



Article

Prediction Model for the Viscosity of Heavy Oil Diluted with Light Oil Using Machine Learning Techniques

Xiaodong Gao^{1,2}, Pingchuan Dong^{1,2,*}, Jiawei Cui^{1,2} and Qichao Gao³

¹ College of Petroleum Engineering, China University of Petroleum (Beijing), Beijing 102200, China; swpugxd@163.com (X.G.); cuijw@163.com (J.C.)

² State Key Laboratory of Petroleum Resources and Prospecting, China University of Petroleum (Beijing), Beijing 102200, China

³ Sinopec Research Institute of Petroleum Engineering, Beijing 100083, China; gaoqch16765.sripe@sinopec.com

* Correspondence: dpcfem@163.com

Abstract: Due to the presence of asphaltene, the flow assurance of high viscosity crude oil becomes more challenging and costly to produce in wellbores and pipelines. One of the most effective ways to reduce viscosity is to blend heavy oil with light oil. However, the viscosity measurement of diluted heavy crude is either time-consuming or inaccurate. This work aims to develop a more accurate viscosity model of diluted heavy crude based on machine learning techniques. A multilayer neural network is used to predict the viscosity of heavy oil diluted with lighter oil. The input data used in the training include temperature, light oil viscosity, heavy oil viscosity, and dilution ratio. In this modeling process, 156 datasets were retrieved from the available literature of various heavy-oil fields in China. Part of the data (80%) is used to train the developed models using Adam optimizer algorithms, while the other part of the data (20%) is used to predict the viscosity of heavy oil diluted with lighter. The performance and accuracy of the machine learning models were tested and compared with the existing viscosity models. It was found that the new model can predict the viscosity of diluted heavy oil with higher accuracy, and it performs better than other models. The absolute average relative error is 10.44%, the standard deviation of the relative error is 8.45%, and the coefficient of determination is $R^2 = 0.95$. The viscosity predicted by the neural network outperformed existing correlations by the statistical analysis used for the datasets available in the literature. Therefore, the method proposed in this paper can better estimate the viscosity of diluted heavy crude oil and has important promotion value.

Keywords: heavy oil dilution; light oil viscosity; heavy oil viscosity; viscosity model; artificial neural network (ANN) model

1. Introduction

Heavy crude oils represent an important energy resource. The need for heavy oil processing has increased in recent years worldwide, backed by higher demands for petroleum products in the face of declining light crude oil resources. Moreover, commercial interest has been developed in heavy oil production systems with the advent of improved drainage areas by drilling multilateral and horizontal wells and highly deviated wells at shallow reservoirs [1].

Citation: Gao, X.; Dong, P.; Cui, J.; Gao, Q. Prediction Model for the Viscosity of Heavy Oil Diluted with Light Oil Using Machine Learning Techniques. *Int. J. Geo-Energy*, 2022, 1(1), doi:

ISSN/© By the Author(s) 2022, under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0>)

Compared to light crude oils, the production and processing of heavy oils in the pipeline and wellbores are challenging due to their very high molecular weight components such as asphaltene, resins, and waxes. Moreover, the presence of heteroatoms and metals makes asphaltene the most polar polycyclic aromatic hydrocarbon, leading to its self-association with the formation of a viscoelastic network of nanoaggregates resulting in a high viscosity of heavy crude oils [2–5]. The viscosity of heavy crude oils is in the range of 100–10,000 cp at room temperature, while the extra-heavy crude oils have viscosities greater than 10,000 cp. Generally, crude oil with a viscosity of 400 cp is the classical maximum desired pipeline viscosity that can be transported through pipelines [6–8]. The most widely used method of reducing heavy crude oil viscosity in the pipeline is by blending heavy crude oils with a lighter oil due to the polarity of lighter oils [9–12]. It is known that reducing the dynamic viscosity of the oil reduces the internal friction resistance in the oil, which can increase the mobility of crude oils [13,14]. Therefore, the prediction of diluted viscosity is important for the pumping, production, and transportation of heavy crude oils.

Many studies have been conducted to predict the viscosity of diluted heavy crude oil. These studies can be classified into two main categories: lab-scale experiments and empirical correlations. Both of these are described and compared in the following.

The diluted viscosity of crude oils is usually measured in the laboratory. Yaghi and Al-Bemani (2002) [15] experimentally studied the effect of blending with light crude on viscosity at 30 and 50 °C. The light crude was added to the heavy crude and sheared at 6000 rpm for 3 min. The result showed that the heavy crudes diluted with light oil behaved as non-Newtonian (27% light oil) and Newtonian (30% light oil) fluid at 30 °C. The viscosity of the Newtonian dropped to approximately 1000 cp and did not meet the need for pipeline transportation, whereas, the heavy behaved as a Newtonian fluid at all the fractions at 50 °C. The viscosity could be reduced to 300 cp with the addition of 30% light oil. The effect of light oil concentration on viscosity behavior has been studied to lower the viscosity of the heavy crude oil by Hasan et al. (2010) using the RheoStress RS100. It was found that the heavy crude blended with a limited amount of lighter crude oil could efficiently lower the heavy crude oil viscosity. The viscosity of heavy crude dropped from 10,000 to 1200 cp with the presence of 10% light oil and from 10,000 to 350 cp with the presence of 20% light oil in the mixture [16]. Although the measured viscosity was more accurate than the calculated viscosity, the experimental viscosity measurement required a higher workload and was more expensive.

Therefore, some authors have researched viscosity calculation models that can predict the viscosity of diluted heavy crude oil, since this method can effectively lower the experimental workload and improve work efficiency. The first correlation for predicting the viscosity of a mixture of two fluids was proposed by Arrhenius [17] in 1887. Although it was suggested for a mixture of two pure liquids, it is worth evaluating its capabilities in predicting the viscosity of heavy crude oil with a high viscosity ratio system. Subsequently, an exponential equation was presented by Kendall and Monroe [18] based on measurements of molar fractions. The comparison results showed it had a good fit with experimental viscosities. However, these models were not accurate for diluted heavy crude [19].

Cragoe [20] introduced a universal equation for binary heavy oil–solvent mixtures, which was a function of the mass fraction and the viscosity of the component. Later, the development of more complex equations, such as the four-parameter model, was developed to predict the viscosities of petroleum fractions blends. There were three methods in Al-Besharah's [21] work: ASTM D341 [22], the Refutas index method [23], and a four-parameter model based on the Ratcliff and Khan [24] equation were compared. The results showed that the four-parameter model had the lowest errors. In 2019, Mohammadi [25] developed a new empirical model for measuring the kinematic viscosity of blends. The parameters of this model were obtained based on a genetic algorithm using 850 data points. In his work, the viscosity of a crude oil blend depended on the mass or volume fraction and viscosity of each component of the blend. However, it ignored the effect of temperature on the viscosity of blends.

Recently, Jing et al. [26] investigated the viscosities of two heavy oils diluted with five light oils from the published literature. The experimental and calculated viscosity in terms of the Lederer

model, the Arrhenius model, the double logarithmic model, and the Cragoe model were compared. It showed that the Cragoe model had the lowest average relative error and the maximum relative error. In addition, Xing [27] evaluated eight kinds of mature oil mixing viscosity models using experimental viscosities. The eight mixing viscosity models consisted of the Cragoe model, double logarithmic model, Lederer model, the Arrhenius model, the Kendal-Monroe model, etc. However, the results calculated by these models showed that the errors were very large. Therefore, the Arrhenius modified model III was proposed and had the lowest average absolute error of 6.9%. From these studies, it is found that existing viscosity models are limited and may not be applicable to other oilfields. Therefore, using artificial neural networks (ANNs) to predict the viscosity of heavy crude oil blended with light oils is the focus of this research.

Many researchers have summarized that ANNs can be a useful tool to model different engineering systems under real-world conditions, such as in solar energy systems [28], solar distiller [29–31], and petroleum engineering [32–34]. Compared with traditional methods, there is no need for ANNs to provide a model function and understand the input/output relationships before forecasting. The learning capability of an ANN has the chance to discover more complex and subtle interactions between the data of input and output. In addition, ANNs also present the potential for high fault tolerance in the wide ranges of variables [35–37].

Thus, this paper aims to develop a robust model using ANNs to predict the viscosity of diluted heavy crude oil sufficiently. To do so, we collected a total number of 156 viscosity datasets from openly published literature. Moreover, the prediction viscosity was also compared with some empirical correlations available in the literature.

2. Data and Methods

2.1. Data Collection

To achieve the objective of this paper, 156 viscosity experimental datasets of diluted heavy crude were collected from several different oil fields in China. Each heavy crude oil was blended with the light oil in different dilution proportions. The mixture was homogenized using a stirrer for 1 h and heated to a specific temperature. Then, the respective mixtures were left for 2 days to reach equilibrium; then, the viscosity of these crude oils was measured using a model VT550 viscometer (Thermo Fischer, Shanghai, China). The related parameters include heavy crude viscosity, light crude viscosity, dilution rate, and temperature. Table 1 summarizes the main information of the parameters used in this study.

Table 1. Characteristics of parameters.

Title 1	Title 2	Minimum Value	Maximum Value
1	Heavy crude viscosity, cp	121	4020.6
2	Light crude viscosity, cp	2.9	35.7
3	Dilution rate	0.2	0.9
4	Temperature, °C	20	60
5	Diluted heavy viscosity, cp	9.3	882

2.2. Artificial Neural Network Theory

A neural network is a series of algorithms that endeavors to perform pattern recognition, predictive evaluation, and evaluation analysis in datasets through a process that mimics the way the human brain operates. It contains layers of interconnected nodes. Each node is known as the perceptron, and a perceptron will take several inputs and produce a single output. Each of those perceptrons makes a decision by weighing the results from the first layer of decision making. The perceptron of the next layer (second, third, etc.) can engage in more complex decision making in this way. By varying the values of weights and the thresholds between elements, we can perform a

specific function and compare the output and the target, until the network output matches. The schematic of the network can be seen in Figure 1.

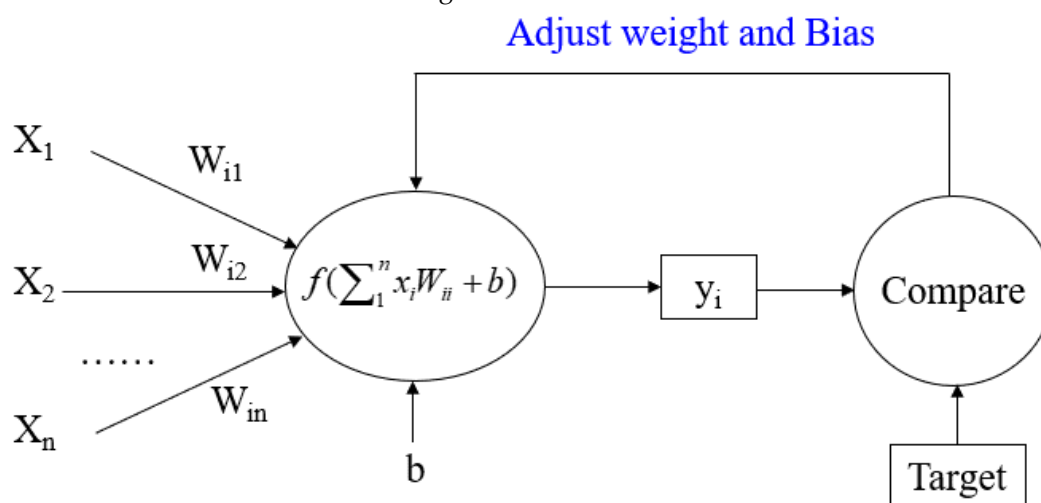


Figure 1. Schematic of the network in an Artificial Neural Network Model.

The steps of the training process for the single-layer neural network are as follows:

- ① The weighted sum of the input signals can be calculated as:

$$v = \sum_1^n X_i W_{ii} + b \quad (1)$$

- ② The output value is calculated using the activation function:

$$y_i = f(v) = f\left(\sum_1^n X_i W_{ii} + b\right) \quad (2)$$

- ③ The error of the output is calculated as:

$$e_i = y_{tar} - y_i \quad (3)$$

- ④ The updated weight is calculated using the delta rule:

$$W_{ii} \leftarrow W_{ii} + \Delta W_{ii} = W_{ii} + \alpha e_i x_i \quad (4)$$

- ⑤ Perform Steps ①–④ for all training samples.

- ⑥ Repeat Steps ①–⑤ until the error meets the requirements and obtains the predicted value.

2.3. Artificial Neural Network Construction and Optimization

To improve the generalization ability of the model, viscosity datasets obtained from published literature were used. They were randomly divided into two parts of which 80% was used for network training and 20% for network prediction accuracy test.

Due to the different ranges of the heavy crude viscosity, light oil viscosity, and diluted viscosity obtained by the experimental method, the first thing to do was to normalize all the datasets and map all the data to the interval $[-1, 1]$. The dimensional differences and magnitude differences between indicators could be eliminated by using the normalization method, so that data information could be mined more accurately and satisfactorily. This work adopted the data normalization method provided by scikit-learn, and its normalization formula is [38]:

$$x_{norm} = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (5)$$

The network in this work used a five-layer neural network structure, that is, three hidden layers, one input layer, and one output layer. It is known that the viscosity decreases sharply with increasing temperature. In addition, the heavy crude viscosity, light oil viscosity, and dilution rate also play an important role in the diluted viscosity. Therefore, four input parameters and an output parameter were designed in our ANN structure. The input parameters included heavy crude viscosity, light crude viscosity, dilution rate, and temperature, while the output parameter was the viscosity of heavy crude oil diluted with light oil, thereby establishing a viscosity prediction model.

2.3.1. Neural Network Training Times

In this paper, the training times for the ANN were set from 100 to 700. The network corresponding to different training times is shown in Figure 2. According to the loss result, the loss decreased with time. When the training times reached 450, the loss was maintained at around 4.7×10^{-4} , and the subsequent training basically stabilized. Therefore, the number of training times was set to 450.

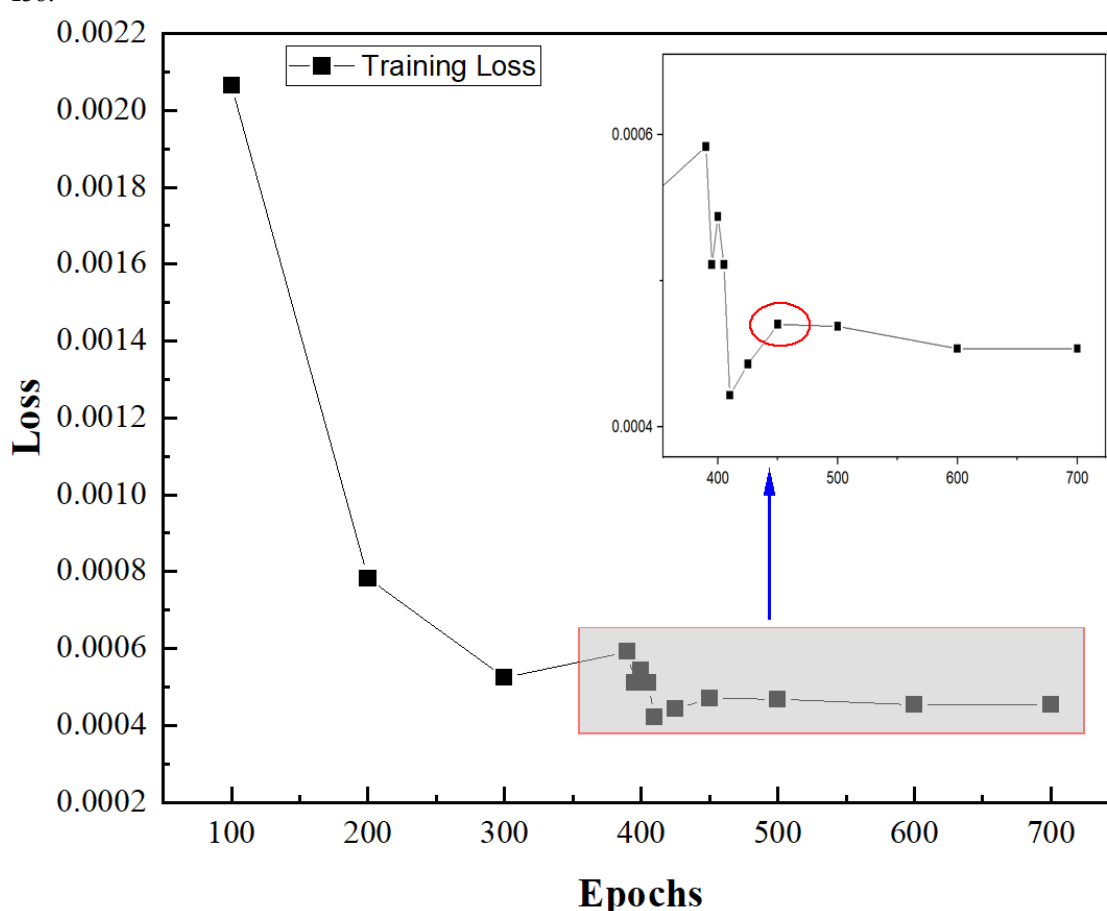


Figure 2. Training set loss value at different training times.

2.3.2. Optimization Model

The error backpropagation was selected as a learning technique in this paper, since it can efficiently calculate gradient vector, computationally speaking. In addition, the adaptive moment estimation (Adam), stochastic gradient descent (SGD), and adaptive gradient algorithm (Adagrad) were used to calculate the loss in the training process. The results of the three optimizer algorithms are shown in Figure 3. It can be seen from the figure that with the change in time, the loss value of the Adam became the smallest, and the curve basically stabilized. Therefore, this paper chose the Adam optimizer algorithm to predict the viscosity of heavy oil diluted with light oils.

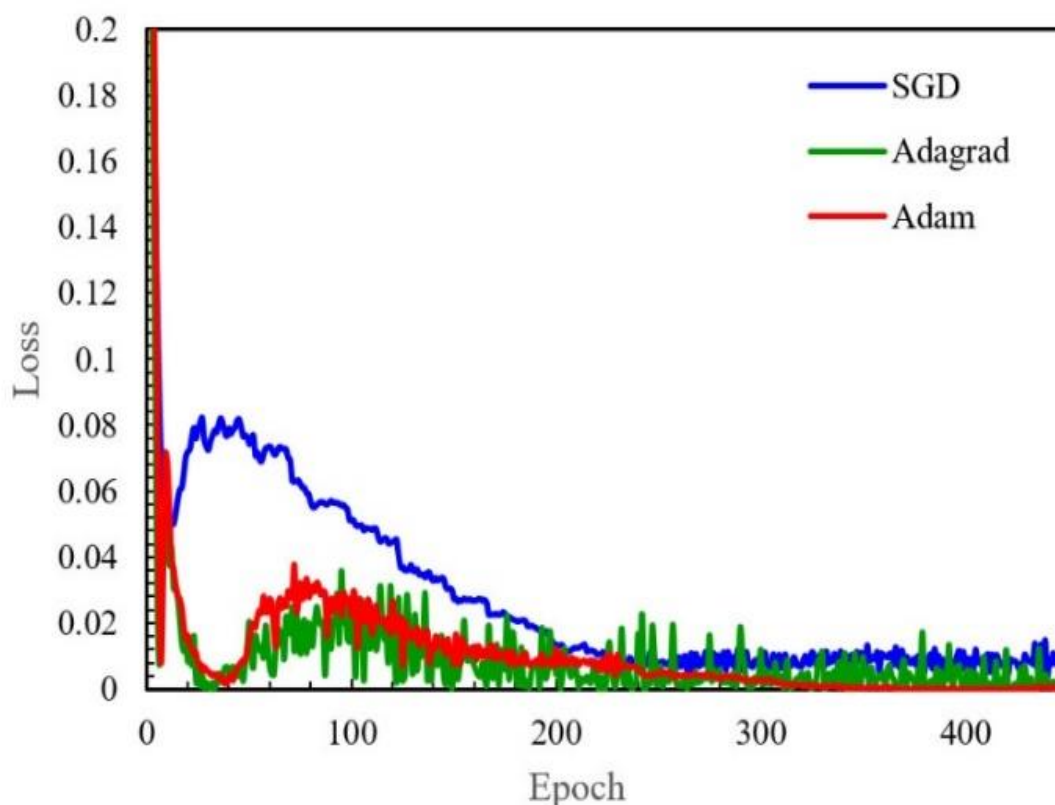


Figure 3. Loss curve of optimizer algorithms.

2.3.3. Execution Procedure

The execution procedures of the proposed ANN model are presented in Figure 4. The dataset was preprocessed and separated into a training and test set. The number of neurons in each layer and MinRMSE were initialized. Then, the ANN model was trained using the training set, and the RMSE was also calculated. The number of neurons of each layer was updated to minimize the RMSE. In addition, this paper assumed a maximum of 50 neurons per layer. Three sets of inner loops were used to screen out the optimal model with the MinRMSE. Once the best model was obtained, we tested the ANNs with the best model. Finally, the optimized ANNs were employed to predict the diluted viscosity of heavy oils.

The above analysis results show that the optimal number of neurons corresponding to the three hidden layers were 11, 11, and 11, respectively. In addition, the corresponding bias and weight are shown in Appendix A.

Therefore, the optimal topology of this work is shown in Figure 5.

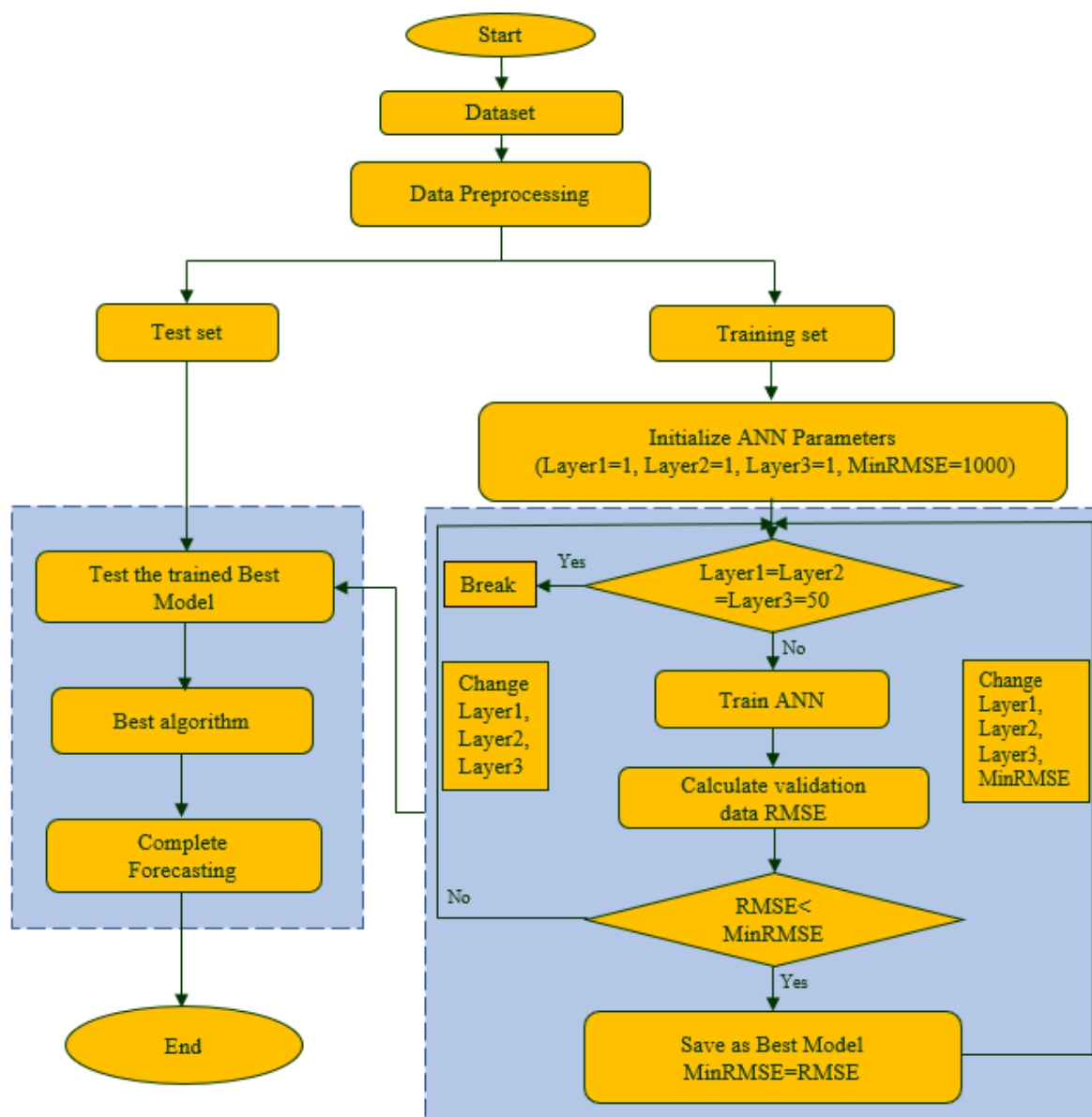


Figure 4. The execution procedures of the ANNs model.

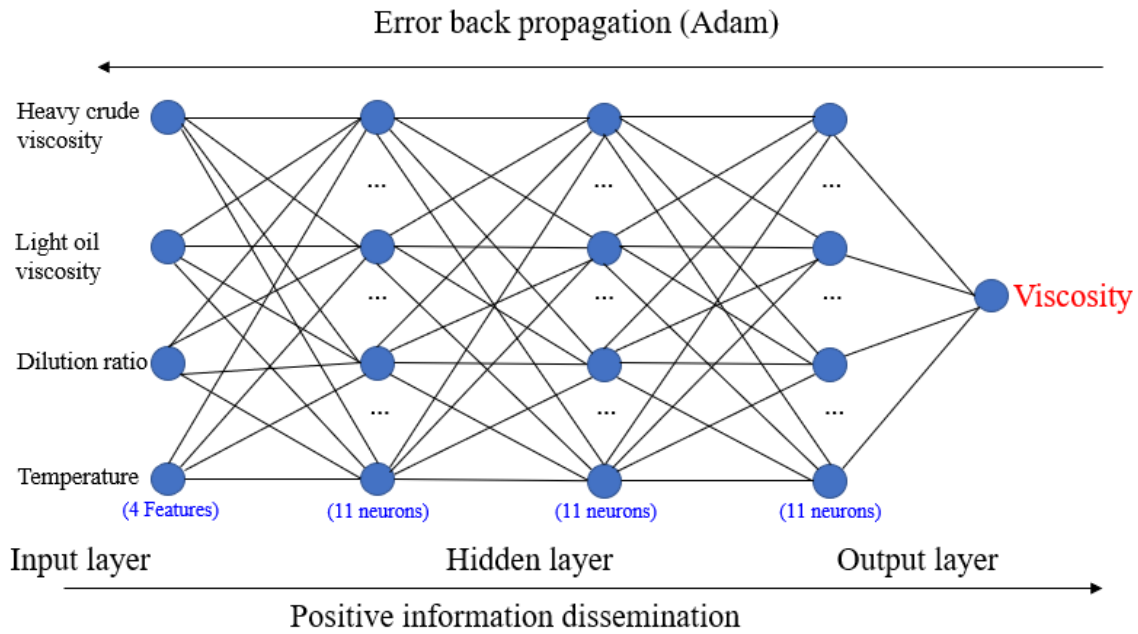


Figure 5. Neural Network Architecture.

2.4. Statistical Error Analysis

To determine the ability and accuracy of the proposed viscosity model, cross plots and statistical methods, including the percentage average relative error (ARE%), the average absolute relative error (AARE), the mean absolute error (MAE), mean relative error (MRE), and the standard deviation (SD) were performed as shown in Equations (6)–(10), respectively. The arithmetic average of the absolute relative errors (AARE) is the difference in the magnitude between the measured and the predicted value of the model. A low value of AARE% means better results and is consistent with real values around the 45° line. The SD represents the deviation of the predicted values from the average result.

$$E_{ARi} = \frac{|\mu_{Experimental})_i - (\mu_{Predicted})_i|}{(\mu_{Experimental})_i} \times 100 \quad (6)$$

$$AARE = \overline{E_{ARi}} = \frac{\sum_{i=1}^n E_{ARi}}{n} \quad (7)$$

$$SD_{ARE} = \sqrt{\frac{\sum_{i=1}^n (E_{ARi} - \overline{E_{ARi}})^2}{n-1}} \quad (8)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |(\mu_{Experimental})_i - (\mu_{Predicted})_i| \quad (9)$$

$$MRE = \frac{1}{n} \sum_{i=1}^n \frac{|(\mu_{Experimental})_i - (\mu_{Predicted})_i|}{(\mu_{Experimental})_i} \quad (10)$$

3. Results and Discussion

3.1. Prediction of Crude Oil Viscosity

Some of the viscosity correlations of heavy crude oil diluted with light oil were used in this paper. These correlations consisted of the Arrhenius's correlation, double log correlation, Cragoe's correlation, Kendal-Monroe's correlation, and Xing's correlation. The viscosity of diluted heavy crude oils can be calculated as follows:

Arrhenius (1887):

$$\begin{aligned} \lg \mu_m &= X_1 \lg \mu_1 + X_2 \lg \mu_2 \\ X_1 + X_2 &= 1 \end{aligned} \quad (11)$$

Double log (1983):

$$\begin{aligned} \lg \lg \mu_m &= X_1 \lg \lg \mu_1 + X_2 \lg \lg \mu_2 \\ X_1 + X_2 &= 1 \end{aligned} \quad (12)$$

Cragoe (1933):

$$\begin{aligned} \mu_m &= 5 \times 10^{-4} \exp(1000 \ln 20 / L_m) \\ L_m &= X_1 L_1 + X_2 L_2 \\ L_i &= \frac{1000 \ln 20}{\ln \mu_i - \ln(5 \times 10^{-4})} \end{aligned} \quad (13)$$

Kendal-Monroe (1917):

$$\mu_m^{1/3} = \sum_{i=1}^n x_i \mu_i^{1/3} \quad (14)$$

Xing (2021)

$$\begin{aligned} \mu_m &= (0.7 - 0.5X_1 + 0.01T)^C \\ C &= \sum_{i=1}^n x_i \lg \mu_i \end{aligned} \quad (15)$$

3.2. Determination of the New Viscosity Model

To determine the accuracy of the ANN training values in predicting the viscosity of diluted heavy crude, the values of training and experimental are plotted in Figure 6. According to the trained value and statistical results, all the training viscosities were closer to the 45° line; in other words, they were close to the experimental viscosities. In addition, the viscosity values had a coefficient of determination R^2 of 0.99 between the training and experimental viscosity. Moreover, the result of the regression plot generated by this model showing the relationship between the tested and experimental viscosity values is shown in Figure 7. It can be seen that the tested viscosities were also closer to the 45° line, and the values had a coefficient of determination R^2 of 0.97. Therefore, it can be concluded that the model had a higher degree of fit and could predict the diluted viscosity satisfactorily.

Figure 8 shows the accuracy plot of the training and validation loss with relation to the number of epochs used for the development of the model.

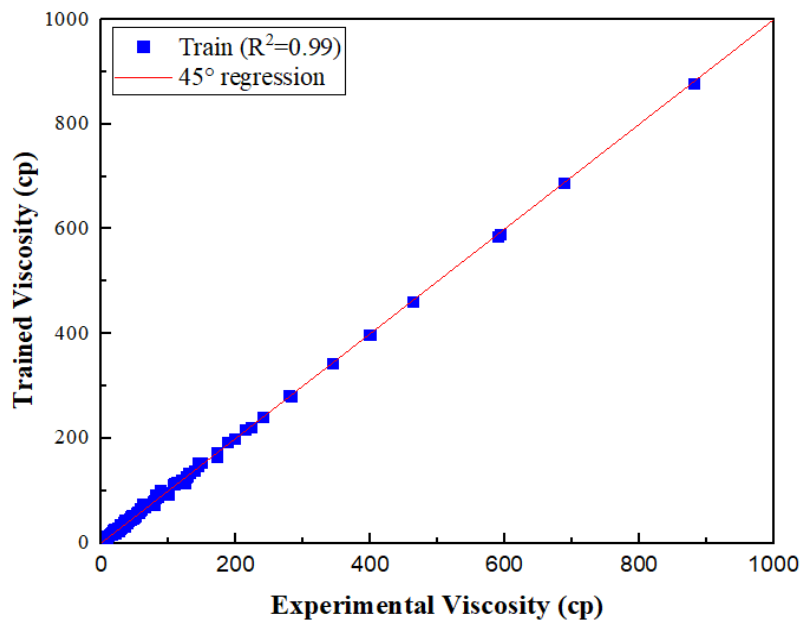


Figure 6. Regression plot of experimental versus trained viscosity.

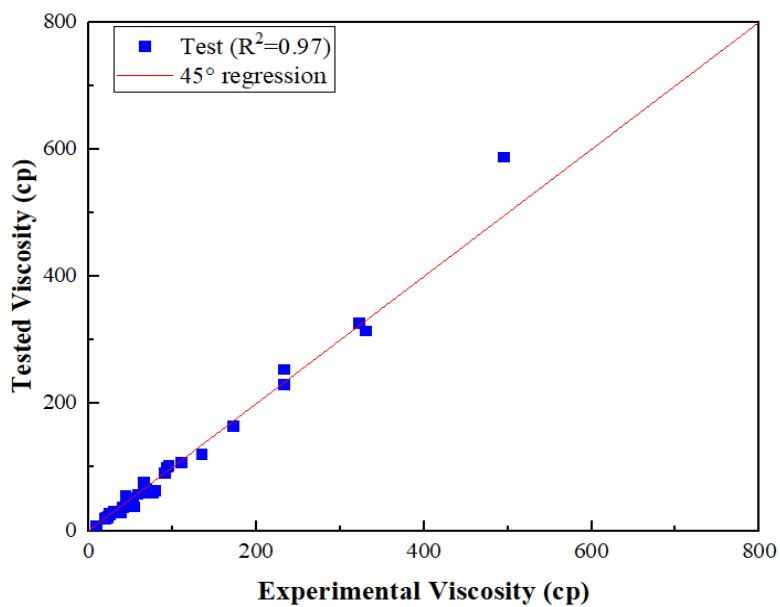


Figure 7. Regression plot of experimental versus tested viscosity.

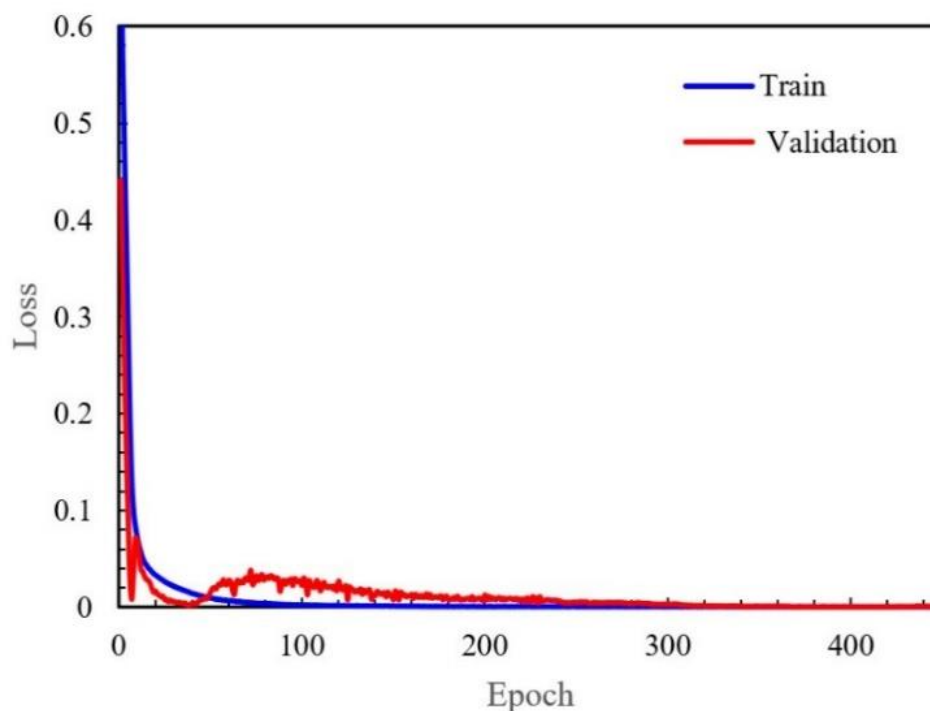


Figure 8. Plot of training model accuracy.

3.3. Comparison of the ANN with Existing Empirical Correlations

This paper cited the experimental viscosity value of the heavy crudes mixed with light oils from the published literature. We compared the values of existing empirical correlations in 3.1, the proposed model in this work, and experimentally, in terms of their performance and accuracy with diluted heavy crude oils.

Figure 9 depicts a comparison between the calculated and experimental viscosities of different models. It can be seen that Kendal-Monroe's and Arrhenius's models significantly overestimated the viscosity of the diluted heavy crude, whereas Xing's model predicted lower viscosities. It should be noted that Cragoe's and double-log's models had a better fit and a closer fitting slope, and the predicted viscosity data of the others were scattered. The reasons for the high prediction accuracy of the ANN model are as follows:

- (1) Compared with the Arrhenius, Double log, Cragoe, and Kendal-Monroe models, the new model considered the effect of temperature on viscosity, which increased the accuracy of the new model;
- (2) As for the Xing model, this model was developed on the basis of the Arrhenius model by introducing a correction coefficient. It can be seen from Figure 9 that the viscosity predicted by the Arrhenius model was larger than the measured viscosity, while the viscosity predicted by the new model was smaller than the experimental viscosity, which may be caused by an inappropriate correlation coefficient of the Xing model.

To determine the degree of prediction bias of the new method, the AARE and SD% of different models were also compared in Figure 10. It is clear from this that the proposed model provided the best prediction with the heavy crude oil diluted with light oils and performed very well against the other models. The newly developed model had the lowest AARE% of 10.45% and an SD% of 8.45%. The closer the errors are to 0, the more accurate the model. Considering that the datasets contained a wide range of viscosities, the model proposed in this paper had better prediction performance on the viscosity of heavy crude diluted with light oils.

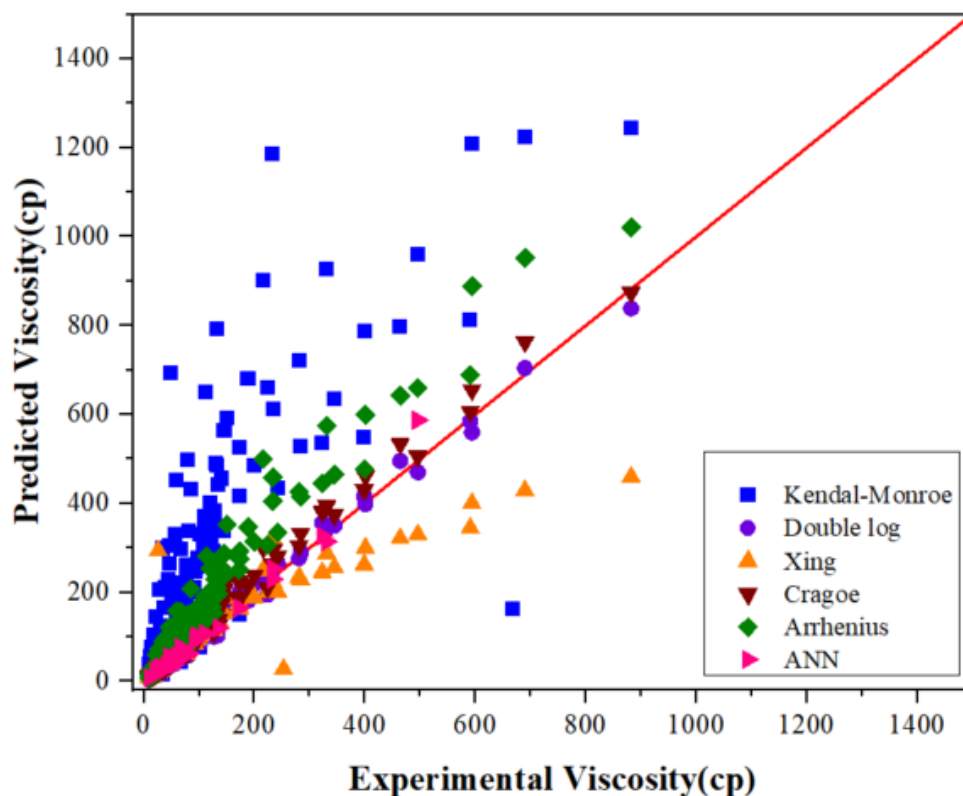


Figure 9. Comparison of the experimental and predicted viscosities of heavy crude diluted with light oils.

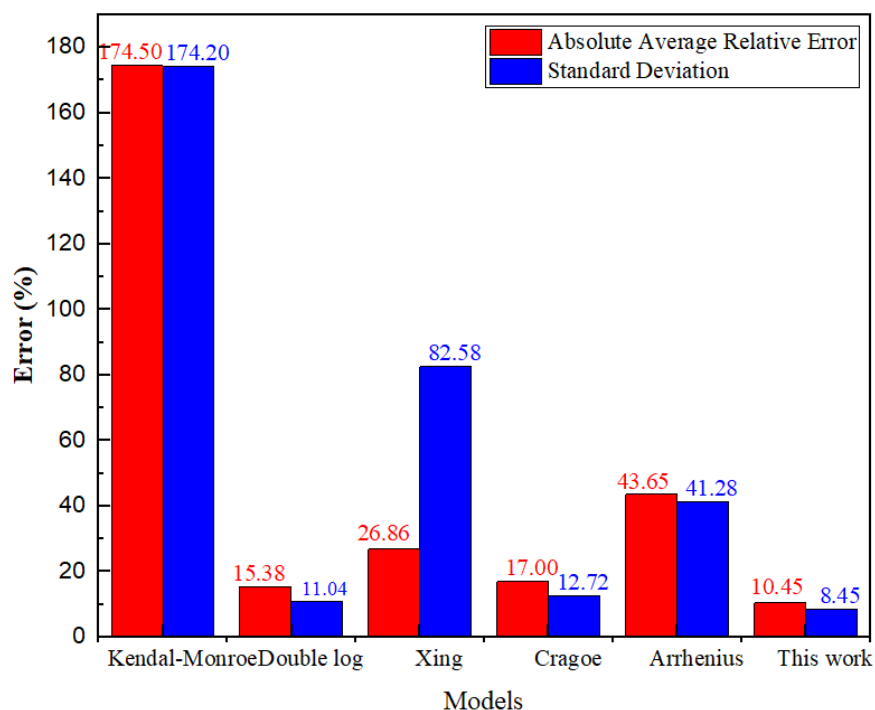


Figure 10. Comparison of the AARE and SD for the various models.

4. Conclusions

In this paper, a novel method for predicting the viscosity of heavy crude oil diluted with lighter oils was proposed based on machine learning. The comparison between the prediction of the new

correlations and the other empirical correlations available in the literature was carried out. Furthermore, the statistical error analysis was used to determine the accuracy of these correlations. The key findings of this novel viscosity model are summarized as follows:

- (1) From the analysis of the results, the new viscosity model had the lowest absolute average relative error values of 10.45%, standard deviation values of 8.45%, and coefficient of determination ($R^2=0.95$).
- (2) The new viscosity showed a good accuracy as well when compared to the conventional viscosity correlations in the published literature. In other words, the new viscosity model outperformed empirical correlations.
- (3) The presence of asphaltene in heavy crude oil is also an important parameter affecting the viscosity of the heavy crude oil, and it can be studied in future research.

Acknowledgments: This section is required for all papers. Here you can acknowledge any support given which is not covered by the author contribution or funding sections. This may include administrative and technical support, or donations in kind (e.g., materials used for experiments). Please do not thank the editors in this section, but you can send an email to express thanks.

Funding: This section is required for all papers. Please add: “This research received no external funding.” or “This research was funded by Name of Funder, grant number XXX.”

Conflict of interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References:

1. Surname Initial(s), Surname Initial(s), Surname Initial(s), *et al.* Article title. Title of journal. Year; volume: pages. Doi:xxx
2. Surname Initial(s), Surname Initial(s), Surname Initial(s), *et al.* Article title. Title of journal. Year; volume: pages. (In Language). Doi:xxx
3. Surname Initial(s), Surname Initial(s), Surname Initial(s), *et al.* Book Title. Edition - if available. Publisher: City, Country. Year.
4. Surname Initial(s), Surname Initial(s), Surname Initial(s), *et al.* Title of chapter. In Surname of the editor(s) Initial(s) (ed(s).) Book title Page. Edition. Publisher: City, Country. Year.
5. Surname Initial(s), Surname Initial(s), Surname Initial(s), *et al.* Title. Year. Available at: URL (Accessed: Day month year).
6. Surname Initial(s), Surname Initial(s), Surname Initial(s), *et al.* ‘Title of paper’, Title of the conference. Place and date of the conference. Publisher: City, Country. Year of Publication.
7. Title of Site. Available online: URL (accessed on Day Month Year)