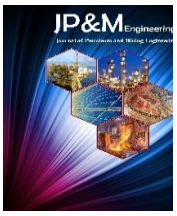
JP&M Engineering
Journal of Petroleum and Mining Engineering



# Hybrid Computing Models to Predict Oil Formation Volume Factor Using Multilayer Perceptron Algorithm

Hazbeh, O.,[a] Alvar, M. A.,[b] Aghdam, S. K.,[c] Ghorbani, H.,[*d] Mohamadian, N.,[e] and Moghadasi, J. [f]

[a] Faculty of earth sciences, Shahid Chamran University, Ahwaz, Iran

[b] Faculty of Engineering, Department of computer Engineering, Shahid Chamran University, Ahwaz, Iran

[c] Department of petroleum engineering, Amirkabir University of Technology, Tehran, Iran

[d] Young Researchers and Elite Club, Ahvaz Branch, Islamic Azad University, Ahvaz, Iran

[e] Young Researchers and Elite Club, Omidiyeh Branch, Islamic Azad University, Omidiyeh, Iran

[f] Petroleum Engineering Department-Petroleum university of technology, Ahvaz, Iran

*Corresponding author e-mail: hamzehghorbani68@yahoo.com

**Article Info**

Received 3 Dec. 2020
Revised 28 Jan. 2021
Accepted 26 Feb. 2021

**Keywords**

Oil formation volume factor;
artificial intelligence;
hybrid model MLP

## Abstract

Achieving important and effective reservoir parameters requires a lot of time and cost, and also achieving these devices is sometimes not possible. In this research, a dataset including 565 datapoints collected from published articles have been used. The input data for forecasting oil formation volume factor (OFVF) were solution gas oil ratio (Rs), gas specific gravity (γg), API gravity (API0) (or oil density γo), and temperature (T). Two hybrid methods multilayer perceptron (MLP) with artificial bee colony (ABC) and firefly (FF) algorithms to predict this parameter have been introduced in that study and their results have been compered after extraction. After essential investigations in this study, the results show that MLP-ABC gives the best accuracy for predicting OFVF. For MLP-ABC model OFVF prediction accuracy in terms of RMSE < 0.002573 bbl/STB and R2 = 0.998 for this test dataset. After comparing the results of the experimental equations, it was concluded that the Dokla and Osman model gives the best results and Based on Spearman's correlation coefficient relationships all input parameters have a positive effect on OFVF prediction, which are as follows: $R_s$> T> API> $\gamma_g$ and these results show that the effect of Rs is more than other input variables and the effect of $\gamma_g$ is the lowest.

## Introduction

Accurate and valuable evaluation of PVT properties (pressure, volume and temperature) is one of the main and most obvious concerns of reservoir engineers for reservoir management and evaluation purposes. These properties include determining and obtaining properties of reservoir fluids' physical characteristics such as bubble point pressure (BPP), solution gas oil ratio an (GORs) and oil formation volume factor (OFVF), which are key development [1-5].

Since estimating the number of hydrocarbons in the reservoir and design is important, one of the most important tasks is to estimate the key parameters of the reservoir that can be used to achieve this importance [42, 44]. For example, liquids undergo fundamental changes in temperature and pressure not only through their production path, but also during normal pressure discharge process. One of the different methods of pressure maintenance and enhance oil recovery (EOR) is injecting gas to increase the pressure of certain chemicals inside reservoir [61]. The optimal design and success of such processes require an accurate understanding of the liquid phase behaviour of the reservoir.

One of the important parameters is determining the concentration of CO2. This parameter is very important in terms of human life. In order to control this parameter, the method of CO2 storage in the subsurface as hydrate is implemented. In 2019 Hassanpouryouzband et al. predicted the solubility of CO2 and N2 in water and brine via three different state equations including, CPA-SRK72, VPT and PC-SAFT. They coupled these equations with binary Interaction Parameters (BIP). Then, they compared results with available experimental data. Acceptable proximity of predictions to experimental results confirms the reliability of the thermodynamic model [62]. A year later in 2020, Hassanpouryouzband et al.

conducted a research on H2 as a substitute for fossil fuels to reduce CO2 emissions, as well as the proper and accurate design of thermodynamic. In order to predict the thermo-physical properties of H2 mixed with CH4, N2, CO2 and a typical natural gas from the North Sea of the GERG-2008, Equation of State (EoS) and SupertRaPP models are used. In addition, a user-friendly software (H2Themobank) was made available to the public.

Due to the high importance of developing and completing oil and gas fields, one of the tasks that has been done in recent years is to use field data to calculate and predict, as well as to determine the parameters used in the oil and gas industry, for example in the following areas have been addressed: reservoirs [6]; formation damage [7], petroleum well blowouts [48], wellbore stability [8], rheology and filtration [9-10], production [11-13]; drilling fluid [14].

Determination of tank properties is conducted through laboratory outputs, which are very costly and time consuming. Also, these tests are not always available and there are not enough samples to determine these properties. Therefore, in order to facilitate the process of determining the characteristics of the reservoir and obtain these characteristics, researchers turned to experimental models. Using previous studies, we conclude that OFVF is a function of: solution gas oil ratio (Rs), gas specific gravity ($\gamma_g$), API gravity ($API^0$) (or oil density $\gamma_o$), and temperature (T) based on the following Eq. (1).

$$OFVF = f\ (R_S, \gamma_g, T, API\ or\ \gamma_o) \quad (1)$$

Previous researchers based on studies and the relationship between the parameters, presented equations in Table 1 that are shown as follows:

In 1947, Standing (1947) proposed an equation for predicting OFVF using 105 data from California oil fields [15]. In 1977, Vazquez and Beggs (1977) proposed an equation for predicting OFVF using 5008 data collected from Worldwide, but in this equation, have a boundary (API=30) and this equation divided two section [16]. In 1980, Glaso (1980) proposed an equation for predicting OFVF using 41 data collected from the North Sea [17]. In 1988, Al-Marhoun (1988) proposed an equation for predicting OFVF using 160 data collected from the Middle East [18]. In 1992, Dokla and Osman (1992) presented an equation for predicting OFVF using 51 data collected from the UAE [19]. In 1993, Petrosky and Farshad (1993) proposed an equation for predicting OFVF using 90 data collected from the Gulf of Mexico [20]. The empirical relationships of BPP and OFVF, along with data and equations for prior researchers are respectively listed in Table 1.

Table 1. Published correlations that predict OFVF for crude oil.

| Authors | Year | Origin | Data No. | Correlation |
|---|---|---|---|---|
| Standing | 1947 | California | 105 | $OFVF = a1 + a2\left[R_s\left(\gamma_g/\gamma_o\right)^{a3} + a4T\right]^{a5}$ a1=0.972, a2=1.472e-4, a3=0.5, a4=1.25, a5=1.175 |
| Vazquez and Beggs | 1977 | World wide | 500 | $OFVF = 1 + a1R_s + (T-520)\left(\frac{API}{\gamma_g}\right)(a2 + a3R_s)$ API<=30: a1=4.677e-4, a2=1.75e-5, a3=-1.811e-8 API>30: a1=4.67e-4, a2=1.1e-5, a3=1.337e-9 |
| Glaso | 1980 | North Sea | 41 | $OFVF = 1 + 10^{[a1+a2\log(G)-a3(\log(G))^2]}$ $G = R_s\left(\frac{\gamma_g}{\gamma_o}\right)^{a4} + a5T$ a1= -6.58511, a2= 2.91329, a3=0.27683, a4=0.526, a5=0.968 |
| Al-Marhoun | 1988 | Middle East | 160 | $OFVF = a1 + a2(T+460) + a3M + a4M^2$ $M = R_s^{a5}\gamma_g^{a6}\gamma_o^{a7}$ a1=0.497069, a2=0.862963e-3, a3=0.182594e-2, a4= 0.318099e-5, a5= 0.74239, a6= 0.323294, a7=-1.20204 |
| Dokla & Osman | 1992 | U.A.E | 51 | $OFVF = a1 + a2(T+460) + a3M + a4M^2$ $M = R_s^{a5}\gamma_g^{a6}\gamma_o^{a7}$ a1=0.431935e-1, a2=0.156667e-2, a3=0.139775e-2, a4=0.380525e-5, a5= 0.773572, a6= 0.404020, a7=-0.882605 |
| Petrosky and Farshad | 1993 | Gulf of Mexico | 90 | $B_o = a1 + a2\left[R_s^{a3}\left(\gamma_g^{a4}/\gamma_o^{a5}\right) + a6T^{a7}\right]^{a8}$ a1=1.0113, a2=7.2046e-5, a3=0.3738, a4=0.2914, a5=0.6265, a6=0.24626, a7=0.5371, a8=3.0936 |

In the experimental correlations, the performance accuracy was very low and unacceptable. Therefore, researchers have started using artificial intelligence in recent years [20-23] for example; Analysis of crude-oil desalting system [51], velocity prediction in sewer pipes [64]; bed load sediment transport estimation in a clean pipe [65]; monthly inflow prediction [66]; predicting sediment transport in clean pipes [67] and estimate velocity at limit of deposition in storm sewers [68].

In this regard, many researchers using artificial intelligence were able to predict the value of OFVF, some of which we report:

A large number of researchers including Gharbi and Elsharkawy (1997) [24], Gharbi and Elsharkawy (1999) [25], Boukadi et al. (1999) [26], Osman et al. (2001) [27], Al-Marhoun and Osman (2002) [28], Goda et al. (2003) [29], Moghadam et al. (2011a) [30] and Asadisaghandi and Tahmasebi (2011) [31] in order to predict OFVF based on artificial neural network (ANN) were able to make a good prediction in the field.

Other researchers have used combinations of algorithms or several algorithms to predicted OFVF, including the following: Elsharkawy (1998) [32] were used radial basis function (RBF)-ANN algorithm, Malallah et al. (2006) [33] were used of alternating conditional expectations (ACE) algorithm, El-Sebakhy et al. (2009) [34] were used support vector regression (SVR) algorithm, Dutta and Gupta were used genetic (GA)-ANN algorithm, Khoukhi (2012) [35] used GA-ANN and GA-ANFIS algorithms, Farasat et al. (2013) [36] were used SVM algorithm, Rafiee-Taghanaki et al. (2013) [37] were used gravitational search algorithm (GSA) – least squares support vector machine (LSSVM) algorithm and Karimnezhad et al. (2014) [38] were used GA algorithm.

Many researchers using ANN, ANFIS, RBF, SVM, SVR, GA, GSA and LSSVM algorithms to predict OFVF have been able to provide models that work better than the experimental correlations. In this paper, we intend to combine MLP-ABC and MLP-FF methods to construct a vigorous model for determining OFVF as a function of input data. Furthermore, the introduction of these recombination algorithms in the field of data forecasting is important for being implemented in crude oil data worldwide.

## Methodology

### Work Flow

Figure 1 shows a workflow diagram that shows all steps in a quick scan to construct a model to determine OFVF. These steps are as follows: data collection, describing of the variables, and data normalization. Through data normalization process all data variables were normalized to range between +1 and -1 by applying Eq. (2).

$$x_i^l = \left(\frac{x_i^l - xmin^l}{xmax^l - xmin^l}\right) * 2 - 1 \quad (2)$$

Where;

$x_i^l$ = the value of attribute l for data records I;

$xmin^l$ = the minimum value of the attribute l among all the data records in the dataset; and,

$xmax^l$ = the maximum value of the attribute l among all the data records in the dataset.

Then we verify the data after normalization and then divide it into two parts: training and testing. At this time, we used 70% of the data for train and 30% of the data for test. Finally, the set of calculated outputs of each method is compared with experimental models using computational error and then the best result is obtained.

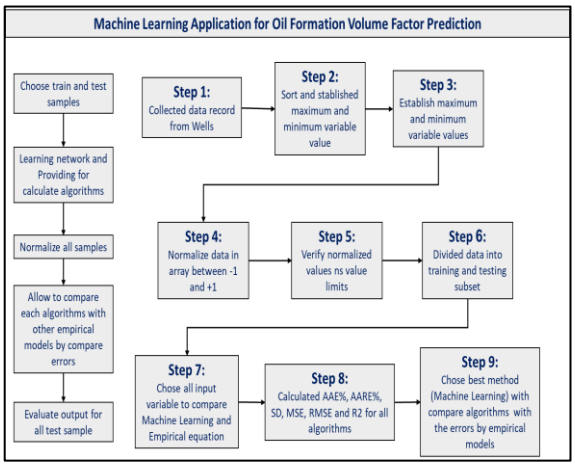



Figure 1. schematic diagram of the workflow sequence

### Machine Learning Algorithm

Today, artificial intelligence is rooted in various industries and sciences. It has found a wide variety of applications in various fields. The oil and gas industry, which has long been the focus of the whole world, and the reason is that the extracted oil and gas has changed the world. Many people did a lot of work to optimize and find important and key parameters in the oil and gas industry, for example in the following areas have been addressed: prediction flow rate of orifice & choke flow [22-23; 39-40, 44]; prediction of casing collapse based on shear modulus & geomechanical approach [41, 43]; prediction of bubble point pressure [42].

#### Artificial Neural Network

#### Multilayer Perceptron (MLP) Algorithm

One of the best up-to-date tools in the world is to create complex nonlinear relationships between sets and create a black box, artificial neural network [45; 49-50]. This ANN algorithm covers a wide range of methods used in various industries. Important factors in choosing ANN type are selecting attributes (i.e., input variables to be considered), network architecture (number of layers and nodes), transferring functions between layers, and selecting training algorithm to optimize their prediction performance [46]. One of the most widely used neural networks is the multilayer perceptron (MLP), which is a versatile and flexible neural network that is suitable for all data sets (large and small) [47]. This network was used to predict OFVF. One of the methods that teaches MLP and causes data aggregation is the Levenberg-Marquardt (LM) algorithm. Using the Levenberg-Marquardt (LM) algorithm, which is implemented in MLP, in conjunction with two optimization algorithms named artificial bee colony (ABC) and Firefly algorithm (FF), two recombinant algorithms can be created. Figure 2 shows the structure of MLP.

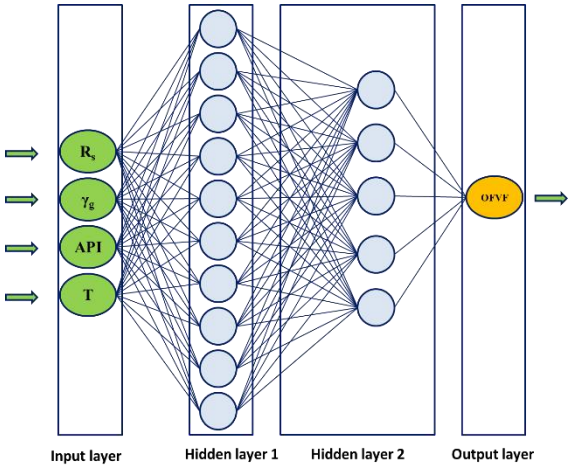



Figure 2. Schematic of MLP algorithm.

**Firefly (FF) Algorithm**

Fireflies are a species of beetles that emit of themselves a green or yellow light. The fireflies have a great tendency to move toward, which makes them brighter than they were. The factors affecting receiving light from a source are the distance between the fireflies, the ability of ambient in light absorption, the type of light source, and the amount of emitting light from the emitting source.

Firefly algorithm (FF) is an optimization method striving to find the optimal solution for the problem by simulating the fireflies' behavior. The flow diagram for FF algorithm is displayed in Figure 3. The procedure of optimization method is described in the following [51-52].

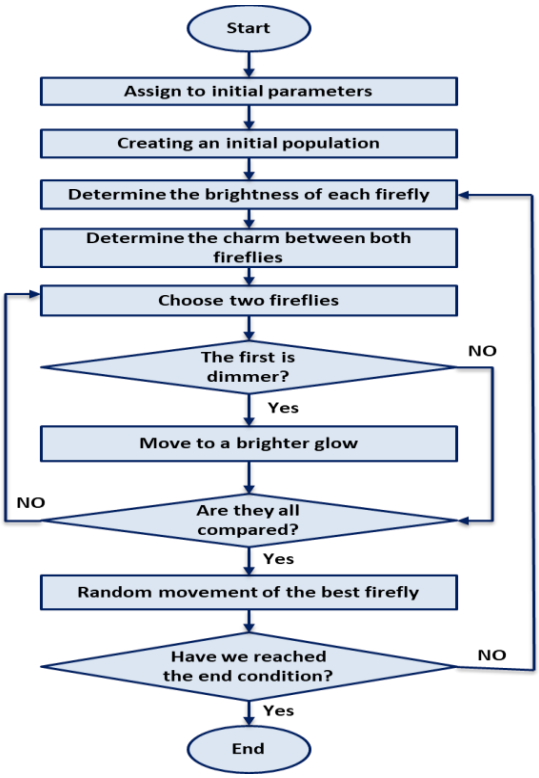



Figure 3. FF algorithm flowchart [69].

The first stage is setting the value for initial parameters in order to the FF algorithm to start. Required parameters for this algorithm are: number of fireflies (n), the number of repetitions (t), random vector coefficient (α), light absorption coefficient (γ), upper line (max) and lower line (min).

After this stage, a population of n numbers of fireflies is created with random values. The brightness of a firefly is considered as its fitness amount [68]. The brightness amount is determined based on the problem type that is going to be optimized and the selected fitness function for that problem. The charm between each two fireflies can be determined using Eq. (3). This equation displays that how charm is the j firefly for the I firefly [68].

$$\beta_{ij} = \beta_j \, . \, e^{-\gamma \, r_{ij}{}^m} \tag{3}$$

Where $\gamma$ is the amount of ambient light absorption, the variable $m$ is the light source that can received one of the three values 0, 1, and 2, and $r_{ij}$ is the Euclidean distance that can be determined using Eq. (4).

$$\mathrm{R_{ij}} = \sqrt{(\mathrm{x_i} - \mathrm{x_j})^2 + (\mathrm{y_i} - \mathrm{y_j})^2} \tag{4}$$

After computation of the distance between fireflies and determination of charm between fireflies' pairs, if a firefly sees that the other firefly in their pair is brighter than itself, then it will move towards that brighter firefly [68]. The movement of dimmer firefly toward the brighter firefly is calculated using Eq. (5).

$$\mathrm{X_i'} = \mathrm{x_i} + \delta_{\mathrm{ij}}(\mathrm{x_j} - \mathrm{x_i}) + \alpha\varepsilon_{\mathrm{i}} \tag{5}$$

Where the variable $\alpha$ is the random vector coefficient and takes a constant value, $\varepsilon$ takes a small value, and $\mathrm{x_i'}$ represents the new position of the firefly i.

All the fireflies in the population move towards the best firefly, while the best firefly moves randomly. The random movement of the best firefly can be obtained using Eq. (6).

$$x'_{Best} = x_{Best} + \alpha\varepsilon \tag{6}$$

**Artificial Bee Colony (ABC) Algorithm**

Artificial bee colony algorithm (ABC) is a simulation of the bee groups' behavior in searching for food. Bees are divided into three categories: i) worker bees which go toward the pre-determined food sources ii) pioneer bees which perform a random search for a food source iii) search bees which stay in the dance area to make a decision on the selection of a source food [53-55]. Figure 4 displays the flowchart of ABC algorithm.

Working with this algorithm, we initially specify the number of initial populations of the worker and pioneer bees as well as the main parameters including cost function, trial index, and problem range and allowable limit for the index trial.

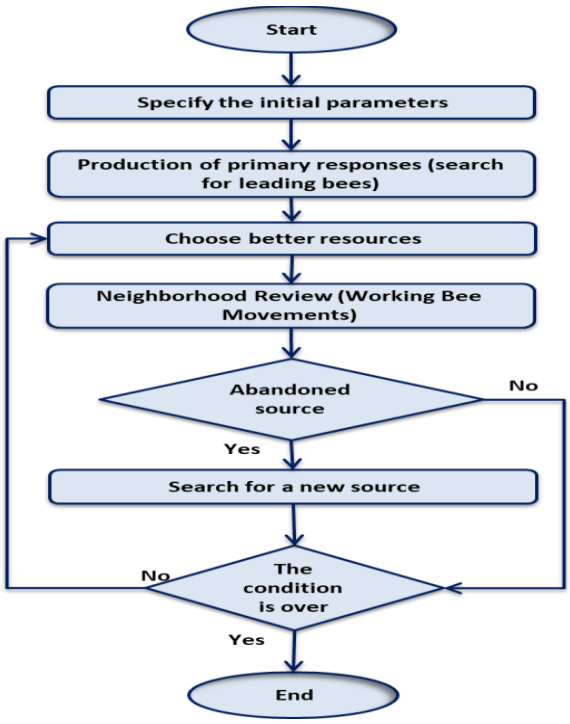



Figure 4. ABC algorithm flowchart.

At this stage, the pioneer bees start searching randomly and achieve the proper sources. The pioneer bees should not go beyond the specified range. The specified range can be obtained using Eq. (7), as follows:

$$X_{ij} = X_{jmin} + r\left(X_{jmax} - X_{jmin}\right) \quad (7)$$

Where in Eq. (7), $X_{ij}$ is the ith response in jth dimension, r represents a random number from o to 1, and $min$ and $max$ are the lower the upper limits, respectively.

In this part, more bees are allocated to the sources with higher suability and eliminate few percent of the sources with lower suitability.

After, the cost function is calculated, and then, based on calculated cost function, the sources' performance is calculated (Eq. (8)).

$$fit(x_i) = \begin{cases} \dfrac{1}{1+f(x_i)} & f(x_i) \le 0 \\ 1+|f(x_i)| & f(x_i) \ge 0 \end{cases} \quad (8)$$

Then, the employed bees move toward the sources found by the pioneer bees. The more efficient the sources, the more bees are allocated to them. The movement of bees is obtained by Eq. (9).

$$X_{ij}(t+1) = X_{ij}(t) + r\left(X_{ij}(t) - X_{kj}(t)\right) \quad (9)$$

Where, $X_{ij}$ represents the position of bee, $X_k$ represents the random selection of an employed bee, $t$ is the t bee, $j$ is the response dimension, and $r$ is a random number selected between 1 and -1.

The sending of employed bees is performed using following methods:

1. Sending more specified bees to better sources and sending specified number of bees to normal sources.
2. Sending bees according to Roulette cycle (on the basis of the performance probability of each source) using Eq. (10).

$$P_i = \frac{fit(x_i)}{\sum_{K=1}^{n} fit(x_i)} \quad (10)$$

Where, $P_i$ is the probability of the $i$ source's selection and $fit(x_i)$ is the suitability value of the $x_i$ source.

Abandoned sources are defined as the sources wasting the computational power, which makes efforts to convince them not to work. To determine the abandoned source, the trial index must be checked for each source. If the trial index is greater or equal to admissible limit and that source is not the best problem's solution, the source is considered as an abandoned source. Indeed, to implement that, a counter need to be set, the value of which increases with each visit to that source, and if the number of visits to that source becomes greater than the specified limit and there is no improvement, then the source will be announce as an abandoned source, and it will no longer considered as a suitable source.

For search a new source instead of abandoned source should be:

- Global search
- The sources having been abandoned must be replaced by new sources
  In fact, the pioneer bees, using Eq. (11), find another initial response once again.

If the number of iterations meets a certain level or reaches a certain value of cost function, the algorithm ends.

## Hybrid Models

### FF-MLP Hybrid Algorithm

The neural network has always been widely used as a problem-solving tool. In this study, we used a combination of MLP with optimization methods to better compare the results [56]. We did this combination with the same two methods FF and ABC that were used in sections before to make a logical comparison of the results. Figure 5 shows the flowchart of the FF-MLP method.

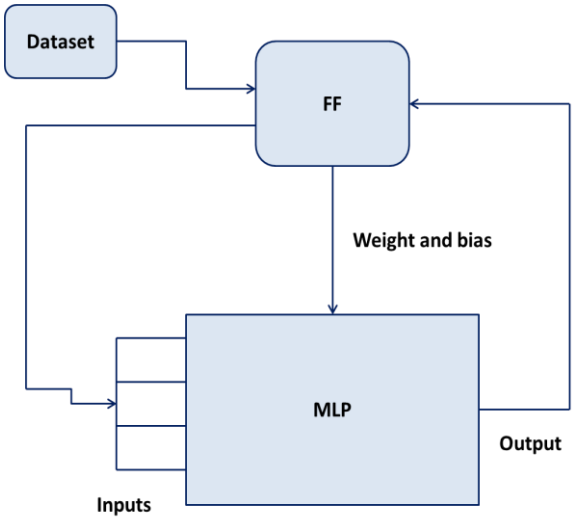



Figure 5. FL-MLP algorithm flowchart.

In this method, we must first normalize the data. In this section, we used exactly the same two sets of training and testing datapoints. In this combined method, in principle, the network training operation is performed by FF, and in fact, the weights and bias of the perceptron's in each layer are obtained by the FF optimization method. As shown in the flowchart, the MLP network has been used as a cost function for the FF algorithm. What is the error of new weights and biases?

After the optimal weights are obtained, we create the network with those weights and this time we give the test data to it to find out how the network has been trained and what is the amount of error.

The interaction between the network and the FF algorithm is such that we consider all the weights of the layers as well as the bias as a presenter or vector, and the FF algorithm creates its population according to the length of this presentation. In fact, each member of the population is a presentation of the length of all weights and bias. The weights of each layer are in the form of a matrix. For example, if the first layer it has 4 inputs and 8 perceptron's, so a 4x8 matrix represents the weights of the first layer and an 8x1 vector represents the first layer bias. The next layers are the same. Table 2 presents the parameters related to the implementation of this method.

Table 2. MLP-FF algorithm parameters

| FF control | Value | MLP algorithm | Value |
|---|---|---|---|
| Fireflies No. | 50 | Activation function | |
| Attraction coefficient | 2 | Hidden layer neuron No. | 10 & 5 |
| Light absorption | 1 | Activation function | tansig |
| Dependent variables No. | 1 | Number of hidden layers | 2 |
| Uniform mutation | 0.05 | Input neurons | 7 |
| input variables No. | 7 | | |
| Iterations No. | 100 | | |
| Fireflies No. | 50 | | |
| Mutation coefficient | 0.98 | | |
| Mutation coefficient | 0.2 | | |

**ABC-MLP hybrid algorithm**

This method is the same as the previous method, only ABC is used instead of FF. Use of this optimization method is only for comparison of the methods. Figure 6 shows the flowchart method. Table 3 presents the parameters related to the implementation of this method.

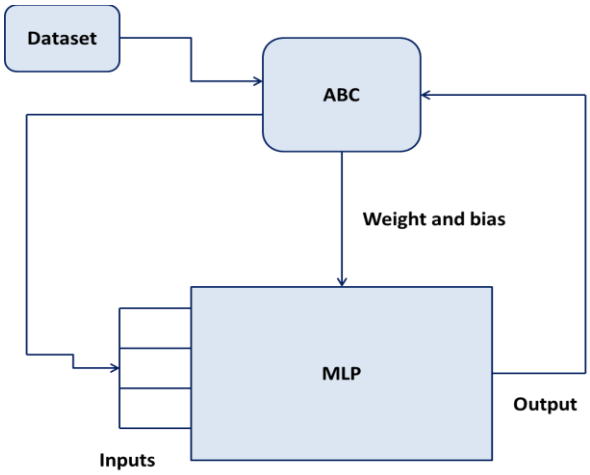



Figure 6. ABC-MLP algorithm flowchart.

Table 3. MLP-ABC algorithm parameters

| ABC parameter | Value | MLP algorithm | Value |
|---|---|---|---|
| Bees No. | 100 | Activation function | purelin |
| Scout bees No. | 50 | Hidden layers neuron No. | 10 & 5 |
| Trial upper limit | 60 | Activation function | tansig |
| Dependent variables No. | 1 | Hidden layers No. | 2 |
| Iterations No. | 100 | Input layers neuron No. | 7 |
| Bees No. | 50 | | |
| Input variables No. | 7 | | |

## Data Collection & Data Analysis

The data used in this study to determine an optimal model for OFVF from Moghadam et al. (2011b) [57], Omar and Todd (1993) [58], Dokla and Osman (1992) [19], Al-Marhoun (1998) [18], Ghorbani et al. (2020a) [42], Gharbi & Elsharkawy (1997) [24], Mahmood and Al-Marhoun (1996) [59] and Ganji et al. (2014) [60] which is a mixture of data samples from different parts of the World. Table 3 summarizes the statistical distributions of these four data variables for the 565 data records compiled.

Table 4. Data record statistical characterization of the variables in this study.

| Parameters | | T | Rs | γg | API | OFVF |
|---|---|---|---|---|---|---|
| Units | | (F) | SCF/STB | - | - | RB/STB |
| N | Valid | 565 | 565 | 565 | 565 | 565 |
| | Missing | 0 | 0 | 0 | 0 | 0 |
| Mean | | 193 | 637 | 1.20 | 35 | 1.44 |
| Std. Deviation | | 52 | 406 | 0.46 | 5.9 | 0.27 |
| Variance | | 2707 | 164926 | 0.21 | 35.6 | 0.07 |
| Minimum | | 74 | 26 | 0.159 | 19.4 | 1.032 |
| Maximum | | 306 | 2496 | 3.4445 | 56.5 | 2.916 |

The parameters that affect the determination of OFVF are: T, Rs, γg and API, which are interpreted in detail for data related to samples collected from worldwide. In order to describe the input data, the contour plot diagram (Figure 3) is used. The description of this diagram is as follows:

In the Figure 7 for the T parameter, about 70% of the data are related to T < 200 F0 and the remaining 30% are related to T > 200 F0. For T <100 F0 (OFVF < 1.3 bbl/STB), the T data contains about 17% of the data and> 300 F0 (2 < OFVF < 2.7 bbl/STB) about 5% of the data. For the Rs parameter, approximately 45% of the data are related to Rs <400 Scf / STB and the remaining 14% are related to Rs > 1200Scf / STB.

For the γg parameter, approximately 70% of the data is related to γg <1.5 and 26% of the data is related to γg> 1.5. of these, 35% of the data are related to 0.5 < γg <1. For the API parameter, approximately 80% of the data is related to 30 <API <40 and of these, 17% of the data is related to 20 <API <30 (OFVF < 1.2 bbl/STB). Histogram plots (Figure 8) confirm the positively skewed character of the Swir and K distributions.

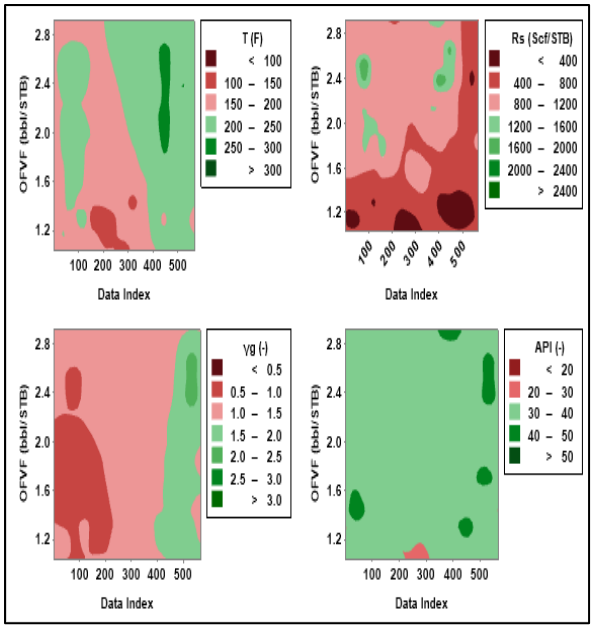



Figure 7. Contour plots of OFVF versus data index for the input variables

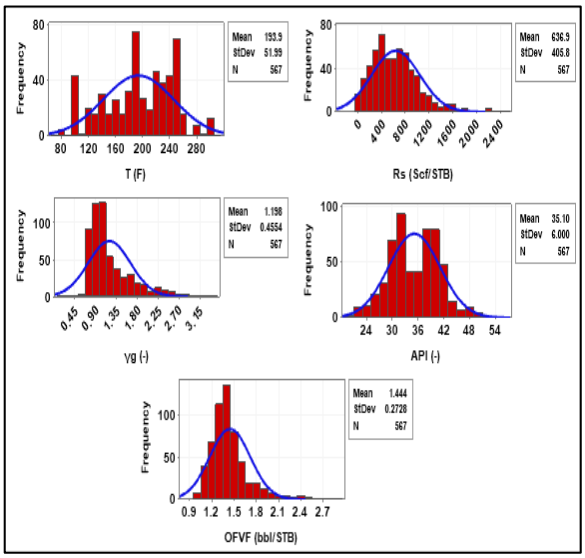



Figure 8. Histogram of OFVF versus data index for the input variables

## Results and Discussion

Performance accuracy assessment of the two-hybrid machine-learning-optimization algorithms and other empirical equations is computational errors between measured and predicted OFVF. The statistical measures of prediction accuracy used are percentage deviation (PDi), average percentage deviation (APD), average absolute percentage deviation (AAPD), standard deviation (STD), mean squared error (MSE), root mean square error (RMSE), and coefficient of determination (R2). The computation formulas for these statistical accuracy measures are expressed in Eq. (12) to Eq. (19).

Percentage difference (PD):

$$\mathrm{PD_i} = \frac{\xi_{(\mathrm{Measured})} - \xi_{(\mathrm{Predicted})}}{\xi_{(\mathrm{Measured})}} x\ 100 \quad (12)$$

Average percent deviation (APD): (13)

$$APD = \frac{\sum_{i=1}^{n} PD_i}{n}$$

Absolute average percent deviation (AAPD):

$$AAPD = \frac{\sum_{i=1}^{n} |PD_i|}{n} \quad (14)$$

Standard Deviation (SD):

$$SD = \sqrt{\frac{\sum_{i=1}^{n} (D_i - Dimean)^2}{n-1}} \quad (15)$$

$$Dimean = \frac{1}{n}\sum_{i=1}^{n}\left(\xi_{Measured_i} - \xi_{Predicted_i}\right) \quad (16)$$

Mean Square Error (MSE):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(\xi_{Measured_i} - \xi_{Predicted_i}\right)^2 \quad (17)$$

Root Mean Square Error (RMSE):

$$RMSE = \sqrt{\text{MSE}} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - y_i)^2}{n}} \quad (18)$$

Where;

n = number of data records;

$x_i$ = measured dependent variable value for the ith data record; and,

$y_i$ = predicted dependent variable value for the ith data record.

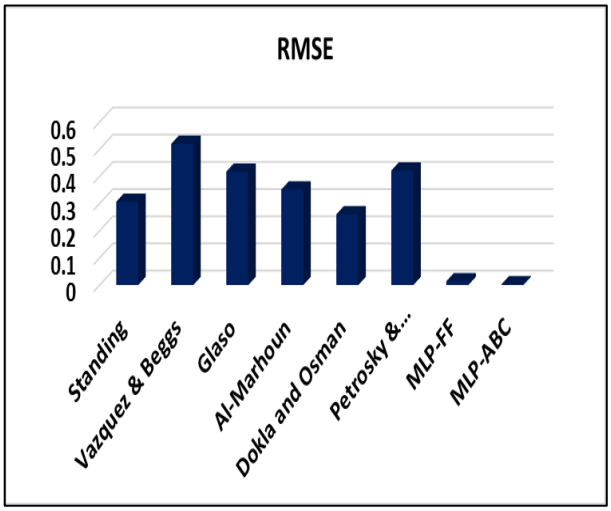
Coefficient of Determination (R2):

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(\xi_{Predicted_i} - \xi_{Measured_i})^2}{\sum_{i=1}^{N}(\varphi_{Predicted_i} - \frac{\sum_{I=1}^{n}\xi_{Measured_i}}{n})^2} \quad (19)$$

Table 5 and Figure 9 reveal that all two-hybrid machine-learning-optimizer models evaluated, MLP-FF, MLP-ABC and plus empirical equation, deliver accurate and credible OFVF prediction for test data. The MLP-ABC model is the least accurate, whereas the models by providing OFVF prediction accuracy in terms of RMSE < 0.002573 bbl/STB and R2 = 0.998 for this test dataset.

Table 5 :OFVF Prediction performance compared for hybrid models

| Performance of the developed regression models based on five statistical error metrics for OFVF (Train, Test and Total Data) | | | | | | | |
|---|---|---|---|---|---|---|---|
| | Authors | APD% | AAPD% | SD | MSE | RMSE | R^2 |
| Train Data (398 Dataset) | Standing | 4.352 | 5.01341 | 4.3141 | 0.0165962 | 0.1288262 | 0.8754 |
| | Vazquez & Beggs | 12.882 | 21.954 | 21.7354 | 0.1554528 | 0.3942750 | 0.8246 |
| | Glaso | 12.882 | 13.0756 | 12.7550 | 0.0778458 | 0.2790087 | 0.6489 |
| | Al-Marhoun | 11.274 | 11.6543 | 11.1639 | 0.0556283 | 0.2358564 | 0.7131 |
| | Dokla and Osman | -8.716 | 8.76535 | 8.6315 | 0.0234548 | 0.1531495 | 0.9333 |
| | Petrosky & Farshad | 13.229 | 13.4035 | 13.0981 | 0.0817121 | 0.2858533 | 0.6052 |
| | MLP-FF | -0.058 | 0.86221 | 0.2257 | 0.0001953 | 0.0139749 | 0.9967 |
| | MLP-ABC | 0.002 | 0.18167 | 0.2186 | 0.0000083 | 0.0028758 | 0.9999 |
| Test Data (167 Dataset) | Standing | 9.788 | 9.78825 | 9.7201 | 0.0938710 | 0.3063838 | 0.6699 |
| | Vazquez & Beggs | 16.669 | 25.9205 | 25.6721 | 0.2698263 | 0.5194480 | 0.6688 |
| | Glaso | 16.669 | 16.6691 | 16.5196 | 0.1739648 | 0.4170909 | 0.6116 |
| | Al-Marhoun | 12.180 | 12.1884 | 12.0819 | 0.1233881 | 0.3512664 | 0.5546 |
| | Dokla and Osman | 1.875 | 6.02602 | 2.0097 | 0.0677033 | 0.2601986 | 0.9119 |
| | Petrosky & Farshad | 16.834 | 16.8341 | 16.6829 | 0.1775154 | 0.4213257 | 0.6278 |
| | MLP-FF | 0.030 | 0.83693 | 0.7604 | 0.0002016 | 0.0142003 | 0.998 |
| | MLP-ABC | 0.005 | 0.14689 | 0.7596 | 0.0000066 | 0.0025733 | 0.998 |
| Test Data (565 Dataset) | Standing | 5.959 | 6.42474 | 5.9049 | 0.0394367 | 0.1985868 | 0.8754 |
| | Vazquez & Beggs | 14.001 | 23.1264 | 22.8964 | 0.1892587 | 0.4350388 | 0.8246 |
| | Glaso | 14.001 | 14.1378 | 13.8637 | 0.1062562 | 0.3259697 | 0.6489 |
| | Al-Marhoun | 11.542 | 11.8122 | 11.4296 | 0.0756564 | 0.2750571 | 0.7131 |
| | Dokla and Osman | -5.585 | 7.95567 | 5.5360 | 0.0365335 | 0.1911375 | 0.9333 |
| | Petrosky & Farshad | 14.294 | 14.4175 | 14.1537 | 0.1100292 | 0.3317065 | 0.6052 |
| | MLP-FF | -0.032 | 0.85474 | 0.2528 | 0.0001972 | 0.0140419 | 0.9999 |
| | MLP-ABC | 0.003 | 0.17139 | 0.2511 | 0.0000078 | 0.0027898 | 0.9967 |



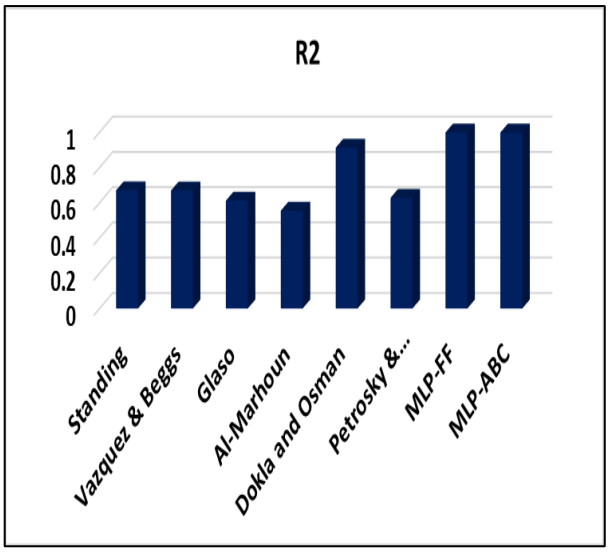




Figure 9. R2 and RMSE for empirical models and hybrid models used to predict OFVF.

Figure 10 reveal that two-hybrid machine-learning-optimizer models evaluated, MLP-FF and MLP-ABC, deliver accurate and credible OFVF prediction for test data. The MLP-ABC model is the least accurate, whereas the models by providing OFVF prediction accuracy in test dataset. The results of Figures 10 and 11 as well as Table 5 show that Dokla and Osman show better results among the experimental models with RMSE = 0.26019 for test data, which is more accurate than other models.

 DOI: **10.21608/jpme.2021.52149.1062**

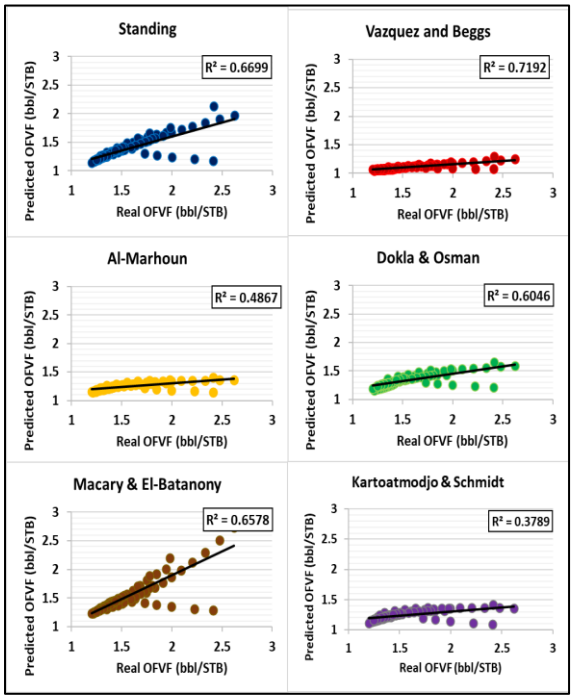
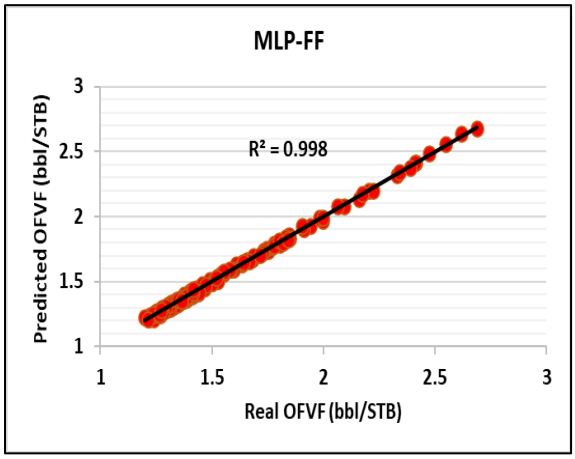



Figure 10. Cross plot of OFVF versus data index for the input variables



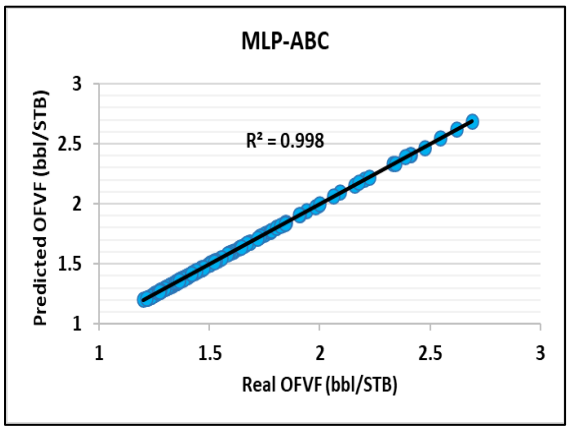




Figure 11. Cross plot of OFVF versus data index for the input variables: hybrid models (MLP-ABC and MLP-FF) with the four independent T, Rs, γg and API for the worldwide sample.

Using Pearson's correlation coefficient, which is in the range -1 (perfect negative correlation) or +1 (perfect positive correlation) with a zero-value indicating a total lack of correlation, the sensitivity of each parameter to the OFVF showed. Spearman's correlation coefficient (ρ) is calculated for ranked data using Eq. (20).

$$\rho = \frac{\sum_{i=1}^{n}(V_i - \bar{V})(Z_i - \bar{Z})}{\sqrt{\sum_{i=1}^{n}(V_i - V)^2 \sum_{i=1}^{n}(Z_i - \bar{Z})^2}} \quad (20)$$

Where;

Vi = the value of data record i for input variable V;

$\bar{V}$ = the average value of the input variable V;

Zi = the value of data record i for input variable Z;

$\bar{Z}$ = the average of the input variable Z; and,

n = the number of data points in the population.

Figures 12 displays the p values for the relationships between OFVF and the seven input variables considered.

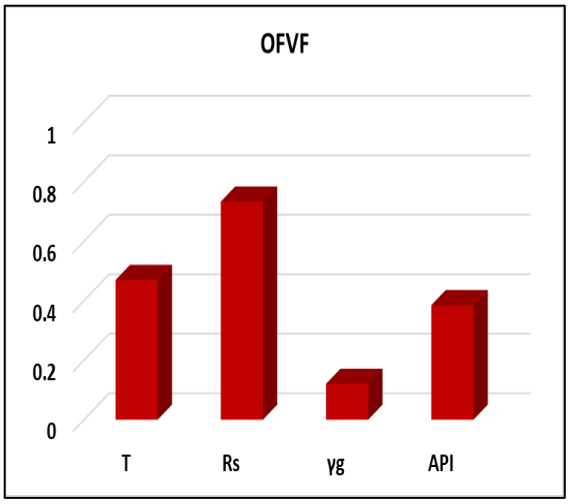



Figure 12. Spearman's correlation coefficient relationships ($\boldsymbol{\rho}$) for OFVF prediction

As shown in the figure, based on 565 available data from around the world and input variables from this data, all input parameters have a positive effect on OFVF prediction, which are as follows: Rs> T> API> γg and these results show that the effect of Rs is more than other input variables and the effect of $\gamma_g$ is the lowest.

## Conclusion

In this research, 565 data collected from worldwide have been used. The input data for forecasting OFVF, solution gas oil ratio (Rs), gas specific gravity (γg), API gravity (API0) (or oil density γo), and temperature (T). These input variables are very important for predicting OFVF because these variables are routinely taken in the oil industry and through this data can be an important parameter that is important for the development of oil and gas reservoirs. Calculated without spending time and money.

In this paper, the combination of MLP method, which is a network of artificial intelligence, with ABC and FF optimization methods has been used, which are the new combination methods for predicting OFVF.

For MLP-ABC model OFVF prediction accuracy in terms of RMSE < 0.002573 bbl/STB and R2 = 0.998 for this test dataset. After comparing the results of the experimental equations, it was concluded that the Dokla and Osman model gives the best results.

Based on Spearman's correlation coefficient relationships all input parameters have a positive effect on OFVF prediction, which are as follows: Rs> T> API> γg and these results show that the effect of Rs is more than other input variables and the effect of $\gamma_g$ is the lowest.

**Nomenclature**

| | | |
|---|---|---|
| ABC | = | Artificial bee colony algorithm |
| ACE | = | Alternating conditional expectations algorithm |
| ANFIS | = | Neuro-fuzzy algorithm |
| ANN | = | Artificial neural network |
| BPP | = | Bubble point pressure |
| EOR | = | Enhance oil recovery |
| FF | = | Firefly algorithm |
| GA | = | Genetic algorithm |
| GSA | = | Gravitational search algorithm |
| ICA | = | Imperialist Competitive Algorithm |
| KNN | = | K-nearest neighbor |
| LSSVM | = | Least squares support vector machine |
| max | = | Upper line |
| min | = | Lower line |
| MLP | = | Distance-Weighted K-nearest neighbor |
| MSE | = | Mean square error |
| MSE | = | Mean square error |
| N1 | = | The number of data points in the population |
| n | = | Number of fireflies |
| OFVF | = | Oil formation volume factor |
| PDi | = | Percentage difference |
| R2 | = | Coefficient of Determination |
| RBF | = | Radial basis function |
| RMSE | = | Route mean square error |
| SD | = | Standard deviation |
| SVM | = | Support vector machine algorithm |
| SVR | = | Support vector regression algorithm |
| t | = | The number of repetitions |
| Vi | = | The value of data record i for input variable V |
| Zi | = | The value of data record i for input variable Z |
| ρ | = | Spearman's correlation coefficient |
| $\hat{y}_i$ | = | Predicted value of $i^{th}$ testing data records |
| $C_p$ | = | Predicted value of the dependent variable for the testing data record |
| $C_t$ | = | Values of dependent variable for the ttb nearest neighbor |
| $D_i$ | = | Euclidean distance |
| $\overline{V}$ | = | The average value of the input variable V |
| $X_{ij}$ | = | Training samples |
| $X_j$ | = | Testing samples |
| $\overline{Z}$ | = | The average of the input variable Z |
| $w_i$ | = | Weight variable of dependent variable |
| $x_i^l$ | = | The value of attribute $l$ for data record i |
| $xmax^l$ | = | The maximum value of the attribute $l$ among all the data records in the dataset |
| $xmin^l$ | = | The minimum value of the attribute $l$ among all the data records in the dataset |
| $y_i$ | = | Measured value of $i^{th}$ testing data records |
| α | = | Random vector coefficient |
| Γ | = | Light absorption coefficient |

Acknowledgments

The authors are grateful to Ms. Kalaei for his technical support and efforts to collecting the data needed for this study.

## Conflicts of interest

"There are no conflicts to declare".

**References**

[1] Adeleke, N., Ityokumbul, M. T., & Adewumi, M. Blockage detection and characterization in natural gas pipelines by transient pressure-wave reflection analysis. Spe Journal, 2012, 18(02), 355-365. Doi: https://doi.org/10.2118/160926-PA.

[2] Dixit, N., Zeng, D. L., & Kalonia, D. S. Application of maximum bubble pressure surface tensiometer to study protein–surfactant interactions. International journal of pharmaceutics, 2012, 439(1-2), 317-323. Doi: https://doi.org/10.1016/j.ijpharm.2012.09.013.

[3] Ikiensikimama, S. S., Ajienka, J. A. J. Impact of PVT correlations development on hydrocarbon accounting: the case of the Niger Delta. Journal of Petroleum Science and Engineering, 2012, 81, 80-85. Doi: https://doi.org/10.1016/j.petrol.2011.12.017.

[4] Ikiensikimama, S. S., & Ogboja, O. New bubblepoint pressure empirical PVT correlation. Paper presented at the Nigeria Annual International Conference and Exhibition, 2009. Doi: https://doi.org/10.2118/128893-MS.

[5] Li, H., & Yang, D. T. Phase behaviour of C3H8/n-C4H10/heavy-oil systems at high pressures and elevated temperatures. Journal of Canadian Petroleum Technology, 2013, 52(01), 30-40. Doi: https://doi.org/10.2118/157744-PA.

[6] Ghorbani, H., Moghadasi, J., Dashtbozorg, A., & Kooti, S. Developing a New Multiphase Model for Choke Function Relation for Iran's Gas Wells. American Journal of Oil and Chemical Technologies, 2017b, 194-202.

[7] Mohammadian, N., & Ghorbani, H. An investigation on chemical formation damage in Iranian reservoir by focus on mineralogy role in shale swelling potential in Pabdeh and Gurpi formations. Advances in Environmental Biology, vol:9(4), 2015, 161-166, Doi: http://www.aensiweb.net/AENSIWEB/aeb/aeb/2015/March/161-166.pdf.

[8] Darvishpour, A., Sefiabad, M. C., Wood, D. A., & Ghorbani, H. Wellbore stability analysis to determine the safe mud-weight window for sandstone layers. Petroleum Exploration and Development, 2019, 46(5), 1031-1038. Doi: https://doi.org/10.1016/S1876-3804(19)60260-0.

[9] Davoodi S, SA AR, Soleimanian A, Jahromi AF. Application of a novel acrylamide copolymer containing highly hydrophobic comonomer as filtration control and rheology modifier additive in water-based drilling mud. Journal of Petroleum Science and Engineering. 2019 Sep 1; 180:747-55. Doi: https://www.sciencedirect.com/science/article/pii/S0920410519304061.

[10] Mohamadian, N., Ghorbani, H., Wood, D. A., & Hormozi, H. K. Rheological and filtration characteristics of drilling fluids enhanced by nanoparticles with selected additives: an experimental study. Advances in Geo-Energy Research, vol:2(3), 2018, 228-236, Doi: https://core.ac.uk/download/pdf/229382392.pdf.

[11] Ghorbani, H., & Moghadasi, J. Development of a New Comprehensive Model for Choke Performance Correlation in Iranian Oil Wells. Advances in Environmental Biology, 2014a, vol:8(17), 877-882, Doi: http://www.aensiweb.net/AENSIWEB/aeb/aeb/September%202014/877-882.pdf.

[12] Ghorbani, H., Moghadasi, J., Mohamadian, N., Mansouri Zadeh, M., Hezarvand Zangeneh, M., Molayi, O., Kamali, A. 2014b. Development of a New Comprehensive Model for Choke Performance Correlation in Iranian Gas Condensate Wells, Advances in Environmental Biology 8(17):308-313. doi: https://www.researchgate.net/publication/308786325_Development_of_a_New_Comprehensive_Model_for_Choke_Performance_Correlation_in_Iranian_Gas_Condensate_Wells

[13] Ghorbani, H., Moghadasi, J., Dashtbozorg, A., & Abarghoyi, P. G. The Exposure of New Estimating Models for Bubble Point Pressure in Crude Oil of One of The Oil fields in Iran. American Journal of Oil and Chemical Technologies, 2017a, 178-193.

[14] Mohamadian, N., Ghorbani, H., Wood, D. A., & Khoshmardan, M. A. A hybrid nanocomposite of poly (styrene-methyl methacrylate-acrylic acid)/clay as a novel rheology-improvement additive for drilling fluids. Journal of Polymer Research, 2019, vol:26(2), 33, Doi: https://link.springer.com/article/10.1007%2Fs10965-019-1696-6.

[15] Standing, M. A pressure-volume-temperature correlation for mixtures of California oils and gases. Paper presented at the Drilling and Production Practice, 1947. Doi: https://www.onepetro.org/conference-paper/API-47-275.

[16] Vazquez, M., & Beggs, H. D. Correlations for fluid physical property prediction. Paper presented at the SPE Annual Fall Technical Conference and Exhibition, 1997. Doi: https://doi.org/10.2118/6719-MS.

[17] Glaso, O. Generalized pressure-volume-temperature correlations. Journal of Petroleum Technology, 1980, 32(05), 785-795. Doi: https://doi.org/10.2118/8016-PA.

[18] Al-Marhoun, M. A. PVT correlations for Middle East crude oils. Journal of Petroleum Technology, 1988, 40(05), 650-666. Doi: https://doi.org/10.2118/13718-PA.

[19] Dokla, M., & Osman, M. Correlation of PVT properties for UAE crudes (includes associated papers 26135 and 26316). SPE formation evaluation, 1992, 7(01), 41-46. Doi: https://doi.org/10.2118/20989-PA.

[20] Petrosky Jr, G., & Farshad, F. Pressure-volume-temperature correlations for Gulf of Mexico crude oils. Paper presented at the SPE annual technical conference and exhibition, 1993. Doi: https://doi.org/10.2118/26644-MS.

[21] AlAjmi, M. D., Alarifi, S. A., & Mahsoon, A. H. Improving multiphase choke performance prediction and well production test validation using artificial intelligence: a new milestone. Paper presented at the SPE digital energy conference and exhibition, 2015. Doi: https://doi.org/10.2118/173394-MS.

[22] Choubineh, A., Ghorbani, H., Wood, D. A., Moosavi, S. R., Khalafi, E., & Sadatshojaei, E. Improved predictions of wellhead choke liquid critical-flow rates: modelling based on hybrid neural network training learning based optimization. Fuel, 2017, 207, 547-560. Doi: https://doi.org/10.1016/j.fuel.2017.06.131.

[23] Ghorbani H, Moghadasi J, Wood DA. Prediction of gas flow rates from gas condensate reservoirs through wellhead chokes using a firefly optimization algorithm. Journal of Natural Gas Science and Engineering, 2017c, 45: 256-271. Doi: https://doi.org/10.1016/j.jngse.2017.04.034

[24] Gharbi, R., & Elsharkawy, A. M. Neural network model for estimating the PVT properties of Middle East crude oils. Paper presented at the Middle East Oil Show and Conference, 1997. Doi: https://doi.org/10.2118/37695-MS.

[25] Gharbi, R. B., Elsharkawy, A. M., & Karkoub, M. Universal neural-network-based model for estimating the PVT properties of crude oil systems. Energy &vFuels, 1999, 13(2), 454-458. Doi: https://doi.org/10.1021/ef980143v.

[26] Boukadi, F., Al-Alawi, S., Al-Bemani, A., & Al-Qassabi, S. Establishing PVT correlations for Omani oils. Petroleum Science and Technology, 1999, 17(5-6), 637-662. Doi: https://doi.org/10.1080/10916469908949738.

[27] Osman, E., Abdel-Wahhab, O., & Al-Marhoun, M. Prediction of oil PVT properties using neural networks. Paper presented at the SPE middle east oil show, 2001. Doi: https://doi.org/10.2118/68233-MS.

[28] Al-Marhoun, M., & Osman, E. Using artificial neural networks to develop new PVT correlations for Saudi crude oils. Paper presented at the Abu Dhabi international petroleum exhibition and conference, 2002. Doi: https://doi.org/10.2118/78592-MS.

[29] Goda, H. M., El-M Shokir, E. M., Fattah, K. A., & Sayyouh, M. H. Prediction of the PVT data using neural network computing theory. Paper presented at the Nigeria annual international conference and exhibition, 2003. Doi: https://doi.org/10.2118/85650-MS.

[30] Moghadam, J. N., Salahshoor, K., & Kharrat, R. Introducing a new method for predicting PVT properties of Iranian crude oils by applying artificial neural networks. Petroleum Science and Technology, 2011a, 29(10), 1066-1079. Doi: https://doi.org/10.1080/10916460903551040.

[31] Asadisaghandi, J., & Tahmasebi, P. Comparative evaluation of back-propagation neural network learning algorithms and empirical correlations for prediction of oil PVT properties in Iran oilfields. Journal of Petroleum Science and Engineering, 2011, 78(2), 464-475. Doi: https://doi.org/10.1016/j.petrol.2011.06.024.

[32] Elsharkawy, A. M. Modeling the properties of crude oil and gas systems using RBF network. Paper presented at the SPE Asia Pacific oil and gas conference and exhibition, 1998. Doi: https://doi.org/10.2118/49961-MS.

[33] Malallah, A., Gharbi, R., & Algharaib, M. Accurate estimation of the world crude oil PVT properties using graphical alternating conditional expectation. Energy & Fuels, 2006, 20(2), 688-698. Doi: https://doi.org/10.1021/ef0501750.

[34] El-Sebakhy, E. A. Forecasting PVT properties of crude oil systems based on support vector machines modeling scheme. Journal of Petroleum Science and Engineering, 2009, 64(1-4), 25-34. Doi: https://doi.org/10.1016/j.petrol.2008.12.006.

[35] Khoukhi, A. Hybrid soft computing systems for reservoir PVT properties prediction. Computers & Geosciences, 2012, 44, 109-119. Doi: https://doi.org/10.1016/j.cageo.2012.03.016.

[36] Farasat, A., Shokrollahi, A., Arabloo, M., Gharagheizi, F., & Mohammadi, A. H. Toward an intelligent approach for determination of saturation pressure of crude oil. Fuel processing technology, 2013, 115, 201-214. Doi: https://doi.org/10.1016/j.fuproc.2013.06.007.

[37] Rafiee-Taghanaki, S., Arabloo, M., Chamkalani, A., Amani, M., Zargari, M. H., & Adelzadeh, M. R. Implementation of SVM framework to estimate PVT properties of reservoir oil. Fluid Phase Equilibria, 2013, 346, 25-32. Doi: https://doi.org/10.1016/j.fluid.2013.02.012.

[38] Karimnezhad, M., Heidarian, M., Kamari, M., & Jalalifar, H. A new empirical correlation for estimating bubble point oil formation volume factor. Journal of Natural Gas Science and Engineering, 2014, 18, 329-335. Doi: https://doi.org/10.1016/j.jngse.2014.03.010.

[39] Ghorbani H, Wood DA, Choubineh A, Tatar A, Abarghoyi PG, Madani M, Mohamadian N. Prediction of oil flow rate through an orifice flow meter: Artificial intelligence alternatives compared, Petroleum, 2018. Doi: https://doi.org/10.1016/j.petlm.2018.09.003.

[40] Ghorbani H, Wood DA, Moghadasi J, Choubineh A, Abdizadeh P, Mohamadian N. Predicting liquid flow-rate performance through wellhead chokes with genetic and solver optimizers: an oil field case study. Journal of Petroleum Exploration and Production Technology, 2019, 9(2): 1355-1373. Doi: https://doi.org/10.1007/s13202-018-0532-6.

[41] Rashidi, S., Mohamadian, N., Ghorbani, H., Wood, D. A., Shahbazi, K., & Ahmadi Alvar, M. Shear modulus prediction of embedded pressurize salt layers and pinpointing zones at risk of casing collapse in oil and gas wells. Journal of Applied Geophysics, 2020, 104205. Doi: https://doi.org/10.1016/j.jappgeo.2020.104205.

[42] Ghorbani H, Wood DA, Choubineh A, Mohamadian N, Tatar A, Farhangian H, Nikooey, A. Performance comparison of bubble point pressure from oil PVT data: Several neurocomputing techniques compared, Experimental and Computational Multiphase Flow, 2020a, 2(4): 225-246. Doi: https://doi.org/10.1007/s42757-019-0047-5.

[43] Mohamadian, N., Ghorbani, H., Wood, D. A., Mehrad, M., Davoodi, S., Rashidi, S., Soleimania, A & Kiani Shahvand, A. A geomechanical approach to casing collapse prediction in oil and gas wells aided by machine learning. Journal of Petroleum Science and Engineering, 2021, 196, 107811. Doi: https://doi.org/10.1016/j.petrol.2020.107811.

[44] Ghorbani H, Wood DA, Mohamadian N, Rashidi S, Davoodi S, Soleimanian A, Kiani Shahvand, A, Mehrad M. Adaptive neuro-fuzzy algorithm applied to predict and control multi-phase flow rates through wellhead chokes. Flow Measurement and Instrumentation. 2020b Nov 12:101849. Doi: https://doi.org/10.1016/j.flowmeasinst.2020.101849.

[45] Ali, J. Neural networks: a new tool for the petroleum industry? Paper presented at the European petroleum computer conference, 1994. Doi: https://doi.org/10.2118/27561-MS.

[46] Maimon, O., & Rokach, L. Introduction to knowledge discovery and data mining. In Data mining and knowledge discovery handbook (pp. 1-15): Springer, 2009. Doi:

https://link.springer.com/chapter/10.1007/978-0-387-09823-4_1.

[47] Bishop, C. M. Pattern recognition and machine learning: springer, 2006. Doi: https://cds.cern.ch/record/998831.

[48] Abdali, M, Mohamadian, N., Ghorbani, H., Wood, D. A. Petroleum Well Blowouts as a Threat to Drilling Operation and Wellbore Sustainability: Causes, Prevention, Safety and Emergency Response, Journal of Construction Materials, 2021. Doi: https://iconsmat.com.au/wp-content/uploads/2021/02/si1.1r.pdf

[49] Rashidi S, Mehrad M, Ghorbani H, Wood DA, Mohamadian N, Moghadasi J, Davoodi S. Determination of Bubble Point Pressure & Oil Formation Volume Factor of Crude Oils Applying Multiple Hidden Layers Extreme Learning Machine Algorithms. Journal of Petroleum Science and Engineering. 2021 Jan 20:108425. Doi: https://doi.org/10.1016/j.petrol.2021.108425.

[50] Farsi, M., Barjouei, H.S., Wood, D.A., Ghorbani, H., Mohamadian, N., Davoodi, S., Nasriani, H.R. and Ahmadi Alvar, M. Prediction of Oil Flow Rate Through Orifice Flow Meters: Optimized Machine-Learning Techniques. Measurement, 2021, p.108943. Doi: https://doi.org/10.1016/j.measurement.2020.108943.

[51] Ranaee, E., Ghorbani, H., Keshavarzian, S., Ghazaeipour Abarghoei, P., Riva, M., Inzoli, F., and Guadagnini, A. 2021. Analysis of the performance of a crude-oil desalting system based on historical data. Fuel. Doi: https://doi.org/10.1016/j.fuel.2020.120046.

[52] Hashmi, A., Goel, N., Goel, S., & Gupta, D. Firefly algorithm for unconstrained optimization. IOSR J Comput Eng, 2013, 11(1), 75-78.
Karaboga, D. An idea based on honey bee swarm for numerical optimization. Retrieved from, 2005. Doi: https://pdfs.semanticscholar.org/015d/f4d97ed1f541752842c49d12e429a785460b.pdf

[53] Karaboga, D., & Akay, B. A comparative study of artificial bee colony algorithm. Applied mathematics and computation, 2009, 214(1), 108-132. Doi: https://doi.org/10.1016/j.amc.2009.03.090

[54] Tereshko, V. Reaction-diffusion model of a honeybee colony's foraging behaviour. Paper presented at the International Conference on Parallel Problem Solving from Nature, 2000. Doi: https://link.springer.com/chapter/10.1007/3-540-45356-3_79

[55] Tereshko, V., & Lee, T. How information-mapping patterns determine foraging behaviour of a honey bee colony. Open Systems & Information Dynamics, 2002, 9(02), 181-193. Doi: https://www.worldscientific.com/doi/abs/10.1023/A:1015652810815

[56] Kröse, B., Krose, B., van der Smagt, P., & Smagt, P. An introduction to neural networks, 1993. Doi: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.18.493

[57] Moghadam, J. N., Salahshoor, K., & Kharrat, R. Introducing a new method for predicting PVT properties of Iranian crude oils by applying artificial neural networks. Petroleum science and technology, 2011b, 29(10), 1066-1079. Doi: https://doi.org/10.1080/10916460903551040.

[58] Omar, M., & Todd, A. Development of new modified black oil correlations for Malaysian crudes. Paper presented at the SPE Asia Pacific oil and gas conference, 1993. Doi: https://doi.org/10.2118/25338-MS.

[59] Mahmood, M. A., & Al-Marhoun, M. A. Evaluation of empirically derived PVT properties for Pakistani crude oils. Journal of Petroleum Science and Engineering, 1996, 16(4), 275-290. Doi: https://doi.org/10.1016/S0920-4105(96)00042-3.

[60] Ganji-Azad, E., Rafiee-Taghanaki, S., Rezaei, H., Arabloo, M., & Zamani, H. A. Reservoir fluid PVT properties modeling using adaptive neuro-fuzzy inference systems. Journal of Natural Gas Science and Engineering, 2014, 21, 951-961. Doi: https://doi.org/10.1016/j.jngse.2014.10.009.

[61] Al-Anazi BD. Enhanced oil recovery techniques and nitrogen injection. CSEG recorder. 2007 Oct;32(8):29-33. Doi: https://memberfiles.freewebs.com/50/69/68186950/documents/Enhanced%20Oil%20Recovery%20Techniques-1.pdf

[62] Hassanpouryouzband A, Farahani MV, Yang J, Tohidi B, Chuvilin E, Istomin V, Bukhanov B. Solubility of flue gas or carbon dioxide-nitrogen gas mixtures in water and aqueous solutions of salts: Experimental measurement and thermodynamic modeling. Industrial & Engineering Chemistry Research. 2019 Jan 23;58(8):3377-94. Doi: https://doi.org/10.1021/acs.iecr.8b04352.

[63] Hassanpouryouzband A, Joonaki E, Edlmann K, Heinemann N, Yang J. Thermodynamic and transport properties of hydrogen containing streams. Scientific Data. 2020 Jul 9;7(1):1-4. Doi: https://www.nature.com/articles/s41597-020-0568-6

[64] Ebtehaj I, Bonakdari H. A support vector regression-firefly algorithm-based model for limiting velocity prediction in sewer pipes. Water Science and Technology. 2016 May 5;73(9):2244-50. Doi: https://doi.org/10.2166/wst.2016.064

[65] Ebtehaj I, Bonakdari H. Bed load sediment transport estimation in a clean pipe using multilayer perceptron with different training algorithms. KSCE Journal of Civil Engineering. 2016 Mar 1;20(2):581-9. Doi:

https://link.springer.com/article/10.1007/s12205-015-0630-7

[66] Moeeni H, Bonakdari H, Ebtehaj I. Integrated SARIMA with neuro-fuzzy systems and neural networks for monthly inflow prediction. Water Resources Management. 2017 May 1;31(7):2141-56. Doi: https://link.springer.com/article/10.1007%2Fs11269-017-1632-7

[67] Ebtehaj I, Bonakdari H, Zaji AH. A new hybrid decision tree method based on two artificial neural networks for predicting sediment transport in clean pipes. Alexandria engineering journal. 2018 Sep 1;57(3):1783-95. Doi: https://doi.org/10.1016/j.aej.2017.05.021

[68] Ebtehaj I, Bonakdari H, Shamshirband S, Ismail Z, Hashim R. New approach to estimate velocity at limit of deposition in storm sewers using vector machine coupled with firefly algorithm. Journal of Pipeline Systems Engineering and Practice. 2017 May 1;8(2):04016018. Doi: https://doi.org/10.1061/(ASCE)PS.1949-1204.

[69] Tjahjono A, Anggriawan DO, Faizin AK, Priyadi A, Pujiantara M, Purnomo MH. Optimal coordination of overcurrent relays in radial system with distributed generation using modified firefly algorithm. International Journal on Electrical Engineering and Informatics. 2015 Dec 1;7(4):691.

**Funding sources**

This research received no external funding

**Conflicts of interest**

There are no conflicts to declare.