International Journal of Scholarly Research in
Engineering and Technology
Journal homepage: https://srrjournals.com/ijsret/
ISSN: 2961-3280 (Online)

(RESEARCH ARTICLE)

Check for updates

# Artificial intelligence–driven prediction on oil formation volume factor for Niger delta region

Mbachu Ijeoma Irene * and Ojonugwa David Micheal

*Petroleum and Gas Engineering Department, University of Port Harcourt, Rivers, Nigeria.*

International Journal of Scholarly Research in Engineering and Technology, 2026, 08(01), 014–026

Publication history: Received on 17 June 2026; revised on 04 August 2026; accepted on 08 August 2026

Article DOI: https://doi.org/10.56781/ijsret.2026.8.1.0017

## Abstract

Oil formation volume factor (OFVF) is very important reservoir fluid property used in the reservoir engineering calculations either directly or indirectly. It indicates the change in the volume of produced oil from the reservoir to surface conditions. The industry standard is to measure this parameter in the laboratory using reservoir samples but the procedures of acquiring these experimental data are very expensive, and time consuming, hence the application of correlations and artificial intelligent. Before now, some empirical correlations existed that determines the oil formation volume factor however, they show high error due to the applied assumptions, and their specification to operate only under a particular range of data. In this study, Super Learner (SL) and Deep Neural Network (DNN) artificial intelligent algorithms were trained to forecast oil formation volume factor at different reservoir condition using Niger Delta reservoir fluid. A total number of 1147 data set was obtained from PVT report from Niger-Delta and validated, out of which, 70% (803) were used to train the models, 15% (172) for cross validation and 15% (172) for test. Quantitative and qualitative statistical analysis were carried out to compare the performance and reliability of the new developed machine learning models with some selected oil formation volume factor empirical correlations. The Super Leaner (SL) model predicted better than Deep Neural Network (DNN) and some of the selected empirical models with the best Rank of 0.05, average absolute percent relative error of 0.054and a correlation coefficient of 0.989. The findings from this research can be applied for estimation of the volumetric properties of hydrocarbon reservoir fluids without the need for conducting routine laboratory analyses.

**Keywords:** Artificial Intelligent; Super Leaner Algorithm; Deep Neural Network; Oil Formation Volume Factor; Niger Delta

## 1 Introduction

The importance of PVT properties in reservoir engineering calculation and management cannot be overemphasized. These reservoir fluid properties include dew point pressure, bubble point pressure, isothermal compressibility factor, formation volume factor (FVF), viscosity and etc. These parameters have a role to play in reservoir engineering such as in material balancing, reserve estimation, reservoir simulation, surface and subsurface facility design, Enhanced Oil Recovery (EOR) process and ultimate recovery, etc [1]. Among the fluid properties of hydrocarbon reservoirs, the oil formation volume factor (OFVF) plays an important role. OFVF shows the change in the volume of produced oil from the reservoir to surface conditions. In fact, the volume of oil that enters the stock tank under surface conditions is less than the volume of oil produced in reservoir conditions that enter the production well. The change in oil volume from reservoir to surface conditions is most affected by the significant pressure reduction below the bubble point and the resultant release of dissolved gases in oil, especially in large amounts of solution gases. Therefore, the oil formation volume factor defined as below is always equal to or greater than 1[2].

* Corresponding author: Mbachu Ijeoma Irene.

The unprecise estimation of this factor will make various processes and calculations challenging for petroleum engineers. Reservoir simulations, inflow performance, fluid flow in porous media, initial- oil-in-place estimation, material balance, well test analysis, and economic analysis are some of the reservoir processes and calculations ([3], [4]). The ideal method to determine OFVF is to use experimental tests, but the laboratory procedures are often costly and time-consuming. Hence, engineers use empirical correlations and machine learning in estimating them. There have been numerous studies on predicting PVT properties using correlations ([5], [4], [6], [7], [8], [9]) and artificial intelligence-based approaches ([10], [11], [12]).

The applications of the Artificial Intelligent in the petroleum industry has been increasing in the last two decades. Artificial Neural Network (ANN) is the first Artificial Intelligent tool applied in the petroleum industry for PVT prediction and it was proposed for the first time by [10]. Afterwards, some authors like [8], [13], [14], [15] and [16] have developed Artificial neural network (ANN) models for estimating PVT properties. The authors reported that their results showed improved performance over the conventional empirical correlations with 40% to 30% reduction in the average error.

Through the literature, it can be found that Artificial Neural Network (ANN) model has gained ground but recently, some authors have started investigating other artificial intelligent as to address the shortcomings of ANN tool in predicting PVT properties. ANN can learn complex non-linear relationships, even when the input information is noisy and less precise. However, the most important problem of ANN is the unexplained behavior of the network; when ANN produces a probing solution, it does not give a clue as to why and how which reduces the trust in the network.

Lately, some researchers ([17], [18], [19], [20], [21], [22], [23]) investigated the use of new machine learning algorithms such as XG Boosts, Neuro-Fuzzy Inference System (ANFIS), random Forest, Supper learn and Lightgbm etc as to address the limitations of Artificial Neural Network. [17] have established adaptive network-based fuzzy inference system (ANFIS) and genetic programming (GP) models for the accurate assessment of reservoir OFVF using 1200 global data sets. Compared to the available correlations in the literature, the new ANFIS and GP models gave a better prediction with an average absolute relative deviation of 1.8% and 2.1%, respectively. [22] used a supervised deep neural network to estimate oil formation volume factor at bubble point pressure using 2994 Niger Delta region data. The author used a multi-layer neural network via the anaconda programming environment as a function of gas-oil ratio, gas gravity, specific oil gravity, and reservoir temperature. Two frameworks for modelling data using deep learning were employed which are TensorFlow and Keras. The result showed that DNN outperformed some of the existing correlations he evaluated with a mean average error of 0.05.

[23] used 2247 PVT data set points from light and heavy oil to forecast dead oil viscosity. The researchers implemented six machine learning algorithms of random forest (RF), lightgbm, XGBoost, MLP neural network, Support Vector machine and Super Learner simultaneously in predicting oil viscosity. Results showed that the Super-Learner algorithm indicated high performance as to compare to other used algorithms with a Mean root error of 0.3387 and $R^2$ of 0.9568. Mbachu and Peter (2024) adopted two machine learning procedures of Artificial Neural Network (ANN) and Support Vector Machine (SVM) to predict GOR for Niger Delta region. The authors used a total number of 852 data set which was obtained from PVT report from Niger-Delta. From the statistical analysis carried out, [20] revealed that the Artificial Neural Network performed better than the Support Vector Machine (SVM) as well as some common selected GOR correlations examined. ANN performed better than other evaluated tool with the best rank of 0.139 with the highest correlation coefficient of 0.98.

PVT data might not be readily available because of cost, time, and difficulty involved in the process of sampling and carrying out laboratory measurements. Many correlations for estimating oil formation volume factor have been published in the past seven decades. Most of these correlations give reasonably accurate results when applied at the bubble-point pressure and above. But, for pressures below the bubble point, the calculated oil formation volume factor yields considerable error. Lately, [24] developed OFVF correlation for at and below bubble point but is limited as to apply to Niger Delta region and also the error associated with empirical correlations. In recent times, Artificial Intelligence (AI) techniques have been able to bridge this gap and has 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; hence, this research centered on developing AL based model to predict oil formation volume factor model that predicts OFVF at and below bubble point pressure for Niger Delta. These models will help in proper management of oil and gas reservoirs in Niger Delta.

## 2 Overview of Artificial Intelligence and Machine Learning

Artificial Intelligence (AL) is a method of making a computer/computer-controlled robot, or a software think intelligently like the human mind. It is accomplished by studying the patterns of the human brain and by analyzing the cognitive process. The examples of cognitive abilities of artificial intelligent are learning, reasoning, problem-solving and perception The outcome of any Artificial intelligence research is to develop intelligent systems and software. The major two ways of implementation are through machine learning and deep learning ([25], [26]).

Machine learning is a core sub-area of Artificial Intelligence (AI) that gives it ability to learn. The learning process is achieved by using algorithms to discover patterns and generate insights from the original or measured data they are exposed to. An algorithm is an estimator that maps a data set of n observations ($W_{I,\ Yi}$), i =1, n, into a prediction function that can be used to map an input W into a predicted value for Y. Some of machine learning algorithms that are frequently adopted in forecasting PVT properties in petroleum and gas engineering are Artificial Neural Network (ANN), support vector machine (SVM), XGBoost, Deep Neural Network, Decision tree, Random Forest, Lightgbm Fuzzy Logic System, Neuro - Fuzzy System and Super Leaner etc. The two algorithms adopted in this study are Super Leaner and Deep Neural Network. These algorithms were carefully selected having reported by many authors about their excellency in predicting PVT parameters ( [23], [22]).

### 2.1 Super Leaner Algorithm (SLA)

Super learning algorithm is a general loss-based learning method that has been proposed and analyzed theoretically in [27]. It was developed by [28]. The super learner is a prediction method designed to find the optimal combination of a collection of prediction algorithms. The super learner algorithm finds the combination of algorithms minimizing the cross-validated risk. The super learner framework is built on the theory of cross validation and allows for a general class of prediction algorithms to be considered for the ensemble. It uses ensemble to stack different base learners together to reach better prediction even compared to the best base learner because super learner absorbs the systematic error of base learners and thus relief the error on final prediction [29]. Super learner has been applied in biology domain including genetics, medicine, epidemiology, and healthcare ([30], [31], [32], [33], [34], [29]). [34] and [29] found super learner more robust than any base learner that the super learner is based on. However, for a super learner, all of its base learners and meta learner have to be trained. Therefore, computational time for super learner will be more than any of the base learners. The detailed super learner algorithm is proposed by [28] as a powerful tool for machine learning.

In order to show the super learner algorithm clearer, a detailed description can be found in [28] and schematic diagram is shown in Fig. 2. Super learner is different from any machine learning algorithm because it is actually a combination algorithm of base learners. Base learners of a super learner algorithm could be any machine learning algorithm such as XGBoost, LightGBM, MLP, and random forest. The base learners are chosen and fitted by train data separately and the whole train dataset is predicted in the form of K-fold cross validation. The data in each group k are predicted using the remaining data as training data. Such process is performed with each base learner and the predictions from all base learners are stacked, which is reason that super learner is also called stacking ensemble. Meta learner, which is usually linear regression model, is used to output final prediction based on base learners stacked prediction., super learner could also be interpreted in a simpler way. For music ensemble, different musical instruments are used by composers for better music than any single instrument. A good conductor coordinates the ensemble orchestra for final music presentation to audience. For machine learning ensemble of super learner, base learners serve as every musical instrument of an orchestra, and the meta learner serves as the conductor who coordinated the final results.

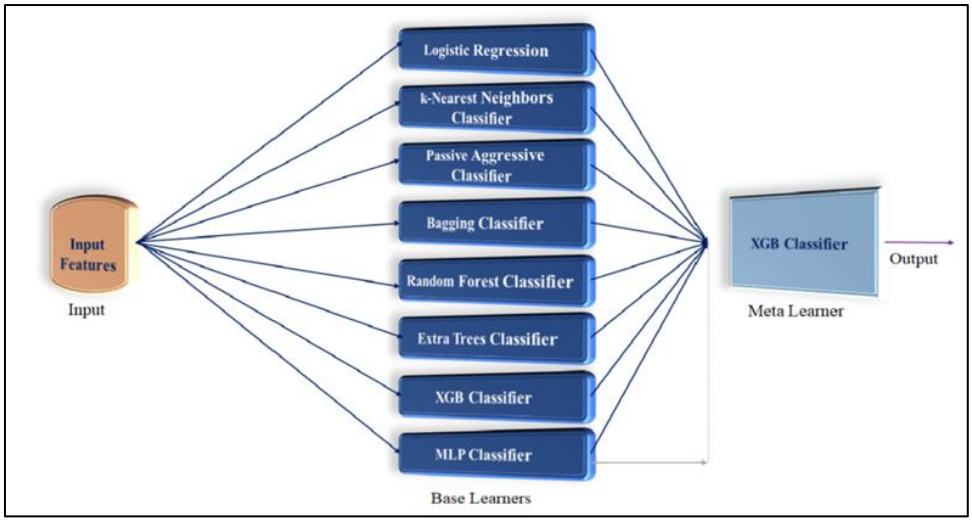


**Figure 1** Super Learner Model Architecture, [28].

### 2.2 Artificial Neural Network (ANN) and Deep Neural Network

Artificial Neural Network (ANN) is based on a collection of connected units or nodes called artificial neurons, which loosely model the neurons in a biological brain. Each connection, like the synapses in a biological brain, can transmit a signal to other neurons. An artificial neuron receives signals then processes them and can signal neurons connected to it (Fig. 3). The "signal" at a connection is a real number, and the output of each neuron is computed by some non-linear function of the sum of its inputs. The connections are called edges. Neurons and edges typically have a weight that adjusts as learning proceeds. The weight increases or decreases the strength of the signal at a connection. Neurons may have a threshold such that a signal is sent only if the aggregate signal crosses that threshold. Typically, neurons are aggregated into layers. Different layers may perform different transformations on their inputs. Signals travel from the first layer (the input layer) to the last layer (the output layer), possibly after traversing the layers multiple times. A network is typically called a deep neural network if it has at least 2 hidden layers.

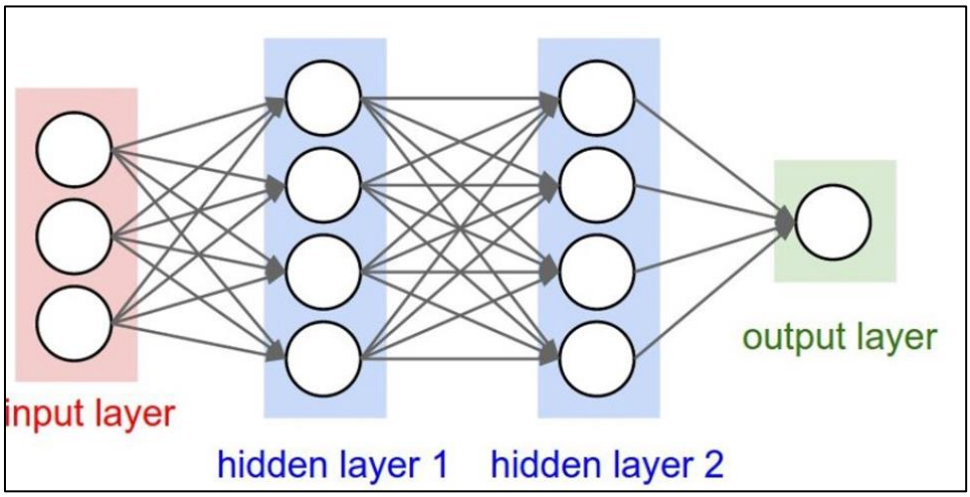


**Figure 2** Deep Neural Network Architecture

The extension of conventional artificial neural networks is deep neural networks. Compared to conventional neural networks, there are two main differences that deep neural networks have. It is shallow for conventional neural networks to have one or two hidden layers. On the other hand, there are many hidden layers in deep neural networks. For instance, a neural network of millions of neurons was used by the Google brain project. There are a wide range of models for deep neural networks, ranging from DNNs, CNNs, RNNs, and LSTMs. Recent studies have even brought us attention-based networks that focus on specific parts of a deep neural network. The larger the network and the more layers it has, the more complex the network becomes and the more resources and more time it needs to train. Deep neural networks work best with GPU-based architectures that take less time to train than classical CPUs, while recent developments have

shortened training times considerably [10]. The biggest advantage of deep learning algorithms is that they try to learn features from data in an incremental manner. The model first learns the basic constituents/low-level features before moving on to the high-level ones, thus eliminating the need of domain expertise and hardcore feature extraction.

## 3 Methodology

### 3.1 Data Description

The data used was obtained from conventional PVT reports that derive the various fluid properties through liberation process from the Niger-Delta Region of Nigeria. The data set comprises of the four (4) input parameters Oil density ($\gamma_o$), Reservoir temperature (°F), Solution Gas-Oil Ratio (Rs-scf/stb), Gas density ($\gamma_g$) and one (1) output parameter of Oil Formation Volume Factor (OFVF). Tables 1 and 2 show the maximum, minimum and mean values of the data set used in this study and test data respectively

**Table 1** Summary of mean, maximum and minimum of data used in this study

| PVT Properties | Maximum | Minimum | Mean |
|---|---|---|---|
| Gas gravity | 1.712 | 0.531 | 0.694 |
| Reservoir Temperature (°F) | 282.24 | 36.50 | 170.0 |
| Oil formation Volume factor (scf/stb) | 4.262 | 1.059 | 1.334 |
| Gas-Oil Ratio $R_s$ (scf/stb) | 5000 | 10.00 | 613.8 |
| Oil gravity | 0.947 | 0.342 | 0.7421 |

**Table 2** Summary of mean, maximum and minimum of data used in testing this study

| PVT Properties | Maximum | Minimum | Mean |
|---|---|---|---|
| Gas gravity | 1.182 | 0.568 | 0.6818 |
| Reservoir Temperature (°F) | 255,0 | 129.0 | 175.9 |
| Oil formation Volume factor (scf/stb) | 4.508 | 1.053 | 1.382 |
| Gas-Oil Ratio $R_s$ (scf/stb) | 6176.01 | 40.50 | 562.0 |
| Oil gravity | 0.923 | 0.468 | 0.7323 |

### 3.2 Data Validation

Before any experimental PVT data are used for design or study purposes, it is very important to ensure that there are no error or major inconsistencies that would render any subsequent work useless. Campbell diagram (Buckley plot) and the Mass Balance Diagram which are otherwise known as cross plot are major two types of data validations. These techniques were used to validate the data set used in this work.

### 3.3 Modelling Technique

For Deep Neural Network, TensorFlow and Keras libraries of python was employed for the training of these deep learning models. The data was imported to the python environment as a Data frame using a machine learning library Pandas. A sequential model which allows us to build in layers and add weight was created and values for the weights are provided. A three-layer network is used in this work. The first layer consists of four neurons representing the input parameters which are oil density ($\gamma_g$), reservoir temperature (T), solution gas- oil ratio (GOR) and gas density. The second layer is the hidden layer, and the third layer contains one neuron representing the output values of oil formation volume factor (OFVF). Out of total 1147 data set used ,70% (803) were used for training, 15% (172) for cross validation and the remaining 15% (172) for testing. The cross-validation process allowed the machine to monitor the performance of the network and prevent overfitting. Rectified Linear (RELU) activation function was applied to train network; it helps the model to count for nonlinearity relationships. stochastic gradient descent (SGD), Adam and mean square error

were used as optimizer and loss function for the development of this model. The model was then trained by fitting onto the training data, target data, validation and number of epochs.

Four machine learning algorithms XGBoost, LightGBM, random forest regressor and MLP neural network were adopted as the base algorithms using Python 3.7 (Fig. 3). for development of super leaner model. The four base learners (XGBoost, LightGBM, random forest regressor, and MLP regressor) used the same 70% data set streamed into the Bayesian ridge regression as meta learner and 15% for cross validation and testing respectively. To avoid bias, it is important to mention that the same pattern and approach adopted for DNN was also adopted for Super Leaner but the test data from DNN was used for validation analysis.

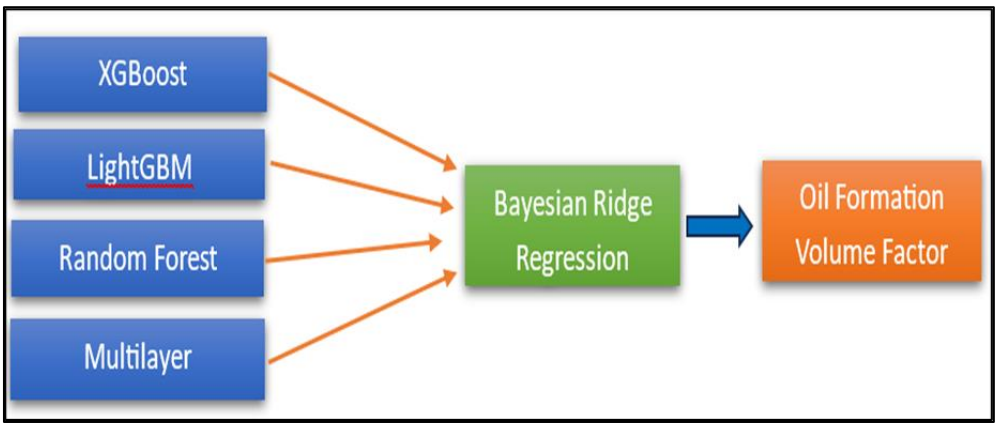


**Figure 3** Schematic of Super Leaner model for oil formation volume factor prediction for this study

### 3.4 Correlation Comparison

To compare the performance and accuracy of the new model to other empirical correlations, two forms of analyses were performed which are quantitative and qualitative screening. For quantitative screening method, statistical error analysis was used. The statistical parameters used for the assessment were percent mean relative error (MRE), percent mean absolute error (MAE), percent standard deviation relative (SDR), percent standard deviation absolute (SDA) and correlation coefficient ($R^2$).

For correlation comparison, a new approach of combining all the statistical parameters mentioned above (MRE, MAE, SDR, SDA and R) into a single comparable parameter called Rank was used [9]. The use of multiple combinations of statistical parameters in selecting the best correlation can be modeled as a constraint optimization problem with the function formulated as;

$$Min\ \ Rank = \sum_{j=1}^{m} S_{i,j} q_{1,j} \tag{1}$$

Subject to $$\sum_{i=1}^{n} S_{i,j} \tag{2}$$

With $$0 \le S_{ij} \le 1 \tag{3}$$

Where $S_{i,j}$ is the strength of the statistical parameter j of correlation i and $q_{ij}$, the statistical parameter j corresponding to correlation i. j = MRE, MAE, …. $R^2$, where $R^2 = (1-R^2)$ and $Z_i$ is the rank, (or weight) of the desired correlation. The optimization model outlined in Equations 1 to 3 was adopted in a sensitivity analysis to obtain acceptable parameter strengths. The final acceptable parameter strengths so obtained for the quantitative screening are 0.4 for MAE, 0.2 for $R^2$, 0.15 for SDA, 0.15 for SDR, and 0.1 for MRE. Finally, Equation 3 was used for the ranking. The correlation with the lowest rank was selected as the best correlation for that fluid property. It is necessary to mention that minimum values were expected to be best for all other statistical parameters adopted in this study except R, where a maximum value of 1 was expected. Since the optimization model (equations 1 to 3) is of the minimizing sense a minimum value corresponding to R must be used. This minimum value was obtained in the form $(1-R^2)$. This means the correlation that has the highest correlation coefficient ($R^2$) would have the minimum value in the form $(1-R^2)$. In this form the parameter

strength was also implemented to 1-$R^2$ as a multiplier. Ranking of correlations was therefore made after the correlations had been evaluated against the available database.

For qualitative screening, performance plots were used. The performance plot is a graph of the predicted versus measured properties with a 45° reference line to readily ascertain the correlation's fitness and accuracy. A perfect correlation would plot as a straight line with a slope of 45°.

## 4 Results and discussion

The study is on the prediction of oil formation volume factor for different reservoir conditions using artificial intelligent approach for Niger Delta region.

The 1147 data set collected from Niger Delta was first used in evaluating some of the best OFVF correlations as to ascertain their weakness in predicting oil formation value factor for Niger Delta region. After training the two (2) specially selected machining learning algorithms (Super Learner (SL) and Deep Neural Network (DNN)) with training data, their performance was tested with 15% (172) of the entire data set. The test data was not used during training and validation, indicating that they were not exposed to the machine. These data were randomly selected by the Machine learning tools to ascertain the accuracy and stability of the model. The performance of the Super Learner (SL) and Deep Neural Network (DNN) artificial intelligent algorithms were evaluated with some selected empirical correlations of [5], [6], [7] and [9]. These predictive models were carefully selected, having been developed specifically for the prediction of oil FVF and [9] model was developed specifically for Niger Delta. [6] and [5] correlations were reported by [29] and [9] as good empirical correlations for forecasting oil formation volume factor.

### 4.1 Statistical Accuracy Result

#### *4.1.1 Evaluation Result*

**Table 3** Statistical Accuracy of Oil Formation Volume Factor Evaluation Using Niger-Delta Data

| CORRELATIONS | MRE | MAE | SRE | SAE | R2 | RANK |
|---|---|---|---|---|---|---|
| Standing (1947) | 2.821 | 2.3304 | 3.416 | 2.135 | 0.885 | 2.06991 |
| Almehaideb, (1997) | 1.176 | 2.261 | 1.878 | 1.283 | 0.916 | 1.51295 |
| Ikiensikimama (2009) | 1.065 | 1.299 | 1.191 | 2.245 | 0.929 | 1.1557 |
| AL- Marhuoun (1985) | 2.691 | 2.933 | 1.153 | 1.336 | 0.922 | 1.83125 |

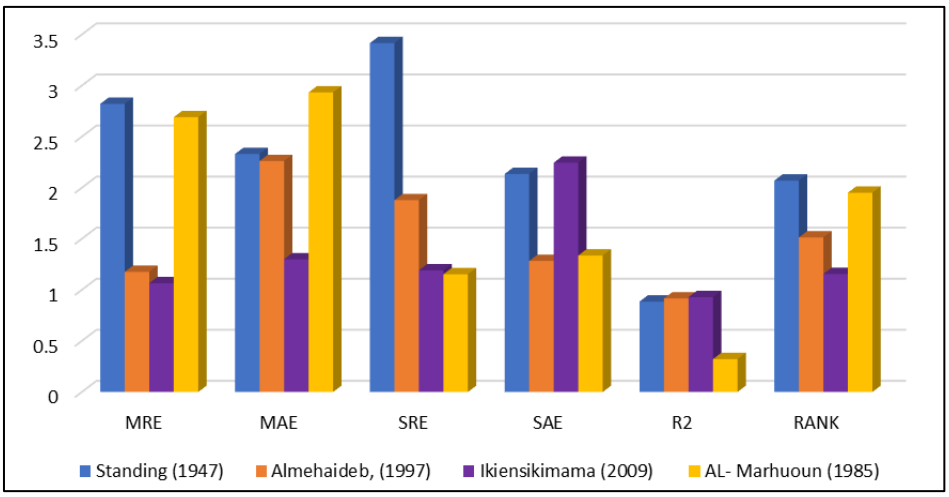


**Figure 4** Comparison for Statistical Accuracy on Evaluation OFVF correlations

Firstly, the four selected empirical correlations were first evaluated with the data set employed (1147) in this study as to confirm their strength statistically in predicting OFVF for Niger Delta for at and below bubble point pressure. It can

be observed from Table 3 and Fig. 2 that the indigenous correlation which is [9] performed better than other correlations which are foreign equations. [9] outperformed other correlations with the rank numerical of 1.155, MAE of 0.099 and $R^2$ of 0.929. The trend is expected since [9] was developed for Niger Delta.

### *4.1.2 Validation Result*

For the validation and performance analysis, test data from DNN was generally used as to avoid biased result. The test data is 15% of the 1174 total data that was initially fed into the DNN machine but was not used during training and validation. These data were randomly selected by the DNN tool as to ascertain the accuracy and stability of the model. The results of the statistical assessment adopted for validation in this research are presented in Table 4 and Fig. 3 for the two intelligent models and empirical correlations examined. The result shows that the two (2) machine learning predicted better than all the empirical correlations investigated. Between the two-machining learning algorithm studied, Super Leaner (SL) emerged the best predictor with the rank value of 0.050, mean absolute error of 0.054 and best correlation coefficient of 0.989. The Deep neural network (DNN) algorithm predicted the oil formation volume factor with the rank value of 0.070, mean absolute error of 0.072 and correlation coefficient of 0.972 for the data set studied. It is evident that the super leaner machine learning models proposed in this study is more reliable, accurate and robust than the Deep Neural Network model in terms of statistical parameters employed. Super Leaner uses the advantages of ensemble to stack different base learners together to reach better prediction because it absorbs the systematic error of base learners and thus relief the error on final prediction.

**Table 4** Statistical Accuracy of Oil Formation Volume Factor Evaluation Using Niger-Delta Data

| **Indicators** | **Super Learner (SL)** | **Deep Neural Network DNN)** | **Ikiensikimama (2009)** | **Almehaideb (1997)** | **Al-Marhoun (1985)** | **Standing (1981)** |
|---|---|---|---|---|---|---|
| MSE | 0.021 | 0.052 | 0.103 | 0.062 | 0.152 | 0.202 |
| MAE | 0.054 | 0.072 | 0.041 | 0.186 | 0.109 | 0.301 |
| SRE | 0.087 | 0.103 | 0.162 | 0.107 | 0.054 | 0.121 |
| SAE | 0.095 | 0.102 | 0.072 | 0.209 | 0.096 | 0.208 |
| $R^2$ | 0.989 | 0.972 | 0.962 | 0.951 | 0.969 | 0.959 |
| Rank | 0.05 | 0.07 | 0.1621 | 0.1936 | 0.2243 | 0.37995 |

Table 3 and Fig. 4 also show the validation results of two machine learning algorithms with the four empirical correlations examined It can be observed from Table 3 and Fig. 4 that all the intelligent models performed better than the empirical correlations examined. The empirical correlation of [9] which is indigenous correlations performed better than the foreign ones which are [5], [6] and [7]. The trend undertaken of the locally developed correlations is expected because correlations performed better in their region of origin. [9] performed better than other three evaluated correlations with the rank value of 0.1621, mean absolute error 0.041and correlation coefficient 0.962 followed by [7]. with the rank value of 0.1936, mean absolute error of 0.186 and correlation coefficient of 0.951. This study recommends [9] correlation to be used in predicting oil formation volume factor for different reservoir conditions in absence of machining learning models trained in this research. This study has once again proven the reliability and efficient performance of Artificial intelligent models (super leaner and Deep Neural Network) in predicting the reservoir PVT properties as reported by various authors. The models developed in this research are very strong powerful tool to forecast oil formation volume factor for Niger Delta region in terms of root mean squared error, absolute average percent error, standard deviation, correlation coefficient and Rank.

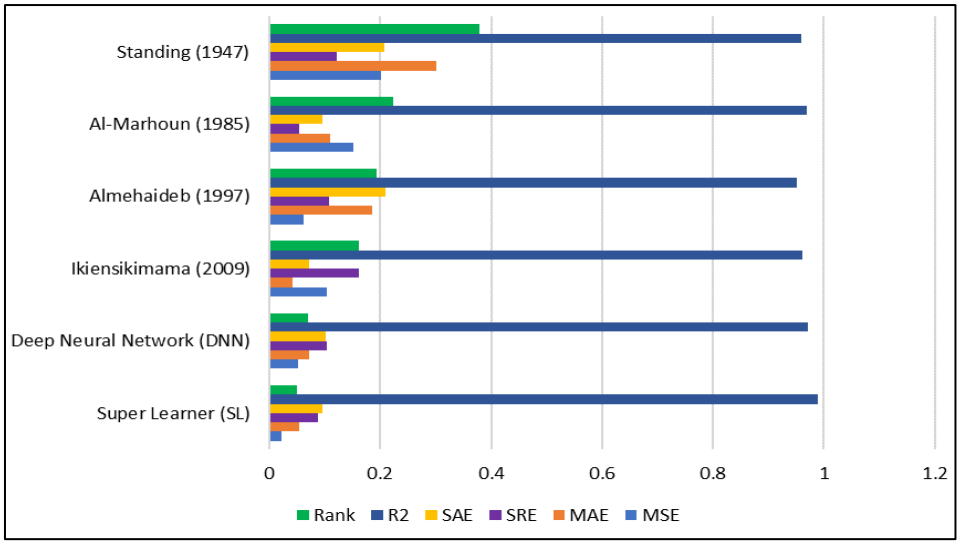


**Figure 5** Comparison on Statistical Accuracy for Validation OFVF correlations

### 4.2 Performance Plot Results

Cross plots of the predicted versus experimental data for oil formation volume factor are illustrated in Figs. 6 to 10. It is a plot of predicted versus measured properties with a 45° reference line to readily ascertain the correlation's fitness and accuracy.

Figs. 10 and 9 which are the Cross plots of Super Leaner Model and Deep Neural Network show the tightest cloud of points around the 45° line with very good clusters at low band, indicating the excellent agreement between the experimental and the predicted data values when compared to Figs. 6 to 8. In addition, this indicates the superior performance of the machine learning model over empirical correlations evaluated. The accuracy of the models indicates that the Machine intelligent model (Super Leaner Model and Deep Neural Network) did not over fit the data, which implies that they were successfully trained.

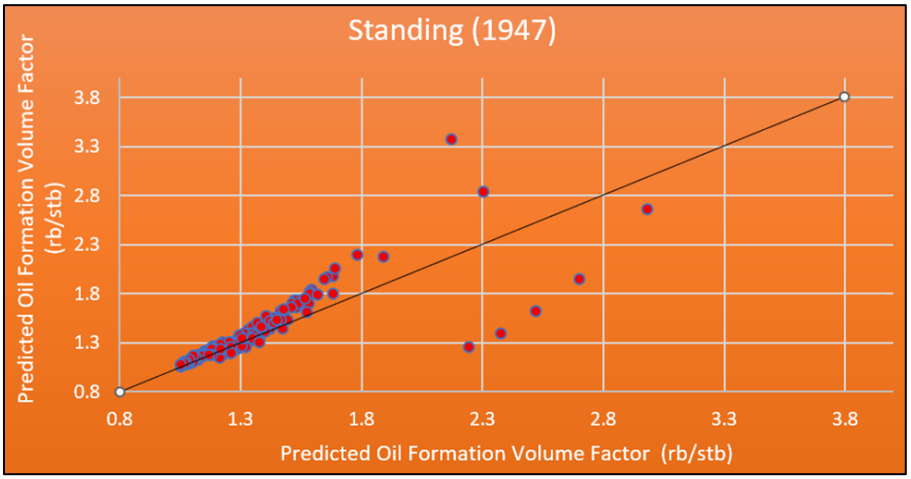


**Figure 6** Cross Plot of Standing [5] Correlation using Niger-Delta Data

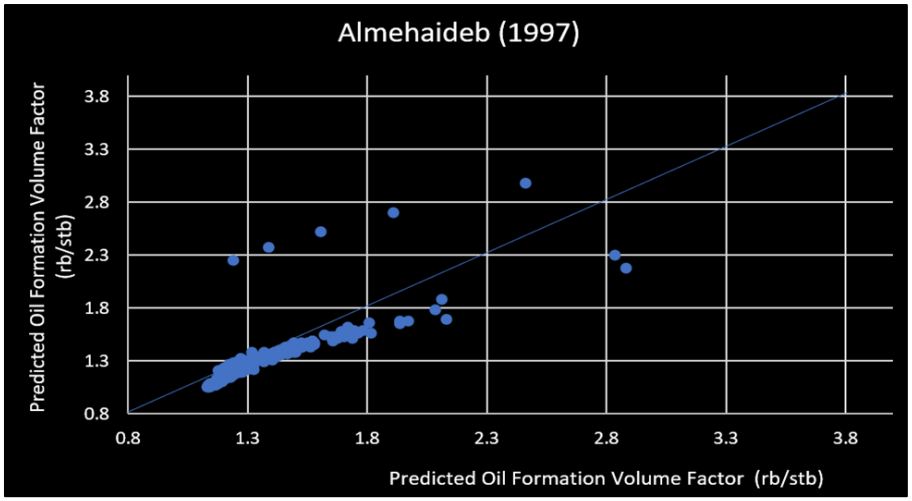


**Figure 7** Cross Plot of Almehaideb [7] Correlation using Niger-Delta Data

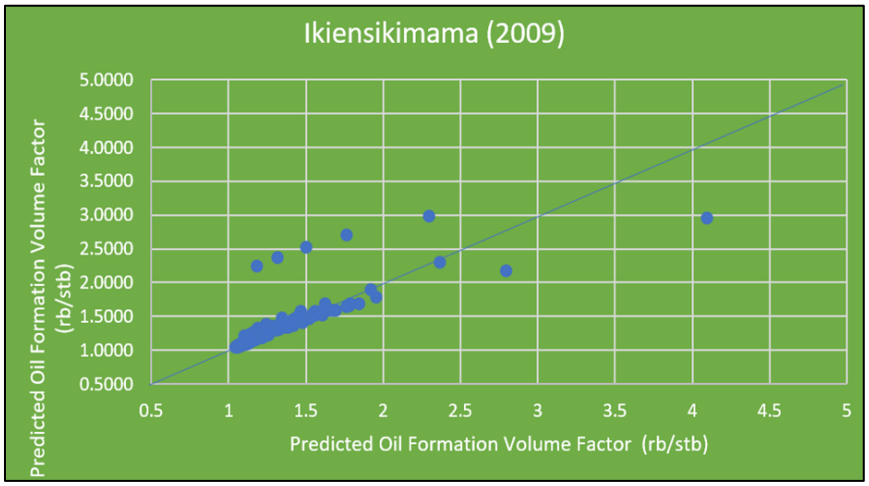


**Figure 8** Cross Plot of Ikiensikimama [9] Correlation using Niger-Delta Data

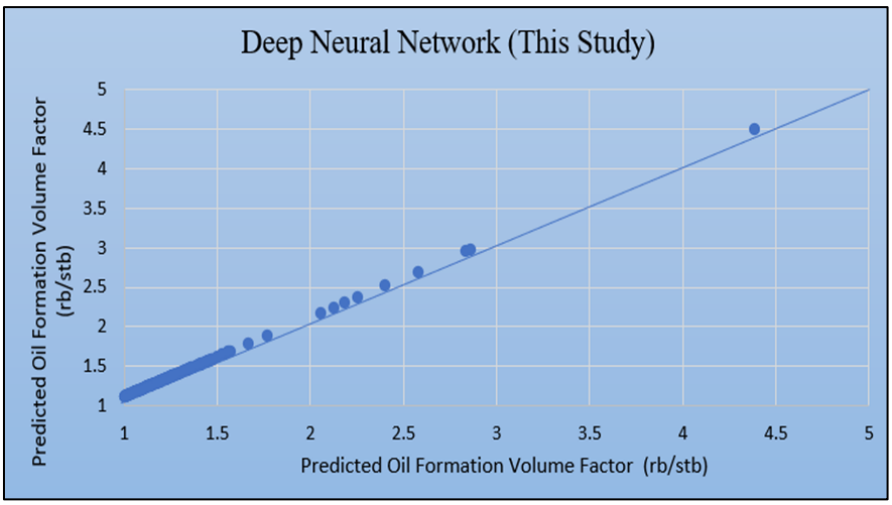


**Figure 9** Cross Plot of Deep Neural Network Model using OFVF Niger-Delta Data

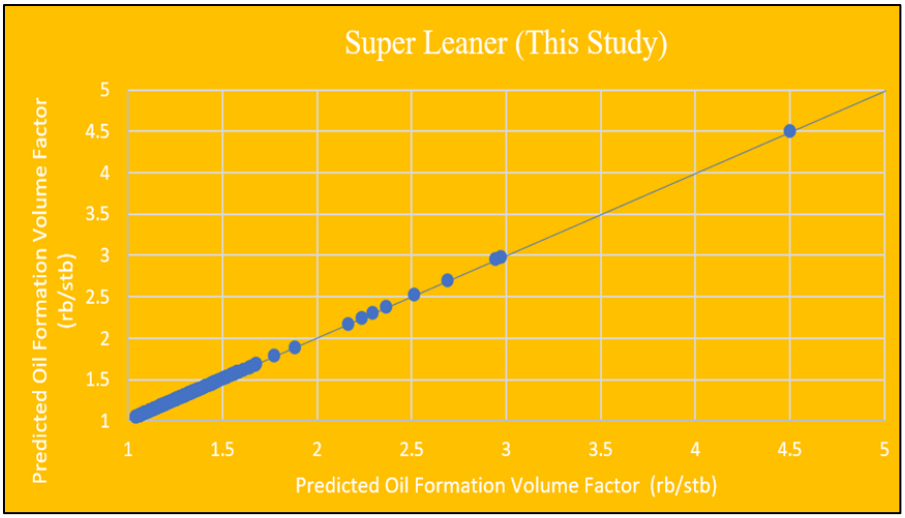


**Figure 10** Cross Plot of Super Leaner Model Correlation using OFVF Niger-Delta Data

## 5 Conclusion

Two machine learning algorithms Super Leaner Model and Deep Neural Network have been adopted in this research for oil formation volume factor prediction using data from Niger Delta of Nigeria. The machine is trained to predict oil formation volume factor for different reservoir conditions. The research revealed that super leaner estimated oil formation volume factor better than Deep Neural Network and empirical correlations examined with the best rank of 0.05, correlation coefficient of 0.989 and better performance plot. Deep Neural Network performed better than the selected empirical correlations with rank value of 0.07 and correlation coefficient of 0.972. Ikiensikimama empirical equation performed better than three other equations evaluated with the rank value of 0.1621 and correlation coefficient of 0.962. Ikiensikimama correlation can be used to predict OFVF for Niger Delta Region in absence of these new developed intelligent tools for Niger Delta region for proper reservoir simulation and management. This study gives a bright light of machine learning modeling and will assist petroleum exploration engineers to predict various reservoir properties with better accuracy, leading to reduced exploration cost and time with increase in productions.

## Compliance with ethical standards

### *Disclosure of conflict of interest*

No conflict of interest to be disclosed.

## References

[1] Ahmed,T.(2013). Equations of State and PVT Analysis; Elsevier.

[2] Dake, L. P. (1983) Fundamentals of Reservoir Engineering, Elsevier

[3] Elsharkawy, A.M. and Alikhan, A.A., (1997). "Correlations for predicting solution gas/oil ratio, oil formation volume factor, and undersaturated oil compressibility". J. Petrol. Sci. Eng. 17, 291–302. https://doi.org/10.1016/S0920-4105(96)00075-7.

[4] Glaso, O., (1980). "Generalized pressure-volume-temperature correlations". J. Petrol. Technol. 32, 785–795. https://doi.org/10.2118/8016-pa

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

[6] Al-Marhoun, M. A. (1992). "New Correlation for formation Volume Factor of Oil and Gas Mixtures," *JCPT,* pp. 22

[7] Almehaideb, R.A., 1997. Improved PVT correlations for UAE crude oils. Proc. Middle East Oil Show. https://doi.org/10.2118/37691-ms.

[8] Gharbi, R.B., Elsharkawy, A.M., Karkoub, M., (1999). Universal neural-network-based model for estimating the PVT properties of crude oil systems. Energy and Fuels 13, 454–458. https://doi.org/10.1021/ef980143v.

[9] kensikimama, S. S. (2009). "Reservoir Fluid Property Correlations" Advances in Petroleum Engineering, Chi Ikoku Petroleum Engineering Series, IPS Publications, Port Harcourt.

[10] Ali, J.K., (1994). Neural Networks: "A New Tool for the Petroleum Industry? EPCC 1994. In: Society of Petroleum Engineers" - European Petroleum Computer Conference, pp. 233–242. https://doi.org/10.2523/27561-ms.

[11] Khamis, M. A. and Fattah, K. A. (2019). "Estimating oil–gas ratio for volatile oil and gas condensate reservoirs: artificial neural network, support vector machines and functional network approach" *Journal of Petroleum Exploration and Production Technology* 9:573–582. https://doi.org/10.1007/s13202-018-0501-0

[12] Jang, J.S. (1993) "ANFIS: Adaptive-Network-Based Fuzzy Inference System". IEEE Trans. Syst. Man Cybern, 23, 665–685.

[13] Ikiensikimama, S., and Azubuike, I. (2012) "Modelling Approach for Niger Delta Oil Formation Volume Factor Prediction Using Artificial Neural Network". Paper Presented at the Nigeria Annual International Conference and Exhibition, Lagos, 04-06 August.

[14] Varotsis, N., Gaganis V., Nighswander, J. and Guieze P. (1999). "A Novel Non-Iterative Method for the Prediction of the PVT Behavior of Reservoir Fluids," paper SPE 56745 presented at the SPE Annual Technic Conference and Exhibition, Houston, Texas.

[15] Osman, E.A., Abdel-Wahhab, O.A., Al-Marhoun, M.A., (2001) "Prediction of oil PVT properties using neural network"s. In: Proceedings of the Middle East Oil Show, pp. 893–906. https://doi.org/10.2118/68233-ms.

[16] Mohaghegh, S. (2011): "Virtual Intelligence Applications in Petroleum Engineering: Part 1 - Artificial Neural Networks", JPT.

[17] Saghafi, H.R., Rostami, A., Arabloo, M., 2019. Evolving new strategies to estimate reservoir oil formation volume factor: smart modeling and correlation development. J. Petrol. Sci. Eng. 181 https://doi.org/10.1016/j.petrol.2019.06.044.

[18] Al-amoudi, L.A., Salem, B., Fahd, K., Patil, S. and Baarimah, S.O., (2019). "Development of Artificial Intelligence Models for Prediction of Crude Oil Viscosity

[19] Shaibu, S. and Mbachu I. I. (2021). "Modeling Approach for Niger Delta Oil Formation Volume Factor Prediction using Support Vector Machining Learning". American Journal of Engineering Research (AJER), vol.10, Is. 12, p. 80 -88.

[20] Mbachu, I. I. and Peter, A. O. (2024). "Gas- Oil Ratio Prediction Using Machine Learning Procedures for Niger Delta Region", International Journal of Scientific Research and Engineering Development— Vol. 7, 743 -753

[21] Origbo, O. A. and Mbachu I. I. (2024). "Forecasting Dead Oil Viscosity Using Machine Learning Processes for Niger Delta Region", International Journal of Current Science Research and Review, Volume 7. p. 820-830

[22] Koffi, I. U. (2021). "A Deep Learning Approach for the Prediction of Oil Formation Volume Factor", SPE-208627-STU.

[23] Hadavimoghaddam, F., Ostadhassan, M., Heidaryan, E., Sadri, M. A., Chapanova, I., Popov, E., Cheremisin, A. and Rafieepour S. (2021). "Prediction of Dead Oil Viscosity: Machine Learning vs. Classical Correlations", https://doi.org/10.3390/en1404930

[24] Omar, M. and Dhifaf, J. S. (2009). " New Correlation for Oil Formation Volume Factor at and Below Bubble Point Pressure crudes", Journal of Engineering, pp 4347-4355.

[25] Duggle, N. (2023). Introduction of Artificial Intelligence A Beginner's Guide Art Material, Simple Learn.

[26] Terra, J. (2023). Exploring Intelligent Agents in Artificial Intelligent

[27] Van der Laan, M. J, Polley, and Hubbard, A. E. (2007). "Super learner. Statistical Applications in Genetics and Molecular Biology", 6(25): Article 25.

[28] Polley, E.C. and Van Der Laan, M.J. (2010). "Super Learner in Prediction"; U.C. Berkeley Division of Biostatistics: Berkeley, CA, USA, p. 266. 80.

[29] Yang, X., Dindoruk, B. and Lu, L. (2020). A comparative analysis of bubble point pressure prediction using advanced machine learning algorithms and classical correlations, Journal of Petroleum Science and Engineering, https://doi.org/10.1016/j.petrol.2019.106598

[30] Sinisi, S.E., Polley, E.C., Petersen, M.L., Rhee, S.Y. and Van Der Laan, M.J. (2007). "Super learning: An application to the prediction of HIV-1 drug resistance". Stat. Appl. Genet. Mol. Biol. 2007, 6, 7.

[31] Wang, H., Rose, S. and van der Laan, M.J. (2011). "Finding quantitative trait loci genes with collaborative targeted maximum likelihood learning". Stat. Probab. Lett. 2011, 81, 792–796.

[32] Sweeney, E.M., Vogelstein, J.T., Cuzzocreo, J.L., Calabresi, P.A., Reich, D.S., Crainiceanu, C.M. and Shinohara, R.T. (2014). A Comparison of Supervised Machine Learning Algorithms and Feature Vectors for MS Lesion Segmentation Using Multimodal Structural MRI, 9, e95753.

[33] Rose, S. A. (2016). "Machine learning framel8 work for plan payment risk adjustment". Health Serv. Res, 51, 2358–2374, PubMed.

[34] Wyss, R., Schneeweiss, S., Van Der Laan, M., Lendle, S.D., Ju, C. and Franklin, J.M. (2018). "Using Super Learner Prediction Modeling to Improve High-dimensional Propensity Score Estimation". Epidemiology 2018, 29, 96–106. (PubMed).

[35] Sleiti, A. K., Wahib, A., Al-Ammari, W. Abdelrazeq, A M. El-Naas, M. Rahman, M. A., Barooah, A. Hasan, R. and Manikonda, K. (2021). "Comprehensive Assessment and Evaluation of Correlations for gas-oil Ratio, Oil Formation Volume Factor, Gas Viscosity, and Gas Density Utilized in Gas Kick Detection", Journal of Petroleum Science and Engineering 2021, https://doi.org/10.1016/j.petrol.2021.109135