PAPER • OPEN ACCESS

Artificial Neural Network Prediction of Sulfur Content of Diesel fuel from its Physical Properties

View the article online for updates and enhancements.

Artificial Neural Network Prediction of Sulfur Content of Diesel fuel from its Physical Properties

Younis Muhsin Younis[1]**, Hakan Kayi**[2]

[1] Material Engineering Department, Technical collage Baghdad, Middle Technical University, Baghdad, Iraq.
[2]Ankara University, Chemical Engineering Department, 06100 Ankara, Turkey.

Abstract. The sulfur content is important in engineering calculations, so this study has two major purposes. The first purpose of the study is to predict the sulfur content from its physical properties by using artificial neural network to decrease time and cost spent on experimental analysis of sulfur content, and the second purpose is to find the simplest formula to predict the sulfur content. Artificial Neural Network is applied as a black-box type modelling for sulfur content prediction of diesel fuel. The experimental data used in this study is obtained from Erbil power station. In this study, the Levenberg-Marquardt training algorithm is used to train the neural network and to predict the sulfur content. It was observed that the ANN model can predict the sulfur content of diesel quite well with correlation coefficient (R) 0.9813. The prediction Mean Square Error was between the targets values and the outputs values were obtained 0.000339 by the matlab software. The findings obtained in this study indicated that the designed neural network performs quite well in the prediction of sulfur content of diesel fuel from its physical properties.

1. Introduction

"*Diesel fuel is a very complex mixture of thousands of different compositions*" [1]. Most of them are with carbon numbers between eleven and twenty-two. A large part of these compositions are constituents of the paraffinic, naphthenic, or aromatic category of hydrocarbons, every category has distinct chemical and physical properties. Artificial Neural Network (ANN) is similarly to a human brain, (artificial) neurons are the main processing elements in the artificial neural network. An ANN comprises of interconnected neurons arranged in the following layers [2].

In recent years, there has been an increasing amount of literature on Artificial Neural Network. Mehdi Mehrpooya et al. (2010) showed that artificial neural network was developed using the most reliable experimental data in the literature. The agreement between the empirical and calculated data was generally acceptable, they are found that the artificial neural network algorithm can be used as an effective tool for predict sulfur content [3].
Artificial Neural Network methods are strong modelling tools to establish diesel properties and they have the ability to recognize very complex relationships or obtain knowledge using data input and output only [4] [5].

In 2017, Fernanda M. de Oliveira et al. predicted the flash point, sulfur content and cetane number. They used ANNs of diesel blends as 7% biodiesel using the many parameter such as distillation curves (ASTM D405), specific gravity at 20 ^{0}C (ASTM D405), sulphur content (ASTM D4924), flash point (ASTM D93) and cetane index (ASTM D4737) [6].A case study approach was used for 162 samples, which were randomly divided into three groups; 15% for testing, 15% for validation and 70% for the samples of training. The aim was to enhance the generalized capability of the models .Therefore a cross

validation process was carried out during training of the network such that the amount of iterations in this step was conditioned by square error in the test samples [6].

Feasibility was shown by ANN in the suggested method to accurately predict cetane index, sulfuer content and flash point, with an aggregate absolute percentage error of 4.6%, 0.4% and 3.3 % correspondingly. These models were proved quite successful as they used the input matrix, data of physicochemical properties such as distillation curves, sulfur content specific gravity at 20 ^{0}C, cetane index and flash point that were already a component of the diesel specification routine. As a result, few of the physicochemical was used [6].

A paper was published by Gatinder Kumar in 2007, in which a comparison of statistical and ANN data was used to predict the physical properties of a mixture of distinct diesel and biodiesel. The author stated that the results showed that ANN is the best option for this particular system. Viscosity, fire point, flash point and density were the tests carried out in this situation [7].

2. Methodology

The methodology was divided into two stages. The first one involves working in a laboratory to measure physical properties of diesel fuel. The second stage introduced the ANN that was used for prediction of sulfur content after collecting the samples.

After collecting the samples, ANN model is used to predict flash point through the network training using Matlab 2013b software.

2.1. Experiments on Sample

To determine the sulfur content and other analysis, diesel fuel samples were collected from trucks that come to the Power Station and 192 diesel samples (approximately 250 ml per sample) were stored in sealed bottles at room temperature until using them in flash point test and other tester. The flow chart in Figure 1 reveals the stages of the experimentally
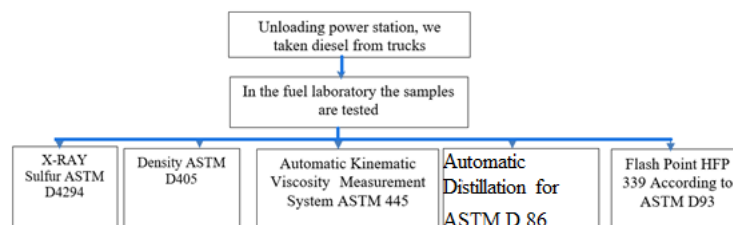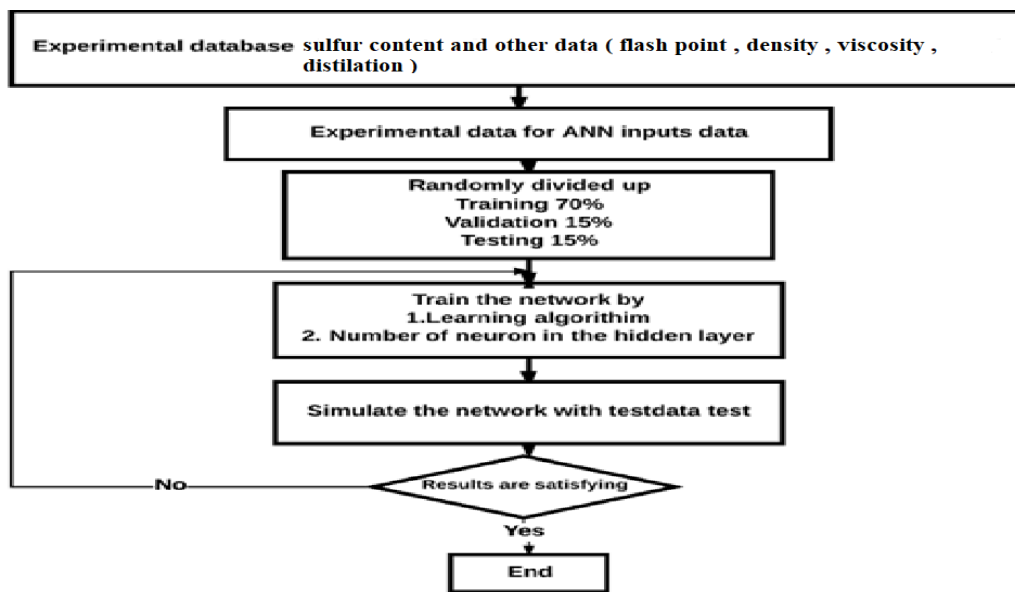

Figure 1. Flow chart of experimental steps

2.2. Artificial Neural Network Modelling (Stage Two)

In developing an artificial Neural Network for any application, The following steps must be followed; collect data, pre-process the data, determine the structure of the network (number of hidden layers and nodes), create the network, initialize weights and biases, train network, validate network and test network [8]. Figure 2 shows the model method artificial Neural Network.

The model to be used a multilayer perceptron; input data is selected from the four possible input of the model. Table 1 below shows the input layer for samples.

Figure 2. Flow chart for artificial neural network

Table 1. Input Layer for samples

No.	Input Samples of Diesel
1.	Flash point
2.	Density
3.	Viscosity
4.	Distillation end point

The output layer is the prediction of sulfur.

Two questions need to be answered when the layer is being designed. Initially, how many hidden layers are required in one neural network? Secondly, how many neurons are required in every layer?

For the answer of the first question, choosing the hidden layer number is quite an easy decision because many problems can be solved only using a single layer.

For the answer of the second question, applying the thumb rules to find the amount of nodes in hidden layers and the most common rule is that the amount of nodes in the hidden layer is the average of the output node and input nodes. To find the correct number of neurons in the hidden layer, these rules are used as a starting point and trial and error approach is used to find the number of nodes.

The following steps were taken in training the network using Matlab R 2013 software to model our network. Matlab has the flexibility to select inputs, outputs and hidden neurons. There is the possibility to easily change the hidden neurons and test the result. On the other hand there are some different methods to divide the dataset such as block or random. The percentage of training, validation and test subsets can be changed easily. There are different tools like performance plot, regression plot and histogram to analyse network performance.

3. Results

3.1. Laboratory results
The samples were collected from trucks (filled with diesel). Therefore, the results of the diesel sample are presented in Table 2. We can start working on the second stage

Table 2. Laboratory Results for Diesel Testing

Parameter	Results between		Total number of results
	The lowest value	The maximum value	
Flash point ^{0}C	57.5	82.5	192
Density 16 °C (g/cc)	0.8231	0.876	192
Viscosity at 40 °C (mm^2/sec)	2.158	2.863	192
Sulfur content wt.%	0.3915	0.829	192
Distillation end point 90%	310	360	192

3.2. Neural Network

In this model, the thumb rules were applied to determine the number of nodes in the hidden layer. The most common rule for the number of nodes in the hidden layer is taking the average of input and output. After trial and error, the best number on node is 10

Then choosing the best method of prediction Neural Network NN.

The function in hidden layer is known as Sigmoid Activation Function. It uses the function to determine activation. The equation of sigmoid function can be given as [9].

$$f(x) = 1/(1 + e^{-x}) \tag{1}$$

Dividing 140 target time steps randomly to three kinds of target time steps.

- Training, the targets that are displayed on the network during training, its error is modified, and the data are divided by up to 70% that means 98 target time steps.
- Validating, the targets that are utilized to measure network generalization. Training should be discontinued when generalization stops improving, and the data are divided by up to 15% that means 21 target time steps.
- Testing, the target that has no effect on training and providing an independent of measuring training network performance that mean the training during and after; the data are divided by up to 15% that means 21 target time steps.

The MSE will be obtained which is the average squared difference between outputs and targets; zero means no error, lower values are better. R-values measure the relationship between outputs and targets that means a close relationship. R-values measure the relationship between outputs and targets that means a close relationship.

3.2.1. The Network Performance

Figure 3 shows the performance plot of the training. Note that the performance of the training reached to the minimum value at the iteration 12 and the training continued to the iteration 18 then halts. In addition, this figure does not show any major problems during training, and the best validation performance is 0.0015695 at epoch 18.
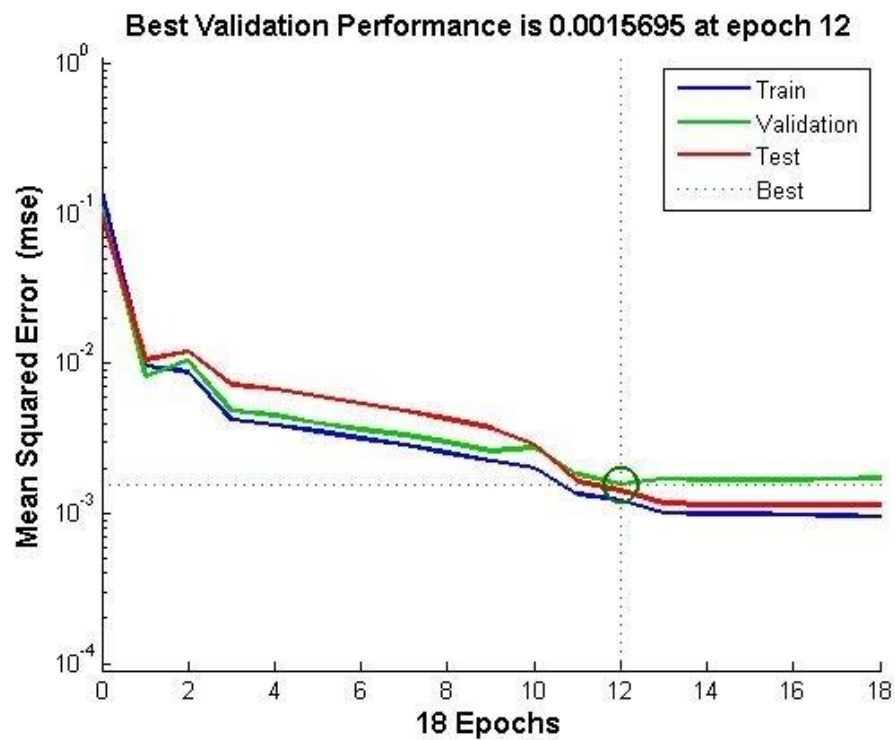
Figure 3. Performance Network

3.2.2. The Regression Plots
Figure 4 shows the regression plots, indicating the relationship between the output, and the target values. If the training is very good, the network output and target will have the same value. The regression all data set is found to be 0.98135; this indicates the suitability of perfect data (fit).
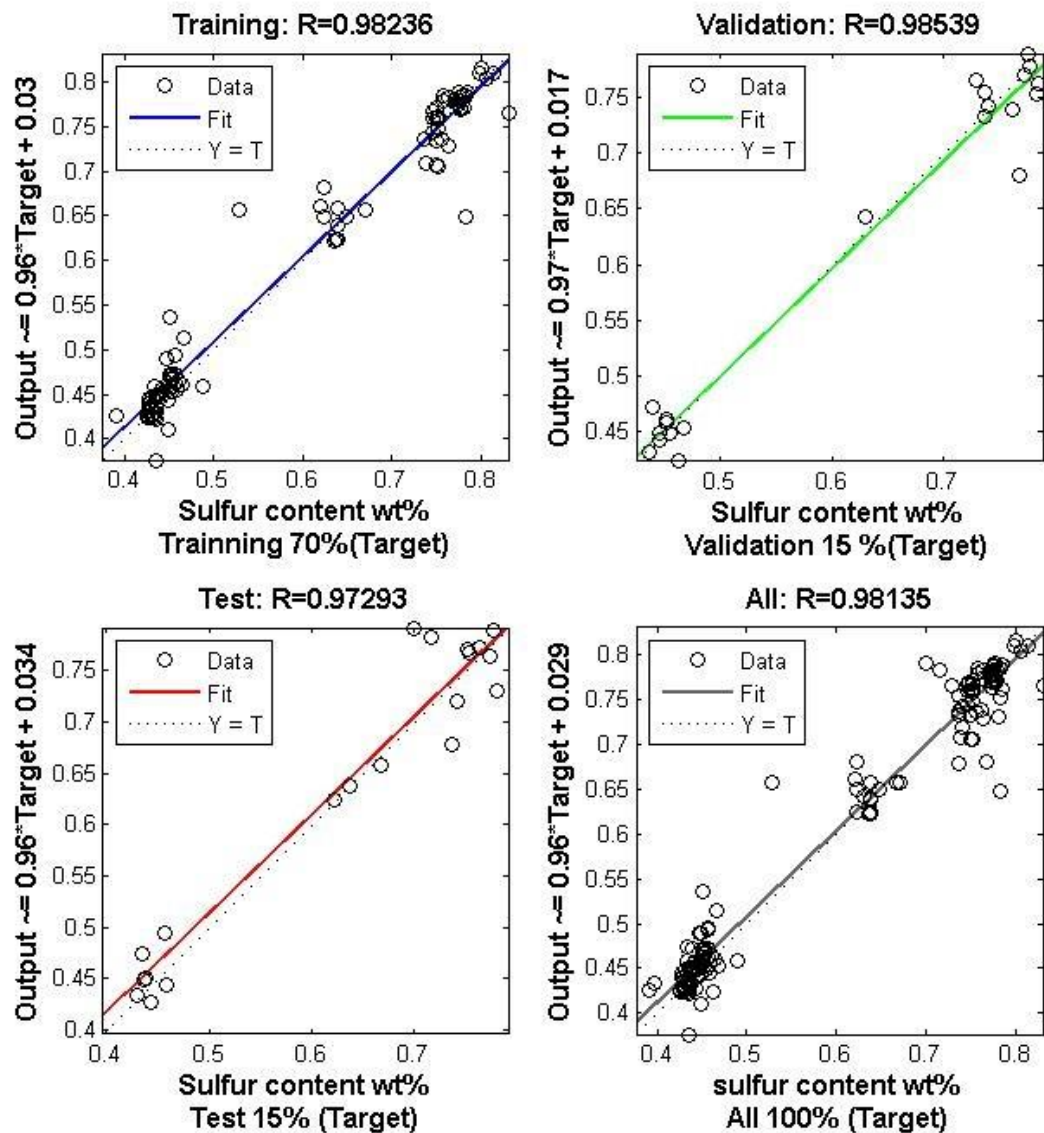
Figure 4. Regression Plots

3.2.3. Error Histogram
Error histogram is another tool used to obtain more information about the neural network. Figure 5 exhibits the distribution of the training. It is clear from the figure that the validation and test results are acceptable and the error distribution results are reasonable.

In addition, the figure gives a good indication of the values of outliers. Nevertheless, these outliers do not affect the results because ten data out of 140 total data and most errors fall 0.000339
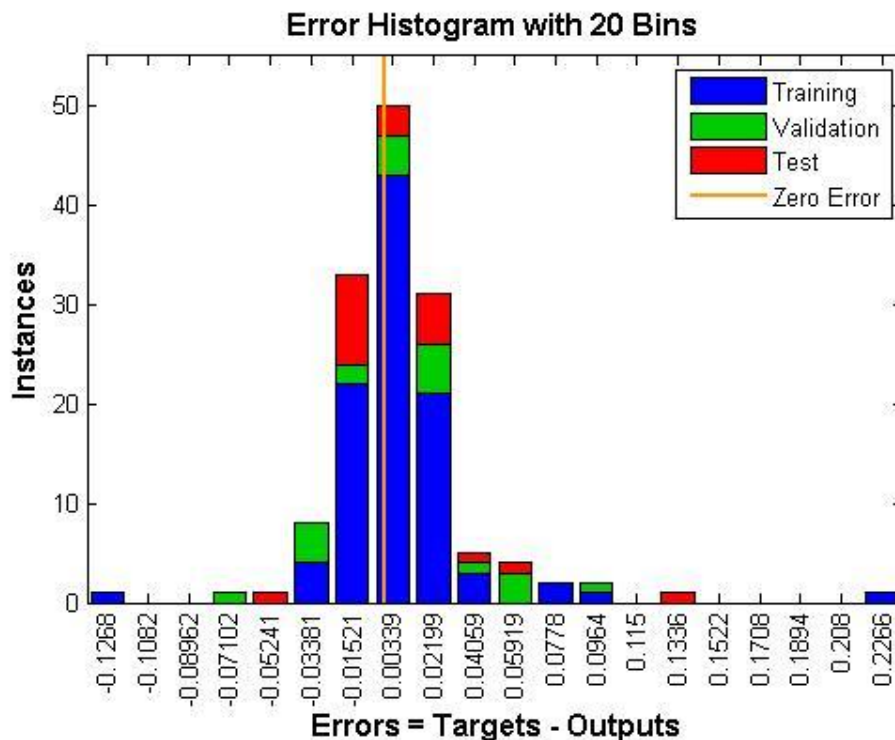
Figure 5. Error Histogram

3.2.4. Test of Network
The regression plots and error histogram explain the situation of the neural network behaviour with data training. The test network is considered the best indicator to see whether the network performance is good enough or no and the database used outside to the training database. Therefore, 52 samples outside the data are used. Satisfactory results are obtained; where, these results are regression 0.97187 and MSE 0.000137.

3.2.5 *The Simplest Formula to Predict sulphur content*
The formula can be described as multiple linear regression (MLR). The analysis of predictive, multiple linear regression is used to interpret the relationship between independent variables two or more and continuous dependent variable one.
The Linear regression is a 0.8572574; root mean square error is 0.0046129

$$sulfur\ content = -0.0010284 * flash\ point + 2.312811 * density - 0.705744 * viscosity + 0.0016081 * distilation\ end\ point\ 90\ \%$$ (2)

4. Conclusions
Neural network is useful tool for predicting the sulfur content. The predicted results have very small errors and they are reasonable, and very close to the actual value.

This study serves laboratories in diesel power stations where a supervisor of the laboratory should use the simplest formula to predict sulfur content, and this equation (2) reduce the time and cost.

References
[1] John B, John F, Andy G, Lew G, Greg H, Kent H, Jerry Horn 2007 *Diesel Fuels Technical Review* Freightliner.
[2] Asadollahi R 2014 Predict the flow of well fluids A big data approach, Faculty of Science and Technology Faculty of Science and Technology , University of Stavanger, *Master thesis*.

[3] Mehrpooya M, Mohammadi A and Richon A 2010 Extension of an Artificial Neural Network Algorithm for Estimating Sulfur Content of Sour Gases at Elevated Temperatures and Pressures *Ind. Eng. Chem. Res* 439–442.

[4] Korres D, Anastopoulos G, Lois E, Alexandridis A, Sarimveis H, Bafas G 2002 A neural network approach to the prediction of diesel fuel lubricity *Fuel* **10** 1243–1250.

[5] Balabin R, Safieva R 2011 Near-Infrared (NIR) Spectroscopy for biodiesel analysis, fractional composition, iodine value, and cold filter plugging point from one vibrational spectrum *Fuel and Energy* **25** 2373-2382.

[6] Oliveira F, Carvalho L, Teixeira L, Sales R 2017 Predicting cetane index, flash point and content sulfur of diesel-biodiesel blend using an artificial neural network (ANN) model *Energy Fuels* **31** 3913–392.

[7] Kumar j, Bansal A, Jha M K 2007 Comparison of statistical and neural network techniques in predicting physical properties of various mixtures of diesel and biodiesel *in Proceedings of the World Congress on Engineering and Computer Science*, San Francisco, USA, October 24-26-2007.

[8] Reza A 2014 Predict the flow of well fluids a big data approach, University of Stavanger, june 2014 , *master thesis*.

[9] Heaton J 2008 Introduction to neural networks with Java (Second Edition), United States of America: Heaton Research.