Manifold Approximation and Projection, and Potential of Heat-diffusion for Affinity-based Trajectory Embedding. Then Feature Relevance-based Unsupervised Feature Selection was evaluated for ranking features. With all features generated, the regression models Ridge Regression, Least Absolute Shrinkage and Selection Operator, k-Nearest Neighbor, Cubist Regression, Random Forest, Light Gradient Boosting Machine, Categorical Boosting, and Deep Forest Regression (DFR) were evaluated. The hyperparameters were tuned with Randomized Search Cross Validation using a 5-fold cross-validation (CV) procedure. Rezazadeh [13] used a non-parametric supervised method known as k-Nearest Neighbor (KNN) to predict NO_x.

Si et al. [14], who employed neural networks to predict NO_x for conventional and Dry Low Emissions (DLE) gas turbines, respectively. Furthermore, Qader et al. [15] showed that neural network nonlinear autoregressive time series and Gaussian Process Regression (GPR) gave the best results in predicting CO₂ emissions from power plants located in Bahrain. Further research on gas turbine emissions prediction using Artificial Neural Networks (ANNs) was carried out by Zhao et al. implement. [16], using a Recurrent Neural Network (RNN) to predict Turbine Exhaust Pressure (TEP) and turbine exhaust temperature (TET). While Kaya et al. [17] and Sun and Huang [18] predicted NO_x and CO exhaust gases using Extreme Learning Machines (ELM). Bhowmik et al. [19] developed an ANN model for mapping ternary blends with engine power performance and exhaust emissions in diesel engines. Adams

et al. [20] showed that Deep Neural Networks (DNN) could effectually reduce the computation time for SO_x and NO_x emissions prediction through data preprocessing and feature selection methods. The ensemble DNN model demonstrated higher prediction ability in predicting NO_x emissions by Wang et al. [21]. A fuzzy inference system with an adaptive network as a type of artificial neural network has been verified by dos Santos Coelho to predict NO_x emissions [2] accurately. Furthermore, AlKheder and Almusalam [22] found a Deep Learning-based FNN model suitable for predicting CO₂ emissions for a specific energy sector in Kuwait; Ağbulut's [23] effectively implemented SVM and DNN models for predicting CO₂ emissions in Turkey.

However, from the above work, it has been noted all the metaheuristics algorithms coupled with the machine learning approaches have never utilized the Particle Swarm Optimization (PSO) algorithm for tuning the machine learning approach weights for achieving more accurate prediction of gas emission in natural-gas power plants and hence, the aim of this work was established. It should not be ignored that gas emissions prediction plays a vital role in environmental management and public health protection. Therefore, a new hybrid model of FNN-based Improved PSO is introduced to predict the emissions of carbon monoxide (CO) and nitrogen oxides (NO_x) from gas turbines.

The contribution of this paper is as follows:

- To develop an objective comparison of different machine learning methods based on characteristics of open gas turbine CO and NO_x emission datasets.
- The CO and NO_x emission prediction error can be minimized by developing a model based on hidden layer FNN and improved PSO algorithm.
- A KNN algorithm is introduced to obtain the optimal weight for FNN and improve model performance by randomly selecting the velocity value in the velocity equation of the PSO algorithm.

Here are the remaining sections of the paper: Section 3 introduces the theoretical background of FNN models, architecture optimization, and the proposed PSO algorithm. Section 4 offers the results and discussions, followed by a comparison study with available hybrid FNN models used in previous works on the same dataset. As a final word, in section 5, we present the conclusions of our analysis and discuss future directions.

3. METHODOLOGY

This section details the methodology, which consists of the FNN model used for emission prediction, the improved PSO used for FNN weights tuning, and the performance matrices to evaluate model performance.

3.1 The FNN prediction model

Feedforward Neural Networks are well-known for their capability to learn complicated problems and provide solutions through learning the (hidden) underlying structural relations of the input strings [24]. For basic single hidden layer FNN (*net*) illustrated in Figure 1, supervised learning would begin by using inputs vector $r = [r_1, r_2, r_3, \dots, r_i]$, and targets vector $t = [t_1, t_2, t_3, \dots, t_i]$. Achieving output with the minimal possible error requires adjusting the weight coefficients (w) between layers [25]. The correlation between

the output and target vectors will be used to identify errors.

$$R = net(r) \quad (1)$$

$$R = w \times r + B \quad (2)$$

R represents the output vector, B is the model bias, and r is a random variable. Therefore, the net can adjust the w coefficient to obtain the highest correlation between output and target vectors [26]. In other words, the main goal of the learning process is to find the minimum of Eq. (4).

$$e = R - t \quad (3)$$

$$MSE = \frac{\sum_{n=1}^i e(n)^2}{i} \quad (4)$$

where, e represents the error vector, MSE is regarded as a measure of training/learning performance [27].

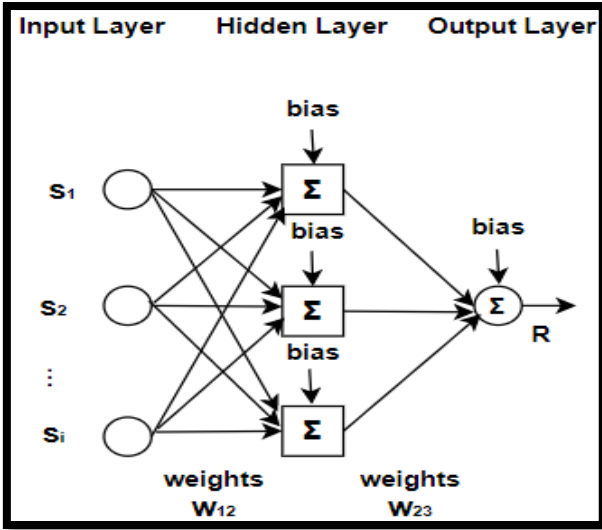


Figure 1. Structure of the FNN layers showing the input, weights/biases, and output

3.2 Improved PSO (IPSO) method

In applications in science and engineering, the PSO algorithm handles multidimensional problems with remarkable efficiency. The PSO's heuristics are modeled after the social and biological ways that birds forage for food [24]. Using the assumption that the population is on the y -axis and consists of N particles, standard PSO iteratively changes the particle's position and velocity to ascertain which particle in the swarm is the best [28]. The i^{th} particle in the population is denoted by m_i , and its position within the population is indicated by p_i . As a result, the following represents the location of a particle moving in the y -dimension:

$$p_i = (p_{i1}, p_{i2}, p_{i3}, \dots, p_{iy}) \quad (5)$$

The m_i Particle updates its position as it moves through the population region with velocity. v_i , which is given by the following vector:

$$v_i = (v_{i1}, v_{i2}, v_{i3}, \dots, v_{iy}) \quad (6)$$

Therefore, PSO can try to evaluate the optimal position of particles m_i in the population and pass them into a vector, as Eq. (7):

$$o_i = (o_{i1}, o_{i2}, o_{i3}, \dots, o_{iy}) \quad (7)$$

Other terms commonly used in PSO include social and cognitive acceleration constants (c_2 and c_1) and inertial weights (W) [29]. To mathematically represent the additional PSO parameters, the inertia weights are first expressed as:

$$W = W_{\min} + \left[\frac{k}{K} \times (-W_{\min} + W_{\max}) \right] \quad (8)$$

$$v_{ix}^{k+1} = r_1^k c_1 (o_{iy} - p_{iy}^k) + r_2^k c_2 (u_{gy}^k - p_{iy}^k) + v_{iy}^k W \quad (9)$$

$$p_{iy}^{k+1} = p_{iy}^k + v_{iy}^{k+1} \quad (10)$$

Among them, K represents the maximum number of iterations, r_1, r_2 , and r_3 are random numbers with a value ranging between (0,1) [27]. Population generation represents the first step in PSO optimization. In order to execute the PSO algorithm, parameters include population size (swarm) (N), random distribution number (r_1, r_2), social and cognitive coefficients (c_1, c_2), global best performance (GP), and inertia weight coefficient (W) must be set [30]. PSO searches for the weight (particles) while maintaining the most accurate approximation of the fitness function [31]. Therefore, the gas leak-based prediction model uses the PSO algorithm to find the optimal weight coefficients in the FNN model to improve the leak prediction accuracy. This study enhances PSO performance by tuning the velocity coefficients using KNN algorithm as a third-party regressor. The velocity of the i^{th} particle is updated using KNN to get the optimal position. Taking v^i to be the velocity of the i^{th} particle at $t = t_0$; therefore, the velocity at $t = t_1$ can be expressed as Eq. (11).

$$v_i^{t1} = W \times v_i^{t0} + c_1 \times R \times p_i^{diff} + c_2 \times R \times p_i^{diff} \quad (11)$$

In order to determine the optimal velocity, the KNN algorithm is used to guess the v_i^{t1} , we use the following:

(1) According to the v_i^{t0} , the v_i^{t1} should be assigned to the optimal velocity among its nearest neighbors. If the k is an even number, then the v_i^{t1} is the optimal velocity of its nearest neighbor.

(2) It is recommended by the random rule that the v_i^{t1} is assigned to the optimal velocity among the nearest neighbors. In that case, the k would be an even number, and the nearest neighbors' v_i^{t1} would be decided at random.

$$p_{diff} = p_i - p_i^b \quad (12)$$

where, R represents a random variable, and p_i^b represents the best position of the i^{th} particle. In order to determine the optimal velocity, the KNN algorithm is used to guess the v_i^{t1} . KNN was used to choose the optimal random variables r_1 and r_2 for updating the velocity. A pseudo-random number generator (PRG) is used to generate a reproducible stream of random numbers directly dependent on the initial seed used to

select r_1 and r_2 . However, most classical random number generators cannot run because they must define a current state and a seed value. In our case, the weights that give the highest accuracy obtained from plain FNN are considered initial seeds. In this case, the other seeds can be obtained from the initial seeds. The applied seed number generator demonstrated as Eq. (13):

$$SeedRG = [length(optimum\ weight), optimum\ weight\ of\ FNN] + PRG(r_1, r_2, [1,1]) \quad (13)$$

Figure 2 below shows the IPSO process, while Table 1 describes the setups of the IPSO algorithm.

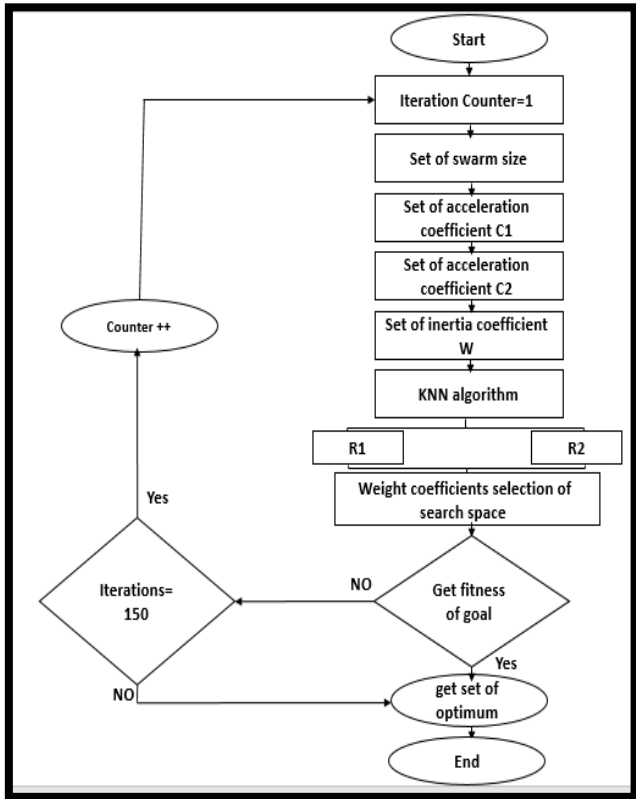


Figure 2. IPSO optimization flow diagram

Table 1. Value of IPSO configuration parameters

Parameter	Value
Swarm (Population) Size (s)	100
Social Constants (c1)	2
Cognitive Constants (c2)	2
Inertia Weight Coefficient (W)	1
No. of Iterations	150

Eq. (14) provides the fitness function for the IPSO mentioned above.

$$[Wiegth]_{FNN} = arg_{\{R1, R2\}} [R1, R2, min\{\frac{\sum_{i=1}^I (net(x[i]) - T)^2}{I}\}] \quad (14)$$

where, i is the total input vectors, $X(i)$ is the i^{th} input array, and T is the predicted target. New configurations, as given in Table 2, are made, and the IPSO algorithm is used to optimize the FNN model results. IPSO is used to generate the best weight

coefficients and apply those coefficients to the FNN model. Figure 3 demonstrates the Improved PSO-FNN paradigm.

Table 2. IPSO-FNN model configurations

Parameter	Values
Number of hidden layers	Single (1)
Goal training performance (MSE)	$1 \times e^{-201}$
Training model	Levenberg-Marquardt algorithm (LM)
Minimum gradience	$1 \times e^{-101}$
Maximum fails	100
Epochs	25
Training time goal (seconds)	30
PSO-search space upper bound	+1.09
PSO-search space lower bound	-1.09

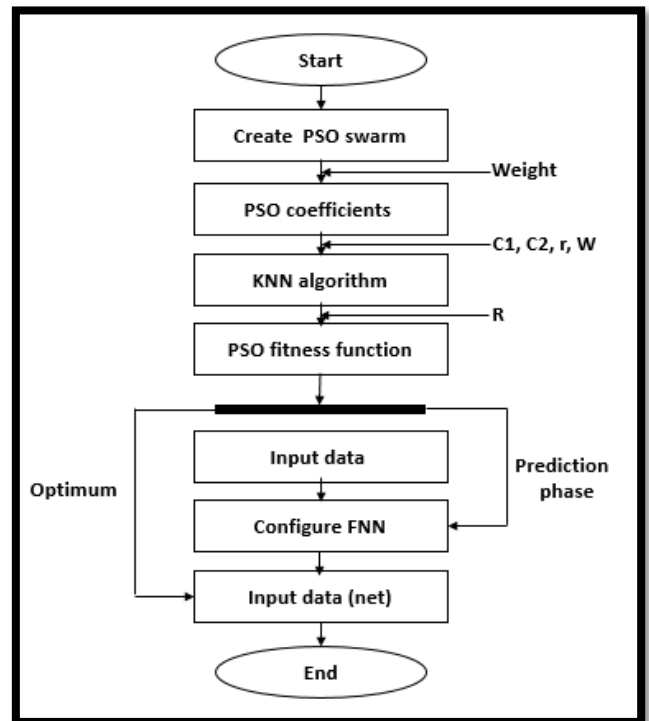


Figure 3. An overview of the proposal IPSO-FNN procedure

3.3 Measurements of performance

The hybrid IPSO-FNN model is used to predict both CO and NO_x gas leakages. For such configurations, performance metrics are determined as Eq. (15):

$$E = T - TR \quad (15)$$

$$MSE = \sum_{n=1}^K \frac{E[k]^2}{K} \quad (16)$$

$$RMSE = \sqrt{\frac{\sum_{n=1}^i e(n)^2}{i}} \quad (17)$$

$$MAE = \sum_{n=1}^K \frac{|E[K]|}{K} \quad (18)$$

If assumed that T is the optimum label of the predictions files and TR is the predicted (classifier results) label, Eq. (15) expresses the error relationship. However, the non-zero values in the error vector represent the number of errors in the prediction results; otherwise, the zero values represent the correct decisions in the prediction results. MSE, RMSE, and MAE can be given the following equations: K is the total number of errors, and the error count propagates from 1 to K.

The accuracy can be calculated as Eq. (19):

$$ACCUR = \frac{\text{Number of correct decisions}}{\text{Total number of outputs}} \quad (19)$$

4. FEATURE SELECTION AND PREPROCESSING OF DATASETS

This section discusses dataset preprocessing methods, FNN, and PSO algorithms for CO and NOx emissions prediction. According to Bhowmik et al. [19], CO and NOx leakage can be tracked by observing the entire gas turbine operation. In addition, gas-fired power generation involves natural air going through five stages, beginning with air filtration and ending with reducing the air temperature. Each stage has inlet and outlet variables such as inlet and outlet pressure, inlet and outlet temperature, ambient temperature, ambient pressure, etc. In order to maintain the power generation performance of the gas turbine, these variables should be kept at known levels [24]. The leakage levels of gases such as CO and NOx are related to the gas turbine mechanical limitations mentioned above. The dataset used to build the predictive model includes the factors that affect the condition of the air and those that affect gas turbine operating conditions. The gas leakage level from the turbine entities is the “amount of gas turbine emissions”.

4.1 Preprocessing methods for datasets

The dataset encountered several preprocessing stages before using it in prediction operations. In order to be familiar with the preprocessing procedures, it is essential to conduct statistical analysis on this dataset. This dataset represents the features from the machine learning point of view, so those features need to be reformed to maintain forecasting with reliable decision-making. After one year of monitoring and data capturing, statistical analysis is performed, and the results are presented in Tables 3 to 6. Firstly, the level of CO leakage is studied Table 3, it was obtained that the maximum leakage of CO gas is 41.097mg/m³, and the minimum leakage of it is 0.2128mg/m³. The median and mean of CO leakage are measured too.

The dataset used to build the predictive model includes the factors that affect the condition of the air and those that affect gas turbine operating conditions. The gas leakage level from the turbine entities is the “amount of gas turbine emissions”. The data presented in this study are openly available [32].

The given values in Tables 3 and 4 reflect the amount of CO and NOx emission in mg/m³. In order to understand the amount of emission during the year in the test data, the frequency of a particular gas emission level every hour for 7385 hours is calculated in Tables 5 and 6.

Knowing that leakage levels of two different gases are being

monitored on this dataset, monitoring is taking place on the biases of nine parameters, e.g., Ambient Temperature (AT), Air Filter Difference Pressure (AFDP), Ambient Humidity (AH), Turbine Inlet Temperature (TIT), Ambient Pressure (AP), Turbine Energy Yield (TEY), Gas Turbine Exhaust Pressure (GTEP), Compressor Discharge Pressure (CDP), Turbine After Temperature (TAT)). NO_x and CO emission ranges are 25.905mg/m³-119.69mg/m³ and 0.2128mg/m³-41.09mg/m³, respectively. From Tables 3 to 6, the NO_x and CO range widely vary between the minimum and maximum ranges. Thus, both targets are represented in more abstract forms, so each gas leakage is labelled according to the emission ranges in Tables 5 and 6.

Table 3. CO emission levels value analysis

Min	0.2128
Max	41.097
Med	2.533
Mean	3.1287867

Table 4. NOx emission levels value analysis

Min	25.905
Max	119.68
Med	56.8385
Mean	59.8905095

Table 5. CO emission distribution over 7385 hours

Emission Range mg/m ³	BINS (label)	Frequency
<1	1	184
1.1-2	2	2342
2.1-5	5	3970
5.1-10	10	740
10.1-15	15	134
15-20	20	7
	More	7

Table 6. NOx emission distribution over 7385 hours

Emission Range mg/m ³	BINS (Label)	Frequency
<30	30	1
31-40	40	5
41-50	50	977
51-60	60	3611
61-70	70	1665
71-80	80	687
	More	438

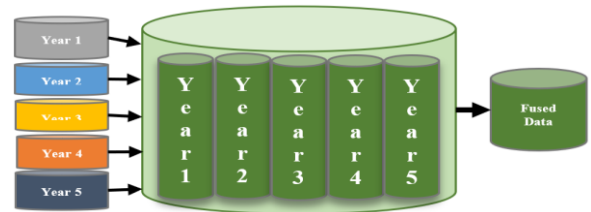


Figure 4. Preprocessing data augmentation

Five years of monitoring data are obtained, and similar features are monitored yearly. Features fusion involves performing the so-called data augmentation. The augmentation process is one of the measures taken to uplift prediction models' accuracy. Features from each year are fused so that essential elements are augmented. Augmentation

provides more possibilities for the same qualities in the prediction model, supporting the so-called training. Figure 4 demonstrates the augmentation process. Furthermore, data post-augmentation is normalized to reduce the variance between data elements. Normalization takes place by identifying the maximum value in each feature and then dividing each weight of that feature by this maximum value.

4.2 Feature selection stage

The so-called features selection will be used to reduce the number of features in the prediction model in order to lessen the burden on it. There are various approaches that can be used in the features selection strategy. The goal of all feature selection techniques is to create a shortlist of features that are more relevant to the target. All other traits, however, are insignificant because they only have a slight effect on the target. As was seen in the previous section, the target vectors' statistical examination of each feature can produce the necessary data for feature selection. The correlation between each feature and the goal (s) is also ascertained using an algorithm such as Neighbor Component Analysis (NCA). Two types of input are accepted by neighbor component analysis: the target and the features to be chosen (abstracted). In order to maximize the output of the correlation calculation, NCA computes the characteristics. The correlation coefficient factor is represented by the algorithm's output. Using our case study as an example, we can feed the Neighbor component analysis algorithm nine columns of features in addition to the first target, which is the CO levels. The algorithm will generate nine coefficients, each of which has a different correlation level with the output target. Procedures are repeated with a different objective, for example, applying NOx levels rather than CO. The important thing to remember is that each time the algorithm runs, the number of characteristics included in the correlation formula is changed, resulting in a random determination of those coefficients. The outputs in Tables 7 and 8 are obtained after three consecutive algorithm runs for CO and NOx gases, taking into account the gas leakage dataset.

Table 7. Weights for three neighbor component analysis trails with CO target

Iteration No.	Iteration 1	Iteration 2	Iteration 3
AP	49.50	57.34	55.12
TIT	40.87	44.91	43.65
TAT	38.61	50.92	49.01
AFDP	23.94	21.25	32.34
GTEP	17.57	28.79	30.84
CDP	16.76	2.34	2.35
AT	10.44	12.73	17.13
AH	7.28	8.93	11.01
TEY	8.31	1.87	1.19

Table 8. Weights for three neighbor component analysis trails with NO_x target

Iteration No.	Iteration 1	Iteration 2	Iteration 3
AP	31.40	28.36	25.41
TIT	23.86	22.08	20.11
TAT	19.63	21.07	20.75
AFDP	17.33	19.51	20.45
CDP	15.20	16.83	17.84
AT	13.62	10.33	8.46
GTEP	11.49	15.35	17.33
AH	7.28	8.74	8.90
TEY	2.72	2.31	1.99

The maximum obtained coefficient features are the best input to train the ANN model. Hence, the highly correlated features with the emission of the two gases in Table 7 (AP, TIT, TAT, AFDP, GTEP, and CDP) and Table 8 (AP, TIT, TAT, AFDP, CDP, and AT) are considered as the number of features column in the dataset.

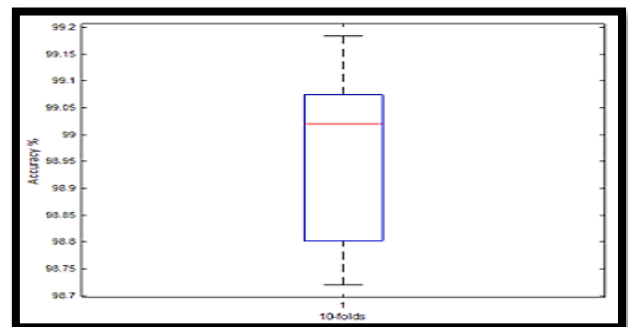
5. ANALYSIS OF RESULTS AND DISCUSSION

IPSO improves the plain model's (FNN, for example) gas leakage forecast ability. As a result of this integration, the model's performance for gas prediction was enhanced. Table 9 presents the accuracy metrics of CO and NO_x gas leakage by an 11-cross-fold validation algorithm. The model is developed using MATLAB/Simulink (R2022b) environment.

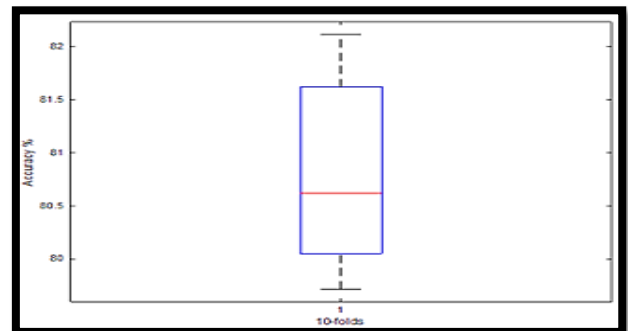
Based on the results shown in Table 9, fold number one indicates the highest potential for prediction accuracy, which is 99.68 percent for CO and 82.62 percent for NO_x. The accuracy measure for CO Figure 5 (a) and NO_x Figure 5 (b) is shown using a box plot for each fold.

Table 9. Accuracy measures per fold for CO and NO_x gas leakage using the IPSO-FNN model

Fold Number	Accuracy (CO)	Accuracy (NO _x)
1	99.68	82.61
2	99.24	81.22
3	99.57	81.23
4	99.46	80.54
5	99.30	80.43
6	99.62	82.62
7	99.30	80.92
8	99.22	80.21
9	99.57	82.12
10	99.58	81.00
11	98.89	80.89



(a)



(b)

Figure 5. An illustration of the accuracy measures per-fold (a) CO and (b) NO_x gas leakage using the IPSO-FNN model

Boxplots, where the red lines within the boxes indicate the median values, the blue boxes represent the lower and upper quartiles, the black lines represent the minimum and maximum accuracy values, and the dotted lines represent outliers, show the differences between the actual and predicted values of 11-fold cross-validation for CO and NOx emission.

6. COMPARISON STUDY

This section's goal is to provide evidence for the suggested hybrid model's superiority over other pertinent, related models. Comparing the outcomes of the several models suggested in earlier studies using the same dataset with the model proposed

in this work, Table 10 presents the findings.

The similarity mentioned above in Table 10 as the acquired performance matrix results show that the upgraded PSO model may greatly increase the plain FNN model's performance, Table 10 shows that the IPSO-FFN model performs relatively better than other models. Our state-of-the-art results show improvements in the MSE, RMSE, and MAE for CO gas leakage prediction of 87.52 to 98.79%, 62.35 to 98.38%, and 46.73 to 99.84%, respectively, when compared to other previously presented models in [17, 33–45]. On the other hand, there is an improvement of 4.25 to 73.33%, 30.95 to 95.71%, and 59.75 to 98.22% in the MSE, RMSE, and MAE for NOx gas leakage prediction.

Table 10. Comparison study

Reference	CO Prediction Performances				NO _x Prediction Performances			
	MSE	RMSE	MAE	Improvement%	MSE	RMSE	MAE	Improvement%
Our Proposal (IPSO-FNN) algorithm	0.13	0.23	0.12	—	0.30	0.54	0.29	—
Kaya et al. [17]	—	—	0.93	MAE by 98.94 MSE by 95.71	—	—	7.91	MAE by 97.72
Dirik [33]	0.265	0.515	—	RMSE by 79.27	—	—	—	—
Etemadi and Khashei [34]	0.670	—	0.532	MSE by 98.29 MAE by 98.15	0.670	—	0.532	MSE by 72.83 MAE by 66.19
Wei et al. [35]	—	0.0751	—	—	—	0.0751	—	—
Tessoni and Amoretti [36]	—	—	0.018	MAE by 46.73	—	—	0.018	—
Dong et al. [37]	0.0846	—	1.367	MSE by 86.52 MAE by 99.34	0.084	—	1.367	MAE by 86.83
Nino-Adan et al. [38]	—	0.284	0.364	RMSE by 62.35 MAE by 97.30	—	1.220	1.559	RMSE by 65.02 MAE by 88.46
Leblanc and Germain [39]	0.190	—	—	MSE by 94.00	0.190	—	—	MSE by 4.25
Wood [40]	—	1.499	0.894	RMSE by 89.53 MAE by 98.90	—	8.921	6.613	RMSE by 95.21 MAE by 97.27
Mahfuz et al. [41]	—	0.475	—	RMSE by 77.53	—	—	—	—
Awasthi et al. [42]	—	1.454	0.774	RMSE by 92.65 MAE by 98.73	—	1.454	0.774	RMSE by 70.65 MAE by 76.75
Berikov and Litvinenko [43]	—	—	—	—	—	—	0.634	MAE by 71.62 MSE by 52.32
Botsas et al. [44]	—	—	—	—	0.382	0.618	0.447	RMSE by 30.95 MAE by 59.75
Potts and Leontidis [45]	—	—	0.50	MAE by 98.22	—	—	2.51	MAE by 92.83

7. CONCLUSIONS AND FUTURE WORKS

In order to predict gas emissions from natural gas power plants, this paper developed a hybrid predictive emissions monitoring model based on Particle Swarm and Neural Network methods. Optimal weights for neural network training were found by combining FNN with the enhanced PSO method to create an integrated predictive model. The velocity generation and, consequently, the selection quality of the FNN weights coefficients are improved in the PSO with the application of the KNN algorithm. Using k-fold cross-validation, the dataset is partitioned elevenfold to assess the model's performance under various feed existence. Statistical performance metrics such as accuracy, MSE, RMSE, and MAE were used to evaluate the prediction ability of the integrated model. Statistical metrics of the IPSO-FNN model for CO gas emissions are Accuracy=99.68, MSE, RMSE, and MAE equal to 0.13, 0.23, and 0.12, respectively. On the other hand, NOx gas emissions had Accuracy=82.62 and MSE, RMSE, and MAE values of 0.30, 0.54, and 0.29, respectively. A Feedforward Neural Network with enhanced PSO performed better on prediction than one without enhanced PSO. In addition, Table 10 shows that hybrid PSO algorithm

performance can be improved when compared with various performance metrics.

Developing a mechanism for assessing the technological viability and economics of the proposed approach for reducing CO, NOx, SOx, and other important emissions will be the focus of future work, according to Islam et al. [46].

REFERENCES

- [1] Terrenoire, E., Hauglustaine, D., Cohen, Y., Cozic, A., Valorso, R., Lefèvre, F., Matthes, S. (2022). Impact of present and future aircraft NO_x and aerosol emissions on atmospheric composition and associated direct radiative forcing of climate. *Atmospheric Chemistry and Physics Discussions*, 2022: 1-39. <https://doi.org/10.5194/acp-22-11987-2022>
- [2] Pachauri, N. (2024). An emission predictive system for CO and NOx from gas turbine based on ensemble machine learning approach. *Fuel*, 366, 131421.
- [3] Ilbas, M., Kumuk, O., Karyeyen, S. (2022). Modelling of the gas-turbine colorless distributed combustion: An application to hydrogen enriched-kerosene fuel.

- International Journal of Hydrogen Energy, 47(24): 12354-12364.
<https://doi.org/10.1016/j.ijhydene.2021.06.228>
- [4] Li, Y., Zhu, Q., Wei, T. (2023). Threshold effects of population aging on carbon emissions: From the perspective of industrial structure and residential consumption. *Science of The Total Environment*, 891: 164654. <https://doi.org/10.1016/j.scitotenv.2023.164654>
- [5] Bakay, M.S., Ağbulut, Ü. (2021). Electricity production-based forecasting of greenhouse gas emissions in Turkey with deep learning, support vector machine and artificial neural network algorithms. *Journal of Cleaner Production*, 285: 125324. <https://doi.org/10.1016/j.jclepro.2020.125324>
- [6] Derafshi, M., Asgari Lajayer, B., Hassani, A., Dell, B. (2023). Effects of acidifiers on soil greenhouse gas emissions in calcareous soils in a semi-arid area. *Scientific Reports*, 13(1): 5113. <https://doi.org/10.1038/s41598-023-32127-0>
- [7] Zhao, Q., Liu, F., Jiao, A., Yang, Q., Xu, H., Liao, X. (2023). Prediction model of NOx emissions in the heavy-duty gas turbine combustor based on MILD combustion. *Energy*, 282: 128974. <https://doi.org/10.1016/j.energy.2023.128974>
- [8] Said, Z., Le, D.T.N., Sharma, P., Dang, V.H., Le, H.S., Nguyen, D.T., Bui, T.A.E. (2022). Optimization of combustion, performance, and emission characteristics of a dual-fuel diesel engine powered with microalgae-based biodiesel/diesel blends and oxyhydrogen. *Fuel*, 326: 124987. <https://doi.org/10.1016/j.fuel.2022.124987>
- [9] Gordon, D., Norouzi, A., Blomeyer, G., Bedei, J., Aliramezani, M., Andert, J., Koch, C.R. (2023). Support vector machine-based emissions modeling using particle swarm optimization for homogeneous charge compression ignition engine. *International Journal of Engine Research*, 24(2): 536-551. <https://doi.org/10.1177/14680874211055546>
- [10] Faqih, M., Omar, M.B., Ibrahim, R. (2023). Prediction of dry-low emission gas turbine operating range from emission concentration using semi-supervised learning. *Sensors*, 23(8): 3863. <https://doi.org/10.3390/s23083863>
- [11] Tuttle, J.F., Blackburn, L.D., Powell, K.M. (2020). On-line classification of coal combustion quality using nonlinear SVM for improved neural network NOx emission rate prediction. *Computers & Chemical Engineering*, 141: 106990. <https://doi.org/10.1016/j.compchemeng.2020.106990>
- [12] Coelho, S., Leandro, Helon, V. H. A., Viviana, C. M. CO and NOx emissions prediction in gas turbine using a novel modeling pipeline based on the combination of deep forest regressor and feature engineering. *Fuel*, 355 (2024): 129366.
- [13] Rezazadeh, A. (2021). Environmental pollution prediction of NOx by predictive modelling and process analysis in natural gas turbine power plants. *Pollution*, 7(2): 481-494. <https://doi.org/10.22059/poll.2021.316327.977>
- [14] Si, M., Tarnoczi, T.J., Wiens, B.M., Du, K. (2019). Development of predictive emissions monitoring system using open-source machine learning library-keras: A case study on a cogeneration unit. *IEEE Access*, 7: 113463-113475. <https://doi.org/10.1109/ACCESS.2019.2930555>
- [15] Qader, M.R., Khan, S., Kamal, M., Usman, M., Haseeb, M. (2021). Forecasting carbon emissions due to electricity power generation in Bahrain. *Environmental Science and Pollution Research*, 1-12. <https://doi.org/10.1007/s11356-021-16960-2>
- [16] Zhao, F., Chen, L., Xia, T., Ye, Z., Zheng, Y. (2019). Gas turbine exhaust system health management based on recurrent neural networks. *Procedia CIRP*, 83: 630-635. <https://doi.org/10.1016/j.procir.2019.04.122>
- [17] Kaya, H., Tüfekci, P., Uzun, E. (2019). Predicting CO and NOx emissions from gas turbines: Novel data and a benchmark PEMS. *Turkish Journal of Electrical Engineering and Computer Sciences*, 27(6): 4783-4796. <https://doi.org/10.3906/elk-1807-87>
- [18] Sun, W., Huang, C. (2022). Predictions of carbon emission intensity based on factor analysis and an improved extreme learning machine from the perspective of carbon emission efficiency. *Journal of Cleaner Production*, 338: 130414. <https://doi.org/10.1016/j.jclepro.2022.130414>
- [19] Bhowmik, S., Paul, A., Panua, R., Ghosh, S.K., Debroy, D. (2018). Performance-exhaust emission prediction of diesosanol fueled diesel engine: An ANN coupled MORSM based optimization. *Energy*, 153: 212-222. <https://doi.org/10.1016/j.energy.2018.04.053>
- [20] Adams, D., Oh, D.H., Kim, D.W., Lee, C.H., Oh, M. (2020). Prediction of Sox-NOx emission from a coal-fired CFB power plant with machine learning: Plant data learned by deep neural network and least square support vector machine. *Journal of Cleaner Production*, 270: 122310. <https://doi.org/10.1016/j.jclepro.2020.122310>
- [21] Wang, Y., Yang, G., Xie, R., Liu, H., Liu, K., Li, X. (2021). An ensemble deep belief network model based on random subspace for NO x concentration prediction. *ACS Omega*, 6(11): 7655-7668. <https://doi.org/10.1021/acsomega.0c06317>
- [22] AlKheder, S., Almusalam, A. (2022). Forecasting of carbon dioxide emissions from power plants in Kuwait using united states environmental protection agency, intergovernmental panel on climate change, and machine learning methods. *Renewable Energy*, 191: 819-827. <https://doi.org/10.1016/j.renene.2022.04.023>
- [23] Ağbulut, Ü. (2022). Forecasting of transportation-related energy demand and CO2 emissions in Turkey with different machine learning algorithms. *Sustainable Production and Consumption*, 29: 141-157. <https://doi.org/10.1016/j.spc.2021.10.001>
- [24] Volponi, A.J. (2014). Gas turbine engine health management: past, present, and future trends. *Journal of Engineering for Gas Turbines and Power*, 136(5): 051201. <https://doi.org/10.1115/1.4026126>
- [25] Benyounes, A., Hafaifa, A., Guemana, M. (2016). Gas turbine modeling based on fuzzy clustering algorithm using experimental data. *Applied Artificial Intelligence*, 30(1): 29-51. <https://doi.org/10.1080/08839514.2016.1138808>
- [26] Zhang, Y., Bingham, C., Garlick, M., Gallimore, M. (2017). Applied fault detection and diagnosis for industrial gas turbine systems. *International Journal of Automation and Computing*, 14: 463-473. <https://doi.org/10.1007/s11633-016-0967-5>
- [27] Dong, M., He, D. (2007). Hidden semi-Markov model-based methodology for multi-sensor equipment health diagnosis and prognosis. *European Journal of Operational Research*, 178(3): 858-878.

- <https://doi.org/10.1016/j.ejor.2006.01.041>
- [28] Lu, F., Jiang, J., Huang, J., Qiu, X. (2018). An iterative reduced KPCA hidden markov model for gas turbine performance fault diagnosis. *Energies*, 11(7): 1807. <https://doi.org/10.3390/en11071807>
- [29] Zhao, D., Wang, Y., Liang, D., Ivanov, M. (2020). Performances of regression model and artificial neural network in monitoring welding quality based on power signal. *Journal of Materials Research and Technology*, 9(2): 1231-1240. <https://doi.org/10.1016/j.jmrt.2019.11.050>
- [30] Matjanov, E. (2020). Gas turbine efficiency enhancement using absorption chiller. Case study for Tashkent CHP. *Energy*, 192: 116625. <https://doi.org/10.1016/j.energy.2019.116625>
- [31] Kwon, H.M., Kim, T.S., Sohn, J.L. (2018). Performance improvement of gas turbine combined cycle power plant by dual cooling of the inlet air and turbine coolant using an absorption chiller. *Energy*, 163: 1050-1061. <https://doi.org/10.1016/j.energy.2018.08.191>
- [32] UC Irvine Machine Learning Repository. (2023). <https://archive.ics.uci.edu/ml/datasets/Gas+Turbine+C+and+NOx+Emission+Data+Set>.
- [33] Dirik, M. (2022). Prediction of NOx emissions from gas turbines of a combined cycle power plant using an ANFIS model optimized by GA. *Fuel*, 321: 124037. <https://doi.org/10.1016/j.fuel.2022.124037>
- [34] Etemadi, S., Khashei, M. (2021). Etemadi multiple linear regression. *Measurement*, 186: 110080. <https://doi.org/10.1016/j.measurement.2021.110080>
- [35] Wei, X.J., Zhang, D.Q., Huang, S.J. (2022). A variable selection method for a hierarchical interval type-2 TSK fuzzy inference system. *Fuzzy Sets and Systems*, 438: 46-61. <https://doi.org/10.1016/j.fss.2021.09.017>
- [36] Tessonni, V., Amoretti, M. (2022). Advanced statistical and machine learning methods for multi-step multivariate time series forecasting in predictive maintenance. *Procedia Computer Science*, 200: 748-757. <https://doi.org/10.1016/j.procs.2022.01.273>
- [37] Dong, J., Chen, Y., Yao, B., Zhang, X., Zeng, N. (2022). A neural network boosting regression model based on XGBoost. *Applied Soft Computing*, 125: 109067. <https://doi.org/10.1016/j.asoc.2022.109067>
- [38] Nino-Adan, I., Portillo, E., Landa-Torres, I., Manjarres, D. (2021). Normalization influence on ANN-based models' performance: A new proposal for Features' contribution analysis. *IEEE Access*, 9: 125462-125477. <https://doi.org/10.1109/ACCESS.2021.3110647>
- [39] Leblanc, B., Germain, P. (2022). Seeking Interpretability and Explainability in Binary Activated Neural Networks. *arXiv preprint arXiv:2209.03450*.
- [40] Wood, D.A. (2023). Long-term atmospheric pollutant emissions from a combined cycle gas turbine: Trend monitoring and prediction applying machine learning. *Fuel*, 343: 127722. <https://doi.org/10.1016/j.fuel.2023.127722>
- [41] Mahfuz, A., Mannan, M.A., Muyeen, S.M. (2022). Enhancement of the HILOMOT algorithm with modified EM and modified PSO algorithms for nonlinear systems identification. *Electronics*, 11(5): 729. <https://doi.org/10.3390/electronics11050729>
- [42] Awasthi, P., Das, A., Sen, R., Suresh, A.T. (2021). On the benefits of maximum likelihood estimation for regression and forecasting. *arXiv Preprint arXiv: 2106.10370*. <https://doi.org/10.48550/arXiv.2106.10370>
- [43] Berikov, V., Litvinenko, A. (2021). Solving weakly supervised regression problem using low-rank manifold regularization. *arXiv Preprint arXiv: 2104.06548*. <https://doi.org/10.48550/arXiv.2104.06548>
- [44] Botsas, T., Mason, L.R., Pan, I. (2022). Rule-based Bayesian regression. *Statistics and Computing*, 32(3): 44. <https://doi.org/10.1007/s11222-022-10100-7>
- [45] Potts, R.L., Leontidis, G. (2023). Attention-based deep learning methods for predicting gas turbine emissions.
- [46] Islam, A., Teo, S.H., Ng, C.H., Taufiq-Yap, Y.H., Choong, S.Y.T., Awual, M.R. (2023). Progress in recent sustainable materials for greenhouse gas (NOx and SOx) emission mitigation. *Progress in Materials Science*, 132: 101033. <https://doi.org/10.1016/j.pmatsci.2022.101033>