Prediction of CO and HC emissions in Otto motors through neural networks

Predicción de emisiones de CO y HC en motores Otto mediante redes neuronales

Rogelio Santiago León Japa[1], José Luis Maldonado Ortega[1],

Rafael Wilmer Contreras Urgilés[1,*]

Abstract

This paper explains the application of artificial neural networks (ANN) for the prediction of pollutant emissions generated by mechanical failures in ignition engines, from which the percentage of CO (% carbon monoxide) and the particulate in parts per millions of HC (ppm of unburned hydrocarbons) can be quantified, through the study of the Otto cycle intake phase, which is recorded through the physical implementation of a Manifold Absolute Pressure (MAP) sensor. A rigorous protocol of sampling and further statistical analysis is applied. The selection and reduction of attributes of the MAP sensor signal is made based on the greater contribution of information and significant difference with the application of three statistical methods (ANOVA, correlation matrix and Random Forest), from which a database that enables training two backpropagation feedforward neural networks, with which a classification error of $5.4061e^{-09}$ and $9.7587e^{-05}$ for CO and HC, respectively, can be obtained.

Keywords: prediction, pollutant emissions, carbon monoxide (CO), non-combustion hydrocarbons (HC), diagnostics, neural networks.

Resumen

En el presente trabajo se explica la aplicación de RNA (redes neuronales artificiales) para la predicción de emisiones contaminantes generadas por fallas mecánicas en motores de encendido provocado, de la cual se puede cuantificar el porcentaje de CO (% monóxido de carbono) y el particulado por millón HC (ppm hidrocarburos sin quemar), a través del estudio de la fase de admisión del ciclo Otto, la cual es registrada por medio de la implementación física de un sensor MAP (*Manifold Absolute Pressure*). Se aplica un riguroso protocolo de muestreo y consecuente análisis estadístico. La selección y reducción de atributos de la señal del sensor MAP se realiza en función del mayor aporte de información y diferencia significativa con la aplicación de tres métodos estadísticos (ANOVA, matriz de correlación y Random Forest), de la cual se obtiene una base de datos que permite el entrenamiento de dos redes neuronales *feed-forward backpropagation*, con las cuales se obtiene un error de clasificación de $5.4061e^{-9}$ y de $9.7587e^{-5}$ para la red neuronal de CO y HC respectivamente.

Palabras clave: predicción, emisiones contaminantes, monóxido de carbono (CO), hidrocarburos no combustionados (HC), diagnóstico, redes neuronales artificiales.

1. Introduction

At present, the area of automotive transportation represents one of the main sources of air pollution; indeed, the discharge of pollutants to the environment has its origin in the accelerated population growth and the development of different urban centers, thus the deterioration of the quality of the air is due to mobile (vehicles), stationary (industry) and areal (domestic activities and services) sources.

The vehicular area (gasoline and diesel vehicles) is one of the main emitters of burnt fossil fuels to the environment, due to the pollutant gases generated during the operation of the automotive transportation, with the primary emissions being: carbon monoxide (CO), carbon dioxide (CO_2), unburnt hydrocarbons (HC) and nitrous oxides (NO_x), such that they affect public health and the equilibrium of the different ecosystems.

In the light of the previous paragraphs, it is necessary to develop new specialized and methodological techniques to obtain assertive diagnoses of mechanical failures; simultaneously, artificial neural networks and computational mathematics are used, due to the complexity of analyzing and interpreting the operational parameters of the ignition engine, and in this way determining the mechanical failures and the emissions that they produce, in short diagnosis times and optimizing resources. The air pollution has harmful effects on the health of all people, as demonstrated by the research study carried out by Ballester [1], who shows that in France, Switzerland and Austria 6 % of the mortality and an important number of new cases of respiratory diseases can be attributed to the air pollution, with half of this impact caused by the pollution emitted by motor vehicles.

The research by Restrepo emphet al. [2] estimates alarming contributions of polluting emissions that generate smog and contribute to the greenhouse effect in the city of Pereira. The study indicates the contribution to pollution of each vehicular category according to a software of the international model of emissions and an extrapolation, and results indicate that particular vehicles contribute more than 80 % of CO emissions, 60 % of CO_2, 65 % of NO_x, 40 % of SO_x, and motorcycles contribute around 65 % of the particulate material (PM).

The use of neural networks is considered a technique of great contribution in the analysis of internal parameters of the ignition engines, as demonstrated by Li *et al.* [3] through the application of a neural network for predicting NO_x emissions; the study utilizes intensity relations of flame radicals, together with flame temperature and NO_x emissions, to train the neural network.

Cortina [4] proposes a model for predicting the concentration of the pollutants in the city of Salamanca (Mexico), with the most critical pollutants being SO_2 and PM_{10}; the model uses artificial neural networks (ANN) in combination with clustering algorithms, and the study uses particular meteorological variables as factors that influence the concentration of pollutants.

It is relevant to reduce the emissions of CO, HC and NO_x of internal combustion engines with ignition start, since they produce different environmental problems, such as air pollution and global warming. Martinez *et al.* [5] used artificial neural networks (ANN) to predict the exhaust emissions of a 1.6 L ignition engine, with the purpose of optimizing such engine, by reducing the CO, HC and NO_x emissions; the inputs of the ANN were six operating parameters of the engine, and the outputs were the three resulting exhaust emissions.

Similarly, Fontes *et al.* [6] apply multilayer perceptron (MLP) neural networks with a hidden layer as a classifier of the impact of the air quality on human health, using as only inputs traffic and meteorological data. Parallel and combined strategies can be used for determining the concentrations of emissions, for example, a hybrid learning of the artificial neural network (ANN) with the non-dominated sorting genetic algorithm – II (NSGAII) to improve the precision and predict the exhaust emissions of an ignition start gasoline four-stroke engine [7].

Different methods can be applied to analyze and predict emissions, such as the multivariate linear regression model to analyze the relationship between atmospheric pollutants and meteorological factors. Lopez and Pacheco [8] show that the source of benzene is the smoke from tobacco, gas stations, industrial emissions, and exhaust pipes of motor vehicles in the urban zone of the city of Cuenca (Ecuador), which has generated a raise in the number of clinical cases, such as asthma (36.34 %), bronchopneumonia (12.19 %), bronchiolitis (16.89 %), bronchitis (6.29 %), pharyngitis (12.41 %), pneumonia (11.73 %) and rhinitis (3.67 %), all this due to the increase of the concentration of benzene in one unit; on the other hand, the PM_{10} exhibits a positive relationship with the venous thrombosis, causing an increment of 3.56 % of the clinical cases per each unit of increase in its concentration. Guadalupe [9] applies a new methodology for modeling the pollutant emissions from terrestrial moving sources in Quito (Ecuador), the international model of vehicular emissions (IMVE), which encompasses a bottom-up type of methodology that gathers a large amount of information to make up the inventory of emissions.

At present, the strategies for predicting the concentrations of gases are diverse. Leon and Piña [10] present a model for predicting emissions (NO_x, CO, CO_2, HC and O_2) applied to gasoline powered vehicles, with the use of artificial neural networks (ANN); the input variables to the ANN are the mean effective pressure (MEP), RPM, load and MAP, and the model also predicts the load of the engine. Similarly, Contreras *et al.* [11] have proposed a diagnosis system that

can detect mechanical failures in Otto cycle engines with ignition start, by means of artificial neural networks (ANN); the system is based on the use of the signals from the MAP and CMP sensors, and has a classification error of $1.89e^{-11}$.

The prediction system proposed in this work can determine pollutant emissions and mechanical failures caused by particular emissions, such as carbon monoxide (CO) and unburnt hydrocarbons (HC); the Engine Control Unit (ECU) does not carry out the diagnosis performed by the prediction system. The system is based on the pressure of the intake manifold, which is recorded through the physical implementation of a MAP sensor, thus the system has the capability of reducing to a minimum the diagnosis time; in addition, the system does not utilize variables related to the quality of the air of the city neither meteorological variables for training the ANN, thus constituting a significant advance for predicting the emissions of exhaust gases and determining mechanical failures; activating this system in centers of automobile service and vehicular technical revision (VTR) is reliable and accessible.

2. Methods and materials

The main topics are developed in this section. These include experimental configuration and minimally invasive instrumentation, conditions for acquiring the samples, methodology for data acquisition, obtaining the matrix for the analysis and reduction of attributes, selection of the attributes for training the ANN, and neural network algorithm in MATLAB for diagnosing and predicting emissions.

2.1. Experimental configuration and minimally invasive instrumentation

The main consideration of the study is to avoid disassembling the elements and systems of the engine of a vehicle to diagnose the failures and predict the polluting emissions; consequently, the depression of the engine is measured through the installation of a MAP sensor in a vacuum connection of the intake manifold, placing it after the throttle valve such that the connection does not affect the operation of the ignition engine.

Table 1 summarizes the characteristics of the engine under test, and Table 2 includes the applied instrumentation.

Table 1. Characteristics of the experimental unit

Characteristic	Value
Model	Hyundai
Number of cylinders	4
Valve train	DOHC
Injection system	MFI
Cylinder	2000 cc
Power	175 CV @ 6000 rpm
Torque	168 N.m @ 4000 rpm
Fuel	Gasoline (RON 95)
Compression index	10.5: 1

Table 2. Applied instrumentation

Characteristic	Value
Personal computer	PC
Automotive scanner	Multiscan (Hanatech) OBDI, OBDII, EOBD y CAN.
Connection	Motor, Transmisión, ABS, Airbag.
Diagnosis	Ni DAQ-6009 8 AI de 14 bits, a 48 KS/s 2 AO de 12 bits
Data acquisition card	
Inputs	
Outputs	
Intake pressure	MAP
Pressure range	20-117 KPa
Voltage	3.2V @ 1600 m.s.n.m
Sensor type	Piezoelectric
Gas analyzer for vehycles	QROTECH / QGA 6000
Measuring range CO	0.0 9.99 %
Resolution	0.01 %
Measuring range HC	0.0 9999 ppm
Resolution	1 ppm

The identification of each cylinder of the engine is carried out using the record of the signal of the camshaft position (CP) sensor.

Figure 1 shows the experimental unit under test, constituted by a Hyundai Sonata 2.0 DOHC engine, a gas analyzer, a personal computer (PC) and an automotive scanner. Figure 2 shows the connection of the MAP type sensor, the vacuum connection in the intake manifold and the data acquisition tool Ni DAQ-6009.
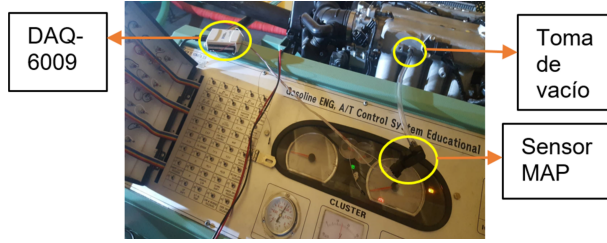

Figure 1. Instrumentation in the engine.

Figure 2. Connection of the MAP sensor.

2.2. Conditions for acquisition of samples

The installation of the intake pressure sensor is carried out, placing it in the intake manifold of the ignition engine, consequently acquiring samples of the emissions of NO_x, CO, CO_2 and HC through the gas analyzer, and recording samples of the signal of the MAP sensor with a Ni DAQ-6009 card together with the software LabVIEW.

The samples of pressure and polluting emissions are acquired at idle condition, at approximately 850 RPM with an engine temperature range between 92 and 99 °C and engine load 35 %; an automotive scanner is used to confirm these conditions. Based on a pre-experimental study carried out in the research, it was determined that the signal of the MAP sensor has peaks of higher frequency, thus the sampling frequency is 10 KHz during a time period of 5 seconds for each of the signals; such frequency is larger than the one established according to Nyquist criterion (1.416 KHz) [11].

2.3. Methodology for data acquisition

Figure 3 presents the physical elements necessary for the corresponding diagnosis of the mechanical failures and prediction of polluting emissions.

Figure 3. Elements necessary for data acquisition.

The procedure represented in the flow diagram of Figure 4, is applied to obtain the signals of the MAP and CP sensors.

The process of obtaining the data starts with the revision of correct operation of the engine or supervised failure, and subsequently the connection of the sensors is inspected. If the connection is correct, the signal is saved with the software LabVIEW and registered in an Excel file, if not, the connection of the sensors

is verified [11]. The previously described procedure is applied to record the signals, both for the cases of engine in good operating condition (Figure 4(a)) and of engine with supervised failure (Figure 4(b)) [11]. The data acquisition process is performed 20 times for each of the engine conditions.

Table 3 indicates the total of six failures that are generated in the ignition engine experimental unit, each with the corresponding identification code; the condition of the engine in optimal operation is also indicated.
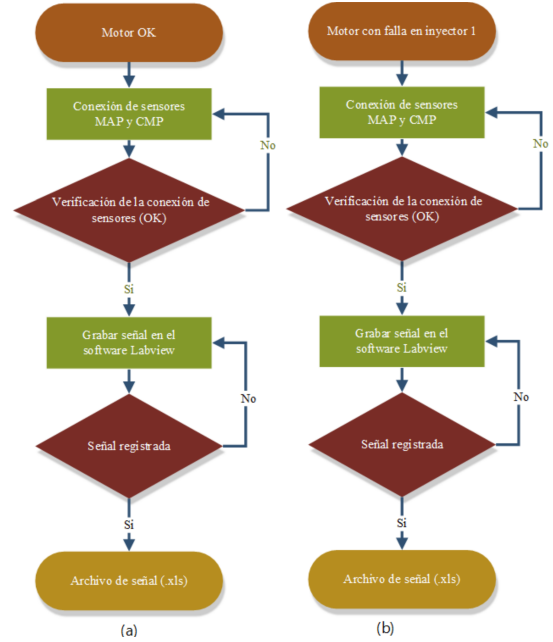

Figure 4. Flow diagram of the procedure for data acquisition (a) engine ok, (b) engine with failure.

Table 3. Operating conditions of the ignition engine experimental unit

N.°	Type of mechanical condition	Identification code
1	Engine in optimal operating condition	100
2	Failure in Injector 1	200
3	Failure in Injector 2	300
4	Failure in Injector 3	400
5	Failure in Injector 4	500
6	Failure in Coil 1-4	1000
7	Failure in Coil 2-3	1100

2.4. Obtaining the matrix of attributes analysis and reduction

It is considered a complete segment of the signal of the MAP sensor, which corresponds to a cycle of the engine (720°±180°) taking into account the timing distribution of advance to intake opening (AIO) and delay

of intake closure (DIC), for each of the cylinders [11]. A windowing of the signal of the MAP sensor is carried out for each cylinder, as can be observed in Figure 5.

Once the time signals have been taken, an algorithm is developed in the software Matlab for reading and obtaining the general matrix with 18 attributes, namely: geometric mean, maximum, minimum, median, covariance, variance, standard deviation, mode, kurtosis factor, skewness coefficient, energy, power, area under the curve, entropy, coefficient of variation, range, mean square root and crest factor [11].
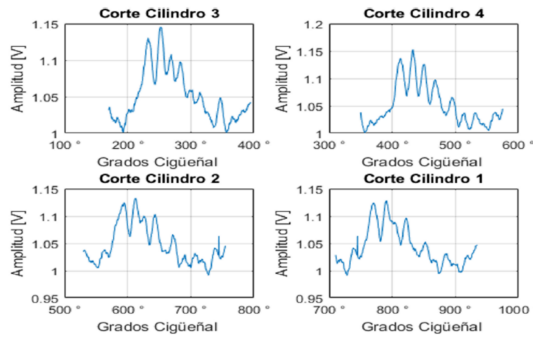
Figure 5. Windowing of the MAP sensor signal for each cylinder.

For the selection and reduction of the number of attributes, the general matrix is analyzed through 3 statistical methods: ANOVA, correlation matrix and Random Forest.

The application of the single factor ANOVA statistical method, allows to determine the best attributes that enter to the general matrix, through the analysis of the 18 attributes considering the greatest value of R^2, since values close to 100 % indicate that there is a correct fit of the model to the data, in other words, the variation between the attributes is determined. In addition, values of p close to 0 are considered, which determine if the attributes are statistically significant [11].

With respect to the correlation matrix, the attributes with coefficients close to -1 or 1 were discarded, since with them there is a strong relationship between the variables, negative or positive, respectively. Indeed, the attributes with coefficients close to zero were chosen, because with those attributes there is no strong correlation between the variables [11].

Regarding the Random Forest method, it allows to obtain the estimation of the importance of the attributes using the Curvature test, Standard CART and Interaction test methods. Afterwards, a Pareto analysis was applied to choose the attributes with greatest priority, considering only the top 95 % of the accumulated distribution [11].

2.5. Selection of attributes for training the ANN

In order to select the attributes that will be considered as inputs of the neural network, it was performed a match analysis of the general matrix from which the most often repeated attributes are chosen among the results of each statistical method applied [11]. The most often repeated attributes are shown in Table 4.

Table 4. Attributes utilized for training the artificial neural network

Statistical attributes	Number of repetitions
Área (v^2)	5
Energy (J)	5
Entropy (J)	4
Maximum (V)	4
Mean (V)	5
Minimum (V)	5
Power (mW)	5
RMS (V)	5

2.6. Matlab algorithm of the neural network for the diagnosis and prediction of emissions

Figure 6 presents the flow diagram of the procedure for creating the artificial neural networks corresponding to CO and HC.

The algorithm initiates reading the matrix of inputs and corresponding responses of the ANN. Then, the input and response vectors are normalized using the maximum value of each matrix, with the purpose of optimizing the creation of the ANN. Once the matrix of attributes was normalized, the ANN was created [11].

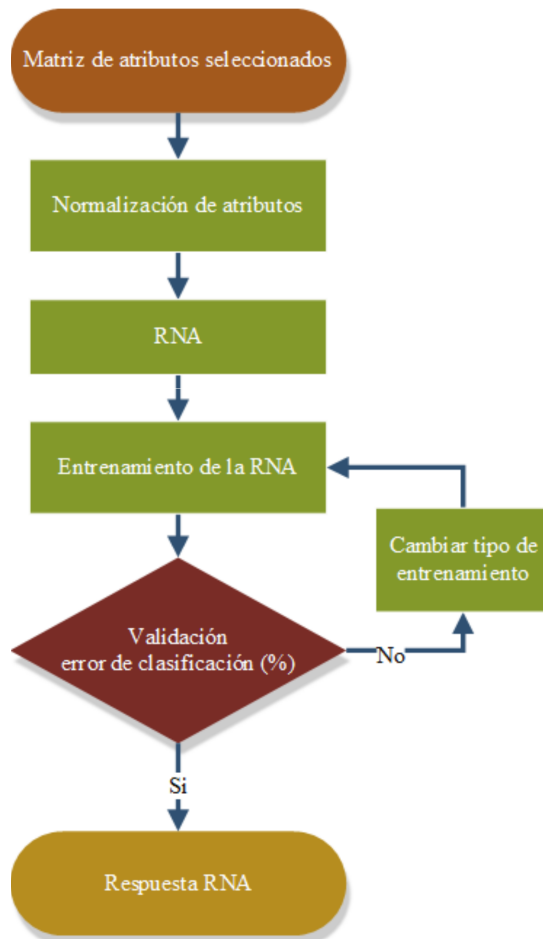The neural networks are established based on the characteristics indicated in Table 5.

Figure 6. Flow diagram of the procedure for creating the ANN.

Table 5. Characteristics of the neural networks models

Characteristic	Tipo
Model	Backpropagation
Typology	4 layers Feedforward
On line/Off line	Off
Learning	Unsupervised
Learning rule	Error correction
Association	Hetero-asociation
Output	Analogous

Figure 7 includes the parameters of the creation of the feedforward backpropagation ANN for predicting the emission of the CO pollutant.

Similarly, Figure 8 includes the parameters of the creation of the feedforward backpropagation ANN for predicting the emission of the HC pollutant.

Once created, the networks were trained considering parameters such as: type of algorithm, number of epochs and maximum error.

The steps and formulas utilized to train the neural network are presented in the following:

1. The weights of the neural network are initialized with small random values.

2. An input pattern is entered to the network with the different conditions of the engine Xp (Xp1, Xp2, ... Xpn) and the target output of the network is specified as Ym, which would be the value of emissions.

3. The actual output of the network is calculated.

The architecture of the network is shown in Figure 9, where subscript p indicates the p-th training vector, j is the number of hidden unit and the index i varies from 1 to the number of units of the input layer.

The classification error was verified for the CO and HC networks previously trained; if the error is greater than 5%, the parameters are changed to reduce such error.
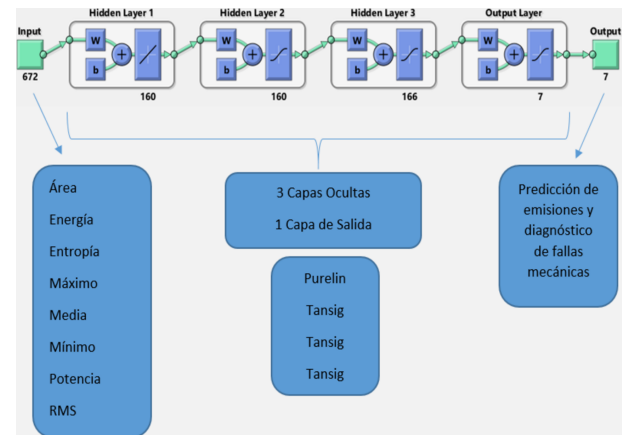

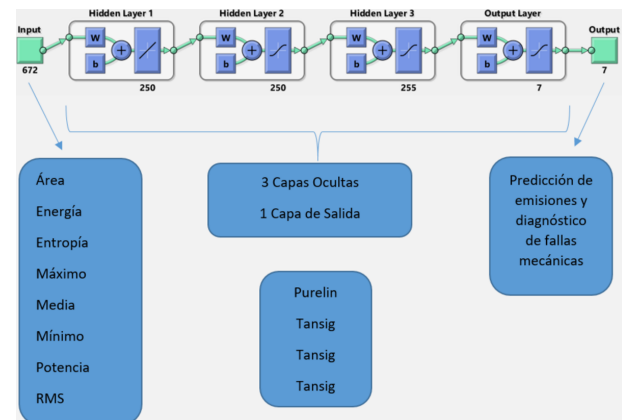
Figure 7. Structure of the CO neural network.

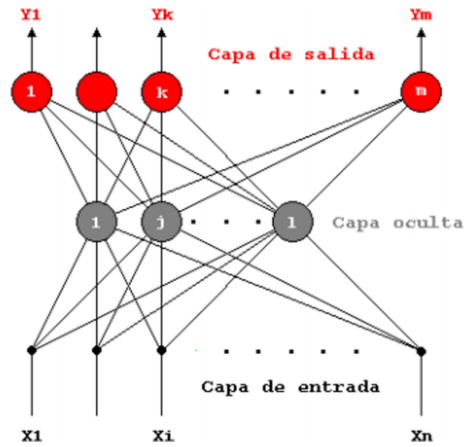**Figure 8.** Structure of the HC neural network.

Figure 9. Architecture of the feedforward network.

The CO neural network utilizes the trainscg (Scaled Conjugate Gradient) training function for the system that predicts emissions and diagnoses mechanical failures, which presented an error of $5.4061\mathrm{e}^{-9}$.

Similarly, the HC neural network utilizes the trainscg training function for predicting emissions and diagnosing mechanical failures, which presented an error of $9.7587\mathrm{e}^{-5}$.

Figures 10 and 11 present the results of the Pearson correlation coefficient R of the CO and HC neural networks, respectively, which is provided by the training code red.trainFcn='trainscg' of the Matlab software.
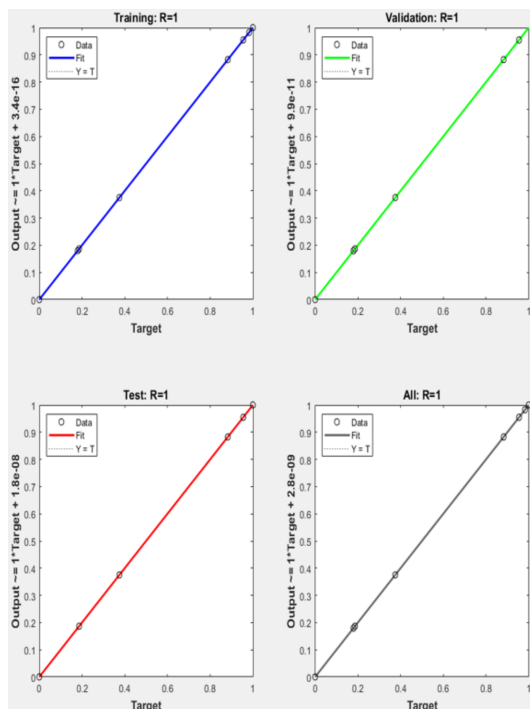
Figure 10. Correlation between the target values and the values predicted by the CO neural network.
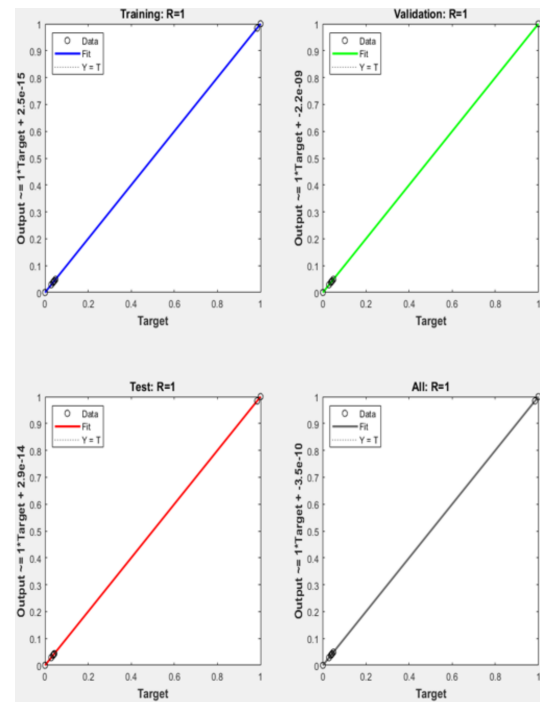
Figure 11. Correlation between the target values and the values predicted by the HC neural network.

The lines indicate the target values and the black circles represent the values predicted by the ANN. The prediction of the neural network is efficient and verifies a good performance, since it yields a global index of 1 in training, validation and testing, which indicates a strong positive linear relationship between the real conditions of the ignition engine and the results provided by the neural network [11].

Figures 12 and 13 show a comparison between the responses of the CO and HC neural network, respectively, and the corresponding target values; observe the seven actual mechanical conditions of the engine identified by the neural networks.
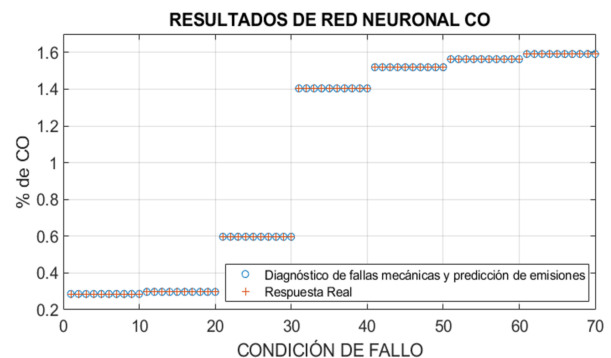

Figure 12. CO neural network with percentage of error $5.4061\mathrm{e}^{-9}$, with training function trainscg.
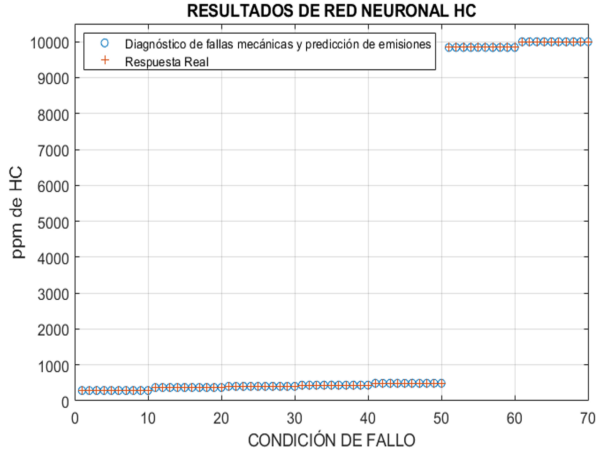
Figure 13. HC neural network with percentage of error $9.7587\mathrm{e}^{-5}$, with training function trainscg.

3. Results and discussion

Various tests were carried out under diverse operating conditions, for the purpose of comparing the correct performance of the system that predicts emissions and diagnoses mechanical failures.

Two specific failure conditions are presented in this section: injector 2 (300) and coil 1-4 (1000).

Figure 14 shows the results of the values obtained by the CO neural network for the operating states when injector 2 fails.
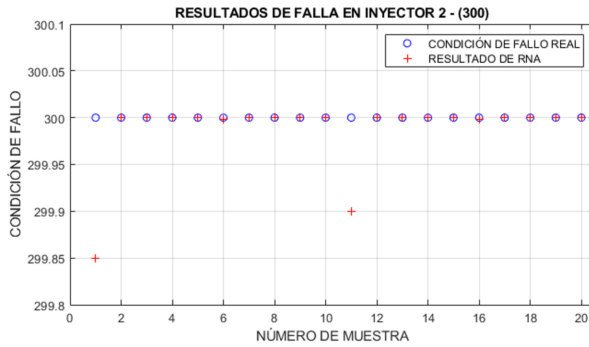
Figure 14. Result of the operating condition of injector 2.

Figure 15 shows the results of the values obtained by the HC ANN for operating conditions when coil 1-4 fails.

After obtaining the results of the operating conditions of the ignition engine, it may be remarked that the differences between the actual responses and the responses given by the CO and HC neural networks have a value close to zero. Therefore, the application of the system that diagnoses mechanical failures and predicts the pollutant emissions is capable of detecting the operational condition of mechanical failure and predicting the pollutant emission.
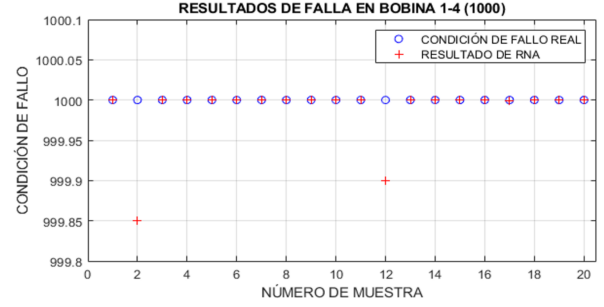

Figure 15. Result of the operating condition of the high voltage ignition coil 1-4.

Indeed, Figure 16 shows that after grouping the data corresponding to the actual engine condition and the responses obtained by the CO ANN, and in Figure 17 the corresponding to the HC ANN, using the Tukey statistical method with a confidence interval (CI) of 95 %, it is determined that the means are equivalent and there is no statistically significant difference, because the means of each of the responses coincide in a value close to zero.

In addition, Figures 18 and 19, which show the intervals of the CO and HC ANNs, respectively, indicate that there is no difference between the averages of the tests in the different operational conditions of the ignition engine.
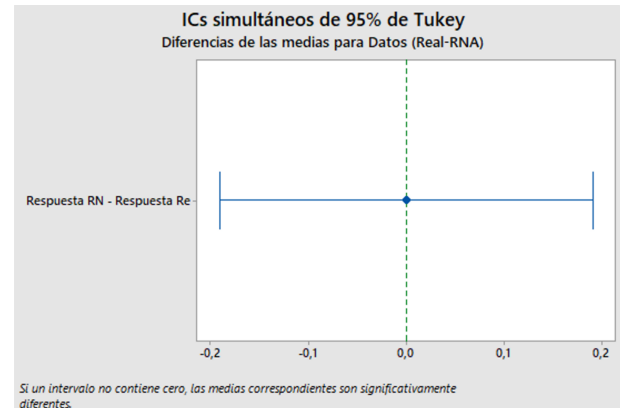

Figure 16. Graph of the differences of the means for data of actual response vs. the CO neural network.
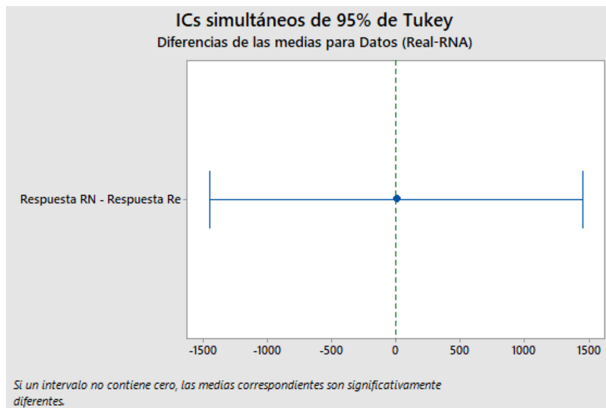
Figure 17. Graph of the differences of the means for data of actual response vs. the HC neural network.
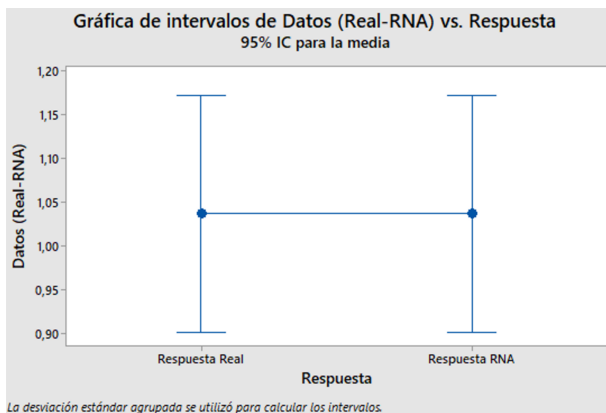

Figure 18. Graph of the data intervals of actual response vs. CO neural network.
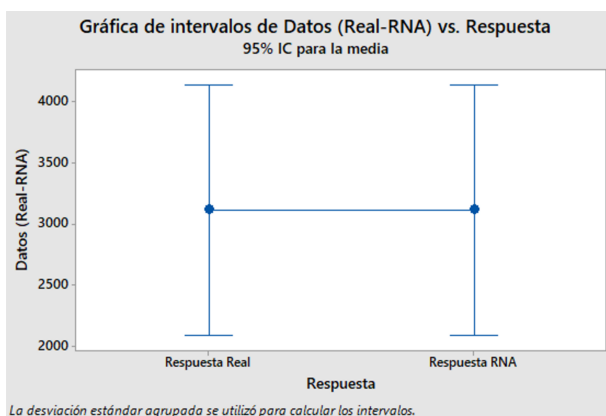

Figure 19. Graph of the data intervals of actual response vs. HC neural network.

Similarly, Figures 20 and 21 corroborate that there is a relationship between the actual response and the response of the neural network, because they share the same grouping letter (A) and the p-value is equal to 1. This results in a reliability value of approximately

100.00 %, which is acceptable for issues of diagnosis of mechanical failures and prediction of polluting emissions of internal combustion ignition engines.

Figure 20. Results of the analysis of variance and comparisons of Tukey pairs of the CO ANN.

Figure 21. Results of the analysis of variance and comparisons of Tukey pairs of the HC ANN.

4. Conclusions

The developed neural network models for the diagnosis and prediction of polluting emissions of both CO and HC, yield classification errors of $5.4061e^{-9}$ and $9.7587e^{-5}$, respectively.

The trainscg training function allows the precise identification of different types of mechanical conditions of the ignition engine and prediction of emissions, thus constituting a viable alternative to be integrated in a diagnosing system such as an automotive scanner or gas analyzer of gasoline powered vehicles, due to the computational speed offered by the artificial neural networks.

By means of the single factor analysis of variance a p-value equal to 1 was obtained, thus demonstrating that the actual response of classification of mechanical failures and prediction of emissions is equivalent to the result obtained through the developed neural networks, such that this value confirms that there is no statistically significant difference.

This work shows that the application of backpropagation feedforward neural networks is valid for the detection of mechanical failure conditions, as well as for the prediction of polluting emissions in gasoline powered vehicles; besides, the applied diagnosis technique has the advantage of avoiding disassembling elements and systems of the engine, by offering a technique which is minimally invasive, reliable and of great precision.

Results show that backpropagation feedforward neural networks with 160 or 250 hidden units and trained with the trainscg (Scaled Conjugate Gradient) function, may yield an average error of $4.87962\mathrm{e}^{-5}$, which demonstrates that the emissions of gasoline powered vehicles can be predicted with high precision.

References

[1] F. Ballester, "Contaminación atmosférica, cambio climático y salud," *Revista Española de Salud Pública*, vol. 79, pp. 159–175, 04 2005. [Online]. Available: https://bit.ly/2piHMQZ

[2] A. Restrepo, S. Izquierdo, and R. López, "Estimación de factores que inciden sobre la contaminación ambiental generada por fuentes móviles en pereira," *Scientia et technica*, vol. 1, no. 37, pp. 267–272, 2007. [Online]. Available: http://dx.doi.org/10.22517/23447214.4053

[3] X. Li, D. Sun, G. Lu, J. Krabicka, and Y. Yan, "Prediction of nox emissions throughflame radical imaging and neural network based soft computing," in *2012 IEEE International Conference on Imaging Systems and Techniques Proceedings*, July 2012, pp. 502–505. [Online]. Available: https://doi.org/10.1109/IST.2012.6295594

[4] M. Cortina, "Aplicación de técnicas de inteligencia artificial a la predicción de contaminantes atmosféricos," Ph.D. dissertation, 2012.

[5] J. D. Martínez-Morales, E. R. Palacios-Hernández, and G. A. Velázquez-Carrillo, "Modeling and multi-objective optimization of a gasoline engine using neural networks and evolutionary algorithms," *Journal of Zhejiang University SCIENCE A*, vol. 14, no. 9, pp. 657–670, Sep 2013. [Online]. Available: https://doi.org/10.1631/jzus.A1300010

[6] T. Fontes, L. M. Silva, S. R. Pereira, and M. C. Coelho, "Application of artificial neural networks to predict the impact of traffic emissions on human health," in *Progress in Artificial Intelligence*, L. Correia, L. P. Reis, and J. Cascalho, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 21–29. [Online]. Available: https://doi.org/10.1007/978-3-642-40669-0_3

[7] J. D. Martínez-Morales, E. R. Palacios-Hernández, and G. A. Velázquez-Carrillo, "Artificial neural network based on genetic algorithm for emissions prediction of a si gasoline engine," *Journal of Mechanical Science and Technology*, vol. 28, no. 6, pp. 2417–2427, Jun 2014. [Online]. Available: https://doi.org/10.1007/s12206-014-0535-2

[8] T. López Ortíz and A. Pacheco González, "Efectos de la contaminación atmosférica en la salud de las personas en la ciudad de cuenca," 2015. [Online]. Available: https://bit.ly/32M8jnz

[9] J. Guadaluoe Almeida, "Modelación de emisiones contaminantes de fuentes móviles terrestres en Quito, Ecuador," 2016. [Online]. Available: https://bit.ly/2qdvtG8

[10] P. León Bacuilima and C. Piña Orellana, "Predicción de emisiones contaminantes de gases de escape a través de la presión media efectiva empleando redes neuronales en motores de encendido provocado, cuenca," 2018. [Online]. Available: https://bit.ly/2q2oTlT

[11] W. Contreras, J. Maldonado, and R. León, "Aplicación de una red neuronal feed-forward backpropagation para el diagnóstico de fallas mecánicas en motores de encendido provocado," *INGENIUS*, 2019. [Online]. Available: https://doi.org/10.17163/ings.n21.2019.03