processes

MDPI

Article

Prediction of Refractive Index of Petroleum Fluids by Empirical Correlations and ANN

Georgi Nikolov Palichev [1], Dicho Stratiev [1,2,*], Sotir Sotirov [3], Evdokia Sotirova [3], Svetoslav Nenov [4], Ivelina Shishkova [2], Rosen Dinkov [2], Krassimir Atanassov [1,3], Simeon Ribagin [1,3], Danail Dichev Stratiev [1], Dimitar Pilev [4] and Dobromir Yordanov [3]

[1] Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences, Georgi Bonchev 105, 1113 Sofia, Bulgaria; georgipalicev@gmail.com (G.N.P.); k.t.atanassov@gmail.com (K.A.); simribagin@gmail.com (S.R.); danail.stratiev@gmail.com (D.D.S.)
[2] LUKOIL Neftohim Burgas, 8104 Burgas, Bulgaria; shishkova.ivelina.k@neftochim.bg (I.S.); dinkov.rosen.k@neftochim.bg (R.D.)
[3] Intelligent Systems Laboratory, Department Industrial Technologies and Management, University Prof. Dr. Assen Zlatarov, Professor Yakimov 1, 8010 Burgas, Bulgaria; ssotirov@btu.bg (S.S.); esotirova@btu.bg (E.S.); dobromirj@abv.bg (D.Y.)
[4] Department of Mathematics, University of Chemical Technology and Metallurgy, Kliment Ohridski 8, 1756 Sofia, Bulgaria; s.nenov@gmail.com (S.N.); pilev@uctm.edu (D.P.)
* Correspondence: stratiev.dicho@neftochim.bg

Abstract: The refractive index is an important physical property that is used to estimate the structural characteristics, thermodynamic, and transport properties of petroleum fluids, and to determine the onset of asphaltene flocculation. Unfortunately, the refractive index of opaque petroleum fluids cannot be measured unless special experimental techniques or dilution is used. For that reason, empirical correlations, and metaheuristic models were developed to predict the refractive index of petroleum fluids based on density, boiling point, and SARA fraction composition. The capability of these methods to accurately predict refractive index is discussed in this research with the aim of contrasting the empirical correlations with the artificial neural network modelling approach. Three data sets consisting of specific gravity and boiling point of 254 petroleum fractions, individual hydrocarbons, and hetero-compounds (Set 1); specific gravity and molecular weight of 136 crude oils (Set 2); and specific gravity, molecular weight, and SARA composition data of 102 crude oils (Set 3) were used to test eight empirical correlations available in the literature to predict the refractive index. Additionally, three new empirical correlations and three artificial neural network (ANN) models were developed for the three data sets using computer algebra system Maple, NLPSolve with Modified Newton Iterative Method, and Matlab. For Set 1, the most accurate refractive index prediction was achieved by the ANN model, with %AAD of 0.26% followed by the new developed correlation for Set 1 with %AAD of 0.37%. The best literature empirical correlation found for Set 1 was that of Riazi and Daubert (1987), which had %AAD of 0.40%. For Set 2, the best performers were the models of ANN, and the new developed correlation of Set 2 with %AAD of refractive index prediction was 0.21%, and 0.22%, respectively. For Set 3, the ANN model exhibited %AAD of refractive index prediction of 0.156% followed by the newly developed correlation for Set 3 with %AAD of 0.163%, while the empirical correlations of Fan et al. (2002) and Chamkalani (2012) displayed %AAD of 0.584 and 0.552%, respectively.

Keywords: petroleum; refractive index; empirical correlation; ANN; intercriteria analysis

Citation: Palichev, G.N.; Stratiev, D.; Sotirov, S.; Sotirova, E.; Nenov, S.; Shishkova, I.; Dinkov, R.; Atanassov, K.; Ribagin, S.; Stratiev, D.D.; et al. Prediction of Refractive Index of Petroleum Fluids by Empirical Correlations and ANN. *Processes* **2023**, *11*, 2328. https://doi.org/10.3390/pr11082328

Academic Editor: Albert Ratner

Received: 5 July 2023
Revised: 28 July 2023
Accepted: 1 August 2023
Published: 2 August 2023

1. Introduction

The refractive index of petroleum fluids is an important characterization parameter that provides information about the composition and identification of petroleum fluids [1–4]. This, along with the molecular weight, can be used to predict the average

number of aromatic rings in the petroleum fluids [5], the volumetric properties, i.e., thermal expansivity and isothermal compressibility for nonpolar hydrocarbon systems [6]; critical temperature, critical pressure, and acentric factor for nonpolar hydrocarbons [7–9]; and viscosity of nonpolar mixtures [10–12]. The refractive index can be used to predict petroleum fluid density [13,14], as well as interaction parameter kij, which takes into account the interactions between unlike molecules in a mixture and surface tension [13], thermal conductivity, and diffusion coefficients of liquid non-polar compounds at various temperatures [12]. The refractive index can also be utilized to quantify the ability of crude oil to disperse its asphaltene fraction because of the relationship of the refractive index to the van der Waals forces between nonpolar molecules and the fact that aggregation and separation of asphaltenes depend mainly on the magnitude of the van der Waals forces between nonpolar species [15]. For that reason, the index is employed to determine the onset of asphaltene flocculation [16–19]. Considering the presence of the almost linear relationship between the solubility parameter and refractive index [15], the index can also be utilized to calculate the solubility parameter [20–23]. Its use has been proposed as an indicator for crude oil stability as well as for controlling and mitigating fouling in petroleum refining [24]. Given the importance of the petroleum fluid refractive index, its correct measurement deserves special attention. The opacity of the heavy petroleum fluids embarrasses their refractive index measurement [25], and the conventional refractometers are typically suitable only for the transparent fractions of crude oils (light fractions) [26,27]. In the case of heavy crude oils, natural bitumen, and heavy fuels, when a conventional refractometer is used, the refractive index is determined by measuring the refractive index of several oil–solvent mixtures, and the results are extrapolated at zero solvent concentration to specify the value for the crude oil [25,27]. To be able to directly measure the refractive index of heavy petroleum fluids, new capillary [26,28] and optical fiber techniques [16,29] were proposed. The employment of these techniques allows for petroleum refractive index measurements in the range 1.33–1.75 [29]. The accuracy of petroleum refractive index measurements by capillary and optical fiber techniques is reported in the range 1×10^{-4}, and 1×10^{-5} [26,29]. The use of direct measurement of the refractive index of opaque petroleum fluids also has its own limitations, which are entailed by the presence of impurities and insoluble particulate matter in petroleum fluids; this matter can block the capillary entrance [29]. In addition, the employment of these methods is expensive, laborious, and time consuming [30]. Using the dilution method and extrapolation at zero diluent concentration to measure the refractive index of opaque petroleum fluids, assuming an ideal binary mixture of the oil and the diluent, returned different values depending on the concentration range of the oil in the mixture [27]. At high oil concentrations, the ideal behavior of the blend oil–diluent was confirmed; however, at low oil concentrations, a strong non-ideality of refractive index was reported, most probably because of asphaltene de-association [27]. These shortcomings of refractive index measurement for petroleum fluids have been an incentive for researchers to develop predictive models [14,30–42]. Vargas and Chapman [14], Yarranton et al. [30], and Stratiev et al. [31] developed correlations to predict the refractive index of petroleum fluids and individual compounds based on their density. Riazi and Daubert [32], Stratiev et al. [33], Hosseinifar and Shahvertdi [34] established correlations to predict viscosity of petroleum fractions from density (specific gravity) and average boiling point or T50%. Dhulesia [35] elaborated a petroleum fraction refractive index correlation that employed specific gravity, average boiling point, and molecular weight. Fan et al. [36] worked out a correlation to predict refractive index of crude oils from SARA fraction composition. Chamkalani [37] developed another SARA fraction composition-based correlation to predict a crude oil refractive index that features better prediction accuracy than the correlation of Fan et al. [36]. Chamkalani et al. [38] employed a hybrid of two heuristic optimizations (which coupled simulated annealing with the Nelder–Mead Simplex method) with a least-square support vector machine (LSSVM) to predict refractive index of crude oils based on SARA fraction composition data that showed a higher prediction accuracy than empirical correlations as well as a neural network model.

Gholami et al. [39] applied a fuzzy logic technique to predict the refractive index of crude oil based on SARA fraction composition data. They reported much higher refractive index prediction accuracy by fuzzy modelling than the empirical correlations of Fan et al. [36] and Chamkalani et al. [37]. Zargar et al. [40] employed an integrated intelligent method coined neuro-fuzzy (NF) with particle swarm optimization (PSO) technique to elaborate a crude oil refractive index prediction based on SARA fraction composition data. They reported a crude oil refractive index with improved accuracy. Tatar et al. [41] used a committee machine intelligent system (CMIS) that utilizes various artificial neural networks, such as multilayer perceptron (MLP), radial basis function (RBF), and least squares support vector machine (LSSVM), to predict a crude oil refractive index based on SARA fraction composition data. They reported that the suggested intelligent system was superior to the classical correlations [41]. Gholami et al. [42] used an intelligent approach based on optimized support vector regression (SVR), hybrid of grid and pattern search (HGP), a genetic algorithm (GA), and an imperialist competitive algorithm (ICA) to predict crude oil refractive index based on SARA fraction composition data. They found that the SVR and ICA provided the highest accuracy for crude oil refractive index prediction, achieving R^2 of 0.9971. To the best of our knowledge, no reports have appeared in the literature yet that compare the available empirical models to predict the refractive index of petroleum fluids on the base of density and boiling point (molecular weight) with an ANN model. For that reason, we have collected data for 254 petroleum fluids with the following range of variation of refractive index ($1.3710 \leq RI \leq 1.6538$), specific gravity ($0.6570 \leq SG \leq 1.0990$) and average boiling point ($53\ ^\circ C \leq BP \leq 510\ ^\circ C$) and tested them using the empirical models available in the current literature. We have also developed a new regression model using computer algebra system (CAS) Maple and NLPSolve with Modified Newton Iterative Method, as described in [43], and developed an ANN model utilizing Matlab.

Another set of data for the refractive index ($1.4283 \leq RI \leq 1.5644$) of 136 crude oils along with specific gravity ($0.7567 \leq SG \leq 0.9813$); molecular weight ($126\ g/mol \leq MW \leq 499\ g/mol$) and SARA composition data ($38.4\% \leq Sat. \leq 79.3\%$); ($11.9\% \leq Aro \leq 29.8\%$); ($0.7\% \leq Res. \leq 33.7\%$); and ($0.1\% \leq Asp \leq 16.5\%$) were taken from the open source literature [43] and employed to develop a new empirical correlation and ANN model.

The aim of this research is to compare the capability of the empirical correlations available in the literature to predict the refractive index of petroleum fluids, develop new empirical correlations with the two sets of data mentioned above, and contrast them with the new developed ANN refractive index models.

2. Materials and Methods

Data regarding boiling point, specific gravity, and refractive index were collected for 254 petroleum fluids and individual hydrocarbons from references [14,30,32,33,44–50], which are shown in Table S1. These data were employed for testing the empirical correlations available in the literature and displayed as Equations (1)–(7).

Equation (1) presents the correlation of Hosseinifar and Shahvertdi [34], published in 2021. They collected a wide range of experimental data regarding normal boiling point, specific gravity, and refractive index for light and heavy petroleum fractions, with refractive index variation in the range 1.36–1.57, to develop their correlation.

$$RI = \left\{ 0.372239 \times T_b^{0.607176} \times \left(\frac{3 - SG}{3 + 2 \times SG} \right)^{0.947982} + \left(\frac{2.032675}{T_b^{0.200525}} \right) \times \left(\frac{3 + 2 \times SG}{3 - SG} \right)^{6.127836} \right\}^{0.089596} \quad (1)$$

where

RI = Refractive index of petroleum fluid at 20 °C;
T_b = Boiling point (average boiling point) of petroleum fluid, K;
SG = Specific gravity of petroleum fluid at 60/60 °F.

Equation (2) presents the correlation from Riazi and Daubert [32], published in 1987, which predicted the refractive index of undefined petroleum fractions. The data base used

by Riazi and Daubert to develop Equation (2) includes predominantly light petroleum fractions having *RI* between 1.4188 and 1.4776.

$$RI = \left(\frac{1 + 2 \times 0.3824 \times T_b^{-0.02269} \times SG^{0.9182}}{1 - 0.3824 \times T_b^{-0.02269} \times SG^{0.9182}} \right)^{1/2} \quad (2)$$

where

T_b = Boiling point (average boiling point) of petroleum fluid, °R.

Equation (3) presents the correlation from Dhulesia [35] that predicted the refractive index of petroleum fractions, which was published in 1986. It was developed for the range of light and heavy petroleum fractions with RI between 1.4459 ÷ 1.5681 and SG variation between 0.8335 and 1.0133.

$$RI = 1 + 0.8447 \times SG^{1.2056} \times T_b^{-0.0044} \times MW^{-0.0044} \quad (3)$$

where

MW = molecular weight of petroleum fluid, g/mol. Where the molecular weight of the petroleum fluids was absent, it was estimated by the new correlation reported in our recent study [51].

Equation (4) presents the correlation from Vargas and Chapman [14] that predicted the refractive index of petroleum fluids and individual hydrocarbons, published in 2010. The correlation of Vargas and Chapman has been validated with over 200 crude oil samples, in a wide range of specific gravities (0.7587–1.000) and temperatures (10–70 °C).

$$RI = \left(\frac{2 \times \left(0.5054 - 0.3951 \times \rho + 0.2314 \times \rho^2\right) + 1}{1 - \left(0.5054 - 0.3951 \times \rho + 0.2314\rho^2\right)} \right)^{1/2} \quad (4)$$

where ρ is density at 20 °C, g/cm^3, and RI is refractive index at 20 °C.

Equation (5) presents the correlation from Yarranton et al. [30] which predicted the refractive index of petroleum fluids and individual hydrocarbons, published in 2015. Yarranton et al. utilizes data from petroleum fluids, including saturate, aromatic, resin, and asphaltene (SARA) fractions that have an RI between 1.3326 and 1.897.

$$RI = \left(\frac{2 \times \left(0.5280 - 0.3784 \times \left(1.2813 - \rho\right)^{0.5}\right) + 1}{1 - \left(0.5280 - 0.3784 \times \left(1.2813 - \rho\right)^{0.5}\right)} \right)^{\frac{1}{2}} \quad (5)$$

Equation (6) presents the correlation from Stratiev et al. [33] that predicted the refractive index of petroleum fluids and individual hydrocarbons, published in 2014. Stratiev et al. used primary and secondary petroleum fractions with the RI variation range between 1.4786 and 1.5695, density at 15 °C between 0.8630 and 1.0971, and T_{50} between 243 and 510 °C.

$$RI = 0.702091 d_{15} - 0.00011 T_{50} + 0.91493 \quad (6)$$

where d_{15} is density at 15 °C, g/cm^3, and T_{50} is boiling point of evaporate at 50%, °C.

Equation (7) presents the correlation from Stratiev et al. [31] that predicted the refractive index of petroleum fluids and individual hydrocarbons, published in 2019. It was developed for petroleum fractions with a range of RI variation between 1.4747and 1.6538 and density at 15 °C between 0.8638 and 1.0971.

$$RI = 0.77887 d_{15} + 0.80065 \quad (7)$$

The data in Table S1 was also used to develop new refractive index empirical correlations to predict the refractive index based specific gravity and the boiling point of

the petroleum fluids and individual hydrocarbons. For that purpose, (CAS) Maple and NLPSolve with Modified Newton Iterative Method were utilized.

In order to use the ANN technique with the data in Table S1 to model the refractive index of these 254 petroleum fluids and individual hydrocarbons more input data were needed than the two properties of specific gravity and boiling point. To this end, the method from Hosseinifar and Shahverdi [52] was applied to generate TBP distillation data from the petroleum fluid $T_{50\%}$ boiling point to obtain more inlet parameters for the ANN. The application of the Hosseinifar and Shahverdi method [52] for making more inlet parameters as required by the ANN method is detailed in our recent papers [51,53]. The data in Table S2 includes the specific gravity, and boiling point, and the TBP generated boiling point at 5, 10, 30, 50, 70, and 90 vol.%, and the Kw-characterization factor of the 254 petroleum fluids and individual hydrocarbons. It should be highlighted here that the method by Hosseinifar and Shahverdi [52] has not been employed to build the real TBP curve; instead, it was used to make more than two inlet parameters (specific gravity and boiling point) to be applied for ANN modeling purposes. The Kw-characterization factor of the studied petroleum fluids was calculated as shown in our recent research [53].

Data from 136 crude oils for specific gravity, molecular weight, Kw-characterization factor, TBP distillation characteristics (generated by the use of the method from Hosseinifar and Shahverdi, [52]), and refractive index, as shown in Table S3, was utilized for refractive index prediction by the employment (CAS) Maple and NLPSolve with Modified Newton Iterative Method and ANN technique. The artificial neural network (ANN) modeling approach used in this study is detailed in our earlier investigation [51]. The data from the 136 crude oils for specific gravity and molecular weight was taken from reference [44].

Data from 102 crude oils for specific gravity, molecular weight, SARA composition, and refractive index, shown in Table S4, was used to develop new empirical correlations, and were then compared with the existing correlations from Fan et al. [36] (Equation (8)) and Chamkalani [37] (Equation (9)).

$$RI_{Crude\ oil} = 0.01452 \times Sat. + 0.014982 \times Aro + 0.016624 \times (Res + Asp) \quad (8)$$

where

$RI_{Crude\ oil}$ = Refractive index of crude oil at 20 °C;
Sat = Saturate content of crude oil, wt.%;
Aro = Aromatic content of crude oil, wt.%;
Res = Resin content of crude oil, wt.%;
Asp = Asphaltene content of crude oil, wt.%.

$$RI_{Crude\ oil} = -0.0008515 \times Sat. - 0.0002524 \times Aro + 0.0016341 \times Res + 0.0013928 \times Asp + 1.524412 \quad (9)$$

The accuracy of the petroleum fluid refractive index prediction has been evaluated by the statistical parameters, shown as Equations (10)–(16).

$$\text{Error } (E): \ E = \left(\frac{RI_{exp} - RI_{calc}}{RI_{exp}} \right) \times 100 \quad (10)$$

$$\text{Standard error } (SE): \ SE_i = \left(\sum \left(\frac{\left(RI_{exp} - RI_{calc}\right)^2}{N-2} \right) \right)^{\frac{1}{2}} \quad (11)$$

$$\text{Relative standard error } (RSE): \ RSE_i = \frac{\text{SE}}{\text{mean of the sample}} \times 100 \quad (12)$$

$$\text{Sum of square errors } (SSE): \ SSE = \sum \frac{1}{RI_{exp}^2} (RI_{exp} - RI_{calc})^2 \quad (13)$$

$$\text{Relative average absolute deviation } (\%AAD):\ \%AAD = \frac{1}{N}\frac{\sum|RI_{exp} - RI_{calc}|}{RI_{exp}} * 100 \quad (14)$$

$$\text{Sum of relative errors } (SRE):\ SRE = \sum\left(\frac{RI_{exp} - RI_{calc}}{RI_{exp}}\right) \times 100 \quad (15)$$

$$\text{Average absolute deviation } (AAD):\ AAD = \frac{1}{N}\sum|RI_{exp} - RI_{calc}| \quad (16)$$

3. Results

3.1. Prediction of Refractive Index of Petroleum Fractions, Individual Hydrocarbons, and Some Hetero Compounds by Empirical Correlations and ANN Using Specific Gravity (Density), and Boiling Point Data

Figure 1 presents parity graphs of measured versus calculated refractive index by the empirical correlations studied in this work (Equations (1)–(7)) for the 254 petroleum fractions, individual hydrocarbons, and hetero compounds shown in Table S1.

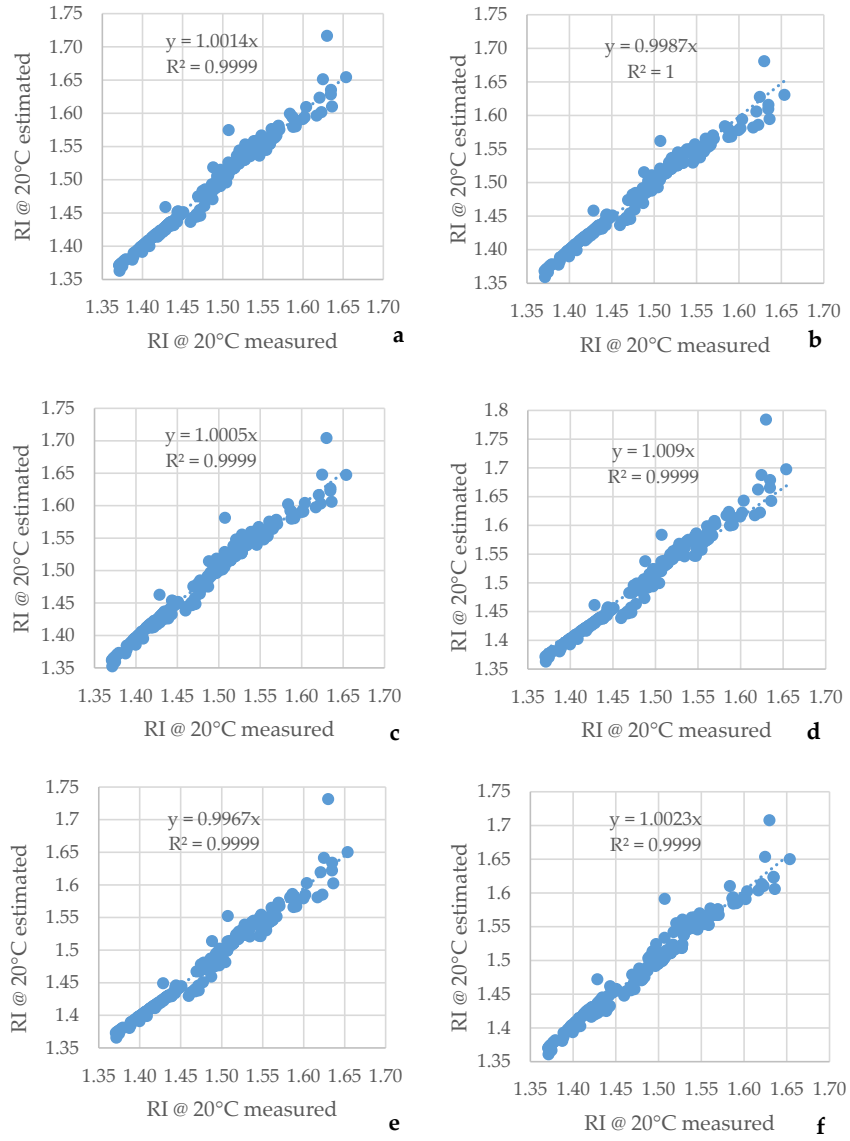

Figure 1. *Cont.*

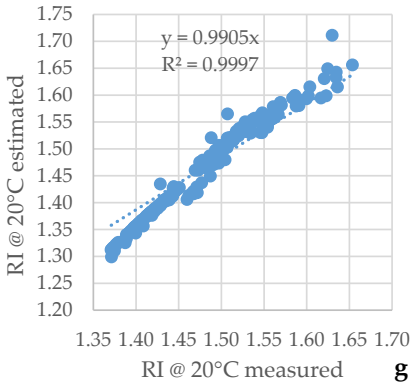

Figure 1. Parity graphs of measured versus calculated refractive index using the empirical correlations of Hosseinifar [34] (Equation (1); (**a**); Riazi and Daubert [32] (Equation (2); (**b**); Dhulesia [35] (Equation (3); (**c**); Vargas and Chapman [14] (Equation (4); (**d**); Yarranton et al. [30] (Equation (5); (**e**); Stratiev et al. [41] (Equation (6); (**f**); Stratiev et al. [33] (Equation (7); and (**g**) for the 254 petroleum fractions, individual hydrocarbons, and hetero compounds whose data is given in Table S1.

It is evident from the data in Figure 1 that the coefficient of determination (R^2) of the evaluated empirical correlations diminishes in the order: Equation (2) (R^2 = 0.9738) > Equation (1) (R^2 = 0.9737) > Equation (3) (R^2 = 0.9732) > Equation (6) (R^2 = 0.9692) > Equation (5) (R^2 = 0.9668) > Equation (4) (R^2 = 0.952) > Equation (7) (R^2 = 0.9147). Among the studied correlations, those of Equations (4) and (7) are characterized with the highest deviation from the value of 1.000, implying that they have the highest bias of predicted refractive index compared to the measured one. While Equation (4) demonstrates a positive bias, that of Equation (7) is negative. One may conclude from this data that Equation (2) is characterized with the best agreement between measured and predicted petroleum fraction refractive index.

The data in Table S1 was used to investigate the possibility of developing a new empirical correlation to demonstrate better prediction accuracy for the petroleum fraction refractive index from the specific gravity and boiling points. Linear and nonlinear models that used both the refractive index itself and a function of the refractive index, FRI, as shown in Equation (17), were employed to obtain the best fit with the experimental refractive index by the use of CAS Maple and NLPSolve with Modified Newton Iterative Method.

$$F_{RI} = \frac{n_{D20}^2 - 1}{n_{D20}^2 + 2} \tag{17}$$

where n_{D20} = refractive index at 20 °C.

No significant difference was observed in the accuracy of refractive index prediction between the linear and nonlinear models using both the refractive index and FRI. Slightly better results were registered with the linear model and FRI, which is shown as Equation (18). A total of 180 data points from the data in Table S1 were used to develop Equation (18), and the remaining 74 points (from the data with Nr.181 to the data with Nr. 254 in Table S1) were applied for verification.

$$\begin{gathered} FRI = 0.324172 \times SG - 0.0000261350 \times T_b + 0.0208779 \\ RI = \left(\frac{1+2\times FRI}{1-FRI}\right)^{1/2} \end{gathered} \tag{18}$$

where T_b is in K.

Along with a regression technique to develop a petroleum fraction refractive index model, an ANN approach was also used. For the data from Table S2, the neural network has a 9:10:10:10:1 structure. This means nine input parameters (*SG*, *Tb*, $T_{5\%}$, $T_{10\%}$, $T_{30\%}$, $T_{50\%}$, $T_{70\%}$, $T_{90\%}$, Kw), 10 neurons in the first neuron layer, 10 in the second hidden neuron

layer, and one output. In this case, the Levemberg–Marquardt algorithm was used. Figure 2 shows the structure of the neural network with four layers.

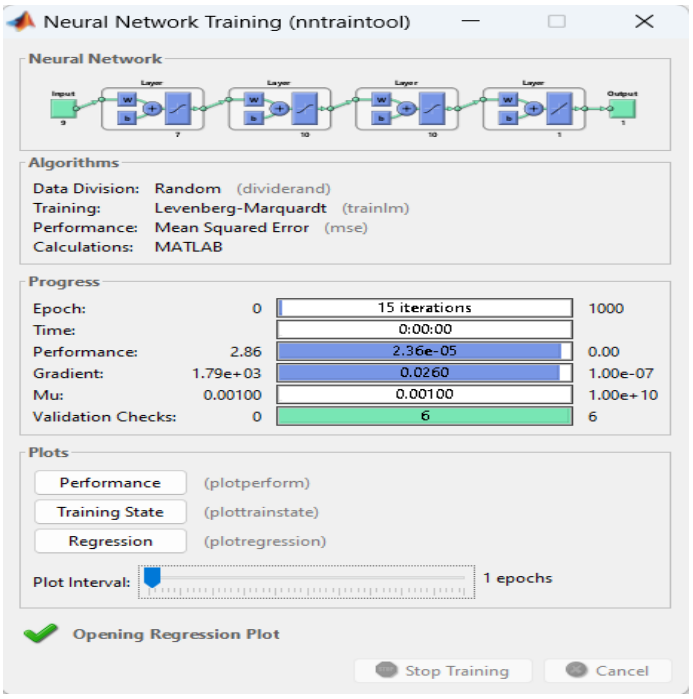

Figure 2. Structure of the neural network employing petroleum fraction specific gravity, simulated TBP distillation data, and Kw-factor.

Figure 3 shows the training process of the neural network. It shows three graphs: training, testing, and verification. Training is usually done with 70% of the training sequence taken at random. Testing is done with 20% of the training sequence, also taken at random, and the remaining 10% of the training sequence is for verification. The verification values are usually not used to train the neural network. Thus, they are considered independent. In our training, the neural network for the particular case of the data in Table S2 was trained for eight epochs, and the smallest mean squared error was obtained at epoch number nine and is 2.4535×10^{-5}.

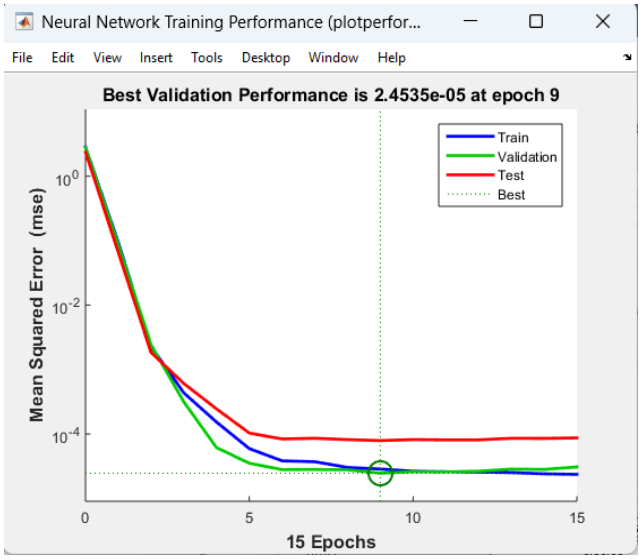

Figure 3. Training process of the neural network.

Figure 4 gives the basic parameters of the training: gradient for each epoch, Mu for each epoch, and the validation of each epoch.

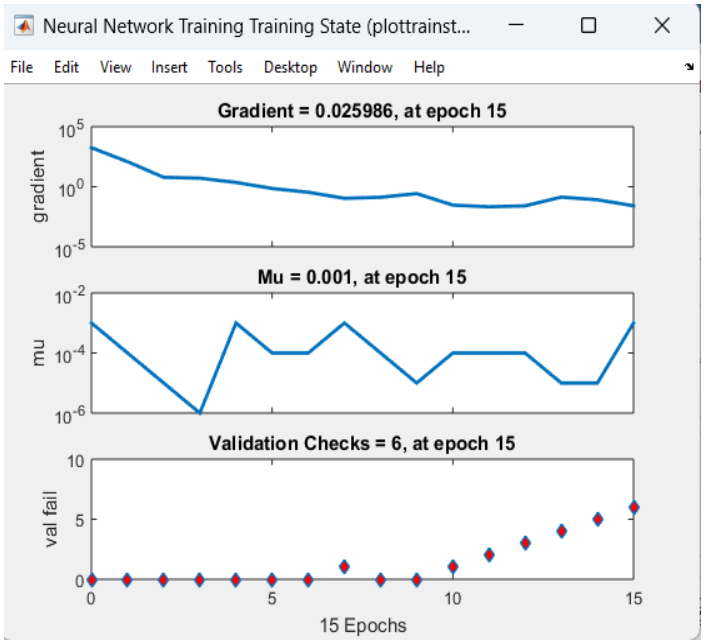

Figure 4. The basic parameters of the training: gradient for each epoch, Mu for each epoch, and the validation of each epoch.

In the actual use of this neural network, the R coefficients of each part of the training sequence are given. The individual coefficients are R = 0.9966 for training, R = 0.99725 for verification, R = 0.98728 for testing, and R = 0.99562 for the entire training process (see Figure 5).

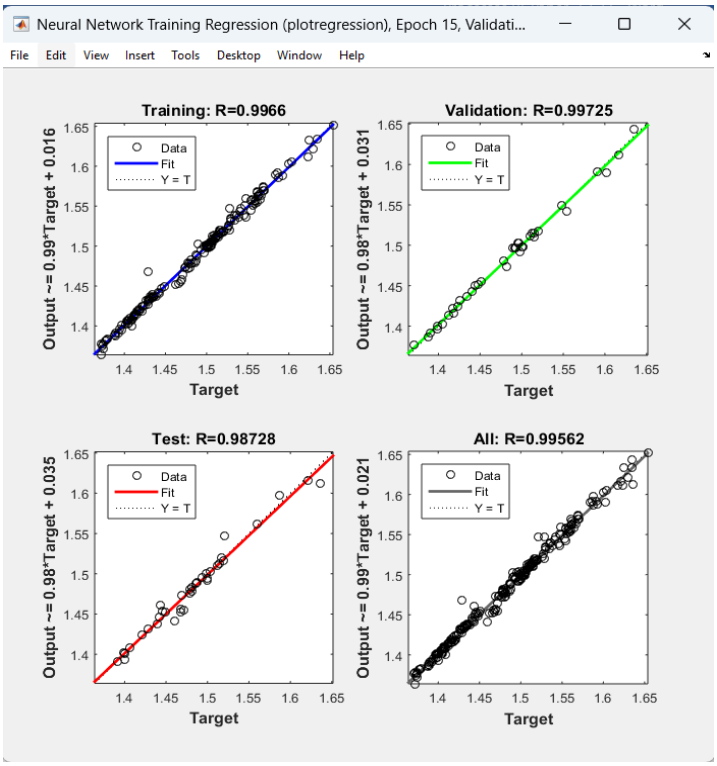

Figure 5. Regression coefficients of the learning process of the ANN employing petroleum fraction refractive index, specific gravity, simulated TBP distillation data, and Kw-factor.

Figure 6 shows the parity graph of measured versus calculated refractive index of petroleum fractions by the newly developed empirical correlation (Equation (18)), and by the ANN.

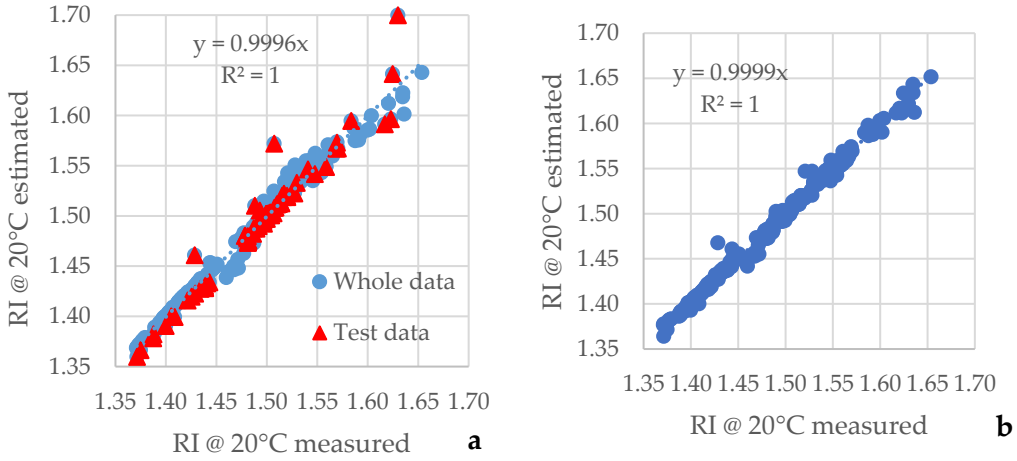

Figure 6. Parity graphs of measured versus calculated refractive index of petroleum fractions by the new empirical correlation (Equation (18) **a**); and the ANN model (**b**).

The data in Figure 6 indicates that the ANN model excels, with a higher coefficient of determination (R^2 = 0.9911) than that of the new empirical correlation (Equation (18); R^2 = 0.9758). These values are higher than those of the tested literature empirical correlations (Equations (1)–(7)). The slopes of regression lines of Equation (18) and the ANN model are almost equal to 1.00, implying a lack of bias in these two models. Based on this comparison, a conclusion could be made that Equation (18) and the ANN model predict with a higher accuracy the petroleum fraction refractive index than the tested literature available empirical correlations.

Table 1 summarizes the statistical analyses for the new empirical correlation, the ANN model, and the seven studied empirical correlations available from the literature.

Table 1. Statistical analysis of studied methods to predict petroleum fraction refractive index for the data in Tables S1 and S2.

RI Petroleum Fraction Models		SE	RSE	SSE	%AAD	SRE	AAD	Max. dev.	Ranking
Hosseinifar and Shahvertdi, 2021 [34]	Equation (1)	0.00012	0.00817	0.0130	0.42	−32.9	0.00629	0.08640	4
Riazi and Daubert, 1987 [32]	Equation (2)	0.00011	0.00718	0.0114	0.40	30.9	0.00607	0.05437	3
Dhulesia, 1986 [35]	Equation (3)	0.00012	0.00831	0.0136	0.45	−9.0	0.00679	0.07411	5
Vargas and Chapman, 2010 [14]	Equation (4)	0.00044	0.02966	0.0465	0.98	−215.4	0.01502	0.15383	8
Yarranton, et al., 2015 [30]	Equation (5)	0.00016	0.01066	0.0169	0.49	81.8	0.00739	0.10135	6
Stratiev, et al., 2014 [33]	Equation (6)	0.00014	0.00959	0.0155	0.51	−56.9	0.00760	0.08361	7
Stratiev, et al., 2019 [31]	Equation (7)	0.00080	0.05375	0.0995	1.48	269.8	0.02129	0.08114	9
New correlation (this work)	Equation (18)	0.00010	0.00679	0.0109	0.37	9.8	0.00555	0.06972	2
ANN		0.00004	0.00239	0.0040	0.26	0.2	0.00395	0.03900	1

The data in Table 1 show that all the estimated statistical parameters characterizing the prediction accuracy of the refractive index have lower values than the new empirical correlation (Equation (18)) and the ANN model, meaning a higher accuracy of prediction for both new models. The ANN model, however, indicates the lowest values of all estimated statistical parameters, implying the best accuracy of the petroleum fraction refractive index forecast. The data in Table 1 replicate the ranking of empirical correlations available in the literature based on the coefficient of determination (R^2) made above. Although the correlation from Riazi and Daubert [32] (Equation (2)) was developed based on light

petroleum fractions, its prediction of the refractive index is characterized by the highest accuracy among the other empirical models available in the literature. The worst performing empirical correlations (Equations (4) and (7)) exhibit the largest bias, as can be seen from the sum of relative errors (SRE) data in Table 1, which are negative for Equation (4) (positive bias) and positive for Equation (7) (negative bias), as seen in Figure 1d,g, respectively.

3.2. Prediction of Refractive Index of Crude Oils by Empirical Correlations and ANN Using Specific Gravity and Molecular Weight Data

Figure 7 presents parity graphs of measured versus calculated refractive index with the empirical correlations: Equations (4), (5) and (7), which employ only density for the 136 crude oils whose characterization data is given in Table S3.

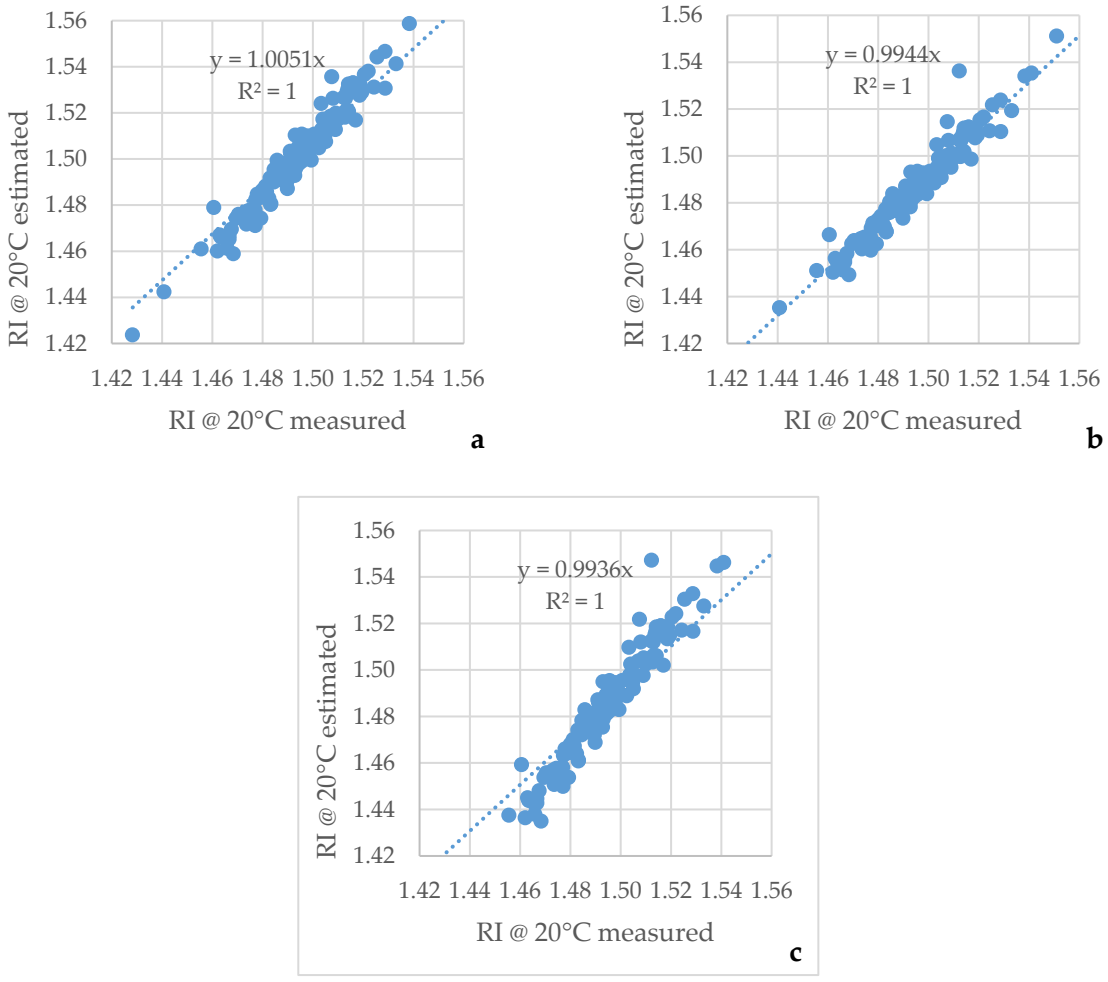

Figure 7. Parity graphs of measured versus calculated refractive index by the empirical correlations of Vargas and Chapman [14] (Equation (4); (**a**); Yarranton et al. [30] (Equation (5); (**b**); Stratiev et al. [31] (Equation (7); (**c**) for the 136 crude oils whose data is given in Table S1.

The data in Figure 7a–c deals with the second set of data of refractive index of crude oils and indicates, similarly to the first set of data of refractive index of petroleum fractions (Figure 1), that Equation (5) outperforms Equations (4) and (7), exhibiting R^2 = 0.9426 (Figure 7b) versus R^2 = 0.9164 (Figure 7a; Equation (4)) and R^2 = 0.8806 (Figure 7c; Equation (7)), respectively. With this data set, Equation (4) continues to show a positive bias (the slope of regression line in Figure 7a 1.0051 is higher than 1.00), while both Equations (5) and (7) exhibit a negative bias (the slope of regression line in Figure 7b,c is lower than 1.00).

The data in Table S3 was used to investigate the possibility of developing a new empirical correlation to demonstrate better prediction accuracy for the crude oil refractive index

from the specific gravity and molecular weight. Linear and nonlinear models employing CAS Maple and NLPSolve with Modified Newton Iterative Method were developed to obtain the best fit with the experimental refractive index.

No substantial difference in the accuracy of refractive index prediction between the linear and nonlinear models was observed. Slightly better results were observed with the nonlinear model, as shown with Equation (19).

$$RI_{crude\ oil} = -0.0557806 + EXP\Big(EXP\Big(-0.372624 + 0.283093 \times (-0.657058 + SG)^{0.766416} \times (-0.294140 + T_b)^{0.0651290} \quad (19)$$

In addition to the new empirical correlation (Equation (19)), a new ANN model was also developed using the data in Table S3. Using this data, the neural network has a 9:7:10:10:1 structure. This means nine input parameters (*SG*, *MW*, $T_{5\%}$, $T_{10\%}$, $T_{30\%}$, $T_{50\%}$, $T_{70\%}$, $T_{90\%}$, Kw), 10 neurons in the first neuron layer, 10 in the second hidden neuron layer, and one output. In this case, the Levemberg–Marquardt algorithm was used. Figure 8 shows the structure of the neural network with four layers.

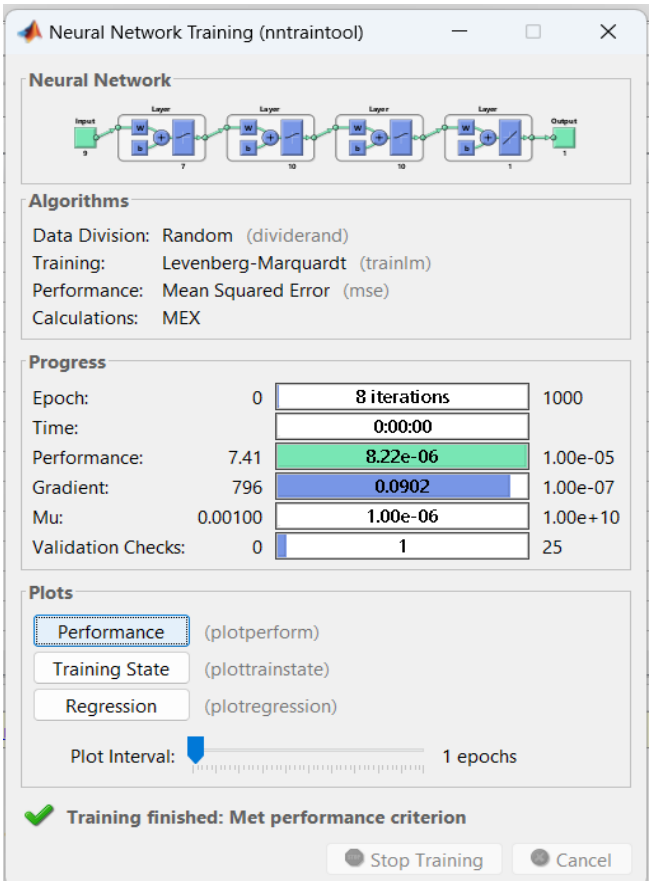

Figure 8. Structure of the neural network employing crude oil specific gravity, molecular weight, simulated TBP distillation data, and Kw-factor.

Figure 9 shows the training process for the neural network. It shows three graphs: training, testing, and verification. Training is usually done with 70% of the training sequence taken at random. Testing is done with 20% of the training sequence, also taken at random, and the remaining 10% of the training sequence is used for verification. The verification values are usually not used to train the neural network. Thus, they are considered independent. In our training, the neural network was trained for eight epochs, and the smallest mean squared error was obtained at epoch number seven and is 3.1344×10^{-5}.

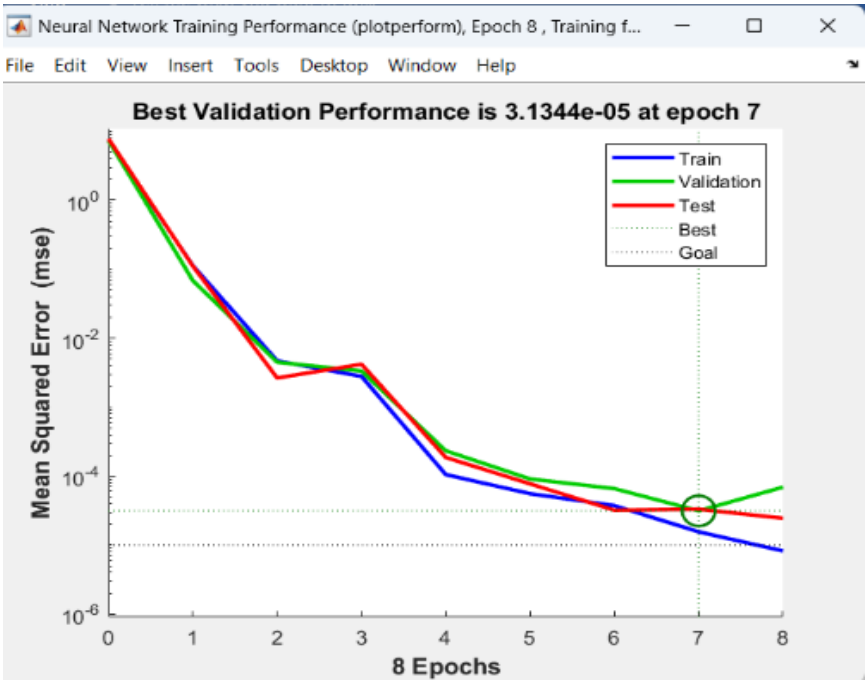

Figure 9. Training process of the neural network for prediction of crude oil viscosity from crude oil specific gravity, molecular weight, simulated TBP distillation data, and Kw-factor.

Figure 10 gives the basic parameters of the training.

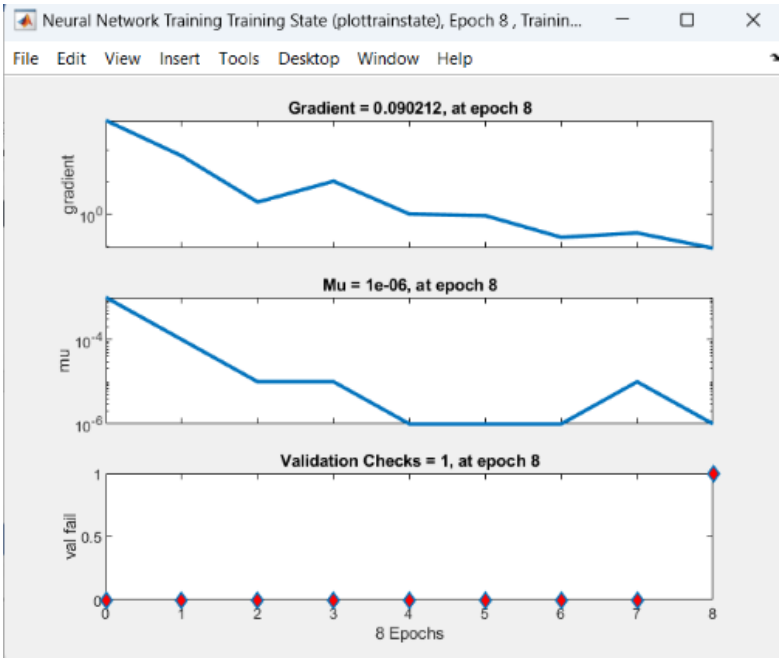

Figure 10. The basic parameters of the training: gradient for each epoch, Mu for each epoch, and the validation of each epoch.

In the actual use of the neural network, the R coefficients of each part of the training sequence are given. The individual coefficients are R = 0.98197 for training, R = 0.98155 for verification, R = 0.95613 for testing, and R = 0.97751 for the entire training process (see Figure 11).

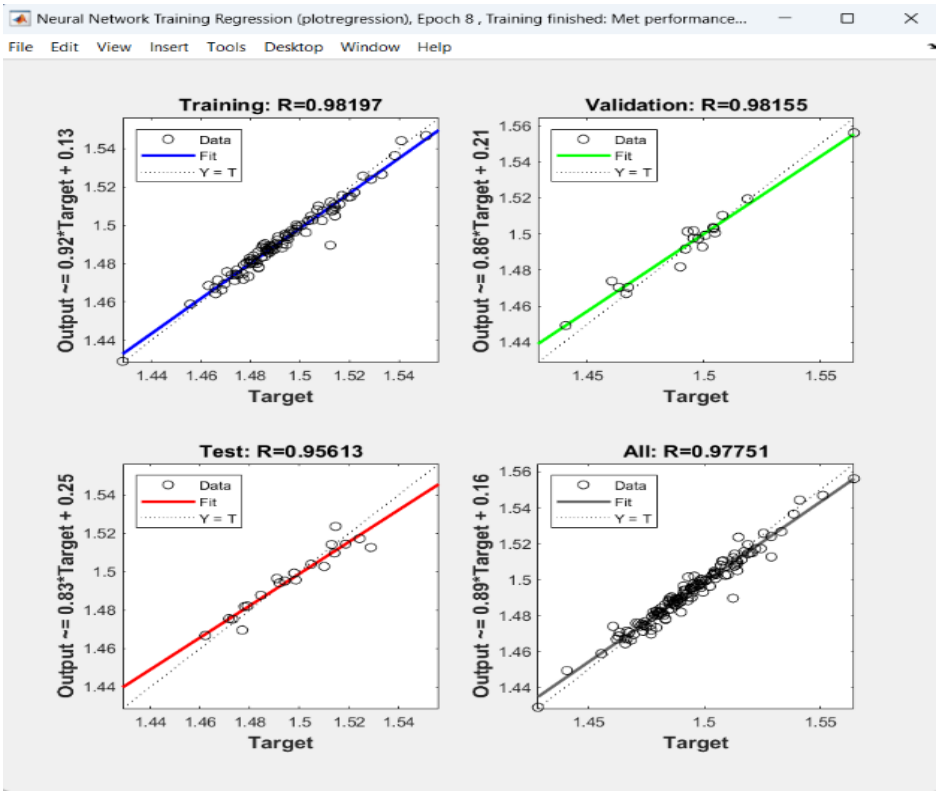

Figure 11. Regression coefficients of the learning process of the ANN model employing crude oil refractive index, specific gravity, molecular weight, simulated TBP distillation data, and Kw-factor.

Figure 12 shows the parity graph of the measured versus calculated refractive index of crude oil by the newly developed empirical correlation (Equation (19)), and by the ANN.

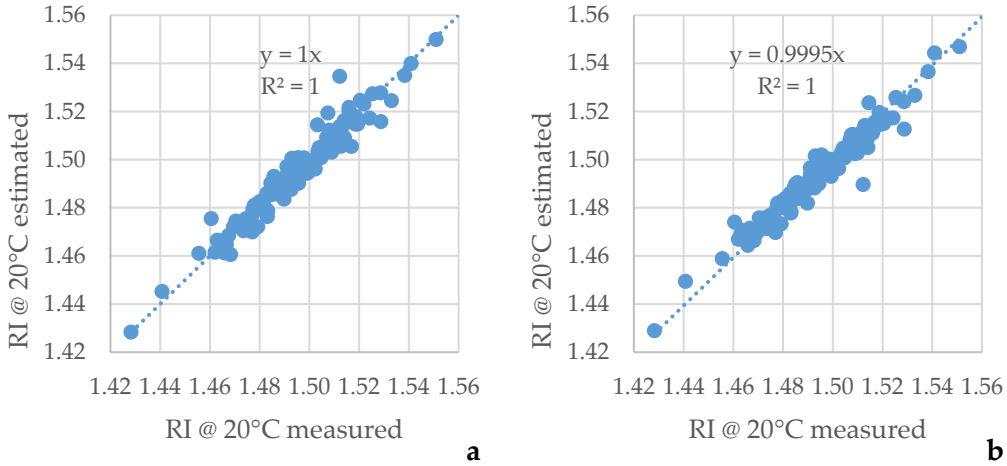

Figure 12. Parity graphs of the measured versus calculated refractive index of crude oils by the new empirical correlation (Equation (19); (**a**); and the ANN model (**b**).

The data in Figure 12 shows slightly better performance of the new empirical correlation (Equation (19); R^2 = 0.9456; Figure 12a) than the ANN model (R^2 = 0.9417; Figure 12a) with the slope of the regression line of Equation (19) (Figure 12a) exactly equal to 1.00, while that of the ANN (Figure 12b) is 0.9995.

Table 2 summarizes the statistical analyses for the new empirical correlation, the ANN model, and the three empirical correlations available from the existing literature that are applicable for the data from 136 crude oils (Table S3).

Table 2. Statistical analysis of studied methods to predict crude oil refractive index from specific gravity, and molecular weight for the data in Table S3.

RI Crude Oil Models		SE	RSE	SSE	%AAD	SRE	AAD	Max. dev.	Ranking
Vargas and Chapman, 2010 [14]	Equation (4)	0.000114	0.0076	0.0067	0.55	−68.58	0.0083	0.0490	4
Yarranton, et al., 2015 [30]	Equation (5)	0.000097	0.0065	0.0058	0.59	75.72	0.0089	0.0240	3
Stratiev, et al., 2019 [31]	Equation (7)	0.000188	0.0126	0.0115	0.76	88.93	0.0113	0.0388	5
New correlation (this work)	Equation (19)	0.000022	0.0014	0.0013	0.22	−0.22	0.0033	0.0222	2
ANN		0.000021	0.0014	0.0012	0.21	5.61	0.0032	0.0227	1

The data in Table 2 indicates that the ANN model and the new empirical correlation (Equation (19)) provide a more accurate prediction for the crude oil refractive index than the empirical models available in the literature (Equations (4), (5) and (7)). Both Equation (19) and the ANN model demonstrate almost the same accuracy of crude oil refractive index prediction, although the ANN model has a tiny advantage. The statistical parameters of the ANN model shown in Table 2 exhibit slightly lower values than those of the new empirical correlation (Equation (19)), suggesting a slightly better performance from the ANN model than Equation (19). This is opposite of what was observed in the data in Figure 12. Li and Heap [54], and Li [55] reported that assessment of a model's prediction accuracy should be conducted on the basis of the statistical parameters because the evaluation made on the basis of the coefficient of determination (R^2) may be biased, insufficient, or misleading. For that reason, in our assessment, we rely more on the statistical parameters than on the coefficient of determination.

3.3. Prediction of Refractive Index of Crude Oils by Empirical Correlations and ANN Using Specific Gravity, Molecular Weight, and SARA Composition Data

Figure 13 presents parity graphs of measured versus calculated refractive index by the empirical correlations. Equations (8) and (9) employ SARA composition data for the 102 crude oils; their characterization data is given in Table S4.

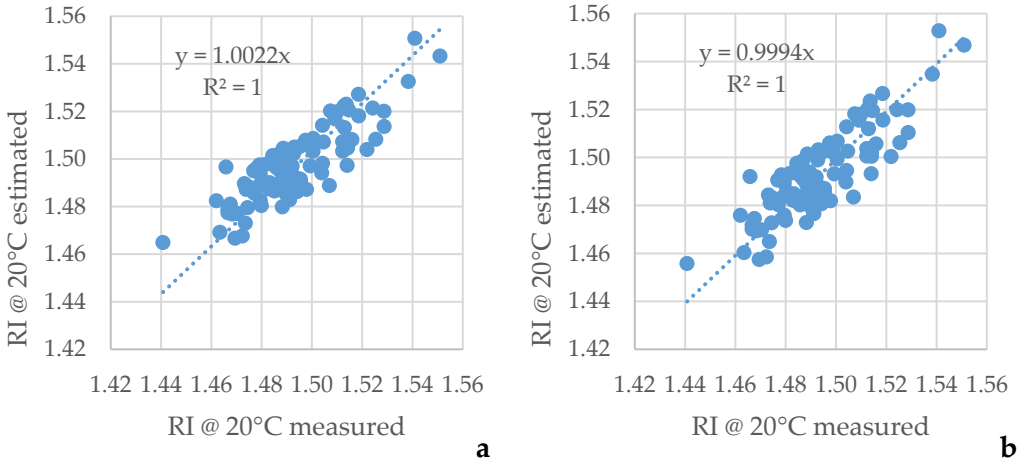

Figure 13. Parity graphs of measured versus calculated refractive index by the empirical correlations of Fan et al. [36] (Equation (8); (**a**) Chamkalani [37] (Equation (9)); (**b**) for the 102 crude oils whose data is given in Table S4.

The data in Figure 13 shows that the empirical correlation of Chamkalani [37] (Equation (9)) outperforms that of Fan et al. [36], but the coefficients of determination for both correlations are much lower than those of the studied empirical models discussed in previous sections and shown in Figures 1 and 7.

The data in Table S4 was also used to develop a new correlation. The data were processed using CAS Maple and NLPSolve with Modified Newton Iterative Method to

determine the specific gravity, molecular weight, and the contents of saturates, aromatics, resins, and asphaltenes of the 102 crude oils, and a new correlation, shown as Equation (20), was developed.

$$RI_{crude\ oil} = -0.375814 \times SG + 0.0000688846 \times MW + 0.0189790 \times Sat + 0.0193168 \times Aro + 0.0193293 \times Res + 0.0195185 \times As - 0.764517 \quad (20)$$

The data in Table S4 were processed using of Matlab, and a new ANN model was developed. For the data from Table S4, the neural network has a 6:7:10:10:1 structure. This means six input parameters (SG, MW, saturates, aromatics, resins, and asphaltenes) seven neurons in the first neuron layer, 10 in the second hidden neuron layer, and one output. In this case, the Levemberg–Marquardt algorithm was used. Figure 14 shows the structure of the neural network with four layers.

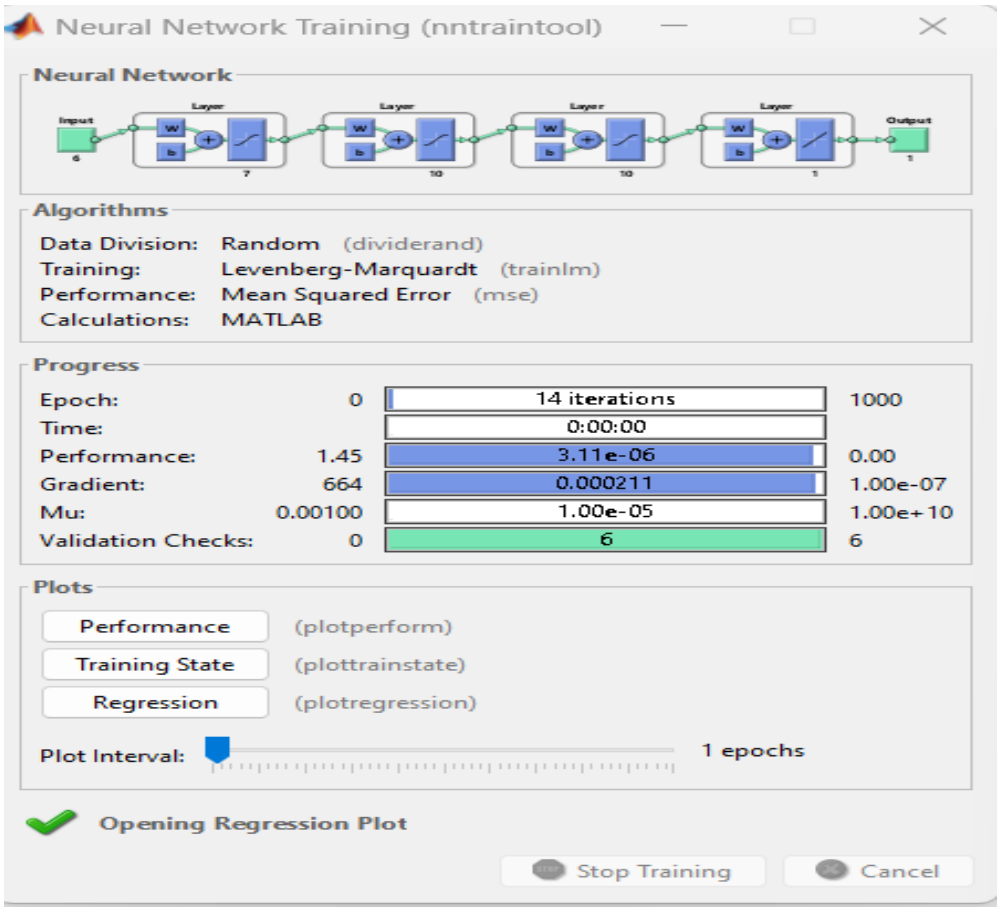

Figure 14. Structure of the neural network employing crude oil specific gravity, molecular weight, and SARA composition data.

Figure 15 shows the training process of the neural network. It shows three graphs: training, testing, and verification. Training is usually done with 70% of the training sequence taken at random. Testing is done with 20% of the training sequence, also taken at random, and the remaining 10% of the training sequence is used for verification. The verification values are usually not used to train the neural network. Thus, they are considered independent. In our training for this particular case, the neural network was trained for 14 epochs, and the smallest mean squared error was obtained at epoch number eight and is 2.3048×10^{-5}.

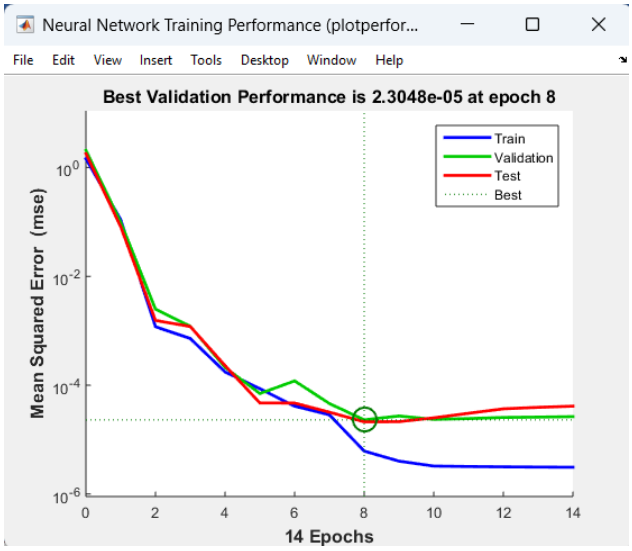

Figure 15. Training process of the neural network for prediction of crude oil viscosity from crude oil specific gravity, molecular weight, and SARA composition data.

In the actual use of the neural network, the R coefficients of each part of the training sequence are given. The individual coefficients are R = 0.99286 for training, R = 0.97194 for verification, R = 0.94342 for testing, and R = 0.98597 for the entire training process (see Figure 16).

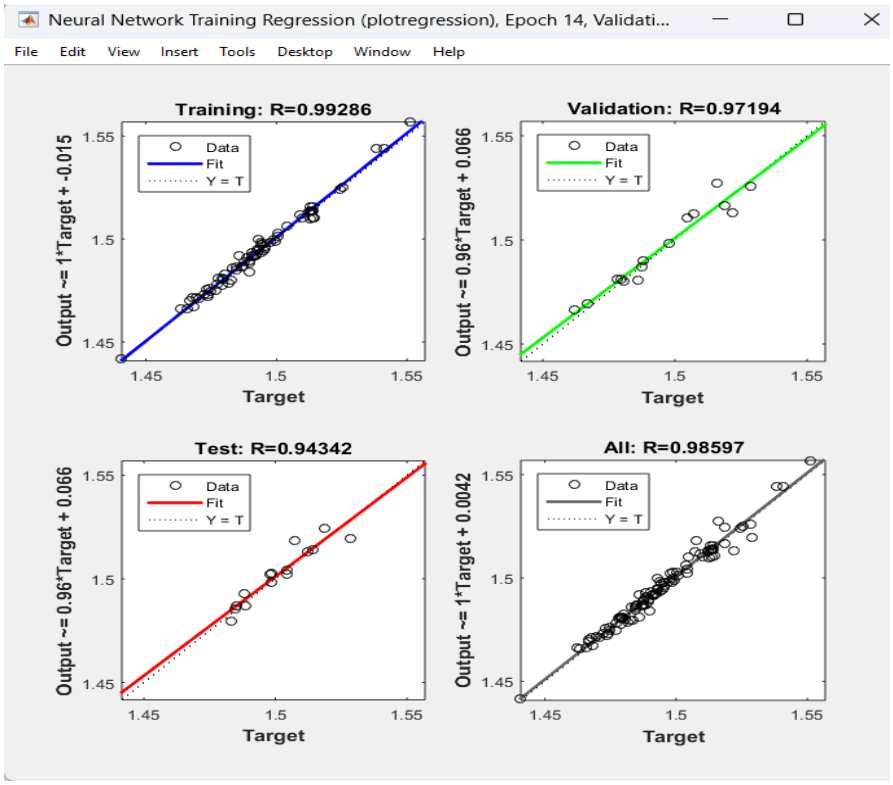

Figure 16. Regression coefficients of the learning process of the ANN model employing crude oil refractive index, specific gravity, molecular weight, and SARA composition data.

Figure 17 displays parity graphs of measured versus calculated refractive index by the new empirical correlation, shown as Equation (20), and the new ANN model employing the data in Table S4.

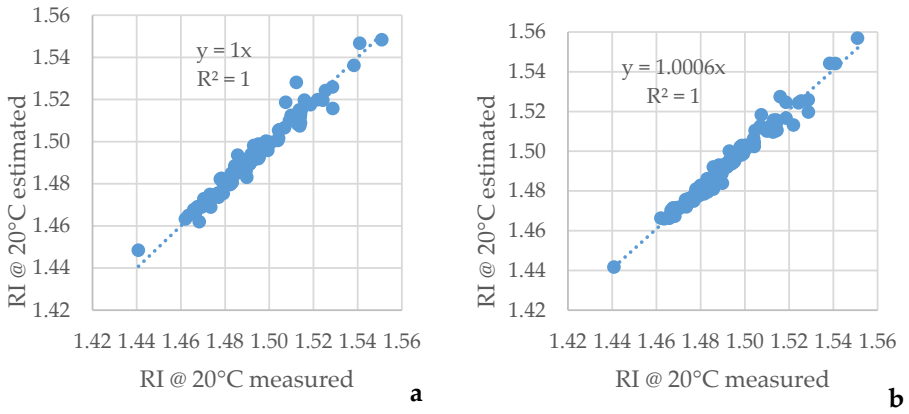

Figure 17. Parity graphs of measured versus calculated refractive index of crude oils by the new empirical correlation (Equation (20); (**a**); and the ANN model (**b**) utilizing crude oil data for SG, MW, and SARA composition.

The data in Figure 17a,b demonstrates much better performance of the new empirical correlation (Equation (20)) and the ANN model than the empirical correlations of Fan et al. [36], and Chamkalani [37] (see Figure 13).

Table 3 summarizes the statistical analyses for the new empirical correlation (Equation (20)), the ANN model, and the two studied empirical models available from the literature that are applicable for the data of 102 crude oils (Table S4).

Table 3. Statistical analysis of studied methods to predict crude oil refractive index for the data in Table S4.

RI Crude Oil Models		SE	RSE	SSE	%AAD	SRE	AAD	Max. dev.	Ranking
Fan, et al., 2002 [36]	Equation (8)	0.00011	0.00731	0.00494	0.584	−23.79	0.0087	0.0307	4
Chamkalaini, 2012 [37]	Equation (9)	0.00010	0.00670	0.00448	0.552	5.58	0.0083	0.0261	3
New correlation (this work)	Equation (20)	0.00001	0.00086	0.00057	0.163	−0.06	0.0024	0.0157	2
ANN		0.00001	0.00074	0.00049	0.156	−6.36	0.0023	0.0114	1

The data in Table 3 indicates that the ANN model outperforms the studied empirical correlations, and that the new empirical correlation demonstrates much better prediction accuracy for RIs than the correlations of Fan et al., [36], and Chamkalani [37].

4. Discussion

4.1. Prediction of Refractive Index of Petroleum Fractions, Individual Hydrocarbons, and Some Hetero Compounds by Empirical Correlations and ANN Using Specific Gravity (Density), and Boiling Point Data

The refractive index strongly correlates with density (specific gravity) as established in the studies of Vargas and Chapman [14], Yarranton et al. [30], and Stratiev et al. [31]. The data in Table S1 (254 petroleum fractions, individual hydrocarbons, and hetero compounds) also indicate a very strong correlation with the specific gravity (R = 0.984). However, as evident from the data in Table 1, the correlations predicting refractive index using only density (Equations (4), (5) and (7)) exhibit lower accuracy in their forecast. Therefore, the inclusion of the boiling point in the correlation leads to improvement in its refractive index prediction. For example, a linear regression of the data predicted by Yarranton et al. [30] (Equation (5)) and the boiling point for the 254 petroleum fractions, individual hydrocarbons, and some hetero compounds from Table S1, decreases the %AAD from 0.49% to 0.43%. This confirms that the inclusion of boiling point in the model to predict refractive index can improve the accuracy of prediction. Among the empirical correlations available in the literature are those from Riazi and Daubert [32]. (Equation (2)) demonstrates the best refractive index prediction, although it has been developed on the basis of light

petroleum fractions ($1.4188 \leq RI \leq 1.4776$). The new empirical correlation developed in this work (Equation (18)), which covers the range of RI between 1.3710 and 1.6538, shows the best prediction accuracy for RIs among all studied empirical correlations. However, the ANN model turned out to have even better predictions of the RI than the new empirical correlation. This finding is in line with other studies which report that the meta-heuristic methods (artificial intelligence models) offer higher prediction accuracy for petroleum fluid properties [51,53,56–58]. Thus, the ANN approach was found to be more suitable for modeling the refractive index of petroleum fractions than the regression methods.

4.2. Prediction of Refractive Index of Crude Oils by Empirical Correlations and ANN Using Specific Gravity and Molecular Weight Data

The prediction of crude oil refractive index from those available in the literature models was possible only by the use of Equations (4), (5) and (7) because, to the best of our knowledge, no correlations were reported in the literature that used both density (SG) and molecular weight. The data in Table 2 indicate that the inclusion of the molecular weight in the new empirical correlation (Equation (19)) has led to a substantial improvement in the prediction accuracy of the crude oil refractive index. For the data set of 136 crude oils from Table S3, the application of the ANN approach did not exhibit as much of an improvement in the prediction accuracy of the crude oil RI as that observed with the petroleum fraction RI prediction that used ANN (Table 1). This finding suggests that the metaheuristic methods may not always be capable of a better prediction of petroleum properties than the empirical correlations; this has also been reported in other studies [59,60].

4.3. Prediction Accuracy of the Refractive Index of Crude Oils by Empirical Correlations and ANN Using Specific Gravity, Molecular Weight, and SARA Composition Data

In several studies dedicated to the prediction of the refractive index of crude oils, SARA composition data was employed instead of specific gravity and molecular weight [36–42]. In order to evaluate the relation of specific gravity, molecular weight, and SARA composition data to the crude oil refractive index, an intercriteria analysis (ICrA) evaluation of the data in Table S4 was performed. Details about the application of ICrA in petroleum chemistry and processing and its meaning can be found in our recent studies [61,62]. $\mu = 0.75 \div 1.00$ and $\upsilon = 0 \div 0.25$ denote a statistically meaningful significant positive relation, where the strong positive consonance exhibits values of $\mu = 0.95 \div 1.00$ and $\upsilon = 0 \div 0.05$, and the weak positive consonance exhibits values of $\mu = 0.75 \div 0.85$ and $\upsilon = 0.2515 \div 0.1525$. The values of negative consonance with $\mu = 0.00 \div 0.25$ and $\upsilon = 0.75 \div 1.00$ indicate a statistically meaningful negative relation, where the strong negative consonance exhibits values of $\mu = 0.00 \div 0.05$ and $\upsilon = 0.95 \div 1.00$, and the weak negative consonance exhibits values of $\mu = 0.15 \div 0.25$ and $\upsilon = 0.75 \div 0.85$. All other cases are considered dissonance.

Tables 4 and 5 summarize the values of μ and υ for the ICrA evaluation of the data in Table S4.

Table 4. μ-Values of the ICrA evaluation of relations between refractive index, SG, MW, and SARA composition data for 102 crude oils.

M	SG	MW	Saturates	Aromatics	Resins	Asphaltenes	RI@20 °C
SG	1.000	0.853	0.234	0.516	0.766	0.704	0.932
MW	0.853	1.000	0.269	0.472	0.736	0.700	0.853
Saturates	0.234	0.269	1.000	0.360	0.203	0.233	0.211
Aromatics	0.516	0.472	0.360	1.000	0.489	0.481	0.529
Resins	0.766	0.736	0.203	0.489	1.000	0.701	0.772
Asphaltenes	0.704	0.700	0.233	0.481	0.701	1.000	0.726
RI@20 °C	0.932	0.853	0.211	0.529	0.772	0.726	1.000

Table 5. υ-Values of the ICrA evaluation of relations between refractive index, SG, MW, and SARA composition data for 102 crude oils.

υ	SG	MW	Saturates	Aromatics	Resins	Asphaltenes	RI@20 °C
SG	0.000	0.135	0.761	0.479	0.230	0.291	0.062
MW	0.135	0.000	0.723	0.519	0.256	0.290	0.137
Saturates	0.761	0.723	0.000	0.638	0.796	0.765	0.786
Aromatics	0.479	0.519	0.638	0.000	0.509	0.516	0.467
Resins	0.230	0.256	0.796	0.509	0.000	0.297	0.226
Asphaltenes	0.291	0.290	0.765	0.516	0.297	0.000	0.270
RI@20 °C	0.062	0.137	0.786	0.467	0.226	0.270	0.000

It is evident from the data in Tables 4 and 5 that the refractive index has stronger relations to SG and MW than to any other data from the SARA composition. This can explain why the new correlation (Equation (20)), which employs SG and MW data along with the SARA composition data, has a much higher prediction accuracy than the correlations of Fan et al. (Equation (8)) [36] and Chamkalani (Equation (9)) [37] (see Table 3). The application of the ANN approach for the data from 102 crude oils (Table S4) again demonstrates a more accurate prediction of crude oil RI than that of all studied empirical correlations. Concerning the meaning of SARA composition data for the prediction accuracy of the crude oil refractive index, a comparison of regression coefficients shown in Figures 11 and 16 indicates higher values of correlation coefficient of validation, and test for the ANN model that employs only *SG* and *MW* (Figure 11) ($R_{validation}$ = 0.98155; R_{test} = 0.95613) than those of the ANN model that employ *SG* and MW along with SARA composition data (Figure 16 ($R_{validation}$ = 0.97194; R_{test} = 0.9432). This comparison confirms the stronger influence of SG and MW than the SARA composition data on the prediction accuracy of the crude oil refractive index.

Another comparison between the prediction accuracy of the ANN model developed in this work, which employs crude oil SG, MW, and the saturates, aromatics, resins, and asphaltenes contents with the metaheuristic models developed by Gholami et al. [30] (support vector regression (SVR) combined with hybrid grid and pattern search (HGP), genetic algorithm (GA), and imperialist competitive algorithm (ICA), expressed by the regression coefficients of the validation shows $R_{validation}$ = 0.97194 of our ANN model versus $R_{validation}$ = 0.9356 (HGP–SVR); 0.9587 (GA–SVR); and 0.9635 (ICA–SVR) of Gholami et al. models. It should be noted here that both our ANN model and the SVR models of Gholami et al. [42] have used the same data based on the open-source literature of Buckley and Morrow [44]. This comparison confirms again that the inclusion of SG and MW in the meta-heuristic model to predict the crude oil refractive index can improve the accuracy of the forecast.

5. Conclusions

By investigating three data sets, i.e., (1) SG, and Tb of 254 petroleum fractions, individual hydrocarbons, and hetero compounds; (2) SG, and MW of 136 crude oils; and (3) SG, MW, and SARA composition data of 102 crude oils, with empirical correlations available in the literature and three newly developed empirical correlations as well as ANN models for the three data sets, the following conclusions can be made:

1. SG is the petroleum fluid property that has the strongest relation to the RI.
2. Predictive methods that use only SG have a lower prediction accuracy for the RI than the methods which use boiling point or molecular weight along with SG.
3. Two of the three ANN models developed in this work exhibit better prediction accuracy for petroleum fluid RIs.
4. The crude oil properties of SG and MW are more informative than the SARA composition data and can provide higher prediction accuracy for both empirical and ANN model RIs.

5. In general, the ANN provides higher prediction accuracy for petroleum fluid RIs than the empirical correlations. However, this could not always be the case, as was reported in other studies dedicated to other petroleum fluid properties. Thus, additional investigations are needed to uncover further improvement opportunities for accurate petroleum fluid property predictions by the use of metaheuristic methods, which seem to have a higher potential than the empirical correlations.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/pr11082328/s1, Table S1: A total of 254 data points for boiling point, specific gravity, and refractive index of petroleum fluids, individual hydrocarbons, and some hetero compounds to be used for regression modeling, and testing of the published empirical correlations to predict refractive index of petroleum fluids [14,30,32,33,45–51]; Table S2: A total of 254 data points for boiling point, specific gravity, Kw-factor, and TBP distillation characteristics (generated by the use of the method of Hosseinifar and Shahverdi, 2022 [52] and refractive index of petroleum fluids, individual hydrocarbons, and some hetero compounds to be used for ANN modeling; Table S3: A total of 136 data points for specific gravity, molecular weight, Kw-characterization factor, TBP distillation characteristics (generated by the use of the method of Hosseinifar and Shahverdi, 2022 [52], and refractive index of crude oils to be used for refractive index prediction by empirical correlations, and ANN [44]; Table S4: A total of 102 data points for specific gravity, molecular weight, SARA composition data, and refractive index of crude oils to be used for refractive index prediction by empirical correlations, and ANN.

Author Contributions: Conceptualization, D.D.S. and I.S.; methodology, G.N.P.; software, S.N., S.S., S.R. and D.S.; validation, E.S., D.Y. and D.P.; formal analysis, R.D.; investigation, G.N.P.; resources, K.A.; data curation, D.S.; writing—original draft preparation, D.D.S. and I.S.; writing—review and editing, D.D.S. and I.S.; supervision, K.A.; project administration, S.S.; funding acquisition, S.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Asen Zlatarov University, Burgas, under the project Center of Excellence UNITE BG05M2OP001-1.001-0004/28.02.2018 (2018–2023).

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Nomenclature

AAD	Average absolute deviation
ANN	Artificial neural network
Aro	Aromatic content of oil, wt.%
Asp	Asphaltene content of oil, wt.%
BP	Boiling point
CAS	Computer algebra system
CMIS	Committee machine intelligent system
d_{15}	Density at 15 °C, g/cm^3
E	Error
FRI	Fraction refractive index
ICrA	Intercriteria analysis
Kw	Watson characterization factor
LSSVM	Least squares support vector machine
Max. dev	Maximum deviation
MW	Molecular weight, g/mol
MLP	Multilayer perceptron
R	Regression coefficient
RBF	Radial basis function
RI	Refractive index
RSE	Relative standard error
Res	Resin content of oil, wt.%
SARA	Saturates, aromatics, resins, asphaltenes
Sat	Saturate content of oil, wt.%

SG	Specific gravity
SE	Standard error
SRE	Sum of relative errors
SSE	Sum of square errors
$T_{5\%}$	Boiling point of evaporate at 5%, °C
$T_{10\%}$	Boiling point of evaporate at 10%, °C
$T_{30\%}$	Boiling point of evaporate at 30%, °C
T_{50}	Boiling point of evaporate at 50%, °C
$T_{70\%}$	Boiling point of evaporate at 70%, °C
$T_{90\%}$	Boiling point of evaporate at 90%, °C
T_b	Average boiling point of petroleum fluid
TBP	True boiling point
%AAD	Relative average absolute deviation
μ	Positive consonance
υ	Negative consonance
ρ	Density at 20 °C, g/cm^3

References

1. Riazi, M.R.; Roomi, Y.A. Use of the Refractive Index in the Estimation of Thermophysical Properties of Hydrocarbons and Petroleum Mixtures. *Ind. Eng. Chem. Res.* **2001**, *40*, 1975–1984. [CrossRef]
2. Khan, M.R. Correlation Between Refractive Indices and Other Fuel-Related Physical/Chemical Properties of Pyrolysis Liquids Derived from Coal, Oil Shale, and Tar Sand. In Proceedings of the Low Temperature Pyrolysis, Toronto, ON, Canada, 5–10 June 1988.
3. Lai, Q.; Xie, Y.; Wang, M.; Wang, C.; Sun, K.; Tan, J. The complex refractive index of crude oils determined by the combined Brewster–transmission method. *Infrared Phys. Technol.* **2020**, *111*, 103515. [CrossRef]
4. Nikolaev, V.F.; Zalaltdinova, N.D.; Vyachkileva, I.O.; Fakhrutdinov, R.Z.; Abakumova, O.O.; Sulaiman, B. Mapping technique for oil refining processes and products. *Fuel* **2022**, *307*, 121870. [CrossRef]
5. Abutaqiya, M.I.L.; AlHammadi, A.A.; Sisco, C.J.; Vargas, F.M. Aromatic Ring Index (ARI): A Characterization Factor for Nonpolar Hydrocarbons from Molecular Weight and Refractive Index. *Energy Fuels* **2021**, *35*, 1113–1119. [CrossRef]
6. Wang, F.; Evangelista, R.F.; Threatt, T.J.; Tavakkoli, M.; Vargas, F.M. Determination of Volumetric Properties Using Refractive Index Measurements for Nonpolar Hydrocarbons and Crude Oils. *Ind. Eng. Chem. Res.* **2017**, *56*, 3107–3115. [CrossRef]
7. AlHammadi, A.A.; Abutaqiya, M.I.L. Predictive Modeling of Phase Behavior of Reservoir Fluids under Miscible Gas Injection Using the Peng–Robinson Equation of State and the Aromatic Ring Index. *ACS Omega* **2023**, *8*, 3270–3277. [CrossRef]
8. Evangelista, R.F.; Vargas, F.M. Prediction of the Temperature Dependence of Densities and Vapor Pressures of Nonpolar Hydrocarbons Based on Their Molecular Structure and Refractive Index Data at 20 °C. *Fluid. Phase Equilibria* **2018**, *468*, 29–37. [CrossRef]
9. Sisco, C.J.; Abutaqiya, M.I.L.; Saade, H.S.; Vargas, F.M. Predicting Phase Behavior of Nonpolar Mixtures from Refractive Index at Ambient Conditions. *Fuel* **2019**, *237*, 637–647. [CrossRef]
10. Zhang, J.; Khemka, Y.; Vargas, F.M. Predicting Viscosity of Nonpolar Mixtures as a Function of Temperature and Pressure from Volumetric and Optical Measurements Performed at Ambient Conditions. *Fluid. Phase Equilibria* **2019**, *487*, 16–23. [CrossRef]
11. Riazi, M.R.; Al-Otaibi, G.N. Estimation of Viscosity of Liquid Hydrocarbon Systems. *Fuel* **2001**, *80*, 27–32. [CrossRef]
12. Riazi, M.R.; Enezi, G.A.; Soleimani, S. Estimation of Transport Properties of Liquids. *Chem. Eng. Commun.* **1999**, *176*, 175–193. [CrossRef]
13. Touba, H.; Mansoori, G.A.; Sarem, A.M.S. New Analytic Techniques for Petroleum Fluid Characterization Using Molar Refraction. In Proceedings of the SPE Western Regional Meeting, Long Beach, CA, USA, 25–27 June 1997.
14. Vargas, F.M.; Chapman, W.G. Application of the One-Third Rule in Hydrocarbon and Crude Oil Systems. *Fluid. Phase Equilib.* **2010**, *290*, 103–108. [CrossRef]
15. Buckley, J.S. Predicting the Onset of Asphaltene Precipitation from Refractive Index Measurements. *Energy Fuels* **1999**, *13*, 328–332. [CrossRef]
16. Castillo, J.; Gutierrez, H.; Ranaudo, M.; Villarroel, O. Measurement of the Refractive Index of Crude Oil and Asphaltene Solutions: Onset Flocculation Determination. *Energy Fuels* **2010**, *24*, 492–495. [CrossRef]
17. Soleymanzadeh, A.; Yousefi, M.; Kord, S.; Mohammadzadeh, O. A Review on Methods of Determining Onset of Asphaltene Precipitation. *J. Pet. Explor. Prod. Technol.* **2019**, *9*, 1375–1396. [CrossRef]
18. Wattana, P.; Wojciechowski, D.J.; Bolaños, G.; Fogler, H. Study of Asphaltene Precipitation Using Refractive Index Measurement. *Pet. Sci. Technol.* **2003**, *21*, 591–613. [CrossRef]
19. Gholami, A.; Asoodeh, M.; Bagheripour, P. Smart Determination of Difference Index for Asphaltene Stability Evaluation. *J. Dispers. Sci. Technol.* **2014**, *35*, 572–576. [CrossRef]
20. Khaleel, A.T.; Sisco, C.J.; Tavakkol, M.; Vargas, F.M. An Investigation of the Effect of Asphaltene Polydispersity on Asphaltene Precipitation and Deposition Tendencies. *Energy Fuels* **2022**, *36*, 8799–8808. [CrossRef]

21. Kuang, J.; Yarbrough, J.; Enayat, S.; Edward, N.; Wang, J.; Vargas, F.M. Evaluation of Solvents for In-situ Asphaltene Deposition Remediation. *Fuel* **2019**, *241*, 1076–1084. [CrossRef]
22. Wang, J.X.; Buckley, J.S. A Two-Component Solubility Model of the Onset of Asphaltene Flocculation in Crude Oils. *Energy Fuels* **2001**, *15*, 1004–1012. [CrossRef]
23. Buckley, J.S.; Wang, J. Crude oil and Asphaltene Characterization for Prediction of Wetting Alteration. *J. Pet. Sci. Eng.* **2002**, *33*, 195–202. [CrossRef]
24. Respini, M.; Salla, G.D.; Medine, G.M.; Sandu, C.L.; Pinnapu, S.R. Process for Predicting the Stability of Crude Oil and Employing Same in Transporting and/or Refining the Crude Oil. US13,924,089, 21 June 2013.
25. Taylor, S.D.; Czarnecki, J.; Masliyah, J. Refractive Index Measurements of Diluted Bitumen Solutions. *Fuel* **2001**, *80*, 2013–2018. [CrossRef]
26. El Ghandoor, H.; Hegazi, E.; Nasser, I.; Behery, G.M. Measuring the Refractive Index of Crude Oil Using a Capillary Tube Interferometer. *Opt. Laser Technol.* **2003**, *35*, 361–367. [CrossRef]
27. Evdokimov, I.; Losev, A. Effects of Molecular De-Aggregation on Refractive Indices of Petroleum-Based Fluids. *Fuel* **2007**, *86*, 2439–2445. [CrossRef]
28. Zamora, V.; Zhang, Z.; Meldrum, A. Refractometric Sensing of Heavy Oils in Fluorescent Core Microcapillaries. *Oil Gas. Sci. Technol. Rev. IFP Energ. Nouv.* **2015**, *70*, 487–495. [CrossRef]
29. Castillo, J.; Canelon, C.; Acevedo, S.; Carrier, H.; Daridon, J.-L. Optical Fiber Extrinsic Refractometer to Measure RI of Samples in a High Pressure and Temperature Systems: Application to Wax and Asphaltene Precipitation Measurements. *Fuel* **2006**, *85*, 2220–2228. [CrossRef]
30. Yarranton, H.W.; Okafor, J.C.; Ortiz, D.P.; van den Berg, F.G.A. Density and Refractive Index of Petroleum, Cuts, and Mixtures. *Energy Fuels* **2015**, *29*, 5723–5736. [CrossRef]
31. Stratiev, D.; Shishkova, I.; Tankov, I.; Pavlova, A. Challenges in Characterization of Residual Oils. A Review. *J. Pet. Since Eng.* **2019**, *178*, 227–250. [CrossRef]
32. Riazi, M.R.; Daubert, T.E. Improved Characterization of Wide Boiling Range Undefined Petroleum Fractions. *Ind. Eng. Chem. Res.* **1987**, *26*, 629–632. [CrossRef]
33. Stratiev, D.S.; Marinov, I.M.; Shishkova, I.K.; Dinkov, R.K.; Stratiev, D.D. Investigation on Feasibility to Predict the Content of Saturate Plus Mono-Nuclear Aromatic Hydrocarbons in Vacuum Gas Oils from Bulk Properties and Empirical Correlations. *Fuel* **2014**, *129*, 156–162. [CrossRef]
34. Hosseinifar, P.; Shahvertdi, H. Development of a Generalized Model for Predicting the Composition of Homologous Groups Derived from Molecular Type Analyses to Characterize Petroleum Fractions. *J. Pet. Sci. Eng.* **2021**, *204*, 108744. [CrossRef]
35. Dhulesia, H. New Correlations Predict FCC Feed Characterizing Parameters. *Oil Gas. J.* **1986**, *84*, 51–54.
36. Fan, T.; Wang, J.; Buckley, J.S. Evaluating Crude Oils by SARA Analysis. In Proceedings of the SPE/DOE Improved Oil Recovery Symposium, Tulsa, OK, USA, 17 April 2002.
37. Chamkalani, A. Correlations Between SARA Fractions, Density and RI to Investigate the Stability of Asphaltene. *Int. Sch. Res. Not.* **2012**, *2012*, 219276. [CrossRef]
38. Chamkalani, A.; Chamkalani, R.; Mohammadi, A.H. Hybrid of Two Heuristic Optimizations with LSSVM to Predict Refractive Index as Asphaltene Stability Identifier. *J. Dispers. Sci. Technol.* **2014**, *35*, 1041–1050. [CrossRef]
39. Gholami, A.; Asoodeh, M.; Bagheripour, P. Fuzzy Assessment of Asphaltene Stability in Crude Oils. *J. Dispers. Sci. Technol.* **2014**, *35*, 556–563. [CrossRef]
40. Zargar, G.; Gholami, A.; Asoodeh, M.; Bagheripour, P.; Vaezzadeh-Asadi, M. PSO-Fuzzy Eliminates Deficiency of Neuro-Fuzzy in Assessment of Asphaltene Stability. *Indian. J. Chem. Technol.* **2015**, *22*, 35–140.
41. Tatar, A.; Shokrollahi, A.; Halali, M.A.; Azari, V.; Safari, H. A Hybrid Intelligent Computational Scheme for Determination of Refractive Index of Crude Oil Using SARA Fraction Analysis. *Can. J. Chem. Eng.* **2015**, *93*, 1547–1555. [CrossRef]
42. Gholami, A.; Ansari, H.R.; Hosseini, S. Prediction of Crude Oil Refractive Index Through Optimized Support Vector Regression: A Competition Between Optimization Techniques. *J. Pet. Explor. Prod. Technol.* **2017**, *7*, 195–204. [CrossRef]
43. Stratiev, D.; Nenov, S.; Nedanovski, D.; Shishkova, I.; Dinkov, R.; Stratiev, D.D.; Stratiev, D.D.; Sotirov, S.; Sotirova, E.; Atanassova, V.; et al. Different Nonlinear Regression Techniques and Sensitivity Analysis as Tools to Optimize Oil Viscosity Modeling. *Resources* **2021**, *10*, 99. [CrossRef]
44. Buckley, J.S.; Morrow, N.R. *Wettability and Imbibition: Microscopic Distribution of Wetting and Its Consequences at the Core and Field Scales*; New Mexico Petroleum Recovery Research Center: Socorro, NM, USA, 2003.
45. Riazi, M.R.; Daubert, T.E. Prediction of Molecular-Type Analysis of Petroleum Fractions and Coal Liquids. *Ind. Eng. Chem. Res.* **1986**, *25*, 1009–1015. [CrossRef]
46. White, C.M.; Perry, M.B.; Schmidt, C.E.; Douglas, L.J. Relationship between Refractive Indices and Other Properties of Coal Hydrogenation Distillates. *Energy Fuels* **1987**, *1*, 99–105. [CrossRef]
47. Stratiev, D.S. Novelties in Thermal and Catalytic Processes at Production of Modern Fuels. Ph.D. Thesis, Burgas University "Assen Zlatarov", Burgas, Bulgaria, 2010.
48. Ancheyta-Juarez, J.; Rodrıguez-Salomon, S.S.; Valenzuela-Zapata, M.A. Experimental Evaluation of Vacuum Gas Oil-Light Cycle Oil Blends as FCC Feedstock. *Energy Fuels* **2001**, *15*, 675–679. [CrossRef]

49. Lappas, A.A.; Iatridis, D.K.; Vasalos, I.A. Production of Reformulated Gasoline in the FCC Unit. Effect of Feedstock Type on Gasoline Composition. *Catal. Today* **1999**, *50*, 73–85. [CrossRef]
50. Bollas, G.M.; Vasalos, I.A.; Lappas, A.A.; Iatridis, D.K.; Tsioni, G.K. Bulk Molecular Characterization Approach for the Simulation of FCC Feedstocks. *Ind. Eng. Chem. Res.* **2004**, *43*, 3270–3281. [CrossRef]
51. Stratiev, D.; Sotirov, S.; Sotirova, E.; Nenov, S.; Dinkov, R.; Shishkova, I.; Kolev, I.V.; Yordanov, D.; Vasilev, S.; Atanassov, K.; et al. Prediction of Molecular Weight of Petroleum Fluids by Empirical Correlations and Artificial Neuron Networks. *Processes* **2023**, *11*, 426. [CrossRef]
52. Hosseinifar, P.; Shahverdi, H. Prediction of the ASTM and TBP Distillation Curves and Specific Gravity Distribution Curve for Fuels and Petroleum Fluids. *Can. J. Chem. Eng.* **2022**, *100*, 3288–3310. [CrossRef]
53. Stratiev, D.; Shishkova, I.; Dinkov, R.; Nenov, S.; Sotirov, S.; Sotirova, E.; Kolev, I.; Ivanov, V.; Ribagin, S.; Atanassov, K.; et al. Prediction of Petroleum Viscosity from Molecular Weight and Density. *Fuel* **2023**, *331*, 125679. [CrossRef]
54. Li, J.; Heap, A. A Review of Spatial Interpolation Methods for Environmental Scientists; Geoscience Australia: Canberra, Australia. 2008. Available online: https://eva.fing.edu.uy/pluginfile.php/142980/mod_resource/content/3/A%20Review%20of%20Spatial%20Interpolation%20Methods%20for%20Environmental%20Scientists%20%28Jin%20Li%2C%20Andrew%20Heap%29.pdf (accessed on 28 July 2023).
55. Li, J. Assessing the accuracy of predictive models for numerical data: Not r nor r2, why not? Then what? *PLoS ONE* **2017**, *12*, e0183250. [CrossRef]
56. Hadavimoghaddam, F.; Ostadhassan, M.; Heidaryan, E.; Sadri, M.A.; Chapanova, I.; Popov, E.; Cheremisin, A.; Rafieepour, S. Prediction of Dead Oil Viscosity: Machine Learning vs. Classical Correlations. *Energies* **2021**, *14*, 930. [CrossRef]
57. Bahonar, E.; Chahardowli, M.; Ghalenoei, Y.; Simjoo, M. New Correlations to Predict Oil Viscosity Using Data Mining Techniques. *J. Pet. Sci. Eng.* **2022**, *208*, 109736. [CrossRef]
58. Ssebadduka, R.; Le, N.N.H.; Nguele, R.; Alade, O.; Sugai, Y. Artificial Neural Network Model Prediction of Bitumen/Light Oil Mixture Viscosity under Reservoir Temperature and Pressure Conditions as a Superior Alternative to Empirical Models. *Energies* **2021**, *14*, 8520. [CrossRef]
59. Sinha, U.; Dindoruk, B.; Soliman, M. Machine Learning Augmented Dead Oil Viscosity Model for All Oil Types. *J. Pet. Sci. Eng.* **2020**, *195*, 107603. [CrossRef]
60. Stratiev, D.; Marinov, I.; Dinkov, R.; Shishkova, I.; Velkov, I.; Sharafutdinov, I.; Nenov, S.; Tsvetkov, T.; Sotirov, S.; Mitkova, M.; et al. Opportunity to Improve Diesel Fuel Cetane Number Prediction from Easy Available Physical Properties and Application of the Least Squares Method and the Artificial Neural Networks. *Energy Fuels* **2015**, *29*, 1520–1533. [CrossRef]
61. Stratiev, D.; Shishkova, I.; Dinkov, R.; Kolev, I.; Argirov, G.; Ivanov, V.; Ribagin, S.; Atanassova, V.; Atanassov, K.; Stratiev, D.; et al. Intercriteria Analysis to Diagnose the Reasons for Increased Fouling in a Commercial Ebullated Bed Vacuum Residue Hydrocracker. *ACS Omega* **2022**, *7*, 30462–30476. [CrossRef]
62. Adams, J.J.; Rovani, J.F.; Planche, J.-P.; Loveridge, J.; Literati, A.; Shishkova, I.; Palichev, G.; Kolev, I.; Atanassov, K.; Nenov, S.; et al. SAR-AD Method to Characterize Eight SARA Fractions in Various Vacuum Residues and Follow Their Transformations Occurring during Hydrocracking and Pyrolysis. *Processes* **2023**, *11*, 1220. [CrossRef]