

Article

# A Deep Learning Method for the Prediction of Pollutant Emissions from Internal Combustion Engines

Federico Ricci <sup>\*</sup>, Massimiliano Avana  and Francesco Mariani 

Department of Engineering, University of Perugia, Via G. Duranti 93, 06125 Perugia, Italy; massimiliano.avana@unipg.it (M.A.); francesco.mariani@unipg.it (F.M.)

\* Correspondence: federico.ricci@unipg.it

**Abstract:** The increasing demand for vehicles is leading to a rise in pollutant emissions across the world. This decline in air quality is significantly impacting public health, with internal combustion engines being a major contributor to this concerning trend. Ever-stringent regulations demand high engine efficiency and reduced pollutant emissions. Therefore, every automobile company requires rigorous methods for accurately estimating engine emissions. The implementation of advanced technologies, including machine learning methods, has proven to be a promising solution. The present work aims to develop an artificial intelligence-based model to estimate the pollutant emissions produced by an internal combustion engine under varying operating conditions. Experimental activities have been conducted on a single-cylinder spark ignition research engine with gasoline port fuel injection under both stationary and dynamic operating conditions. This work explores different artificial intelligence architectures and compares their performance in order to determine the best approach for the presented task. These structures have been trained and tested based on data obtained from the engine control unit and fast emission analyzer. The main target is to evaluate the possibility of applying the presented artificial intelligence predictive model as an on-board virtual tool in the estimation of emissions in real driving conditions.

**Keywords:** pollutant emissions; air quality; internal combustion engine; port fuel injection–spark ignition engine; artificial intelligence; machine learning; virtual sensor



**Citation:** Ricci, F.; Avana, M.; Mariani, F. A Deep Learning Method for the Prediction of Pollutant Emissions from Internal Combustion Engines. *Appl. Sci.* **2024**, *14*, 9707. <https://doi.org/10.3390/app14219707>

Academic Editor: Roberto Carballedo

Received: 9 September 2024

Revised: 26 September 2024

Accepted: 22 October 2024

Published: 24 October 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

To address the critical need for reducing air pollution from vehicles and improving air purity, increasingly stringent guidelines on pollutants and greenhouse gasses (GHGs) are pushing the advancement of greener and higher-performing internal combustion (IC) engines [1,2]. Cutting-edge after-treatment technologies, including optimized diesel/gasoline particulate filters (D/GPFs), selective catalytic reduction (SCR) systems incorporating injection of urea, and innovative techniques for catalyst warm-up, are successfully reducing the release of pollutants such as NO<sub>x</sub>, CO, unburned hydrocarbons (HCs), and particulate matter, to target even more stringent regulations [3,4]. Traditional spark ignition (SI) engines struggle to balance high performance with low emissions [5,6]. For modern SI engines, lowering fuel consumption entails implementing high boost levels with engine downsizing [7], as well as using water injection [8], de-NO<sub>x</sub> converters [9], lean mixtures and/or mixtures diluted with exhaust gas recirculation (EGR) [10]. It is essential to investigate modern combustion techniques, such as low-temperature combustions (LTCs) [11], enhance hybrid electric vehicle technologies to align with sustainable mobility specifications [12], and encourage the utilization of alternative and renewable fuels [13], such as methanol M100, ethanol E85 [14], and hydrogen H<sub>2</sub> [15,16]. Nevertheless, the adoption of these advanced technologies adds to the complexity of the engine and amplifies the volume of data that must be gathered from various physical sensors during both engine calibration and operational phases [17]. Consequently, considerable computational resources are necessary,

leading to extended operating times and higher costs. Furthermore, their performance progressively declines over time due to challenging conditions [18].

Sophisticated technologies, including machine learning (ML) methodology [19], are currently under investigation to efficiently track the parameters of the SI engine, with the aim of addressing the aforementioned challenges. The improvement of engine design depends on attaining optimal efficiency, which is essential in boosting engine performance while concurrently minimizing fuel consumption and reducing pollutant emissions [20,21]. Incorporating advanced technologies such as machine learning can enhance this optimization process. The ability of ML models to examine sophisticated dataset structures derived from input data renders them particularly suitable in predicting pollutant emissions in internal combustion engines. Data-driven approaches can enhance combustion strategies, leading to improved efficiency and lower emissions [22,23]. However, the direct on-board implementation of these measurements using physical instruments is challenging. Therefore, preliminary characterizations on specific test benches are necessary for proper engine calibration.

Moradi et al. [24] proposed the modeling of  $\text{NO}_x$  and HC raw emissions in a six-cylinder gasoline engine that operates under highly transient conditions with the utilization of machine learning approaches. The regression accuracy metric ( $R^2$ ) for the optimal model predicting  $\text{NO}_x$  is 0.98 for the training data and 0.97 for the test data. In contrast, the best model for predicting HC achieves values of 0.90 for the training data and 0.89 for the test data. Khac et al. [25] proposed models based on artificial neural networks (ANNs) for estimating  $\text{NO}_x$  and  $\text{CO}_2$  emissions from the in-cylinder pressure of a maritime diesel engine. The models utilize the Multi-Layer Perceptron (MLP) and Radial Basis Function (RBF) network architectures. The results demonstrate that the MLP model exhibits greater accuracy, with a low mean average percentage error (MAPE), in estimating both  $\text{NO}_x$  (MAPE = 4.39%) and  $\text{CO}_2$  (MAPE = 1.08%) compared to the RBF network (MAPE = 11.8% for  $\text{NO}_x$  prediction and MAPE = 14.2% for  $\text{CO}_2$  prediction). Godwin et al. [26] demonstrated the use of machine learning techniques to predict combustion, performance, and emission parameters in a dual-fuel SI engine operating on neat gasoline and E20 ethanol. While the ANN model exhibited reliable performance and accuracy, the ensemble least-squares boosting (ELSB) technique achieved an even higher degree of agreement with the experimental data. Cui et al. [27] developed a back propagation (BP) neural network model to predict ignition delays for three-component surrogates using the pressure, ambient temperature, and molar fractions of n-heptane, iso-octane, and toluene as inputs. Their model, trained with data from single- and two-component surrogates, successfully predicted ignition delays for Toluene Primary Reference Fuel (TPRF) surrogates. The neural network composed of two hidden layers outperformed one with a single layer, and optimization improved accuracy with a correlation coefficient above 0.9996. The model also significantly reduced the computation time and accurately predicted the Motor Octane Number (MON) and Research Octane Number (RON). It enabled the precise matching of real fuels to surrogate advanced combustion engines. Wright et al. [28] introduced a hybrid algorithm called physics-aware training that enables the use of backpropagation to train physical systems in situ. This approach integrates deep-learning techniques into controllable physical systems, allowing the training of physical neural networks made from optics, mechanics, and electronics. Their method addresses the challenge of applying backpropagation to unconventional hardware, demonstrating its effectiveness for tasks like audio and image classification. This innovation promises faster and more energy-efficient machine learning and opens up new avenues for automatically designed functionalities in robotics, materials, and smart sensors. Wang et al. [29] developed a model to predict transient  $\text{NO}_x$  emissions for heavy-duty diesel vehicles, addressing the issue of emission cheating during real-world driving. By conducting road tests with portable emission measurement equipment and employing an innovative feature engineering approach combining principal component analysis and gray correlation analysis, they improved the data processing efficiency. They used a double-hidden-layer BP neural network optimized with an advanced Grey Wolf

algorithm. Their model achieved a root mean square error of 1.91 mg/s and  $R^2$  equal to 0.87, demonstrating superior accuracy in predicting actual road  $\text{NO}_x$  emissions compared to simpler models.

#### *Present Contribution*

This research investigates the possibility of integrating predictive models into the engine control unit (ECU), with the aim of eliminating the need for cumbersome portable devices during certification cycles, monitoring combustion quality in real time, and enabling remote emission monitoring from any location. Tests were carried out on an SI single-cylinder research engine fueled with conventional gasoline E5, considering a relative air excess index ( $\lambda$ ) equal to 1.0 and a wide range of different operating conditions. The experimental setup involved collecting data coming from ECU and a raw gas analyzer. These data served as training data for the ML architecture, which was developed to forecast the composition of exhaust gases based on the observed behavior of the engine. To accomplish this mission, different Feed-Forward neural networks with a back propagation (BP) optimizer [30] were tested.

In the current study, research was conducted on fine-tuning parameters such as the number of neurons, hidden layers, and input variables in ANN structures to enhance the prediction performance. By conducting a series of experiments and validations iteratively, the ML architecture ideal design was established.

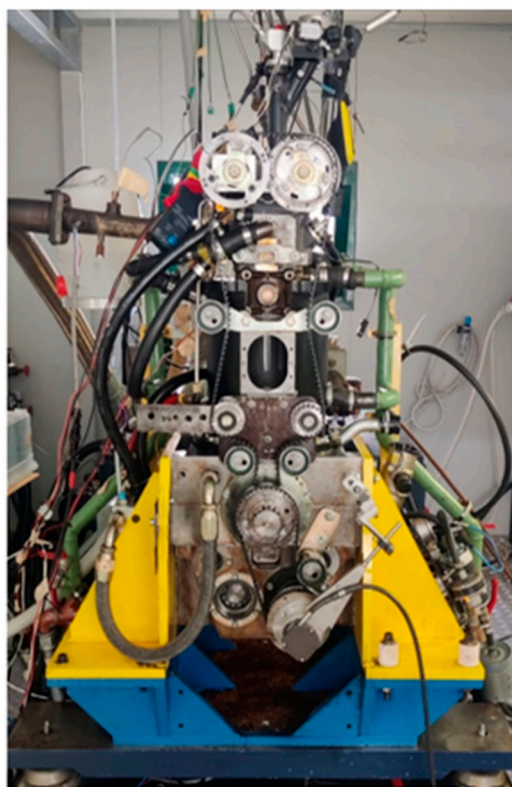
The current study extensively tested different ANN architectures, focusing on configurations optimized for pollutant emission prediction in a spark-ignition engine. The models were trained on dynamic engine operating cycles, allowing them to predict emissions under varying real-world conditions. This dynamic approach contrasts with earlier studies that focused more on static engine conditions. This research created and tested five unique dynamic cycles designed to stress the predictive capabilities of the network. Each cycle represented different engine speeds, torques, and throttle conditions, providing a more thorough evaluation of the model's robustness across various operating conditions. The capacity of the optimized model in generalizing over various cycles (engine speeds, throttle openings, and torque settings) demonstrates its superior robustness. Previous models, by comparison, may have shown limited performance outside of specific test conditions.

The findings indicate that the proposed model achieves convergence throughout the training process while avoiding overfitting. This demonstrates its proficiency in efficiently extracting knowledge from the dataset and generating accurate predictions. The results indicate that the model holds considerable promise for predicting pollutant emission concentrations in SI engines. Indeed, for each of the aforementioned dynamic cycles, the prediction of the analyzed pollutant emissions yields an average root mean square error (RMSE) of less than 5%, with a maximum of 5.87% and a minimum of 1.57%. Throughout, it consistently remains below the 10% acceptability threshold, thereby ensuring compliance with high quality standards [31]. Therefore, the BP neural network model has potential applications as an onboard virtual tool for estimating emissions in real driving conditions through the presented ANN predictive model.

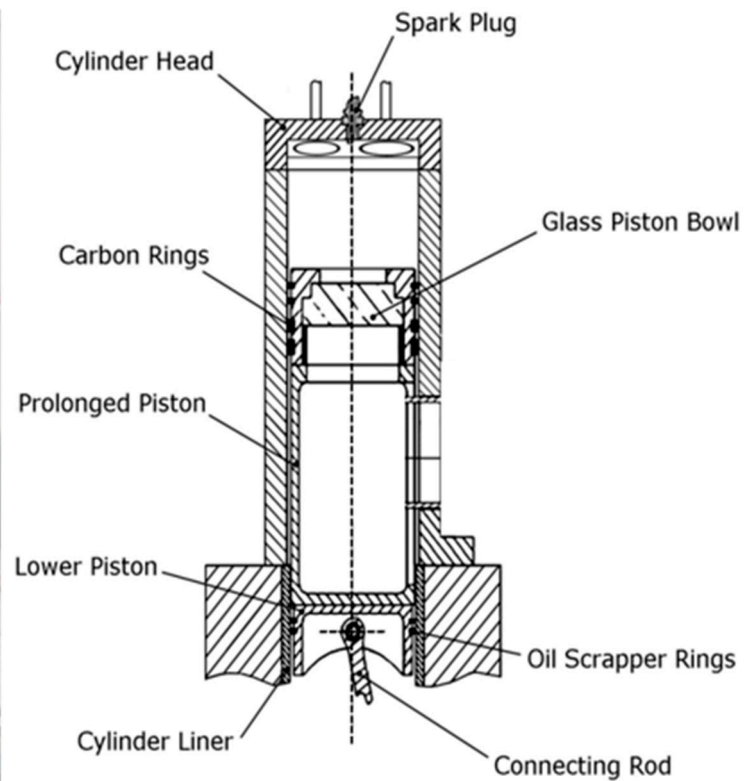
## **2. Materials and Methods**

### *2.1. Experimental Setup*

The measurements were conducted on a 500-cc single-cylinder research engine, shown in Figure 1, characterized by a pent-roof combustion chamber with four valves and a reverse tumble intake port system precisely crafted for operation in port fuel injection (PFI) [32] mode. Moreover, the engine features a stroke of 88 mm, a bore of 85 mm, a connecting rod length of 139 mm, and a compression ratio of 8.8:1. Additional detailed information about the test engine can be found in [33,34] and in Table 1.



(a)



(b)

Figure 1. Test motor: (a) real-world depiction and (b) layout illustration.

Table 1. Engine's main characteristics.

Feature	Value	Unit
Displaced volume	500	cm <sup>3</sup>
Stroke	88	mm
Bore	85	mm
Connecting rod length	139	mm
Compression ratio	8.8:1	-
Number of valves	4	-
Exhaust valve open	-13	CAD aBDC
Exhaust valve closed	25	CAD aBDC
Intake valve open	-20	CAD aBDC
Intake valve closed	-24	CAD aBDC

The air-flow rate was regulated by a throttle valve positioned upstream of the intake manifold.

An AVL 5700 dynamic brake was mechanically linked to the engine crankshaft, providing speed control for the engine via National Instruments hardware and custom LabVIEW software (LabVIEW version number: 12.0.1f5—32 bit), applicable in both motored and firing conditions.

A European-market gasoline (E5, alcoholic fuel made of 95% gasoline and 5% ethanol), with a Research Octane Number (RON) equal to 95 and a Motor Octane Number (MON) equal to 85, was used as fuel and was injected by a Weber IWP092 port fuel injector at 4.8 bar absolute.

The injector's energizing time and ignition timing (IT) were controlled by utilizing an Athena GET HPUH4 research ECU by sending an activation signal to the control unit of the igniter.

A Horiba MEXA-7100D with an OVN-723A was used to conduct the exhaust gas analysis. The general schema of the experimental setup employed, including the inputs and outputs for the analyzed nets, is presented in Figure 2, with further details and specifications provided in Table 2.

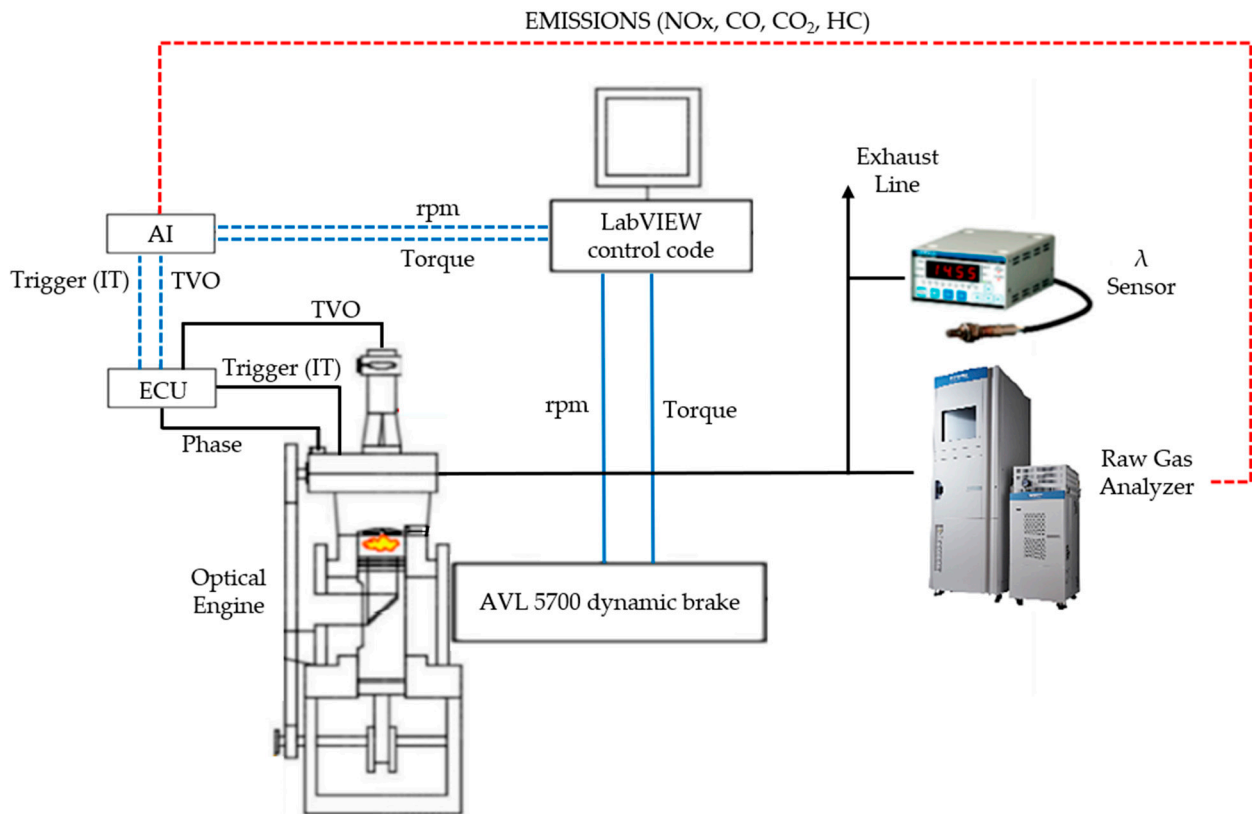


Figure 2. Experimental apparatus.

Table 2. Measurements and apparatus details.

Device	Description	Specifications
Kistler Kibox	Indicating analysis system for signal acquisition and combustion analysis	10 analog input channels and 2 encoder input channels
Kistler 6061B	In-cylinder pressure piezoelectric sensor	Sensitivity: 25.9 pC/bar Range: 0–250 bar
Kistler 5011B	Charge amplifier	Scale: 10 bar/V
Kistler 4075A5	Piezoresistive pressure sensor, used for the intake line, downstream of the throttle; reference for in-cylinder pressure pegging	Sensitivity: 25 mV/bar/mA Range: 0–5 bar
AVL 365C	Optical encoder for crankshaft angular position and engine speed measurement	Resolution up to 0.1 CAD
AVL 5700	Dynamic brake, mechanically coupled with the engine crankshaft	Ensures the engine speed control through National Instruments hardware and in-house LabVIEW code

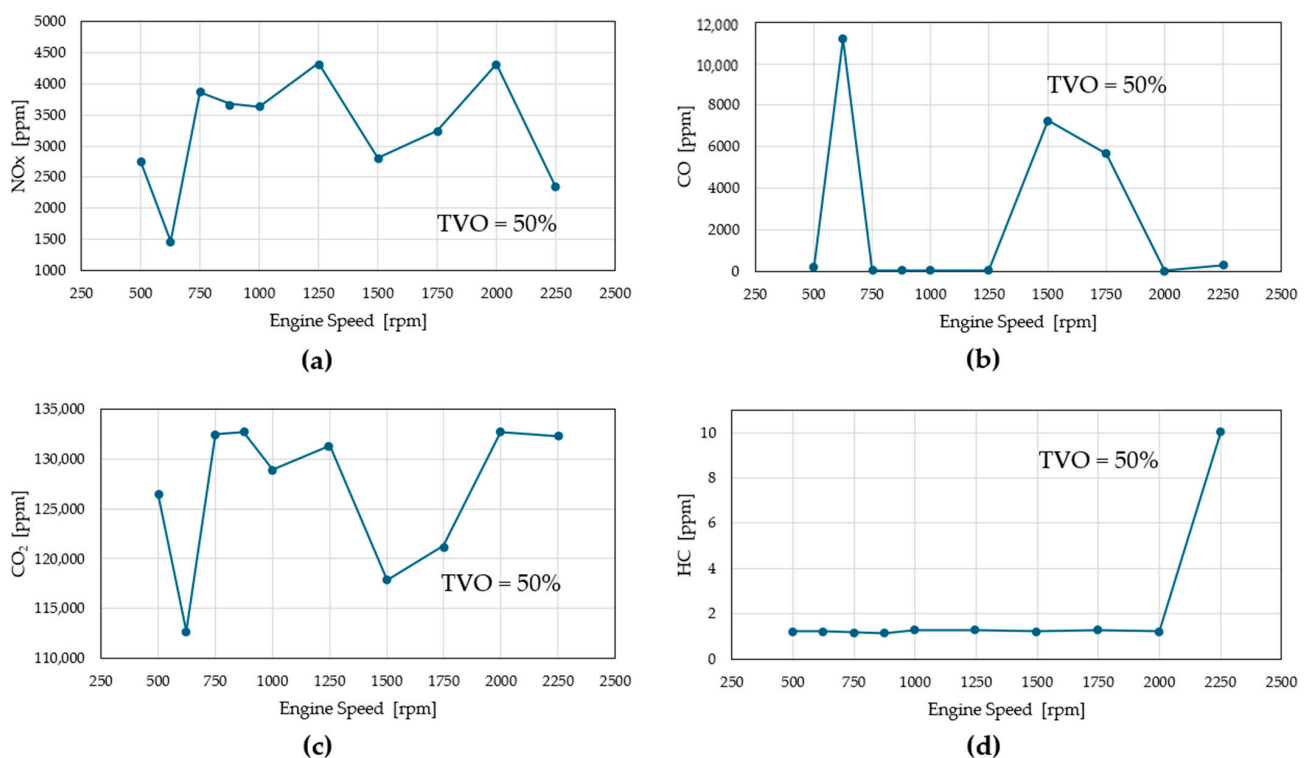
Table 2. Cont.

Device	Description	Specifications
Athena GET HPUH4	Engine control unit	Control the injector energizing time and IT by sending a trigger signal to the igniter control unit Output: AFR, $\lambda$ , and [O <sub>2</sub> ]
Horiba Mexa 720	Fast lambda probe	Adjustable for various fuels through setting the O/C and H/C ratios
Horiba Mexa 7100D	Exhaust gas analyzer	Output: HC, CO, CO <sub>2</sub> , NO <sub>x</sub> , SO <sub>2</sub> , O <sub>2</sub> , and THC

## 2.2. Artificial Neural Network Setup

### 2.2.1. Description of the Initial Dataset

This study evaluates the effectiveness of the back propagation neural architecture in predicting pollutant concentrations, in particular, NO<sub>x</sub>, CO, CO<sub>2</sub>, and HC emissions, in ppm (Figure 3) over 5 different dynamic cycles, at the exhaust port of a spark ignition engine. The initial dataset, for training and validating the architecture, comprises experimental data collected under a constant  $\lambda$  value (i.e.,  $\lambda = 1.0$ ), with varying throttle valve openings (TVOs) ranging from 5% to 100% and engine speeds ranging from 500 rpm to 2250 rpm (Figure 4).



**Figure 3.** The trend in the (a) NO<sub>x</sub>, (b) CO, (c) CO<sub>2</sub>, and (d) HC concentration at the exhaust pipe, for different engine speeds, with one particular operating condition taken as an example (i.e., TVO = 50%).

The dataset, detailed in Table 3, includes 100 operating cases, with each case featuring 100 consecutive combustion events. The machine learning model utilized the following four parameters as inputs for each combustion event:

- Engine speed [rpm].
- Ignition timing [CAD aTDC].
- Throttle valve opening [%].

- Torque [Nm].

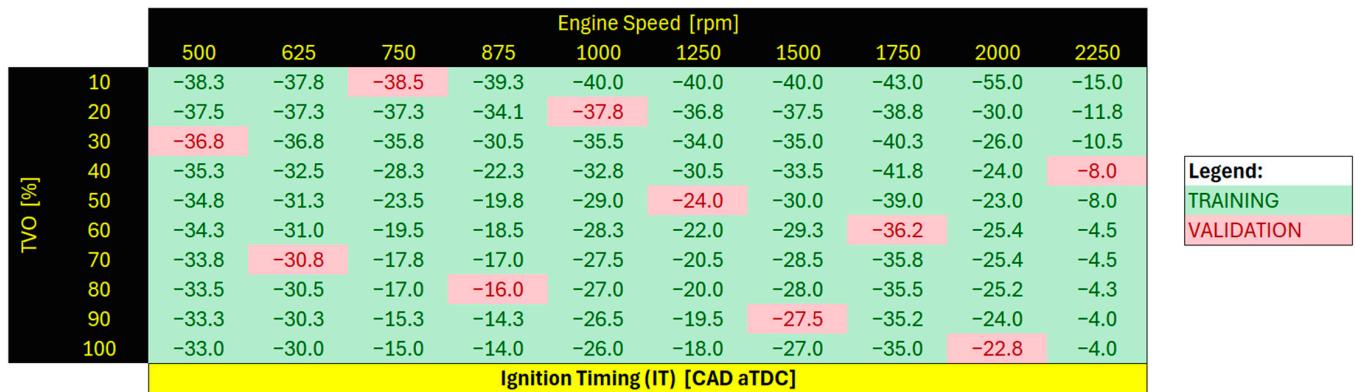


Figure 4. Engine map. Green operating points were used for training, and red operating points were used for validation.

Table 3. General description of the initial dataset. With respect to each operating case, the mean value across 100 combustion events has been considered for each parameter.

Case Number [-]	Engine Speed [rpm]	IT [CAD aTDC]	TVO [%]	Torque [Nm]	NO <sub>x</sub> [ppm]	CO [ppm]	CO <sub>2</sub> [ppm]	HC [ppm]
1	500	-38.3	10	9.63	3545.63	35.13	131,387.39	1.20
2	625	-37.8	10	11.37	4121.68	26.74	126,021.47	1.21
3	750	-38.5	10	12.62	4146.88	32.21	130,625.34	1.22
...	...	...	...	...	...	...	...	...
99	2000	-22.8	100	36.50	4321.81	24.59	132,692.25	1.22
100	2250	-4.0	100	28.04	1972.78	610.42	131,912.61	9.48

Based on the observations previously outlined and presented in Table 3 and Figure 4, Figure 5 shows the division of the primary dataset into training and validation subsets, comprising [100 × 4] input variables and [100 × 4] output variables. The dataset is composed of 100 experimental cases, as detailed in Table 3, with each case defined by 8 variables (Figure 5a). The input parameters (i.e., engine speed, IT, TVO, and torque) are organized into a [4 × 100] matrix, while the output parameters (i.e., NO<sub>x</sub>, CO, CO<sub>2</sub>, and HC) are represented by another [4 × 100] matrix; see Figure 5b. The dataset was partitioned so that 90% of the data were utilized for training and the residual 10% were reserved for validation (Figure 5c).

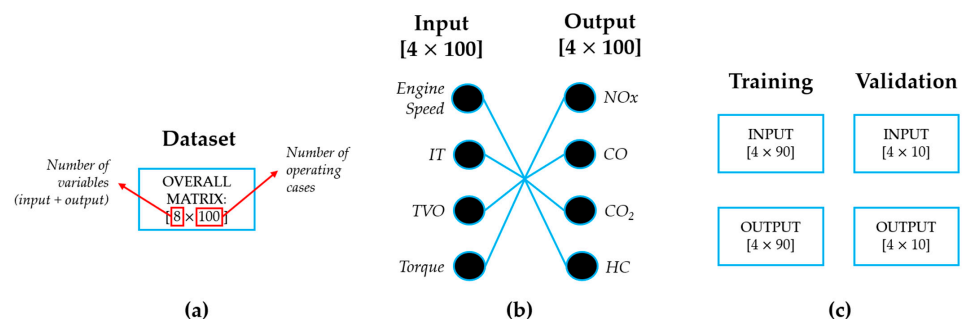
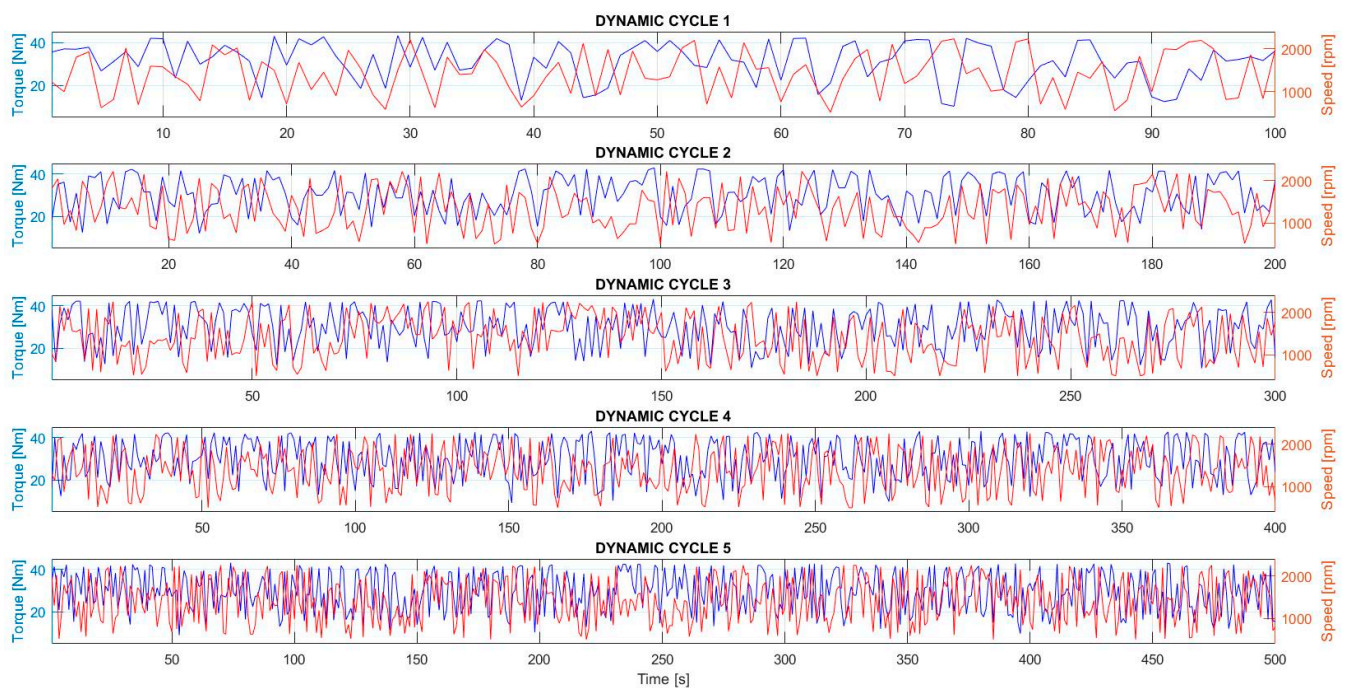


Figure 5. (a) Dataset comprehensive summary, encompassing the total count of the operational points examined and the inputs/outputs included; (b) a detailed analysis of the input and output parameters for each case, based on the findings from the initial sensitivity assessment; and (c) dataset partitioning into training and validation sets.

### 2.2.2. Description of the Dynamic Cycles

Building on the dataset previously detailed in Table 3 and illustrated in Figure 4, the network has been subsequently trained using the specified data. Following the training phase, dynamic cycles were defined, which constitute the basis for predicting pollutant emissions.

Specifically, as depicted in Figure 6, 5 distinct dynamic cycles have been defined, each characterized by a unique road path, particularly concerning variations in engine speed (red curves) and torque (blue curves) and different durations, as specified in Table 3. These differences have been carefully considered to ensure a comprehensive analysis of the network's predictive capabilities for the pollutant concentrations at the exhaust pipe across varying operating conditions.



**Figure 6.** Engine speed (red curves) and torque (blue curves) variations for the 5 dynamic cycles.

Therefore, based on these dynamic cycles, additional experimental tests were conducted for each cycle utilizing the torque and engine speed values previously presented in Figure 6 as inputs for the test bench, described in Section 2.1. A dedicated system enabled control of the engine's speed, and torque, together with the throttle valve opening in order to obtain, from these tests, the corresponding pollutant emissions. These emissions data have been used as critical reference points for evaluating the accuracy and effectiveness of the BP architecture's predictive capabilities.

## 3. Developing and Optimization of the Artificial Neural Network

### 3.1. Back Propagation Structure

The back propagation artificial neural network (BPANN) is a fundamental type of ANN primarily used in supervised learning tasks. It is designed to learn from data by minimizing the discrepancy between actual and predicted results through the backpropagation process. The network (Figure 7) consists of layers, including an input layer that receives raw data, one or more hidden layers that process these data by applying weights and activation functions, and an output layer that produces the network's prediction. Neurons in these layers take inputs, multiply them by weights, apply an activation function (for example, Sigmoid function, ReLU function, ...) to introduce non-linearity, and pass the result to the next layer [27].



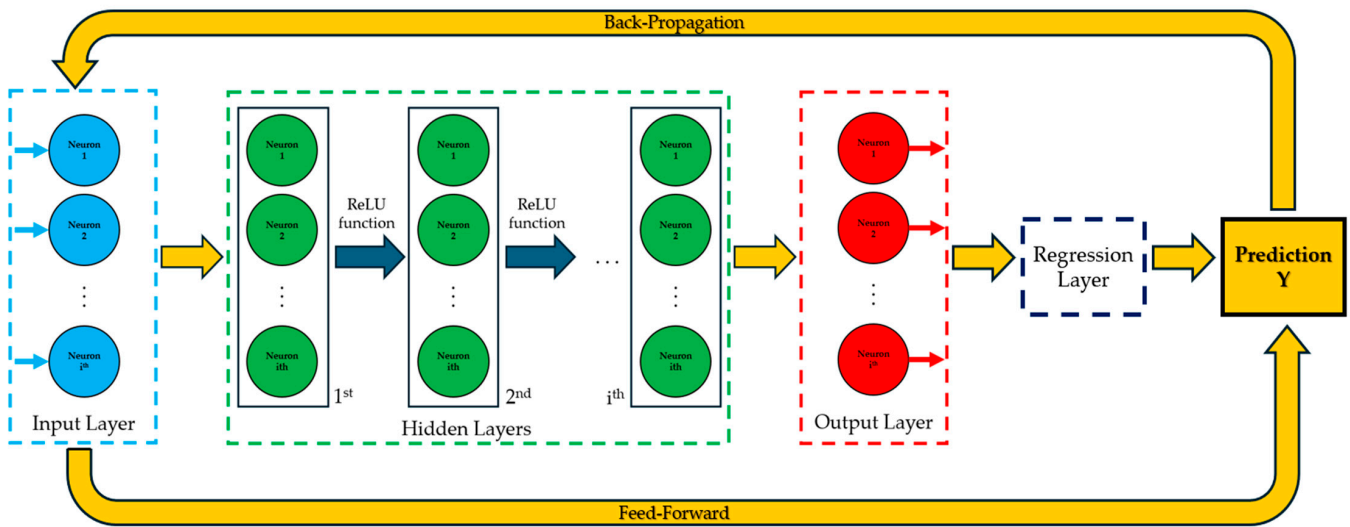


Figure 7. Complete BPANN general scheme.

The network operates through a forward pass, where data flow through the network layer by layer and an output is generated based on the processing in the hidden layers. After the forward pass, the network's output is compared to the actual target using a loss function, which measures the error between the predicted output and the target. The process of Back Propagation follows, during which the error is transmitted from the output layer back to the input layer. Gradients of the loss function with respect to each weight are calculated, and the weights are updated in the direction that minimizes the loss, typically using optimization algorithms like gradient descent [28].

This process is repeated over many iterations, known as *epochs*, allowing the network to gradually learn the optimal weights that minimize the error and thus improve its performance on the training data. Key concepts include the following:

- Learning rate, which controls the pace of learning.
- Overfitting, in which the network exhibits strong performance on training data but struggles with unseen data, a challenge that can be mitigated using techniques such as regularization.
- Convergence, where training stabilizes as the network learns optimal weights.

### 3.2. Overview of the Procedures for Establishing the Structural Parameters of the Proposed Model

A preliminary comparative analysis of the BP neural architectures was conducted, focusing on optimizing their internal structures to achieve high prediction quality and faster convergence times. This optimization involved experimenting with various numbers of hidden layers, neurons, and activation functions. All activities, including architecture development and optimization, were executed within the MATLAB environment, using a single CPU (central processing unit) and 16 GB of RAM (Random Access Memory).

To assess the precision and performance of the model's parameters, the root mean square error (RMSE), which represents the average squared differences between predicted values and observed values, has been employed as the primary metric for measuring loss (Equation (1)). In regression problems, it is common practice to also employ additional metrics, such as the mean squared error (MSE) and the mean absolute error (MAE), to further evaluate model performance.

$$RMSE [\%] = 100 \times \sqrt{\frac{1}{N} \sum_{i=1}^N (Y_{predicted}^i - Y_{target}^i)^2} \quad (1)$$

where

- $N$  = cycle duration.

- $i = i$ th temporal instant.
- $Y_{predicted}^i =$  predicted value.
- $Y_{target}^i =$  target value (experimental results).

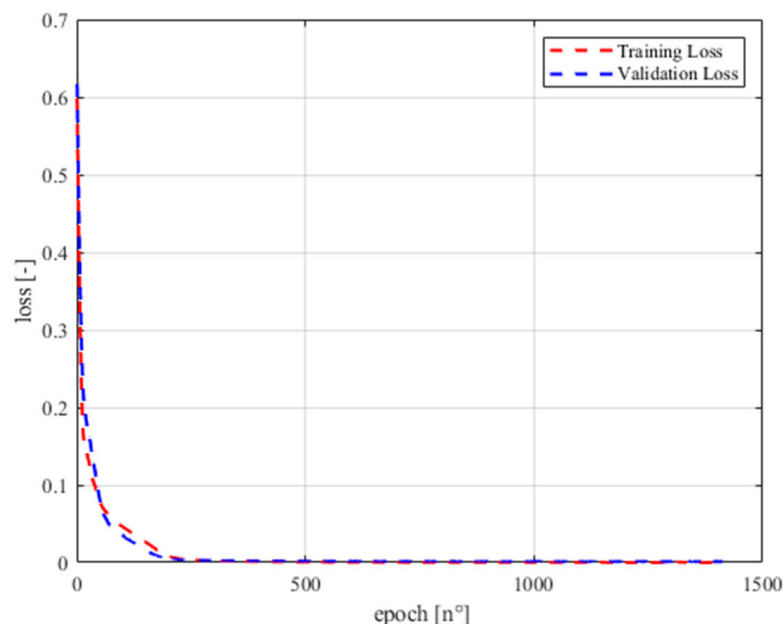
As mentioned in the introduction, an upper threshold of 10% is established for these calculated RMSEs to guarantee precise predictions.

For the BPANN architecture, the structural parameters examined to optimize network performance include the number of hidden layers, varying from 2 to 6, and the number of neurons per hidden layer, ranging from 20 to 100.

The network has been trained for 10,000 epochs, enabling the calculation of the final loss function value for each prediction model once the maximum learning iteration has been achieved. The most effective configuration of the BPANN consists of an input layer with 4 neurons, corresponding to the number of input variables. This is followed by 4 hidden layers, each containing 100 neurons. These hidden layers are separated by ReLU (Rectified Linear Unit) activation functions, which improve the network's capability to recognize intricate patterns from the input strongly non-linear data. Finally, the network features an output layer with 4 physical neurons, the associated values of which are synchronized with the input physical quantities. It is important to emphasize that the alignment between 4 input and 4 output variables is merely coincidental and should not be misconstrued as a structural limitation of the model.

This architecture, fine-tuned through a thorough preliminary analysis, is designed to take advantage of deep learning through multiple hidden layers and ReLU activation to improve pattern recognition and predictive performance. The Adam optimizer is utilized in this study, as it incorporates an adaptive learning rate along with momentum adjustments throughout the training process. This advanced optimization algorithm is specifically designed to enhance the efficiency of the weight matrix and bias adjustments within the BPANN model, ultimately improving its overall performance and convergence speed.

To ensure completeness, the validation loss and training loss curves of the most effective BPANN structure are illustrated in Figure 8:



**Figure 8.** The trend in loss values for the BPANN architecture indicated optimal performance throughout the training session.

#### 4. Results and Discussion

Figure 9 illustrates the predictions generated by the previously presented BPANN concerning pollutant emissions concentration, i.e., NO<sub>x</sub>, CO, CO<sub>2</sub>, and HC, across the 5 dynamic cycles earlier depicted; see Figure 6.

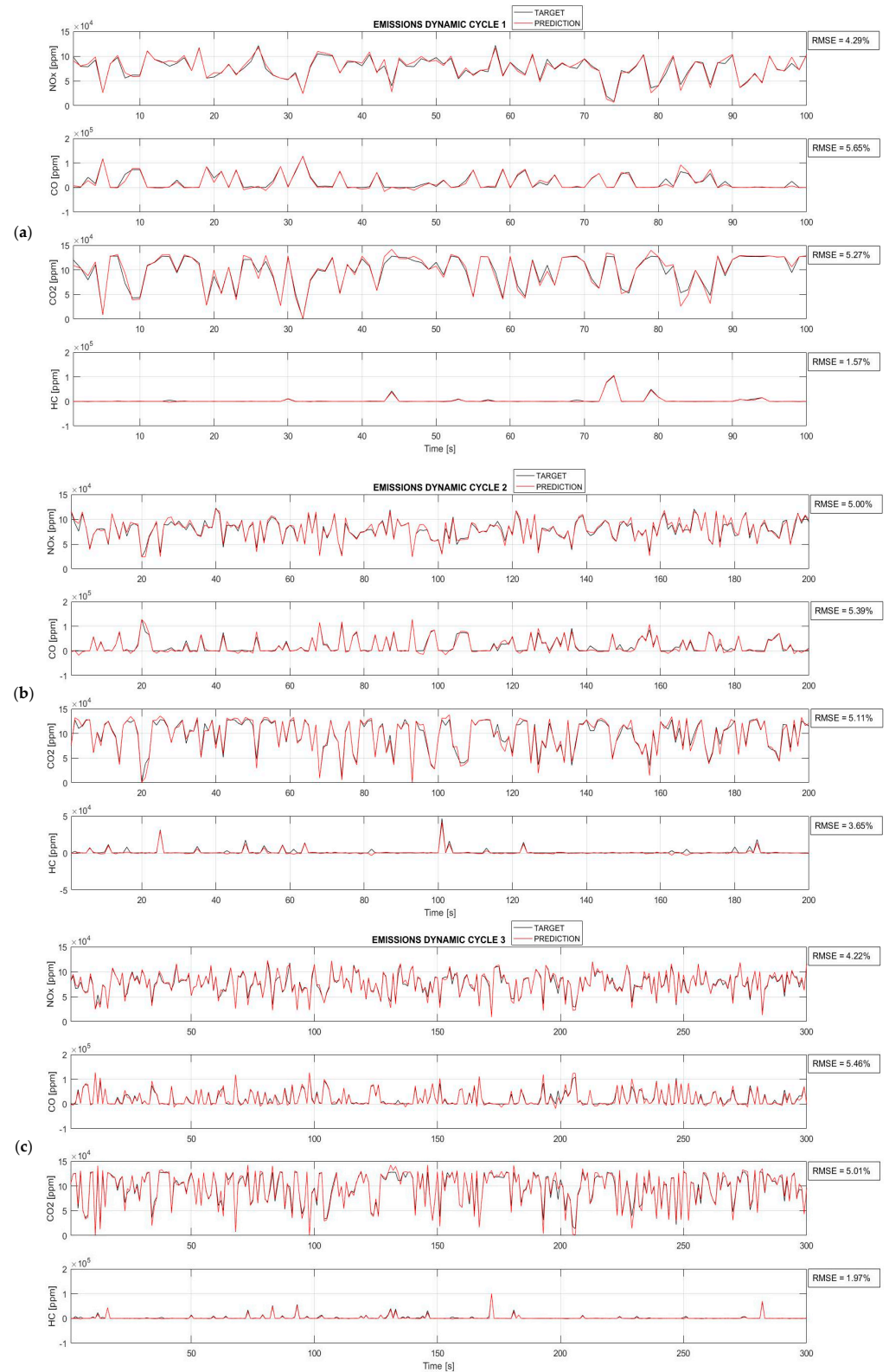
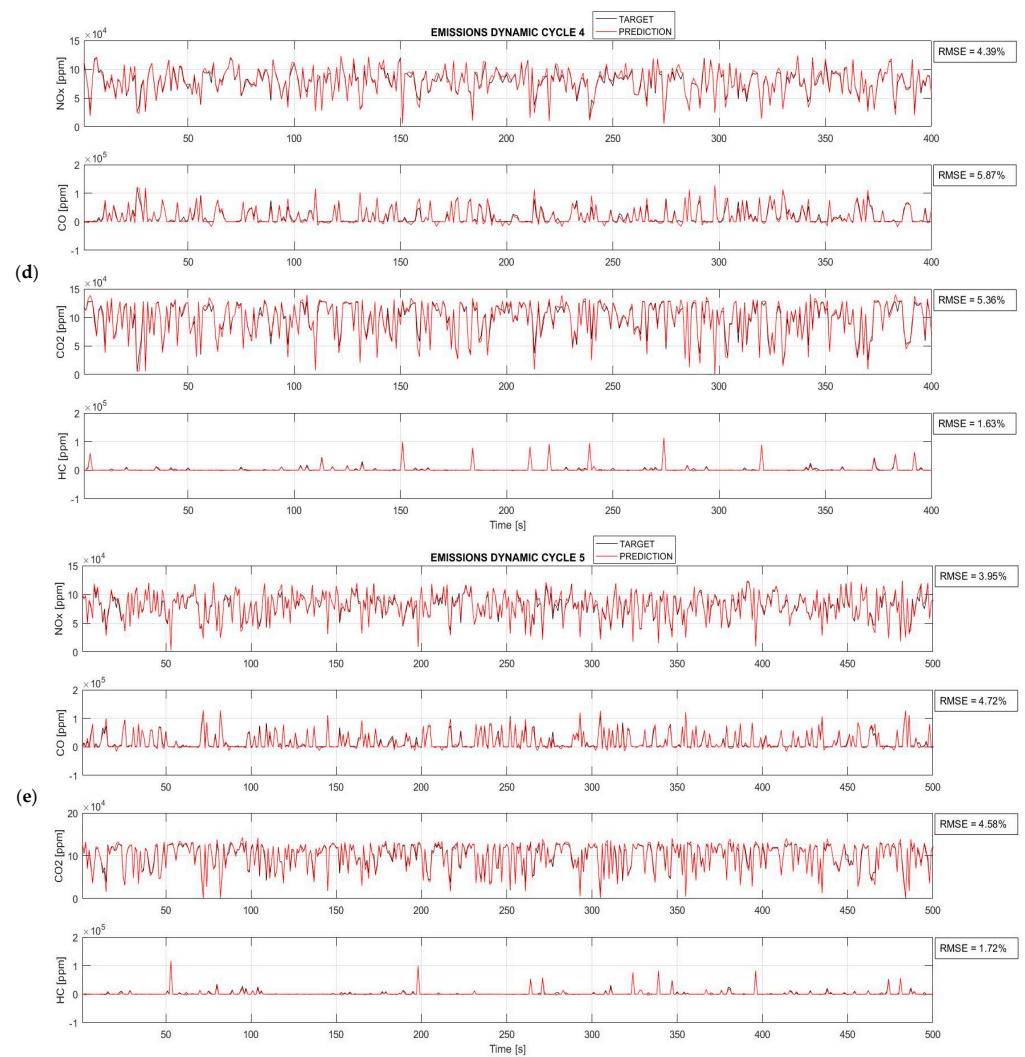


Figure 9. Cont.



**Figure 9.** BPANN prediction of the pollutant emission concentration for (a) dynamic cycle 1, (b) dynamic cycle 2, (c) dynamic cycle 3, (d) dynamic cycle 4, and (e) dynamic cycle 5.

As illustrated in Figure 9, where the red curves depict the predictions generated by the network and the black curves represent the experimentally obtained target values, the BPANN architecture demonstrates a remarkable ability to accurately replicate the emission trends for each dynamic cycle and every pollutant analyzed. Upon closer examination, it is evident that for all the cases studied, the RMSE remains significantly below the acceptable threshold of 10%, never exceeding 6%. Specifically, the highest RMSE, observed in the prediction of CO for dynamic cycle 4 (Figure 9d), is 5.87%, while the lowest, at 1.57%, is observed in the prediction of HC for dynamic cycle 1 (Figure 9a).

The results previously presented in Figure 9 are comprehensively summarized in Table 4.

The prediction's regression accuracy ( $R^2$ ) performed by the optimal tested architecture, as shown in Figure 10, was determined using Equation (2):

$$R^2 = 1 - rMSE = 1 - \frac{MSE}{Var(Y_{target})} \quad (2)$$

where

- $rMSE$  is the relative mean squared error.
- $MSE$  is the mean squared error.

- $Var(Y_{target})$  refers to the variance of the random variable  $Y_{target}$ . Variance is a statistical measure that represents the dispersion or spread of the values of  $Y_{target}$  around its mean. It quantifies how much the values of  $x$  deviate from the mean value  $Y_{target}$ .

As is possible to observe, in all the charts, the data points are represented with predicted values plotted on the y-axis and target values on the x-axis. The closeness of the points to the diagonal dashed line reflects the accuracy of the predictions. Each graph is scaled according to the data range specific to each case.

In accordance with the observations made in Figure 9 and Table 4, the BPANN demonstrates a consistent alignment along the interpolation line, with no significant deviations, for all pollutants analyzed across each dynamic cycle, thereby achieving exceptionally high accuracy. It is noteworthy that this architecture displays very little dispersion, with  $R^2$  values becoming close to unity, definitely  $> 0.93$ , for all pollutants analyzed across each dynamic cycle: the highest  $R^2$  value, 0.9871, is observed for HC emissions in dynamic cycle 1, while the only exception is the HC emission value in dynamic cycle 2, with an  $R^2 = 0.8814$ .

The results previously presented in Figure 10 are comprehensively summarized in Table 5.

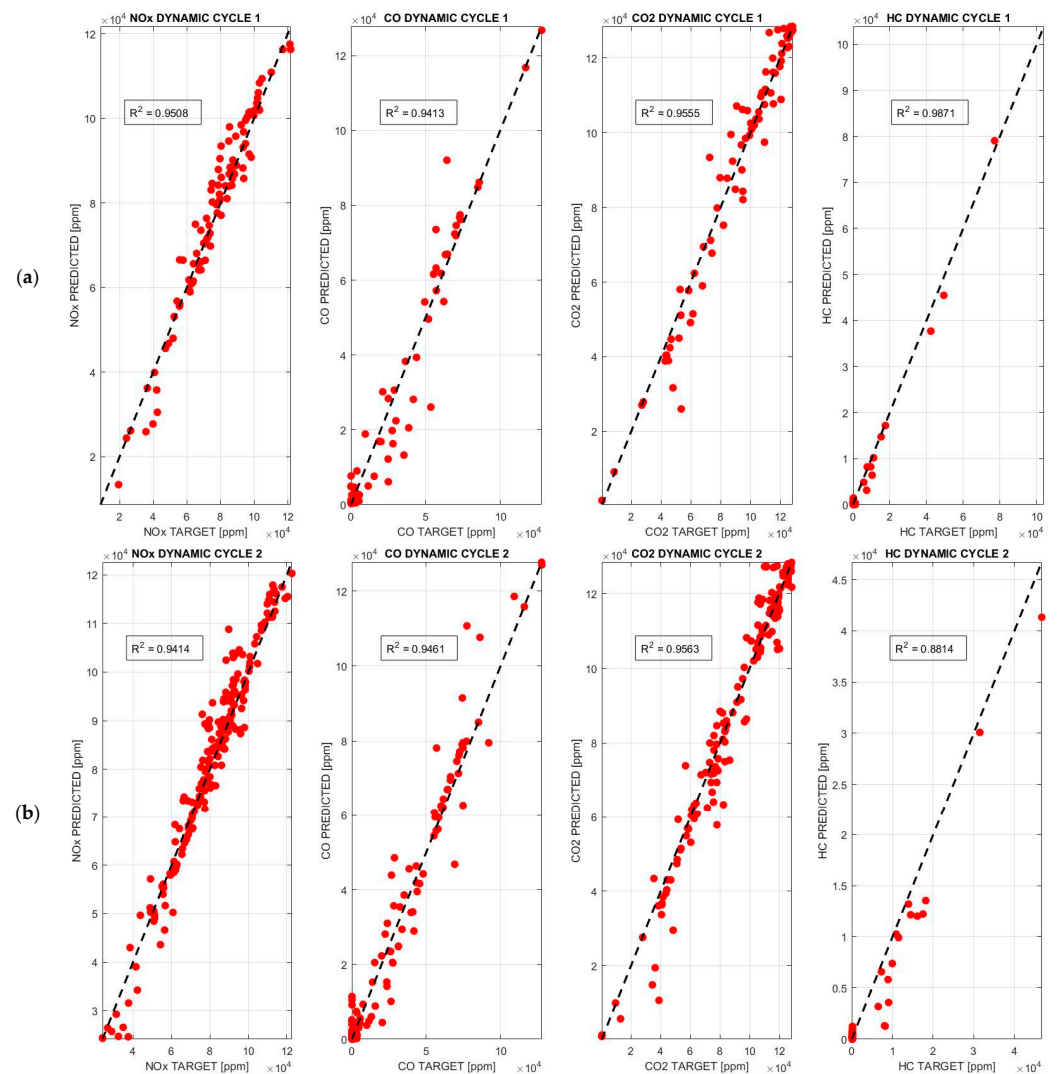


Figure 10. Cont.

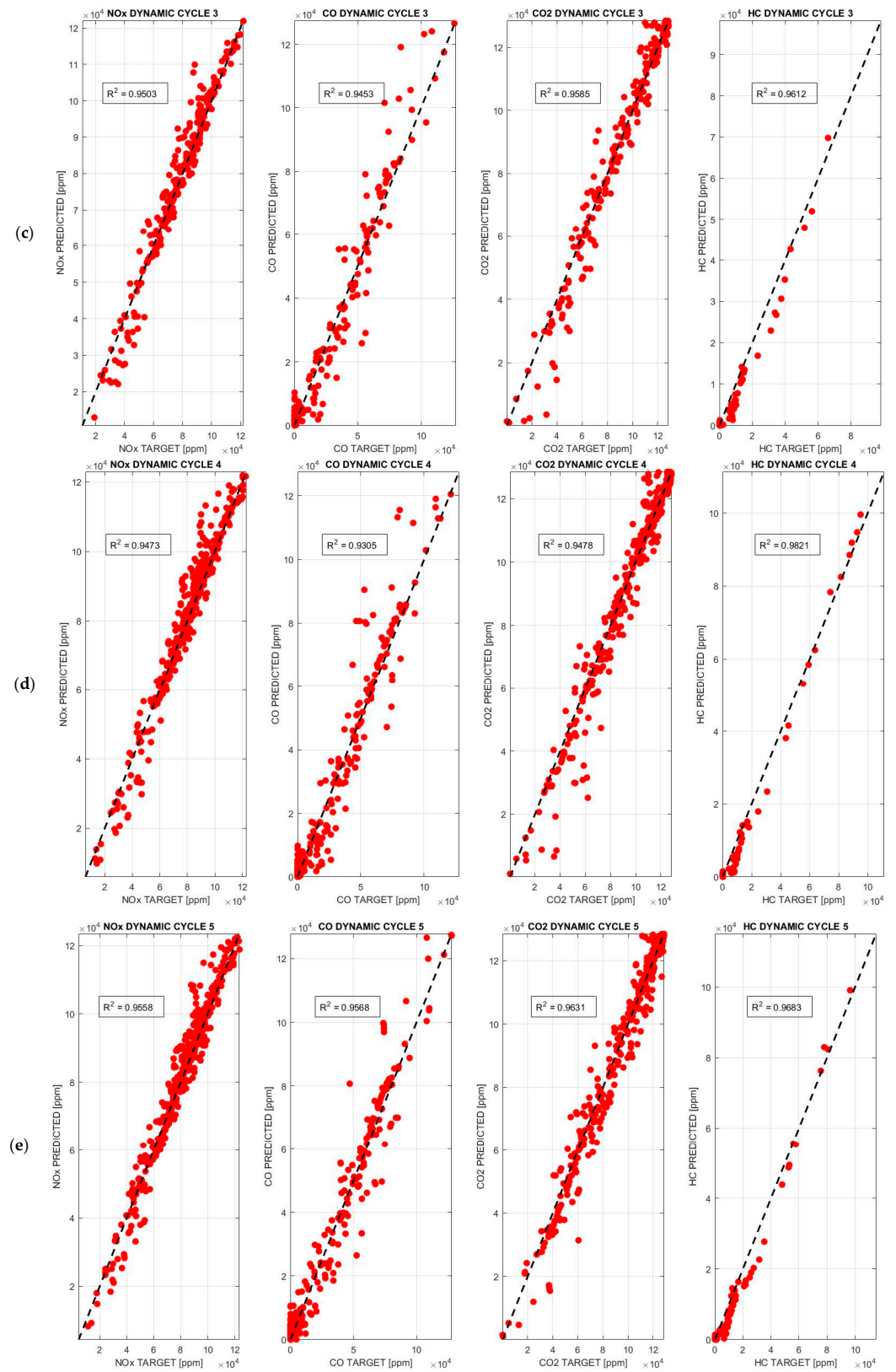


Figure 10. BPANN regression prediction chart for (a) dynamic cycle 1, (b) dynamic cycle 2, (c) dynamic cycle 3, (d) dynamic cycle 4, and (e) dynamic cycle 5.

**Table 4.** RSME results.

Dynamic Cycle Number [-]	Duration [s]	RMSE NO <sub>x</sub> [%]	RMSE CO [%]	RMSE CO <sub>2</sub> [%]	RMSE HC [%]
1	100	4.29	5.65	5.27	1.57
2	200	5.00	5.39	5.11	3.65
3	300	4.22	5.46	5.01	1.97
4	400	4.39	5.87	5.36	1.63
5	500	3.95	4.72	4.58	1.72

**Table 5.** R<sup>2</sup> results.

Dynamic Cycle Number [-]	Duration [s]	R <sup>2</sup> NO <sub>x</sub> [-]	R <sup>2</sup> CO [-]	R <sup>2</sup> CO <sub>2</sub> [-]	R <sup>2</sup> HC [-]
1	100	0.9508	0.9413	0.9555	0.9871
2	200	0.9414	0.9461	0.9563	0.8814
3	300	0.9503	0.9453	0.9585	0.9612
4	400	0.9473	0.9305	0.9478	0.9821
5	500	0.9558	0.9568	0.9631	0.9683

The results highlight the robust learning abilities of the BPANN architecture, showcasing its effectiveness in accurately capturing and replicating the target trend throughout the training process.

Nonetheless, certain uncertainties persist. The observed variability in RMSE values and the somewhat lower R<sup>2</sup> value of 0.8814 for HC in dynamic cycle 2 indicate that despite the BPANN's generally strong performance, there are factors affecting the accuracy of its predictions. These factors may include the model's sensitivity to specific operating conditions, the representativeness of the training data, and the inherent complexity of emission processes. To address these uncertainties, it is essential to improve data quality, refine the BPANN model, and perform sensitivity analyses to better understand and mitigate sources of prediction variability. Although the BPANN exhibits commendable performance, recognizing and addressing these uncertainties will be vital in enhancing its reliability in practical applications.

## 5. Conclusions

This study proposed a deep learning methodology utilizing back propagation artificial neural networks to predict pollutant emissions in a single-cylinder spark-ignition engine across various operating conditions. The network architecture was refined by optimizing the number of hidden layers, neurons, and activation functions to ensure optimal performance. The model demonstrated robustness and reliability, achieving accurate predictions, with RMSE values below 6%, alongside strong regression accuracy, with R<sup>2</sup> values exceeding 0.93 in most instances.

The back propagation artificial neural network's capacity to generalize across different dynamic cycles, encompassing variations in torque and engine speed, underscores its potential as an effective tool for real-time emission monitoring. By integrating this model into engine control units, it may be possible to conduct on-board emission estimation, thereby facilitating compliance with environmental regulations and reducing emissions in real-world driving scenarios. The implementation of this predictive architecture could significantly streamline emission monitoring, eliminating the need for cumbersome portable devices during certification cycles while supporting proactive maintenance strategies.

Future research should focus on experimentally validating and deploying trained back propagation artificial neural network within engine control units, addressing challenges such as compatibility with open and closed engine control unit architectures and fostering collaborations with stakeholders to ensure the seamless integration of the neural network into existing hardware systems.

**Author Contributions:** Conceptualization, F.R. and F.M.; methodology, F.R. and M.A.; software, F.R. and M.A.; validation, F.M.; formal analysis, F.R. and M.A.; investigation, F.R. and M.A.; resources, F.R. and F.M.; data curation, F.R. and M.A.; writing—original draft preparation, F.R. and M.A.; writing—review and editing, F.R., M.A. and F.M.; visualization, F.R. and M.A.; supervision, F.M.; project administration, F.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The raw data supporting the conclusions of this article will be made available by the authors on request.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Nomenclature

ANN	artificial neural network
aBDC	after bottom dead center
aTDC	after top dead center
BP	back propagation
BPANN	back propagation artificial neural network
CAD	crank angle degree
CO	carbon monoxide
CO <sub>2</sub>	carbon dioxide
CoV <sub>IMEP</sub>	coefficient of variance of IMEP
CPU	central processing unit
DI	direct injection
DPF	diesel particulate filter
E5	gasoline
E20/E85	ethanol
ECU	engine control unit
EGR	exhaust gas recirculation
ELSB	ensemble least-squares boosting
GHG	greenhouse gasses
GPF	gasoline particulate filter
H <sub>2</sub>	hydrogen
HC	hydrocarbons
IC	internal combustion
IMEP	indicated mean effective pressure
IT	ignition timing
$\lambda$ ( $1/\varphi$ )	air excess coefficient
LTC	low-temperature combustion
M100	methanol
MAPE	mean average percentage error
MLP	Multi-Layer Perceptron
ML	machine learning
MON	Motor Octane Number
NO <sub>x</sub>	nitrogen oxides
O <sub>2</sub>	oxygen
PFI	port fuel injection
R <sup>2</sup>	coefficient of determination



RAM	Random Access Memory
RBF	Radial Basis Function
SCR	selective catalytic reducer
RMSE	root mean square error
RON	Research Octane Number
TPRF	Toluene Primary Reference Fuel
TVO	throttle valve opening

## References

- Joshi, A. Review of vehicle engine efficiency and emissions. *SAE Int. J. Adv. Curr. Pract. Mobil.* **2020**, *2*, 2479–2507. [[CrossRef](#)]
- Suresh, D.; Porpatham, E. Influence of high compression ratio and hydrogen addition on the performance and emissions of a lean burn spark ignition engine fueled by ethanol-gasoline. *Int. J. Hydrogen Energy* **2023**, *48*, 14433–14448. [[CrossRef](#)]
- Reitz, R.D.; Ogawa, H.; Payri, R.; Fansler, T.; Kokjohn, S.; Moriyoshi, Y.; Agarwal, A.; Arcoumanis, D.; Assanis, D.; Bae, C.; et al. IJER editorial: The future of the internal combustion engine. *Int. J. Engine Res.* **2020**, *21*, 3–10. [[CrossRef](#)]
- Wallington, T.J.; Anderson, J.E.; Dolan, R.H.; Winkler, S.L. Vehicle emissions and urban air quality: 60 years of progress. *Atmosphere* **2022**, *13*, 650. [[CrossRef](#)]
- Berggren, C.; Magnusson, T. Reducing automotive emissions—The potentials of combustion engine technologies and the power policy. *Energy Policy* **2012**, *41*, 636–643. [[CrossRef](#)]
- Kalghatgi, G.; Johansson, B. Gasoline compression ignition approach to efficient, clean and affordable future engines. *Proc. Inst. Mech. Eng. Part D J. Automob. Eng.* **2018**, *232*, 118–138. [[CrossRef](#)]
- Dernotte, J.; Najt, P.M.; Durrett, R.P. Downsized-Boosted Gasoline Engine with Exhaust Compound and Dilute Advanced Combustion. *SAE Int. J. Adv. Curr. Pract. Mobil.* **2020**, *2*, 2665–2680. [[CrossRef](#)]
- Arabaci, E.; İcingür, Y.; Solmaz, H.; Uyumaz, A.; Yilmaz, E. Experimental investigation of the effects of direct water injection parameters on engine performance in a six-stroke engine. *Energy Convers. Manag.* **2015**, *98*, 89–97. [[CrossRef](#)]
- Lutkemeyer, G.; Weinowski, R.; Lepperhoff, G.; Brogan, M.S.; Brisley, R.J.; Wilkins, A.J.J. *Comparison of De-Nox and Adsorber Catalysts to Reduce NOx Emission of Lean Burn Gasoline Engine*; SAE Technical Paper 962046; SAE International: Warrendale, PA, USA, 1996; p. 962046.
- Dadam, S.R.; Jentz, R.; Lenzen, T.; Meissner, H. *Diagnostic Evaluation of Exhaust Gas Recirculation (EGR) System on Gasoline Electric Hybrid Vehicle*; SAE Technical Paper 2020-01-0902; SAE International: Warrendale, PA, USA, 2020.
- Singh, A.P.; Kumar, V.; Agarwal, A.K. Evaluation of comparative engine combustion, performance and emission characteristics of low temperature combustion (PCCI and RCCI) modes. *Appl. Energy* **2020**, *278*, 115644. [[CrossRef](#)]
- Pielecha, J.; Skobieć, K.; Kurtyka, K. Exhaust emissions and energy consumption analysis of conventional, hybrid, and electric vehicles in real driving cycles. *Energies* **2020**, *13*, 6423. [[CrossRef](#)]
- Anika, O.C.; Nnabuife, S.G.; Bello, A.; Okoroafor, E.R.; Kuang, B.; Villa, R. Prospects of low and zero-carbon renewable fuels in 1.5-degree net zero emission actualization by 2050: A critical review. *Carbon Capture Sci. Technol.* **2022**, *5*, 100072. [[CrossRef](#)]
- Martinelli, R.; Ricci, F.; Zemi, J.; Battistoni, M.; Grimaldi, C.; Papi, S. *Lean Combustion Analysis of a Plasma-Assisted Ignition System in a Single Cylinder Engine fueled with E85*; SAE Technical Paper 2022-24-0034; SAE International: Warrendale, PA, USA, 2022.
- Ricci, F.; Zemi, J.; Avana, M.; Grimaldi, C.N.; Battistoni, M.; Papi, S. Analysis of Hydrogen Combustion in a Spark Ignition Research Engine with a Barrier Discharge Igniter. *Energies* **2024**, *17*, 1739. [[CrossRef](#)]
- Duan, X.; Xu, L.; Xu, L.; Jiang, P.; Gan, T.; Liu, H.; Ye, S.; Sun, Z. Performance analysis and comparison of the spark ignition engine fueled with industrial by-product hydrogen and gasoline. *J. Clean. Prod.* **2023**, *424*, 138899. [[CrossRef](#)]
- Cervantes-Bobadilla, M.; García-Morales, J.; Saavedra-Benítez, Y.I.; Hernández-Pérez, J.A.; Adam-Medina, M.; Guerrero-Ramírez, G.V.; Escobar-Jiménez, R.F. Multiple fault detection and isolation using artificial neural networks in sensors of an internal combustion engine. *Eng. Appl. Artif. Intell.* **2023**, *117*, 105524. [[CrossRef](#)]
- Antinyan, V. Revealing the complexity of automotive software. In Proceedings of the 28th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering, Virtual Event, 8–13 November 2020; pp. 1525–1528.
- Khurana, S.; Saxena, S.; Jain, S.; Dixit, A. Predictive modeling of engine emissions using machine learning: A review. *Mater. Today Proc.* **2021**, *38*, 280–284. [[CrossRef](#)]
- Huang, G.; Fukushima, E.F.; She, J.; Zhang, C.; He, J. Estimation of sensor faults and unknown disturbance in current measurement circuits for PMSM drive system. *Measurement* **2019**, *137*, 580–587. [[CrossRef](#)]
- Abu-Nabah, B.A.; ElSoussi, A.O.; Abed, E.K.; Alami, A.I. Virtual laser vision sensor environment assessment for surface profiling applications. *Measurement* **2018**, *113*, 148–160. [[CrossRef](#)]
- Bai, S.; Li, M.; Lu, Q.; Fu, J.; Li, J.; Qin, L. A new measuring method of dredging concentration based on hybrid ensemble deep learning technique. *Measurement* **2022**, *188*, 110423. [[CrossRef](#)]
- Le Cornec, C.M.A.; Molden, N.; van Reeuwijk, M.; Stettler, M.E.J. Modelling of instantaneous emissions from diesel vehicles with a special focus on NOx: Insights from machine learning techniques. *Sci. Total Environ.* **2020**, *737*, 139625. [[CrossRef](#)] [[PubMed](#)]
- Moradi, M.H.; Heinz, A.; Wagner, U.; Koch, T. Modeling the emissions of a gasoline engine during high-transient operation using machine learning approaches. *Int. J. Engine Res.* **2022**, *23*, 1708–1716. [[CrossRef](#)]

25. Khac, H.N.; Modabberian, A.; Zenger, K.; Niskanen, K.; West, A.; Zhang, Y.; Silvola, E.; Lendormy, E.; Storm, X.; Mikulski, M. Machine Learning Methods for Emissions Prediction in Combustion Engines with Multiple Cylinders. *IFAC-PapersOnLine* **2023**, *56*, 3072–3078. [[CrossRef](#)]
26. Godwin, D.J.; Varuvel, E.G.; Martin, M.L.J. Prediction of combustion, performance, and emission parameters of ethanol powered spark ignition engine using ensemble Least Squares boosting machine learning algorithms. *J. Clean. Prod.* **2023**, *421*, 138401. [[CrossRef](#)]
27. Cui, Y.; Liu, H.; Wang, Q.; Zheng, Z.; Wang, H.; Yue, Z.; Ming, Z.; Wen, M.; Feng, L.; Yao, M. Investigation on the ignition delay prediction model of multi-component surrogates based on back propagation (BP) neural network. *Combust. Flame* **2022**, *237*, 111852. [[CrossRef](#)]
28. Wright, L.G.; Onodera, T.; Stein, M.M.; Wang, T.; Schachter, D.T.; Hu, Z.; McMahon, P.L. Deep physical neural networks trained with backpropagation. *Nature* **2022**, *601*, 549–555. [[CrossRef](#)]
29. Wang, Z.; Feng, K. NOx Emission Prediction for Heavy-Duty Diesel Vehicles Based on Improved GWO-BP Neural Network. *Energies* **2024**, *17*, 336. [[CrossRef](#)]
30. Petrucci, L.; Ricci, F.; Mariani, F.; Cruccolini, V.; Violi, M. *Engine Knock Evaluation Using a Machine Learning Approach*; SAE Technical Paper 2020-24-0005; SAE International: Warrendale, PA, USA, 2020.
31. Petrucci, L.; Ricci, F.; Mariani, F.; Mariani, A. From real to virtual sensor, an artificial intelligence approach for the industrial phase of end-of-line quality control of GDI pumps. *Measurements* **2022**, *199*, 111583. [[CrossRef](#)]
32. Petrucci, L.; Ricci, F.; Mariani, F.; Discepoli, G. A development of a new image analysis technique for detecting the flame front evolution in spark ignition engine under lean condition. *Vehicles* **2022**, *4*, 145–166. [[CrossRef](#)]
33. Irimescu, A.; Tornatore, C.; Marchitto, L.; Merola, S.S. Compression Ratio and Blow-by Rates Estimation Based on Motored Pressure Trace Analysis for an Optical Spark Ignition Engine. *Appl. Therm. Eng.* **2013**, *61*, 101–109. [[CrossRef](#)]
34. Merola, S.S.; Irimescu, A.; Tornatore, C.; Valentino, G. Effect of the Fuel-Injection Strategy on Flame-Front Evolution in an Optical Wall-Guided DISI Engine with Gasoline and Butanol Fueling. *J. Energy Eng.* **2016**, *142*, E4015004. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.