Revista de Chimie
https://revistadechimie.ro
https://doi.org/10.37358/Rev. Chim.1949

Comparison of an Artificial Neural Network and a Multiple Linear Regression in Predicting the Heat of Combustion of Diesel Fuel Based on Hydrocarbon Groups

YOUNIS M. YOUNIS[1], SALMAN H. ABBAS[1], FARQAD T. NAJIM[2], FIRAS HASHIM KAMAR[1], GHEORGHE NECHIFOR[3]*

[1]Engineering Technical College, Middle Technical University, Baghdad, Iraq
[2]Electrical Engineering Department, College of Engineering, Al-Iraqia University, Baghdad, Iraq
[3]University Poltehnica of Bucharest, Department of Analytical Chemistry and Environmental Engineering, 1-7 Gheorghe Polizu, Str. 011061, Bucharest, Romania

Abstract*. A comparison between artificial neural network (ANN) and multiple linear regression (MLR) models was employed to predict the heat of combustion, and the gross and net heat values, of a diesel fuel engine, based on the chemical composition of the diesel fuel. One hundred and fifty samples of Iraqi diesel provided data from chromatographic analysis. Eight parameters were applied as inputs in order to predict the gross and net heat combustion of the diesel fuel. A trial-and-error method was used to determine the shape of the individual ANN. The results showed that the prediction accuracy of the ANN model was greater than that of the MLR model in predicting the gross heat value. The best neural network for predicting the gross heating value was a back-propagation network (8-8-1), using the Levenberg–Marquardt algorithm for the second step of network training. R = 0.98502 for the test data. In the same way, the best neural network for predicting the net heating value was a back-propagation network (8-5-1), using the Levenberg–Marquardt algorithm for the second step of network training. R = 0.95112 for the test data.*

Keywords*: Artificial neural network, Gross heat combustion, Net heat combustion, Multiple linear regression, Diesel fuel*

1. Introduction

Diesel fuel technology has greatly improved over the last decade, with high-pressure direct injection being commonly used, with the addition of a new process for supercharging internal combustion engines to increasing power [1,2]. Diesel engines are mostly used in trade vehicles, such as buses and trucks, because of their high fuel efficiency, but high-strength particulate formation in the combustion processes is a major worry due to the health and environmental effects of particulates [3]. Combustion is a major source of power generation, significantly so in automotive applications, so operating systems that reduce fuel consumption and high-efficiency direct-injection diesel engines are used [4-6].

Diesel is one of numerous petroleum products that is used as a fuel in all types of compression-ignition engines [7] [8]. It is produced from crude oil, extracted from oil wells. To produce diesel fuel, crude oil goes through fractional distillation at atmospheric pressure and temperatures between 250 and 350˚C [7]. Diesel fuel can broadly be defined as "... a very complicated blend of thousands of different organic compounds" [9]. In general, its carbon numbers are between 11 and 22 [9-11], and such compounds can be classified as paraffin, naphthene or aromatics. These families can play a significant role in the chemical and physical properties of diesel, with different proportions of these being one of the factors that distinguishes diesel fuel from other diesel compounds, and affects the properties of diesel fuel performance and combustion [12]. Other elements found in diesel fuel include low amounts of sulphur, oxygen and nitrogen. These elements are known as heteroatoms [10].

**email: doru.nechifor@yahoo.com*

Diesel's heat of combustion (also known as the gross heat of combustion [GHC] or gross heating value [GHV] is fundamental in controlling and planning diesel engines. During combustion, the enthalpy of the entire combustion of the fuel converts carbon into carbon dioxide, whilst all the hydrogen is converted into water [13]. The GHC is the amount of energy released when a unit mass of fuel is burned at constant volume, producing gases, with the water being condensed into its liquid state [14]. The net heat of combustion (NHC) is the amount of energy released when a unit mass of fuel is burned at constant pressure [14]. The GHC is greater than the NHC. The term 'higher heating value' refers to the GHV and the lower heating value, and is also known as the net heating value (NHV). In Europe, the NHC is generally used, but the GHC is used in the United States. Compared with the GHC and NHC values, they are approximately 8 or 9% [9].

One of the greatest challenges in combustion is the many chemical reactions that occur. It is often necessary to know the energy generated so that the most suitable fuel for a particular purpose can be identified. Currently, global warming, created by greenhouse gases in the atmosphere, poses a major climatic problem for our planet. Thus, environmental protection is important for our world's future habitability.

Powerful modelling techniques are available that can be used to identify highly complex, nonlinear relationships between input and output data and neural networks. Artificial neural networks (ANNs) describe such relations by choosing network weights, using a trial-and-error-based calculation method and a training algorithm, such as the Levenberg-Marquardt algorithm [15-17].

Saldana *et al.* [18] predicted the melting points and NHC of diesel fuel. The compounds proposed by their models are oxygenated compounds and hydrocarbons. The models were explained using two groups (molecular and functional), with different numerical algorithms producing nonlinear and multilinear predictive models. In determining the effects of fuel compounds on predicting the heat of combustion, Merckel *et al*. [19] observed that the heat of combustion is more sensitive to variations in the percentage of oxygen in the fuel than to similar changes in the content of carbon and hydrogen.

Yang *et al*. [20] examined a back-propagation neural network model, using validation and training data, and established a multiple linear regression (MLR) for predicting the cetane index (CI) of diesel fuel, based on its chemical composition, which matched with a model developed using an ANN. Ramón *et al.* [21] compared an ANN with a MLR in order to predict the CI of biodiesel from a fatty acid methyl ester. The model attained an accuracy of >92%.

Pan *et al.* [22] predicted the NHC of organometallic compounds containing carbon silicon bonds based on molecular structures, and they applied a MLR equation to predict the NHC. The results explained that the model is powerful for predictive use, and that it could provide a sensible prediction for most organic silicon elements.

The standard NHC of pure hydrocarbons was predicted by Albahri [23], based on their atomic structures, and using a least-squares-method-based multionlinear regression. He stated that the method is very straightforward and does not require empirical data, and can predict heat combustion just by knowing the molecular structure.

As noted by Tareq [24] , ANNs are far more accurate than MLRs and least squares. In that work, a quantitative structure-property relation method was used to break down a molecule into a series of numerical values. The outcome predicted the NHC of the chemical compound based on molecular structure. The ANN in the proposed system accurately predicted the correlation coefficient (R^2 = 0.999) and an average relative error of 0.89%.

Similarly, Zhou *et al.* [25] predicted the heat of the interaction of organic peroxides, and investigated the effects of the molecules on the model, finding a strong relationship between the linear relationship and the heat of reaction. Erdi *et al.* [26] compared an ANN with a linear regression model, using different types of fuel (biodiesel/alcohol mixtures). They suggested that the ANN model could predict the performance and emission of diesel fuel by using different types of fuel. The ANN results were far more accurate than other models at predicting some of the performance using an emission-data test. They pointed out that ANNs are few cost-effective, and thus avoid wasting both time and

money. In the same vein, Samet [27] suggested that ANN applications could help in estimating the performance of engines and their exhaust emissions, thus also reducing time and cost .

The aim of this work was to predict the heat of combustion of diesel, based on eight parameters that are present in diesel fuel. The methodological approach taken was a mixture of experimental data and software programming. The parameters cetane number and cetane index were determined based on the fuel's density, cetane improver-2-ethylhexyl nitrate, aromatics, polynuclear aromatics, aromatic jet fuel and naphthalene. This method presented some practical advantages that reduced costs and saved time when choosing the appropriate modelling.

2. Materials and methods

2.1. Diesel experimental results

A total of 150 samples of diesel fuel were obtained from 150 trucks that stopped at the Bismayah Gas Power Station in Iraq. About 10 mL of each sample was taken and stored in bottles at room temperature, following which they were subjected to fuel gas chromatography analysis, which provided details on the chemical composition of each sample.

2.2. Statistical analysis

An ANN and a MLR were used to predict the heat of combustion of the diesel samples (GHC, NHC) using chemical composition as the input parameter. ANN is often obtained in three steps-an input layer, an output layer and hidden layers. The next ANN training steps involved the application of MATLAB 2017. The MATLAB programme has well-known hidden inputs, outputs and neurons, and the ability to easily change hidden neurons. Also, it includes various ways of dividing the data, such as block or random, and the percentage of training, validation and test subsets can easily be changed. Previous studies have mostly defined the neurons as all layers having a particular number of small individuals. The neurons are bonded to each other by joint connections that are related to the link weights. The neuron in a certain layer receives information from all the neurons in the previous layer. The information is summed, and weighted by a factor linking to a connection. The network is based on a signal that is transported to the neurons by the weights of the connections [28]. Several reports have shown that an activation or transfer function determines the output of such networks. The transfer functions generally used are logsig, tansig and purelin, and the general formula is shown in Equation (1) [29].

$$f(x) = \frac{1}{1+\exp(-x)} \quad (1)$$

The tan-sigmoid (tansig) is used in a multilayer network to transport the function; the general formula is shown in Equation (2).

$$f(x) = tanh = \left(\frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}\right) \quad (2)$$

If there are sigmoid neurons in the last layer of a multilayer network, the network input will be limited to a small scale. The general formula for purelin is given by:

$$f(x) = x \quad (3)$$

In this study, the best ANN for predicting the heat of combustion was a back-propagation model, the steps for using this are as follows [30, 31] . First, start the weights matrix using random values. Then, predict the output for the input parameter. Next, calculate the errors at the output layer for the output neurons and the correlation coefficient (R, Equation 4). Following that, calculate the mean bias error (MBE, Equation 5) between the output and the hidden layer. Then, calculate the mean absolute error (MAE, Equation 6) and the root mean squared error (RMSE, Equation 7) between the input layer and the hidden layer. Finally, repeat the second step to the final step until the minimum error is reached.

$$R = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - Y_i^-)^2 - \int_{i=1}^{n}(Y_i - Y_i^\wedge)}{\sum_{i=1}^{n}(Y_i - Y_i^-)}} \quad (4)$$

$$MBE = \frac{1}{n}\sum_{i=1}^{n}(Y_i^\wedge - Y_i) \quad (5)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|Y_i^\wedge - Y_i| \quad (6)$$

$$RMSE = \sqrt{\frac{1}{n} * \sum_{i=1}^{n}(Y_i^\wedge - Y_i)} \quad (7)$$

Where the Levenberg–Marquardt algorithm was used for the learning. The training is automatically discontinued when the generalisation process stops improving. This is explained by increasing the mean square error (MSE) that is used to validate the samples. The correlation coefficient values measure the relationship between the outputs and targets. If R = 1, that indicates a close relationship. The MSE obtained is the average squared difference between the outputs and targets, with lower values being better and 0 meaning no error [32, 33].

MLR is commonly employed in predictive studies, and is used to describe the relationship between two or more independent variables and a continuous dependent variable. Here, it was used to describe the relationship between each of the independent variables (cetane number, cetane index, cetane improver-2-ethylhexyl nitrate, fuel density, aromatics, polynuclear aromatics, aromatic jet fuel, naphthalene) and the dependent variables of the GHC and NHC of the diesel samples. In MLR, the adjusted coefficient of determination, R^{adj}, gives an idea of how well a multiple regression equation fits the sample data. In a perfect case, its value will be as close as possible to the value of R^2, which means that R^{adj} is closely related to R^2. Variance is a measure of how the observed values differ from the average of the expected values; the lower the variance, the better.

3. Results and discussions

3.1 Determination of fuel properties

The fuel properties are shown in Table 1. These are used in the statistical analysis.

Table 1. Properties of diesel

Parameter	Results between		Total number of results
	The lowest value	The maximum value	
Cetane number	51.4	54.7	150
Cetane index	51.3	53.9	150
Cetane improver (2-EHN)	266	430	150
PNA-DISEL-HPLC wt%	3.1	4.6	150
ARO-DIESEL-HPLC wt%	17.5	20.4	150
Density g/cc	0.8199	0.84	150
Naphthalene wt%	3	5.5	150
ARO-JET-HPLC wt%	16.1	20.5	150
Gross heating value MJ/Kg	45.372	45.611	150
Net heating value MJ/Kg	42.921	43.093	150

3.2 ANN results

Using the Neural Network Toolbox in MATLAB 2017, the data were divided into two groups, according to the training and testing data presented in Table 1. Data selection criteria were chosen. One hundred diesel fuel samples were used as a neural network training dataset, and 50 were used to test the dataset. Figure 1 shows the main structure of the ANN. The input layer represents eight neurons, and the output layer represents one parameter (GHV or NHV). The trial-and-error method determined the number of neurons in the hidden layer. For the multiple input, single output (MISO), first, the GHV was obtained without using the net heating value. Second, the same data was used to obtain the NHV without using the GHV.

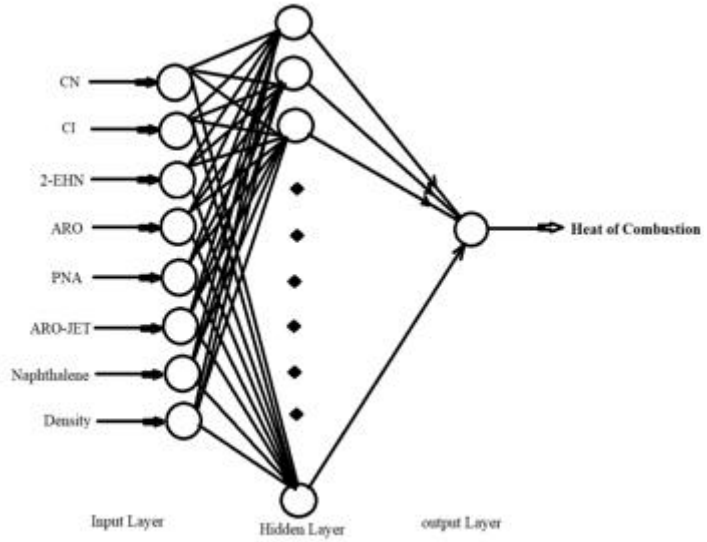

Figure 1. Structure of the ANN – MISO

The learning algorithm and ANN structure are presented in Table 2 for all parameters. The Levenberg–Marquardt training algorithm was applied to train the ANN.

The results obtained from the preliminary analysis of the linear correlation coefficient (R = 0.90967) of the training dataset (GHV) are shown in Figure 2. A positive correlation was found between the experimental GHV and the output. This resulted in a lower MSE for the training dataset of 0.00006951. Interestingly, there were also differences in the error for the training data, which was very small. Figure 3 presents the results obtained from the target and output (error histogram of diesel) data.

The regression and error schemes describe the state of the ANN performance relating to the data training. Thus, the test network is the best indicator for understanding if the network performance was suitable or not, and the test data is used outside the training. Accordingly, 50-sample test dataset was used (Table 1). This dataset test gave an unexpected result. Figure 4 presents an overview of the test ANN, where the results were MSE = 0.00005616 and R = 0.98502. However, the observed difference in the linear correlation coefficient (R) between the training and testing set, in this study, was significant, but satisfactory, which means there was a positive network response when adding new data.

To predict the NHV, the same methodology was used as in the previous simulation.

The results of the linear correlation coefficient are shown in Figure 5, where R = 0.97073 and MSE = 0.00002757. It can be seen from the data in Figure 6 that the error is very small. From Figure 7, the linear correlation coefficient of the test is R = 0.95112 and MSE = 0.000059157.

Table 2. Learning algorithms and ANN structures

Output Parameter	Learning Algorithm	ANN	Hidden Layer Transfer Function	Output Layer Transfer Function
Gross heating value	L-M	8- 8- 1	$f(x) = \left(\frac{2}{1+e^{-2x}}\right) - 1$, Sigmoid Symmetric Transfer Function , (logsig)	$f(x) = x$,Linear Transfer Function (Purelin)
Net heating value	L-M	8- 5- 1	$f(x) = \left(\frac{2}{1+e^{-2x}}\right) - 1$, Sigmoid Symmetric Transfer Function , (logsig)	$f(x) = x$, Linear Transfer Function (Purelin)

Revista de Chimie
https://revistadechimie.ro
https://doi.org/10.37358/Rev. Chim.1949

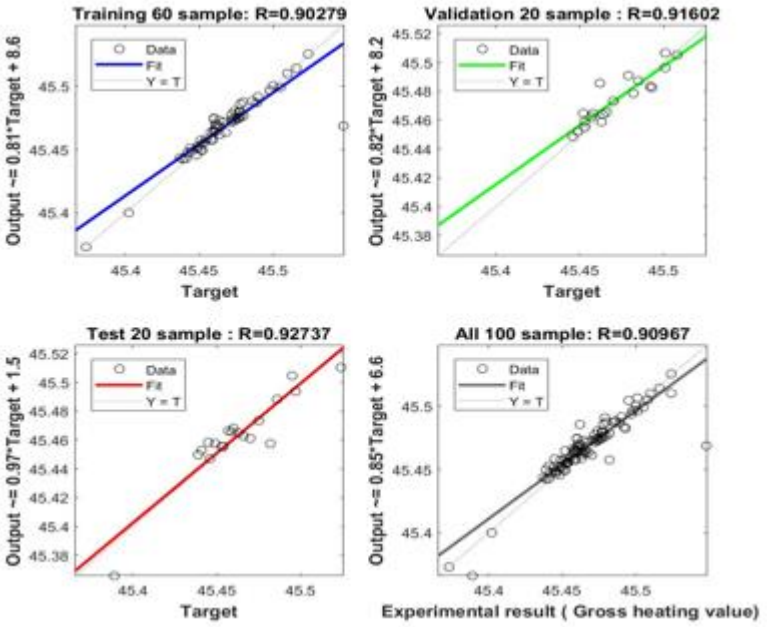

Figure 2. ANN prediction results for GHV

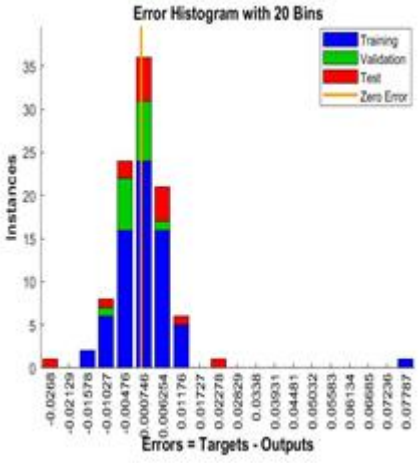

Figure 3. Error histogram of GHC

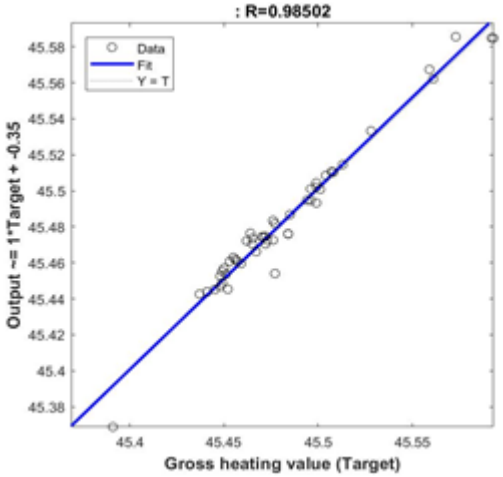

Figure 4. ANN prediction results for GHC (test data)

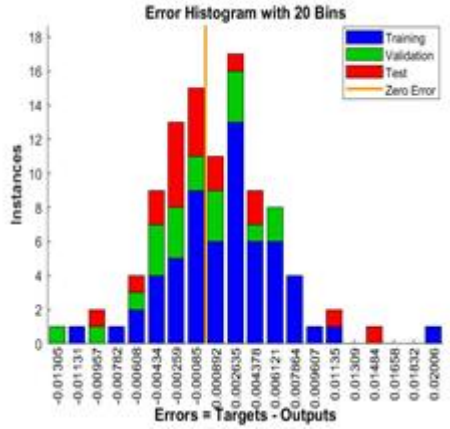

Figure 5. ANN prediction results for net heating value

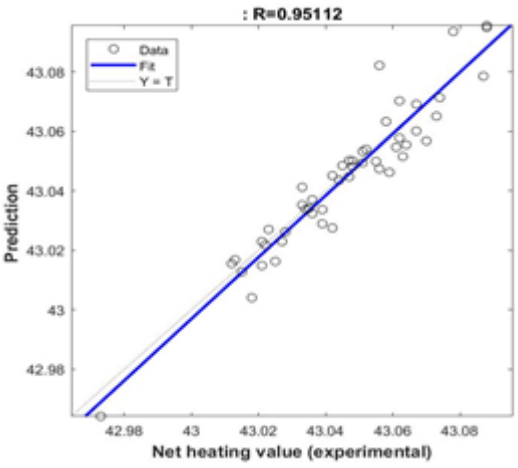

Figure 6. Error histogram for GHC

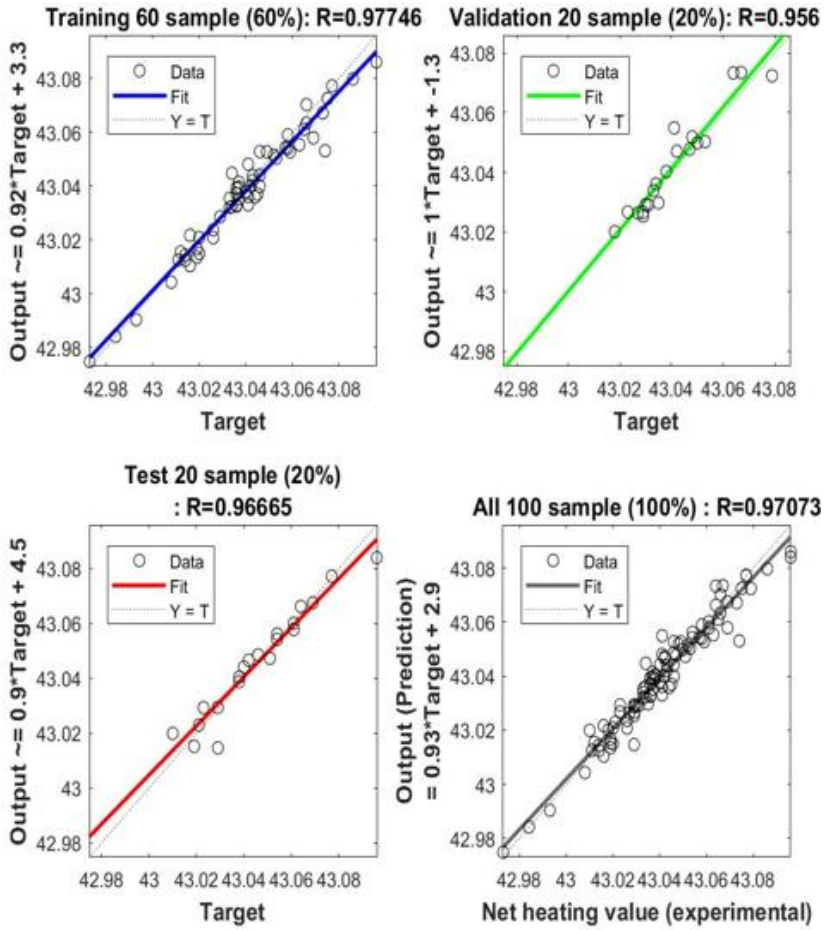

Figure 7. ANN prediction results for NHC (test data)

3.3 Multiple linear regression

Table 3 shows the summary MLR statistics for the GHV and NHV of diesel fuel. It can be seen that the equation was easy to use to estimate the heat combustion of the diesel fuel. This data shows that the GHV resulted in R^2, with the R^{adj} being the lowest NHV. The results in this table can be compared with the results from the ANN that appear in Figures 3-7.

Both the R^2 and R^{adj} give an idea of the number of data points that fall within the regression line, and this helps us to evaluate the number of predictors in this model. The RMSE is an absolute measure of fit, the lower values indicating a better fit.

The results of the R^2 GHC and NHC are shown in Figures 8 and 9.

In general, the models were successful in being able to predict the heat combustion of diesel from the hydrocarbon groups using the ANN and MLR. In summary, a model for predicting the combustion heat in diesel fuel was obtained using an ANN with an accuracy of >90% for the GHC and >95% for the NHC. The best ANN prediction for GHC by the learning algorithm was 8-8-1, and R = 0.98502 for the sample test. The best ANN prediction for NHC by the learning algorithm was 8-5-1, and R = 0.9511 for the sample test.

In the MLR model, $R^2 = 0.8414$ and $R^{adj} = 0.8274$ for the GHC, and $R^2 = 0.9518$ and $R^{adj} = 0.9476$ for the NHC. It was found that the ANN modelling approach produced more accurate results for all expected parameters for the GHC. In general, it seems that the RMSE values of the ANN model were less than those of the MLR. Both models (ANN, MLR) produced satisfactory results. These results indicate that the ANN is more accurate and acceptable for use in prediction than MLR.

Table 3. Summary statistics of the MLR

Output	Equation (MLR)	Statistics	
Gross heating value[a]	Gross heating value=42.46417 -0.0024592 CN +0.0487904 CI -0.0001833 2EHN -0.01682 PNA -0.0287311 ARO + 0.9677149 Density -0.0484407 Naphthalene+ 0.0388306 ARO_JET	R^2	0.8414101
		R^{adj}	0.8274681
		RMSE	0.0010054
		Variance	0.0001111
Net heating value[b]	Net heating value = 39.93309 - 0.0051329 CN + 0.051869 CI - 0.0001503 2EHN - 0.0262604 PNA + 0.0110892 ARO + 0.8240147 Density -0.0237033 Naphthalene + 0.0027313 ARO_ JET	R^2	0.9518979
		R^{adj}	0.9476691
		RMSE	0.0004752
		Variance	2.481E-05

[a,b](Number of independent variables = 8, Regression including a free parameter, Number of observations= 100)

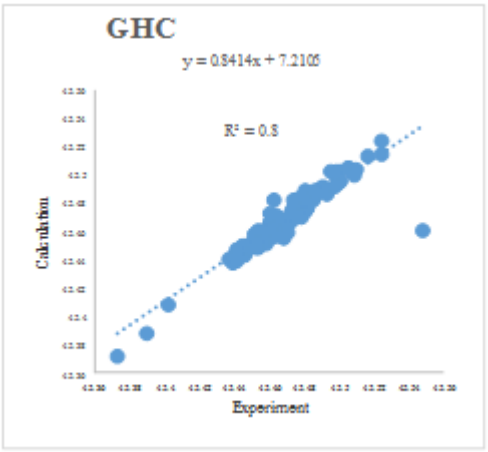

Figure 8. MLR for GHC

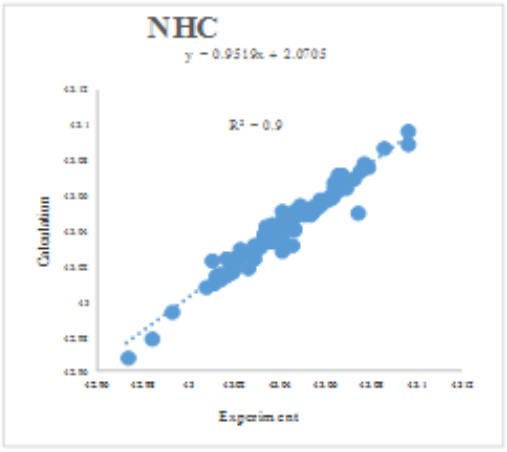

Figure 9. MLR for NHC

4. Conclusions

The purpose of this study was to predict heating combustion using empirical results by employing two different models an ANN and MLR. The created ANN had a three-layered structures one input layer with eight neurons, one hidden layer and one output layer. Logistic sigmoid and linear transfer functions were used in the hidden and output layers as linear activation functions.

The ANN model predicted the combustion heat more accurately than the MLR, confirming that the ANN model is an effective tool for correlating and simulating the parameters of GHC and NHC.

In the MLR model, the R^2 = 0.8414 of the GHV was found to be insufficiently effective in predicting GHC, perhaps because of the nonlinear relationship between the experimental and calculated data. On the other hand, the MLR of the NHV was more accurate than for GHC because the R^2 = 0.951, except for one outlier.

The heat of combustion of diesel fuel is an important parameter to measure in the performance and emissions of generators, where the diesel fuel in every truck needs to be monitored particularly for chemical composition and physical properties. Overall, this study supports the idea that ANN modelling can predict the heat of combustion of diesel fuel, thus avoiding costly and time-consuming experiments.

Reference

1.CRETON, B., DARTIGUELONGUE, C., DE BRUIN, T., TOULHOAT, H., Energy and Fuels, **24,** no. 10, 2010, p. 5396.
2. LI-DA, L. I., CHENG-ZHI, X. U., MEI-QIN, Z., XIAO-HUI, C., J. FUEL Chem. Technol., **43**, no. 8, 2015 , p. 990.
3. XUE-MIN, Y. A. N., GAO-SHEN, S. U., LIN, X., J. Fuel Chem. Technol., **37**, no. 3, 2009, p. 318.

4. ALLAWI, M. K., OUDAH, M. H., MEJBEL, M. K., Journal of Mechanical Engineering Research and Developments (JMERD), **42**, no. 5, 2019, p. 211.
5. SINGH, D., SUBRAMANIAN, K. A., GARG, M. O., Renew. Sustain. Energy Rev., **81**, no. 2, 2018, p. 2947.
6. PI-QIANG, T. A. N., JIAN-YONG, Z., ZHI-YUAN, H. U., DI-MING, L. O. U., AI-MIN, D. U., J. FUEL Chem. Technol., **41,** no**.** 3, 2013, p. 347.
7. SRIVASTAVA, A., PRASAD, R., Renew. Sustain. Energy Rev., **4**, no. 2, 2000, p. 111.
8. JIE, X., QINGHONG, X., FENG, Z., ZHIQIANG, W., China Pet. Process. Petrochemical Technol., **20**, no. 2, 2018, p. 71.
9. AL-GHOUTI, M., AL-DEGS, Y., MUSTAFA, F., Fuel, **89**, no. 1, 2010, p. 193.
10.YOUNIS, Y. M. Neural Network Prediction of Flash Point of Diesel Fuel From its Chemical Composition and Physical Properties, Master Thesis , Atılım University, 2017.
11. LÜ, Y., SHI, J., SUN, L., J. FUEL Chem. Technol., **36,** no. 3, 2008, p. 297.
12. WANG Y., HOU M., LI L., HUANG C., TAN P., LOU D., Advanced Technology in Teaching - Proceedings of the 2009 3rd International Conference on Teaching and Computational Science (WTCS 2009), Berlin, Heidelberg, 2012, p. 891.
13. GHARAGHEIZI, F. A., Chemom. Intell. Lab. Syst., **91**, no. 2, 2008 , p. 177.
14. ASTM International. Standard Test Method for Heat of Combustion of Liquid Hydrocarbon Fuels by Bomb Calorimeter. ASTM Standard D240-02, 2009.
15. GUO, Z., LIM, K. H., CHEN, M., THIO, B. J. R., LOO, B. L. W., Fuel, **207**, 2017, p. 344.
16. MOHAMMAD HASSANI, J., KHALILARYA, S., SOLIMANPUR, M., DADVAND, A., Model. Simul. Eng., **2012**, 2012, p. 1.
17. SANDRU S., ONUTU I., *Rev. Chim*., **70**, (7), 2019, 2367.
18.SALDANA, D. A., STARCK, L., MOUGIN, P., ROUSSEAU, B., CRETON, B., SAR and QSAR in Environmental Research., **24**, no. 4, 2013, p. 525.
19. MERCKEL, R. D., LABUSCHAGNE, F. J. W. J., HEYDENRYCH, M. D., Appl. Energy, **235**, 2019 , p. 1041.
20. YANG, H., RING, Z., BRIKER, Y., MCLEAN, N., FRIESEN, W., FAIRBRIDGE, C., Fuel **81**, no. 1, 2002 , p. 65.
21.PILOTO-RODRÍGUEZ, R., SÁNCHEZ-BORROTO, Y., LAPUERTA, M., GOYOS-PÉREZ, L., VERHELST, S., Energy Convers. Manag., **65**, 2013, p. 255.
22.PAN, Y., JIANG, J., ZHANG, Y., Ind. Eng. Chem. Res. **51**, no. 40, 2012 , p. 13274.
23.ALBAHRI, T. A., Energy Convers. Manag., **76** , 2013 , p.1143.
24.ALBAHRI, T. A., Process Saf. Environ. Prot., **93**, 2015, p. 182.
25.ZHOU, L. L., JIANG, J. C., PAN, Y., WANG, Z. R., J. Loss Prev. Process Ind., **38**, 2015, p. 254.
26.TOSUN, E., AYDIN, K., BILGILI, M., Alexandria Eng. J., **55**, no. 4, 2016 , p. 3081.
27.USLU, S., CELIK, M. B., Eng. Sci. Technol. an Int. J. **21**, no. 6, 2018, p. 1194.
28.YOUNIS, Y. M., KAYI, H., 2nd International Conference on Sustainable Engineering Techniques (ICSET 2019), Baghdad, Iraq, 6–7 March 2019, **518,** 062008, doi:10.1088/1757-899X/518/6/062008.
29.RAMADHAS, A. S., JAYARAJ, S., MURALEEDHARAN, C., PADMAKUMARI, K., Renew. Energy, **31**, no.15, 2006 , p. 2524.
30.HAYKIN, S., Neural Networks and Learning Machines, New York, NY, USA, Prentice-Hall, 2009.
31.NITA I., OSMAN S., IULIAN O., SANDU R. G., *Rev. Chim*., **69** (12), 2018, 3410.
32.GARDNER, M. W., DORLING, S. R., Atmos. Environ., **34**, no.1, 2000 , p. 34.
33.KAMAR F. H., NIAMAT F. E., FAISAL A. A. H., MOHAMMED A. A., NECHIFOR A. C., NECHIFOR G., *Rev. Chim*., **69** (8), 2018, 1919

Manuscript received: 20.04.2020