Available online: http://journal.uir.ac.id/index.php/JEEE/index

Journal of Earth Energy Engineering

Publisher: Universitas Islam Riau (UIR) Press

Oil Formation Volume Factor Prediction Using Artificial Neural Network: A Case Study of Niger Delta Crudes

Chiebuka M. Okoro[1]*, Angela N. Nwachukwu[1]
[1]Department of Petroleum Engineering, School of Engineering and Engineering Technology, Federal University of Technology, Owerri, P.M.B 1526, Owerri, Imo State, Nigeria
*Corresponding Author: chiebuka.martins.okoro@gmail.com

Article History:

Received: June 16, 2021
Receive in Revised Form: July 19, 2023
Accepted: July 19, 2023

Keywords:

Oil Formation Volume Factor, Artificial intelligence, Machine Learning, Artificial Neural Network, Principal Component Analysis, Niger Delta

Abstract

Artificial intelligence techniques provide an alternative to conventional empirical correlation methods when experimentally determined oil formation volume factors (OFVF) are lacking. A new mathematical model is proposed using an artificial neural network (ANN) for estimating the OFVF for the Niger Delta crude oils. The method consists of two stages: data decorrelation through principal component analysis (PCA) and OFVF estimation through ANN. Data decorrelation was used to reduce redundancy in the data which decreased the number of neurons in the hidden layer needed for an ANN to achieve high accuracy. In the development of the model, 316 data points were obtained from the Niger Delta region of Nigeria. Application of data cleaning, outliers' elimination and PCA analysis reduced the data to 243 points. 213 data points were used to develop the model of which 75% was used for training, 15% for validation and 10% for testing. The remaining 30 data points were used to test the predictive capability of the proposed model. The results obtained were compared with widely accepted empirical correlations of Standing, Glaso, Vazquez, Ikiensikimama & Ajienka, and Al-Marhoun. The proposed new model performed better than all of them in terms of coefficient of correlation, AAPE and RMSE. Hence the ANN model will reduce cost, save time, and also predict the OFVF of Niger Delta crudes with higher precision.

INTRODUCTION

The importance of PVT properties in reservoir engineering calculation and management cannot be overemphasized. These properties which include dew point pressure, bubble point pressure, isothermal compressibility factor, formation volume factor (FVF), viscosity, etc. have a role to play in material balancing, reserve estimation, reservoir simulation, surface, and subsurface facility design, Enhanced Oil Recovery (EOR) process, ultimate recovery, etc. (Baziar & Shahripour, 2015)

PVT data can be gotten directly or indirectly (Baziar & Shahripour, 2015). The ideal source of reliable PVT data is the direct method which involves carrying out laboratory measurements on sampled reservoir fluids (either surface or subsurface) using PVT cells (Baziar & Shahripour, 2015; Osman, Abdel-Wahhab, & Al-Marhoun, 2001). Nevertheless, PVT data might not be readily available because of cost, time, and difficulty involved in the process of sampling and carrying out laboratory measurements (Osman, Abdel-Wahhab, & Al-Marhoun, 2001; Abooali & Khamehchi, 2016). These disadvantages of the direct method led to the birth of the indirect method which is cheaper and faster than the direct method. The indirect method involves the use of developed empirical correlation and Equation of State (EOS) to estimate PVT properties.

There are numerous empirical correlations for estimating PVT properties, most of which were developed using a certain range of measured readily available data from various sources. However, these correlations have not lived up to the expectation in estimating the reservoir fluid properties over a wide range of conditions. This deficiency could be due to the fluid composition complexity, various fluid characteristics in different regions, and incomplete information. Nevertheless, these correlations have proved useful in

DOI: 10.25299/jeee.2022.7121
eISSN: 2540-9352 pISSN: 2301-8097

predicting PVT properties where experimental data are not available or where the need for validation of laboratory results is required (Osman, Abdel-Wahhab, & Al-Marhoun, 2001; Alimadadi, Fakhri, Alimadadi, & Dezfoulian, 2011; Ramirez, Valle, Romero, & Jaimes, 2017). One important PVT parameter that is very important to industry practitioners is the Oil Formation Volume Factor (OFVF). OFVF is the ratio of oil volume at reservoir condition to the same oil volume at standard condition (i.e. 14.7 psi and 60°F). If OFVF is measured at bubble point pressure, it is called bubble point OFVF. It gives an idea of the volume of oil that will be produced in the stock tanks. Accurate estimation of OFVF is very important in reserve estimation, material balance calculations and many other reservoir and production engineering calculations.

OFVF may vary from region to region and there is the need for each region to develop correlations for OFVF that is suitable for their region. Taking regional characteristics variation into account, PVT empirical correlations need to be modified before being applied (Baziar & Shahripour, 2015). The EOS is largely dependent on the knowledge of the fluid composition which is both expensive and time-consuming to obtain and also involves complex mathematical calculations (Kamari, 2016; Oloso, 2018; Ramirez, Valle, Romero, & Jaimes, 2017). Most of these empirical correlations and EOS have their limitations and range of applicability. The existing corrections are usually accurate within their range of applicability and may require corrections to achieve accurate prediction of OFVF for other regions due to variations in fluid compositions. Standing was the first to propose a correlation for determining OFVF of a gas saturated oil and used a total of 105 experimentally determined data points on 22 different crude oil mixtures from California fields. Glaso presented another correlation for OFVF with 41 data points mostly from North Sea Region. Al-Marhoun also presented an OFVF correlation using a total of 160 data points acquired from Middle East Region. Crude oil samples from different regions have been known to vary due to regional characteristics arising from varying amounts of paraffinic oil components and also varying amounts of nonhydrocarbons particularly CO_2, N_2 and H_2S. None of these correlations were developed using Niger Delta crude oil samples and may not give accurate predictions of OFVF for the region's crude oil mixtures due to these regional variations.

In recent times, Artificial Intelligence (AI) techniques have been able to bridge this gap and have been applied to the petroleum industry with good results. AI have helped in utilizing large data that have been collected over time to improve oil and gas operations. This has also led to the saying "data is the new oil". Several AI techniques have been used to estimate PVT properties and have helped to overcome the challenges of both the direct and indirect method of getting PVT data as well as giving accurate results when compared to them.

There are many AI techniques available but, in this study, only Artificial Neural Network (ANN) was used. ANN is a form of Machine Learning (ML) that mimics the biological neuron of the human brain. The advantage of ANN over the conventional correlations is that neural networks have larger degrees of freedom for fitting parameters which enables them to capture the systems' non-linearity than regression techniques. Another advantage of ANN is that it could be trained further and refined when additional data becomes available potentially improving their prediction accuracy. ANN has been used in developing a more accurate PVT properties model (Ramirez, Valle, Romero, & Jaimes, 2017), for other regions and can also be used for the Niger Delta region. In the Niger Delta, few correlations have been presented using ANN such as Ikiensikimama & Ajienka**.** However, improvements to existing techniques are always needed.

METHOD

Data Acquisition

The PVT data used in this study were obtained from various fields in the Niger Delta data region. The obtained data comprises reservoir temperature (Tr), gas specific gravity (GG), stock tank oil gravity (API), solution gas-oil ratio (Rs), and the saturated OFVF.

Data Processing

Before processing the data, 316 data points were available but after cleaning (i.e. removing rows with zero value and rows with incomplete information), it was reduced to 256 data points. Outliers are a statistical sample that does not fit a pattern that describes most other data points; specifically, a value that lies 1.5 times the Inter Quartile Range (IQR) beyond the upper or lower quartile. The outliers were identified using Microsoft Excel by:

1. Computing for the lower quartile and upper quartile using 1.5 times the IQR
2. Using the 'OR' function to set a limit for the individual data between the upper and lower quartile.
3. Each data turns in a result of FALSE or TRUE. If TRUE for any of the data in a row, the row is deleted.

Journal of Earth Energy Engineering
Vol. 11 No. 3, December 2022, pp 125-136

After carrying out the above procedure, the data was reduced to 247 data points. Below is a table presenting briefly the statistical information about the processed data:

Table 1. Statistical Data Description for the Niger Delta Crude Used in Training and Verification of the Model

Data Type	PVT Properties	Maximum	Minimum	Mean
INPUT	Temperature (°F)	219.739	126.1665	166.2433
	Gas Specific Gravity	1.018657	0.560134	0.701442
	Stock Tank Oil Gravity (°API)	59.78139	18.30432	28.96114
	Solution Gas-Oil Ratio (SCF/STB)	1325.071	11.57112	471.276
OUTPUT	Oil Formation Volume Factor (BBL/STB)	1.923	1.04	1.223798

Development of OFVF model for Niger Delta crude

The model to be developed is divided into two steps: Linear decorrelation of the processed data using PCA and Nonlinear regression using ANN. The structure of the proposed model is shown in Figure 1 below.

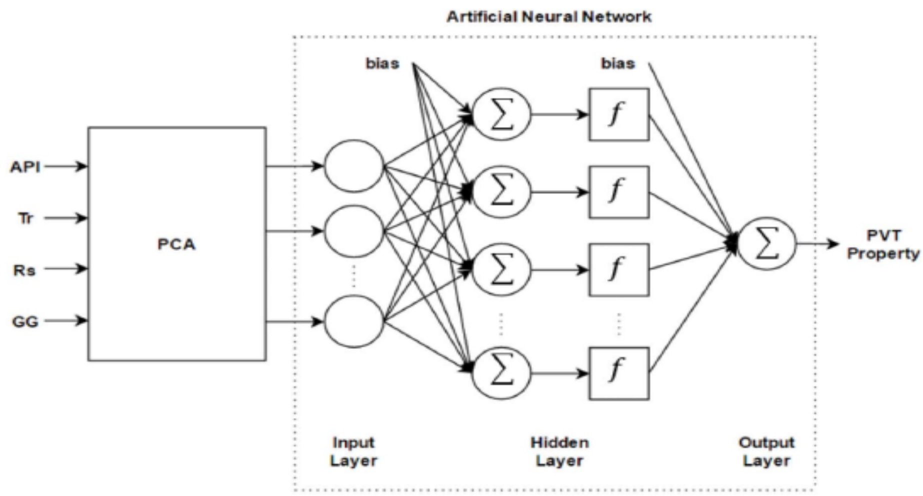

Figure 1. Structure of the proposed prediction model (Ramirez, Valle, Romero, & Jaimes, 2017)

Principal Component Analysis

Principal Component Analysis (PCA) was carried out on the 247 data points using Numerical Analysis for Excel (NumXL). NumXL is an econometric and time series analysis add-in for Microsoft Excel which provides a wide variety of statistical and time series analysis techniques, including linear and nonlinear time series modeling, statistical tests, and others. The PCA falls under the factor analysis function category of NumXL. The input data was transformed through PCA into four PCs. The PCA function further reduced the data to 245 data points and linearly reduced redundancy in the data.

Neural Network Architecture

The ANN was developed using MATrix LABoratory software popularly known as MATLAB. MATLAB neural network tool was used to develop the network using a feedforward backward propagation algorithm with Levenberg-Marquardt transfer function which obtains the weights and biases for optimization. Tan-sigmoidal type activation function was used as a transfer function between the input and hidden layer and the linear type activation function was used between the hidden and output layer. A backpropagation neural network algorithm was implemented to model OFVF.The proposed ANN model was based on the specific gravity of gas (GG), the solution gas to oil ratio (Rs), the oil specific gravity (API), and the temperature of the reservoir (Tr) which served as the input layer. The model was built using one hidden layer and one output parameter (i.e. OFVF). Seven neurons were used in developing the model because it gave the best coefficient of correlation after many variations in the number of neurons in the hidden layer. Therefore, the neural network (NN) has an architectural structure of 4:7:1.

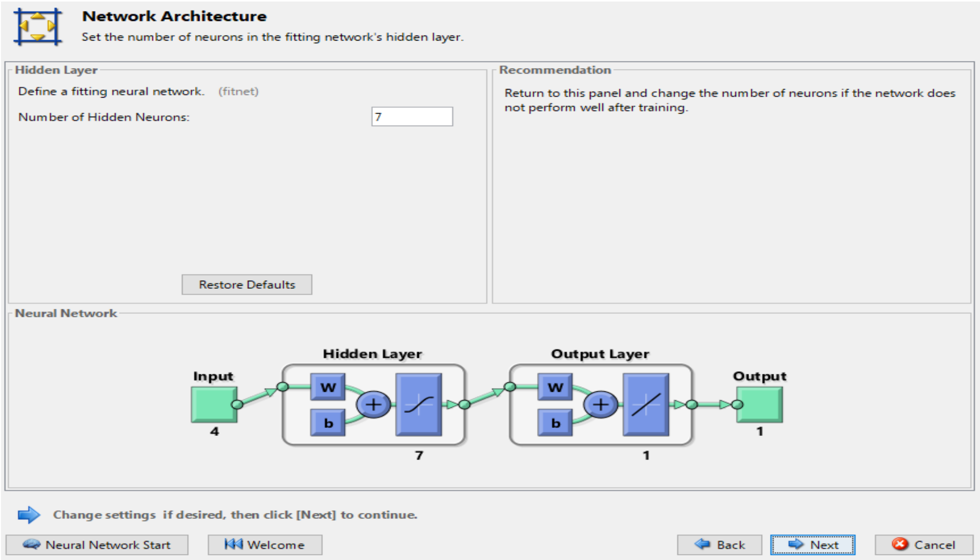

Figure 2. Neural Network Architecture of the developed model

Before developing the model, the processed data was checked and two rows which were having negative solution GOR were removed. Thereafter, the data was divided into two: Learning Data (213 data points) and Verification data (30 data points). During the training, the learning data was used as the ANN database was divided into training, validation, and testing i.e. 75%, 15%, and 10% of the total input parameter respectively. The software provides a better means of monitoring the performance of the ANN database (i.e. training, validation, and testing) simultaneously. After the training, the verification dataset which was not seen by the NN during training was used to test the validity of the model.

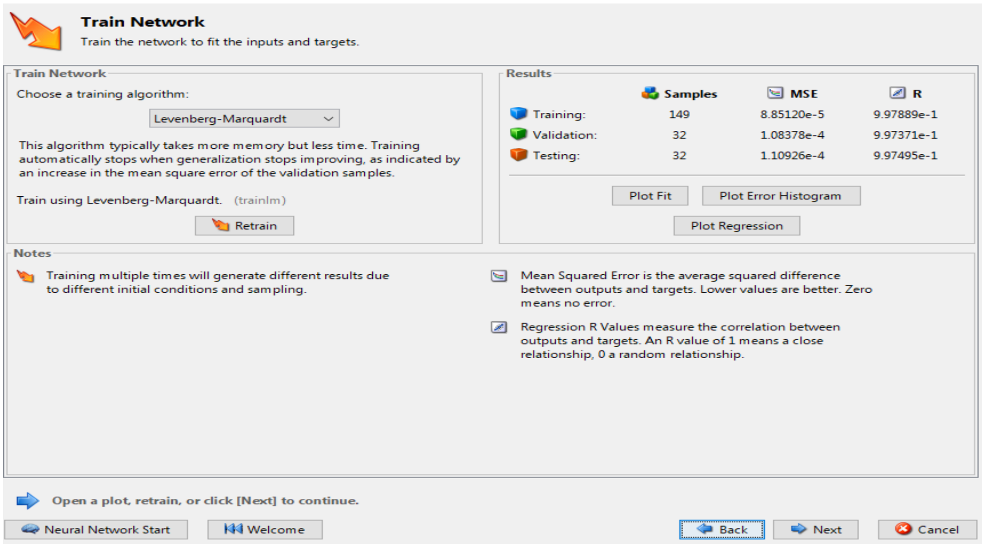

Figure 3. A MATLAB interface showing the division of the ANN database.

Journal of Earth Energy Engineering
Vol. 11 No. 3, December 2022, pp 125-136

RESULTS AND DISCUSSION

Results

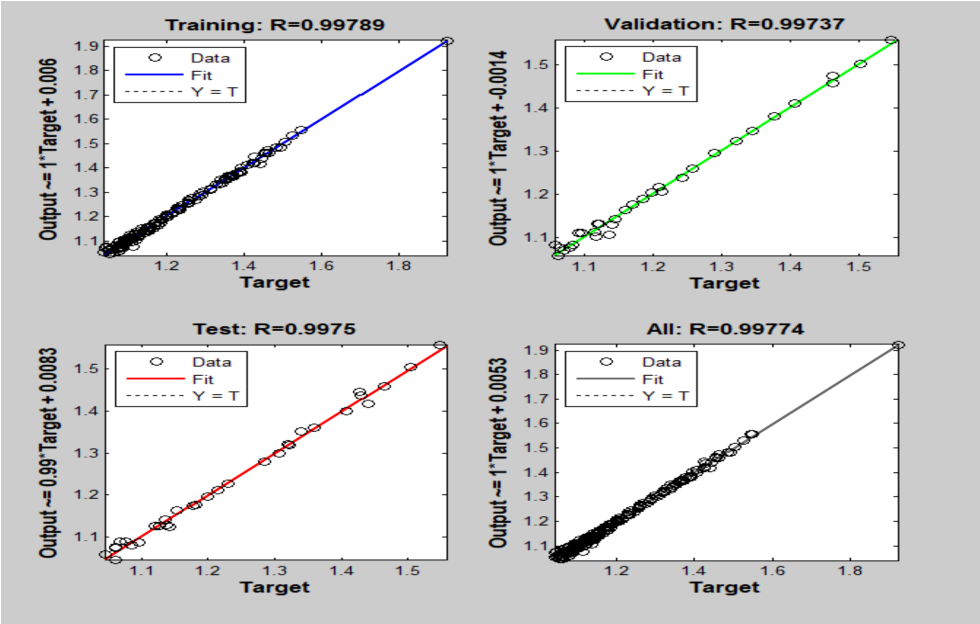

Figure 4. ANN database registration plot

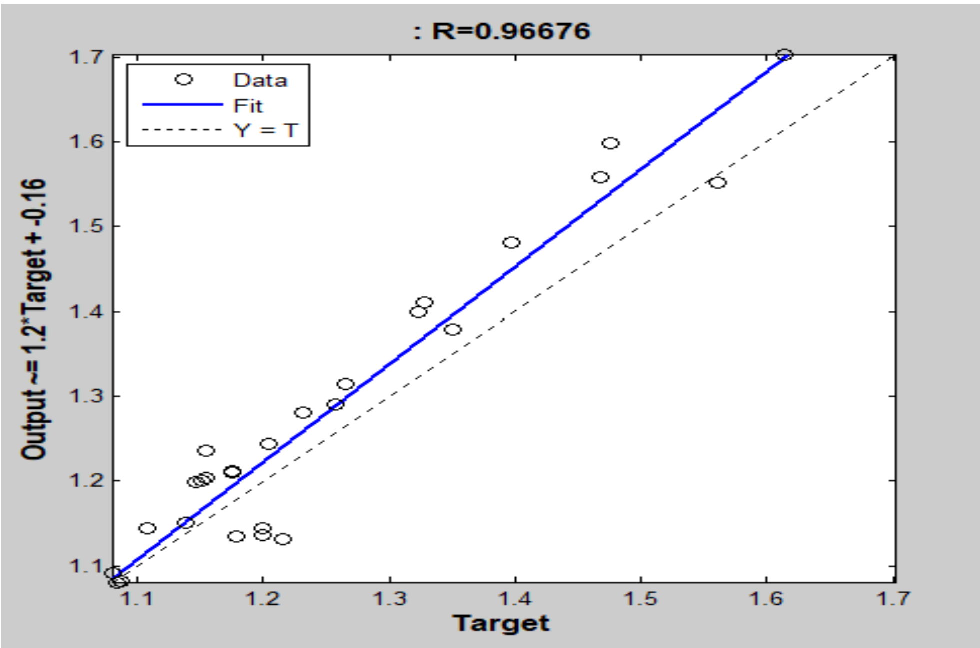

Figure 5. Verification dataset regression plot

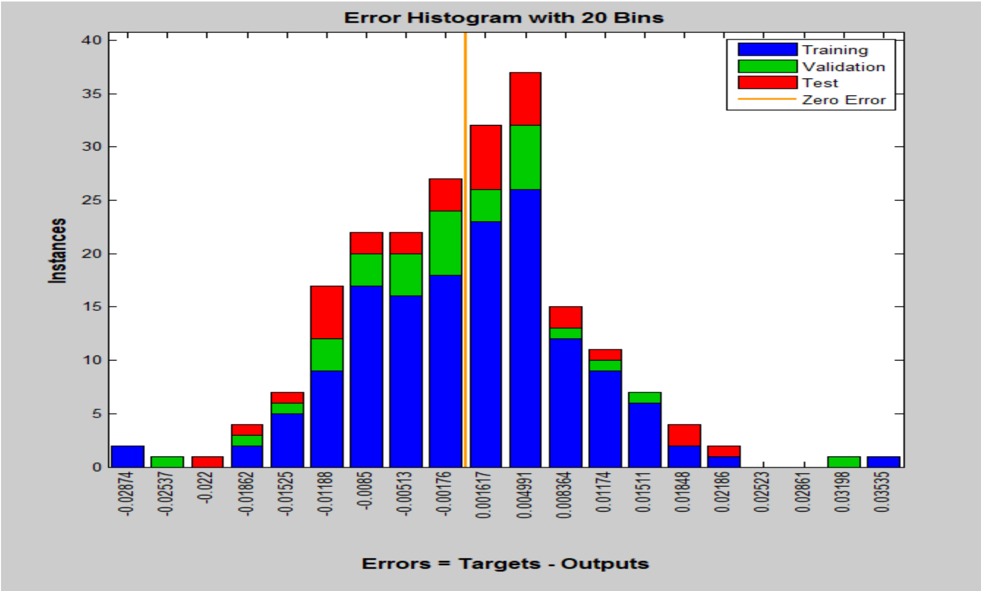

Figure 6. Error Histogram of the Training dataset

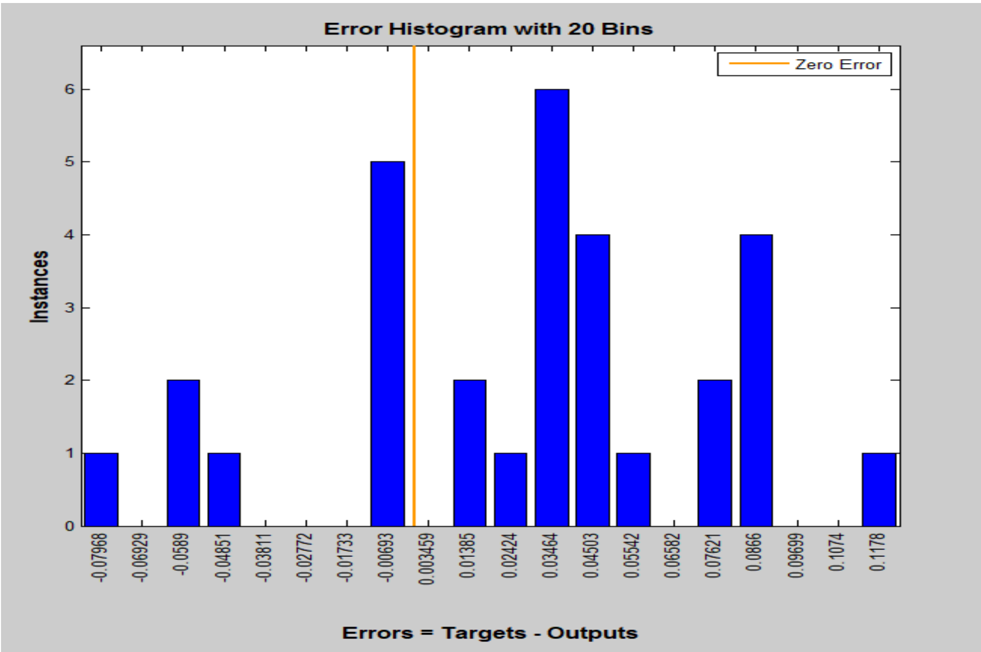

Figure 7. Error Histogram for verification datasets

Journal of Earth Energy Engineering
Vol. 11 No. 3, December 2022, pp 125-136

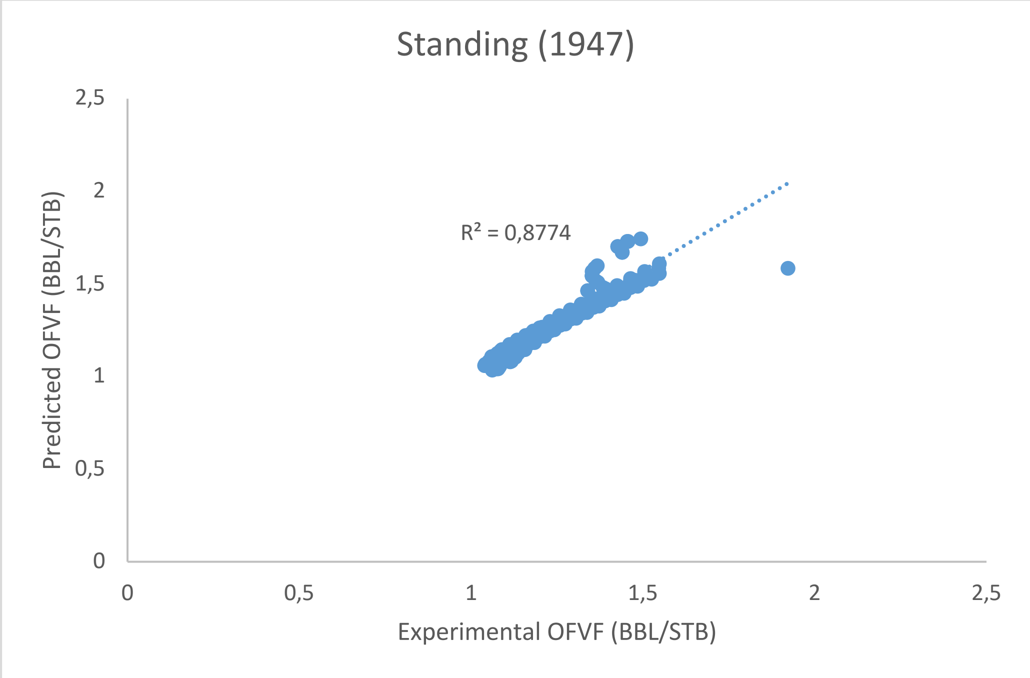

Figure 8. Cross-plot of OFVF correlation (Standing, 1947)

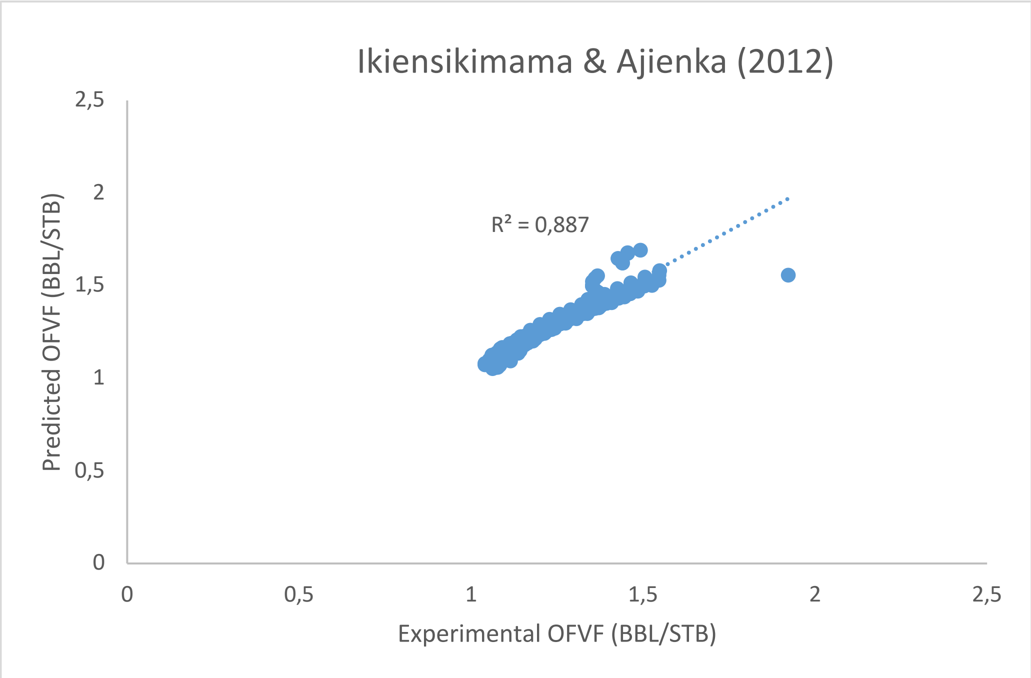

Figure 9. Cross-plot of OFVF correlation (Ikiensikimma & Ajienka, 2012)

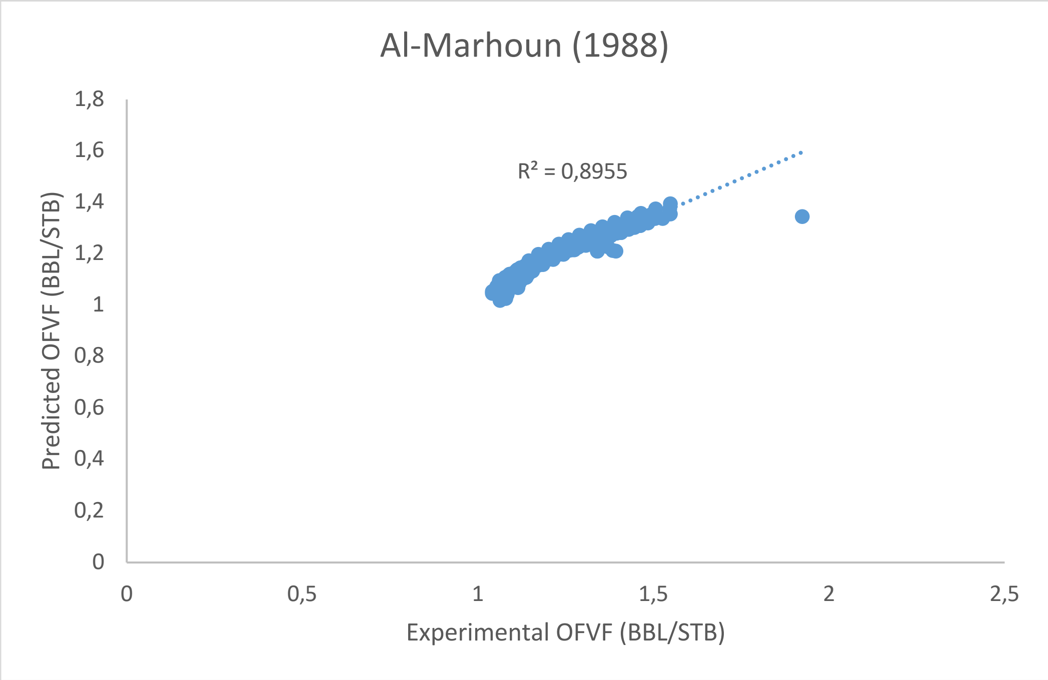

Figure 10. Cross-plot of OFVF correlation (Al-Marhoun, 1988)

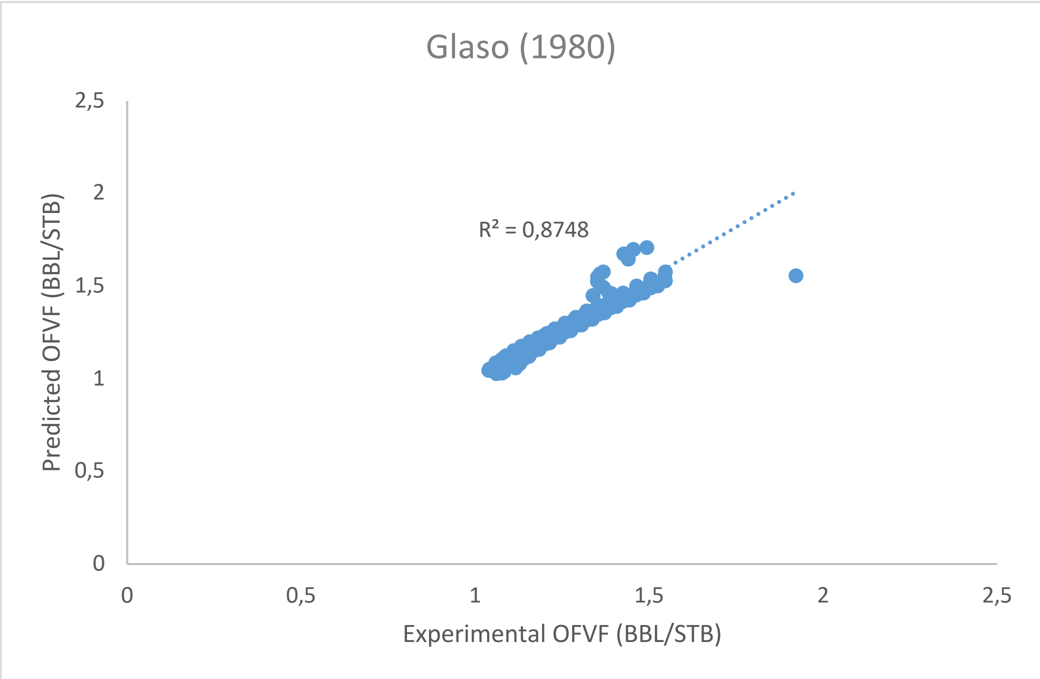

Figure 11. Cross-plot of OFVF correlation (Glaso, 1980)

Journal of Earth Energy Engineering
Vol. 11 No. 3, December 2022, pp 125-136

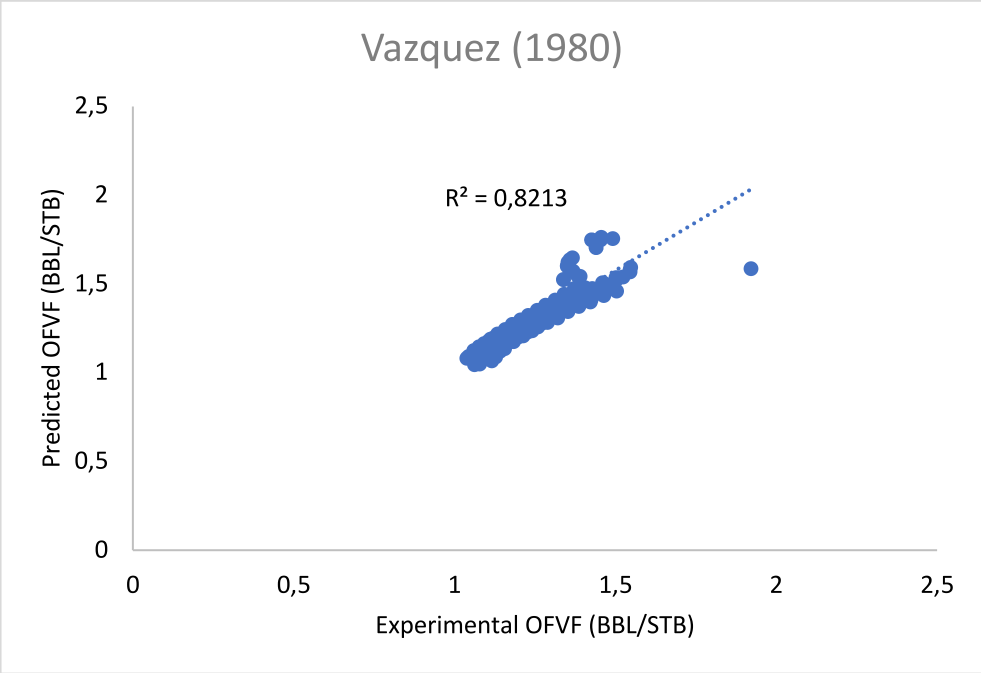

Figure 12. Cross-plot of OFVF correlation (Vazquez, 1980)

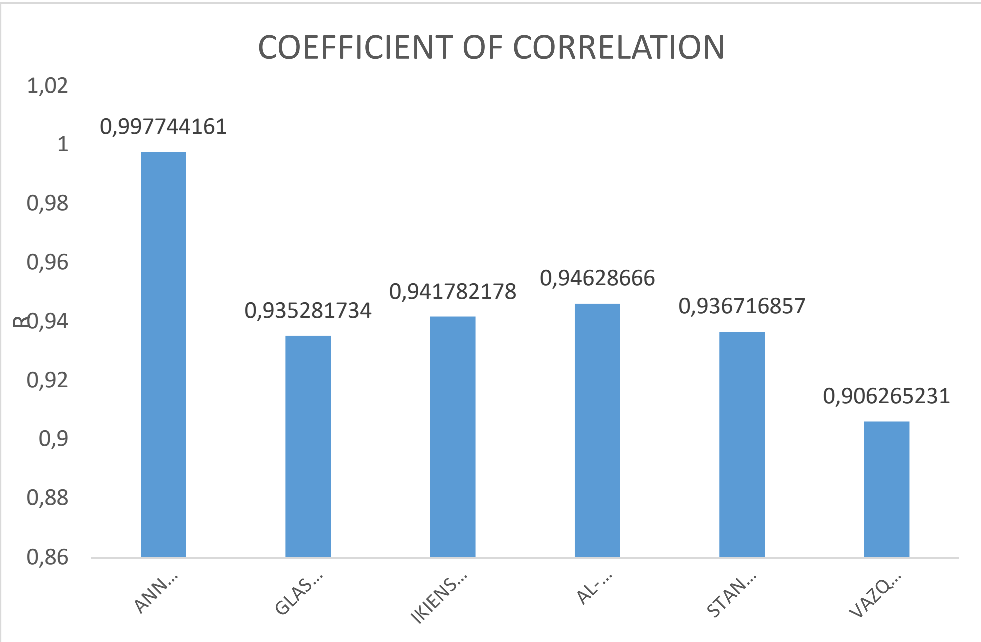

Figure 13. Coefficient of correlation comparison for different correlations

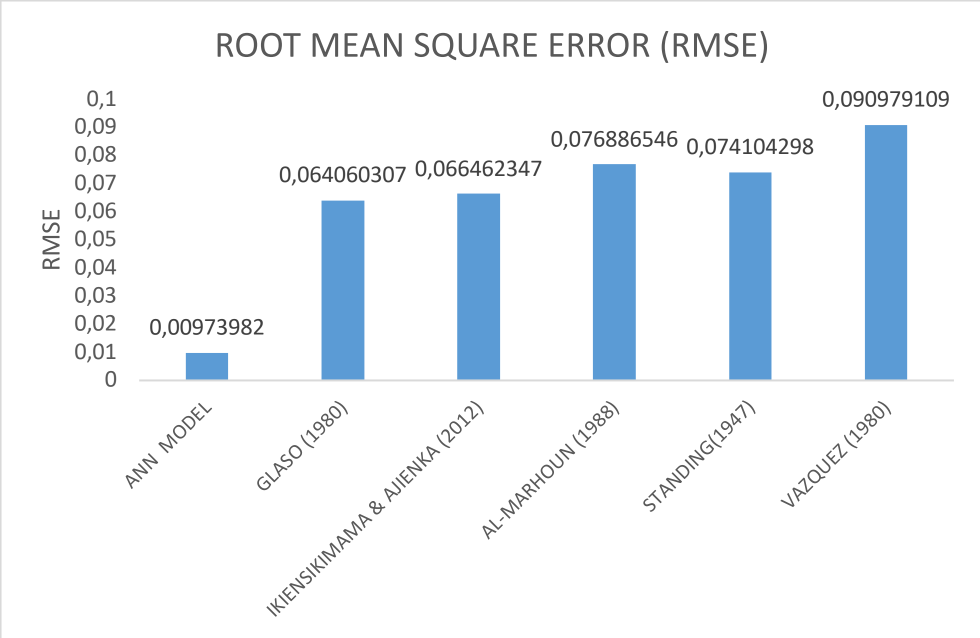

Figure 14. Root means square error comparison for different correlations.

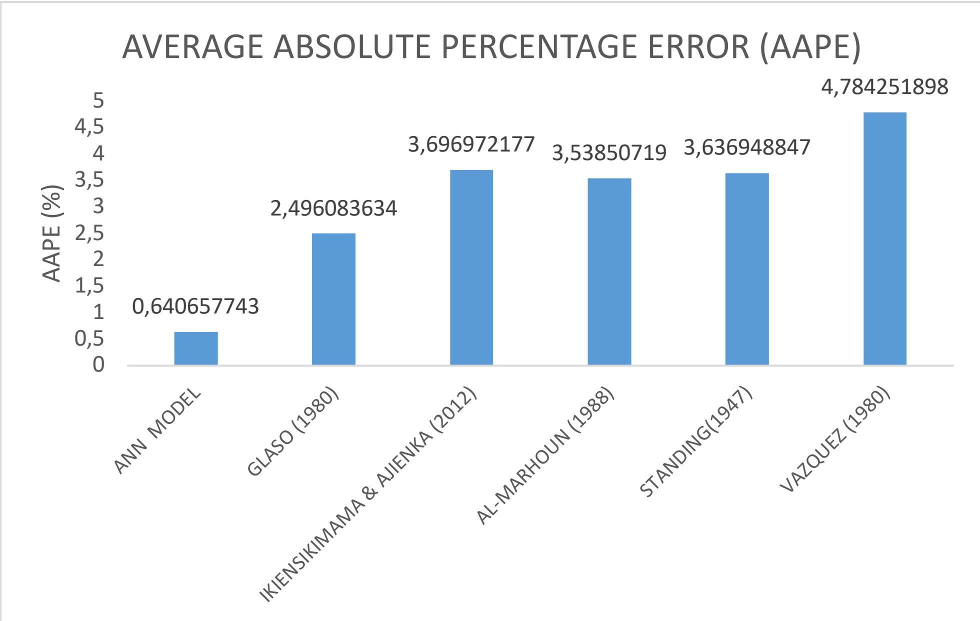

Figure 15. Average absolute percentage error comparison for different correlation

Discussion

Figure 4 and 5 shows the relationship between the outputs of the network and the targets for the training, validation, testing and when all these parameters are combined. A good match was observed. Figures 6 and 7 present the histogram of the errors between target values and predicted values after training and generally indicate how predicted values differ from the target values. The data fitting errors are distributed within a reasonably good range around zero for training and verification respectively.

Figure 8 to 12 show the cross-plots of predicted and experimental values of OFVF correlation for this study and the other existing correlations. The new ANN model performed better than the other correlations compared with it. It is worthy to note that most of these existing correlations are usually incorporated in most petroleum engineering PVT simulation software currently used in petroleum industry. The accuracy of correlations relative to experimental values is determined by various statistical means. The criteria used in this study were correlation coefficient (R), root mean square error (RMSE) and average absolute

Journal of Earth Energy Engineering
Vol. 11 No. 3, December 2022, pp 125-136

percentage error (AAPE). The statistical error analyses are shown in Figure 13, 14 and 15. A value of R close to one, a low value of RMSE and low value of AAPE are preferable. This study's ANN correlation obtained higher accuracy than that of Standing, Glaso, Vazquez, Al-Marhoun and even that from Niger Delta Ikiensikimama & Ajienka.

CONCLUSION

From this study, it can be concluded that:

1. A new artificial neural network (ANN) correlation for predicting oil formation volume factor suitable for Niger Delta crude oil mixture has been developed.
2. The developed model can be used to accurately predict the OFVF of the Niger Delta crude oil when experimental data is not readily available.
3. The correlation coefficients of this study that were based on Niger Delta oil samples are closer to 1 than those of the existing correlations considered in this study.
4. The developed model performed better than the existing correlations of Standing, Ikiensikimama & Ajienka, Glaso, Vazquez and Al-Marhourn.
5. Data cleaning and the removal of redundant data helped to improve the accuracy of the new ANN model and could be applied to any ANN modeling of PVT properties.

ACKNOWLEDGMENTS

The authors did not receive funding for the project from anybody or organization.

REFERENCES

Baziar, S., & Shahripour, H. B. (2015, April). A novel correlation approach to predict total formation volume factor, using artificial intelligence. *AshEse Journal of Engineering, 1(X)*, 001-007.

Osman, E. A., Abdel-Wahhab, O. A., & Al-Marhoun, M. (2001). Prediction of Oil PVT Properties Using Neural Networks. *Society of Petroleum Engineers (SPE)*.

Abooali, D., & Khamehchi, E. (2016, October-December). Towards Predictive Models for Estimation of Bubble Point Pressure and Formation Volume Factor of Crude Oil Using an Intelligent Approach. *Brazilian Journal of Chemical Engineering, 33*(4), 1083-1090. doi:dx.doi.org/10.1590/0104-6632.20160334s20150374.

Alimadadi, A., Fakhri, A., Alimadadi, F., & Dezfoulian, M. (2011, October 13-14). Predicting the PVT Properties of Iran Crude Oil by Neural Network. *1st International eConference on Computer and Knowledge Engineering (ICCKE)*.

Ramirez, A. M., Valle, G. A., Romero, F., & Jaimes, M. (2017, May 18-19). Prediction of PVT Properties in Crude Oil Using Machine Learning Techniques MLT. *Society of Petroleum Engineers*.

Kamari, A. (2016). *The Determination of Petroleum Reservoir Fluid Properties: Application of Robust Modeling Approaches.* Ph.D. Dissertation, University of KwaZulu-Natal, Durban, South Africa, Department of Chemical Engineering, School of Engineering, College of Agriculture, Engineering and Science.

Oloso, M. A. (2018, June). *PREDICTION OF RESERVOIR FLUID PROPERTIES.* Ph.D. Dissertation, University of Portsmouth.

Oman E.A, Abdel-Wahhab, O.A & Al-Marhoun, M.A (2001). Prediction of Oil Properties using Neural Network. *Society of Petroleum Engineers (SPE).*

Glaso, O. (1980). Generalized Pressure-Volume-Temperature Correlations. *Journal of Petroleum Technology*, 785-795.

Standing M. B. (1947). A Pressure-Volume-Temperature Correlation for Mixtures of California Oils and Gases. *Drill & Prod. Pract., API* (1947), 275-87.

Al-Marhoun, M. A. (1988). PVT Correlations for Middle East Crude Oils. *Journal of Petroleum Technology*, 650-666. doi: 10.2118/13718-PA.

Azubike,I.I., Ikiensikimama, S. S., & Oyinkepreye, D.O. (2016). Forecasting Gas Compressibility Factor Using Artificial Neural Networks Tool for Niger-Delta Gas Reservoir. *Society of Petroleum Engineers (SPE)*.

Ikiensikimama, S. S., & Ajienka, J. A. (2012). Impact of PVT correlations development on hydrocarbon accounting: The case of the Niger Delta. *Journal of Petroleum Science and Engineering*, 80-85. doi:10.1016/j.petrol.2011.12.017 .

Vazquez, M. (1980). Correlations for Fluids Physical Property Prediction. *Journal of Petroleum Technology*, 968-970.