ACS OMEGA

http://pubs.acs.org/journal/acsodf Article

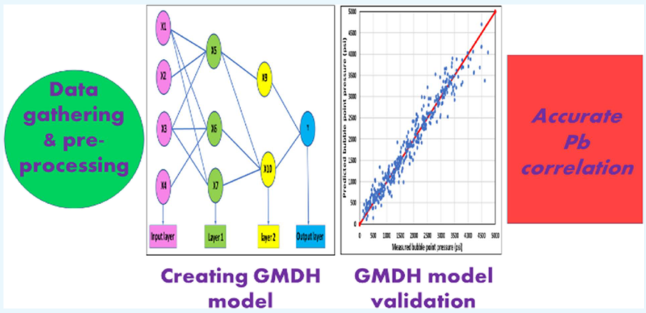
An Accurate Reservoir's Bubble Point Pressure Correlation

Fahd Saeed Alakbari,* Mysara Eissa Mohyaldinn,* Mohammed Abdalla Ayoub, Ali Samer Muhsan, and Ibnelwaleed A. Hussein

Cite This: *ACS Omega* 2022, 7, 13196−13209 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Bubble point pressure (P_b) is essential for determining petroleum production, simulation, and reservoir characterization calculations. The P_b can be measured from the pressure−volume−temperature (PVT) experiments. Nonetheless, the PVT measurements have limitations, such as being costly and time-consuming. Therefore, some studies used alternative methods, namely, empirical correlations and machine learning techniques, to obtain the P_b. However, the previously published methods have restrictions like accuracy, and some use specific data to build their models. In addition, most of the previously published models have not shown the proper relationships between the features and targets to indicate the correct physical behavior. Therefore, this study develops an accurate and robust correlation to obtain the P_b applying the Group Method of Data Handling (GMDH). The GMDH combines neural networks and statistical methods that generate relationships among the feature and target parameters. A total of 760 global datasets were used to develop the GMDH model. The GMDH model is verified using trend analysis and indicates that the GMDH model follows all input parameters' exact physical behavior. In addition, different statistical analyses were conducted to investigate the GMDH and the published models' robustness. The GMDH model follows the correct trend for four input parameters (gas solubility, gas specific gravity, oil specific gravity, and reservoir temperature). The GMDH correlation has the lowest average percent relative error, root mean square error, and standard deviation of 8.51%, 12.70, and 0.09, respectively, and the highest correlation coefficient of 0.9883 compared to published models. The different statistical analyses indicated that the GMDH is the first rank model to accurately and robustly predict the P_b.



1. INTRODUCTION

Bubble point pressure (P_b) is one of the critical reservoir fluid pressure−volume−temperature (PVT) properties. It is the pressure at which the first bubble of the gas comes out of the oil solution.[1] An accurate reservoir P_b is critical for performing proper material balance, reservoir, and petroleum production calculations.[2−5] Therefore, it is significant to predict the P_b accurately. However, PVT measurements are costly and time-consuming. Therefore, many studies have used different methods such as regressions, machine learning, and deep learning (long short-term memory (LSTM)) to predict the P_b with varying degrees of success. The P_b prediction methods are developed based on different approaches.

Standing[6] and Lasater[7] used graphical methods to predict the P_b. Standing[6] utilized more than 100 datasets from USA crude oil to develop a correlation for the prediction of P_b. Lasater[7] published an equation to predict the P_b using 158 datasets from Canada and the United States. In the 1980s to 1990s, some researchers used linear and nonlinear regressions to determine the P_b. Vasquez and Beggs[8] employed linear regression analysis to determine the P_b, operating 6004 datasets from different fields. Glaso[9] used linear and nonlinear regressions to determine the P_b and stated that his correlation has a standard deviation (SD) of 6.98. Other authors used linear and nonlinear regression to predict P_b using data from different parts of the globe.[5,10−13]

Al-Marhoun, Kartoatmodjo and Schmidt, Dokla and Osman, Petrosky and Farshad, Macary and El-Batanoney, De Ghetto et al., and Hanafy et al.[14−20] utilized multiple regressions to predict the P_b. Different equations have been developed for the prediction of the P_b using 160 Al-Marhoun,[14] 5392 Kartoatmodjo and Schmidt,[15] 51 Dokla and Osman,[16] 90 Petrosky and Farshad and Macary and El-Batanoney,[17,18] 3700 De Ghetto et al.,[19] and 324 Hanafy et al.[20] datasets from different parts of the world.

In 2001−2016, Dindoruk and Christman[21] built their model to obtain the P_b by applying solver tool built-in MS-Excel, and their correlation has an average absolute error (AAE) of 5.7%. The correlation was based on 100 PVT laboratory reports from

Received: February 1, 2022
Accepted: March 16, 2022
Published: April 8, 2022

https://doi.org/10.1021/acsomega.2c00651
ACS Omega 2022, 7, 13196−13209

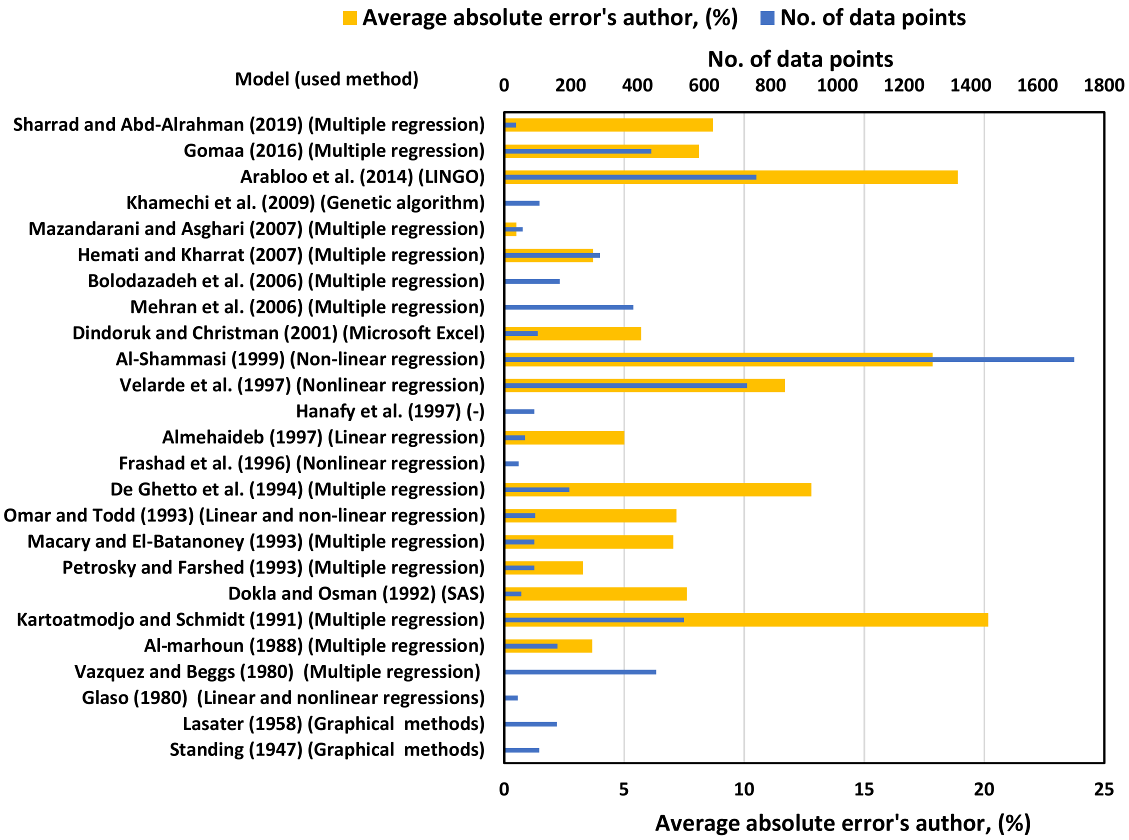


Figure 1. Comparison of the error % for different P_b correlations.

the Gulf-of-Mexico. Bolondarzadeh et al.,[22] Mehran et al.,[23] Hemmati and Kharrat,[24] and Mazandarani and Asghari[25] also used multiple regressions and 387, 166, 287, and 55 datasets, respectively, from Iran fields to estimate the P_b. Arabloo et al.[26] applied 750 global datasets and LINGO software to develop their P_b correlation with an AAE of 18.9%. Ahmadi et al.[27] applied gene expression programming to estimate the P_b. Gomaa[28] utilized multiple-regression methods and 441 datasets from the Middle East to build his P_b.

Artificial intelligence, machine learning, and deep learning, such as LSTM methods, are widely used in engineering applications, and the P_b is one of these applications. Artificial neural network and fuzzy-logic approaches were utilized to estimate the P_b.[29] Sharrad and Abd-Alrahman[30] showed the P_b correlation with an AAE of 8.7% based on 35 Libyan datasets and using EViews software. An artificial neural network, an adaptive neuro-fuzzy inference system, and a support vector machine were operated to determine a P_b.[31,32] Yang et al.[33] applied python packages of machine learning algorithms such as scikit-learn.ensemble to obtain a P_b prediction correlation. Tariq et al.[34] introduced a P_b model using a functional network with particle swarm optimization. A deep learning (LSTM) approach was used to determine the P_b.[35] The comparison of the previous models' used methods, no. of data points, and average absolute errors is shown in Figure 1.

The Group Method of Data Handling (GMDH) outperformed other artificial intelligence, machine learning, and deep learning (LSTM) methods, namely, no requirement for the pre-structure of the networks, no overfitting issues, and few trials to obtain the convergence. Moreover, the GMDH method can identify features with important relevancy to the target to remove unnecessary input parameters.[36] The GMDH represents the accurate and easy-to-use correlation. Ayoub and Mohamed[37] applied GMDH to determine the missing log interval of well log properties, such as a gamma-ray log. They proved that GMDH could be effectively used in building a model to obtain the missing well logs with high accuracy.[37] Menad et al.[38] utilized gene expression programming and GMDH to get temperature dependency on oil−water relative permeability. Their models were used as promising tools utilizing a broad set of data.[38] The GMDH was successfully employed to detect lithofacies present in the South Yellow Sea.[39] Ayoub et al.[40] applied GMDH to obtain oil compressibility below P_b. However, different statistical error analyses were deployed to show the GMDH model's robustness.[40] Mathew Nkurlu et al.[41] used GMDH to obtain permeability utilizing well log data in their calculations. They presented that the GMDH model outperformed the neural network.[41]

Ayoub et al.[36] used the GMDH method to predict the oil formation volume factor at the bubble point pressure and proved that their correlation is the most accurate compared to all studied models. Hemmati-Sarapardeh et al.[42] determined the asphaltene precipitation based on temperature, type of solvent, and solvent to crude oil dilution ratio using the GMDH technique. They showed an explicit and simple correlation to find the asphaltene precipitation with an average absolute percent relative error (AAPRE) of less than 4%.[42] Dargahi-Zarandi et al.[43] presented the GMDH model to obtain the minimum miscibility pressure of pure and impure CO_2 crude oil and proved that their model surpassed other existing models.[43] Amar et al.[44] applied the GMDH method to find the viscosity of CO_2 at high conditions, i.e., pressure and temperature using 1124 datasets. The viscosity of the CO_2

https://doi.org/10.1021/acsomega.2c00651
ACS Omega 2022, 7, 13196−13209

Table 1. Comparison of Independent Parameters Used in the Published Correlations and the GMDH Model

No.	Model	Independent parameters				
		Oil formation volume factor at the Bubble point pressure (Bob) (bbl/STB)	R_s (scf/STB)	γ_g	API (°API)	T_f (°F)
1	Standing (1947) [6]		√	√	√	√
2	Lasater (1958) [7]		√	√	√	√
3	Glaso (1980) [9]		√	√	√	√
4	Vazquez and Beggs (1980) [8]		√	√	√	√
5	Al-marhoun (1988) [14]		√	√	√	√
6	Kartoatmodjo and Schmidt (1991) [15]		√	√	√	√
7	Dokla and Osman (1992) [16]		√	√	√	√
8	Petrosky and Farshed (1993) [17]		√	√	√	√
9	Macary and El-Batanoney (1993) [18]		√	√	√	√
10	Omar and Todd (1993) [10]	√	√	√	√	√
11	De Ghetto et al. (1994) [19]		√	√	√	√
12	Frashad et al. (1996) [11]		√	√	√	√
13	Almehaideb (1997) [12]	√	√	√	√	√
14	Hanafy et al. (1997) [20]		√			
15	Velarde et al. (1997) [5]		√	√	√	√
16	Al-Shammasi (1999) [13]		√	√	√	√
17	Dindoruk and Christman (2001) [21]		√	√	√	√
18	Mehran et al. (2006) [23]		√	√	√	√
19	Bolodazadeh et al. (2006) [22]		√	√	√	√
20	Hemmati and Kharrat (2007) [24]	√	√	√	√	√
21	Mazandarani and Asghari (2007) [25]		√	√	√	√
22	Khamechi et al. (2009) [46]		√	√	√	√
23	Arabloo et al. (2014) [26]		√	√	√	√
24	Gomaa (2016) [28]		√	√	√	√
25	Sharrad and Abd-Alrahman (2019) [30]		√	√	√	√
26	**Proposed GMDH method**		√	√	√	√

model was built based on temperature and density obtained at the given conditions. They stated that their model could determine the viscosity of CO_2 accurately with an R of 0.9995.[44] Rezaei et al.[45] predicted the interfacial tension of CO_2 paraffin as a function of the pressure, temperature, and molecular weight of paraffin applying the GMDH method. They used 879 datasets gathered from the literature and showed that their model has an AAPRE of 4.42%.[45]

In the literature, some methods were used to determine the P_b. However, the previous models and correlations have some limitations, such as lack of accuracy and a need for computer software such as machine learning and deep learning (LSTM) approaches. Therefore, this research aims to build, validate, and test a highly accurate correlation to obtain P_b directly from the equation using the GMDH method. There is no need to use any computer software using the GMDH approach. The GMDH model depends on R_s, γ_g, API, and T_f. A trend analysis technique was achieved to study the relationships between the features and targets to represent the actual physical behavior. In addition, the performances of GMDH and all studied models were compared using statistical error analysis, namely, the correlation coefficient (R), absolute average percentage relative error (AAPRE), maximum absolute percent relative error ($E_{max.}$), minimum absolute percent relative error ($E_{min.}$), standard deviation (SD), and root mean square error (RMSE). In addition, crude oils from different regions are different in chemical properties. This study uses worldwide data on different types of crude oils to obtain a more general correlation.

2. METHODOLOGY

2.1. Data Collection and Pre-processing. More than 700 datasets were collected from the literature to develop the GMDH model. The P_b is highly dependent on R_s, γ_g, API, and

https://doi.org/10.1021/acsomega.2c00651
ACS Omega 2022, 7, 13196−13209

T_f, as most previous studies have stated (see Table 1). In this research, the data set was split into three subparts, 50, 25, and 25%, for training, validating, and testing the model, respectively. The used data's statistical description is shown in Table 2.

Table 2. Statistical Description of the Used Data

parameters	minimum	maximum	SD
bubble point pressure (P_b), psi	126	7127	1118
gas to oil ratio (R_s), SCF/STB	9	2637	415
gas specific gravity (γ_g), fraction	0.5890	1.367	0.1619
oil specific gravity (API), °API	15.30	59.50	7.122
reservoir temperature (T_f), °F	74	294	48.462

2.2. GMDH Method. GMDH or polynomial neural networks combine the unique attributes of neural networks with improved statistical methods to provide a faster and more accurate model. GMDH could be a multi-layer algorithm amalgamation. GMDH has been used to identify and recognize nonlinear relationships between the dependent and independent variables.[47,48] GMDH also performs by connecting neurons linked to quadratic polynomials, which results in neurons in the successive layers.[49] The polynomial neural network has some benefits. Initially, the polynomial neural network can choose the most appropriate inputs in a set of applicants. Such networks can be used to minimize the impact of the user on the modeling results by arranging different solutions. The optimal structure of the proposed model and the laws acting on the system are automatically obtained using the computer. The polynomial neural network has a significant feature that enables achieving that. The conventional neural network design includes challenges like getting suitable topology and weights. Consequently, the robustness of the model is remarkably impacted by the developer.[50]

GMDH conducts a series of consecutive layers with the connection. The successive layers include polynomial terms, which can be obtained utilizing regressions. The first layer can be obtained through the regressions of input parameters before it chooses the best one. On the next layer, the values found from the previous layer are regressed along with the input parameter, and the process continues until no better result is reached.[40] Figure 2 shows the self-organizing GMDH algorithm.[51]

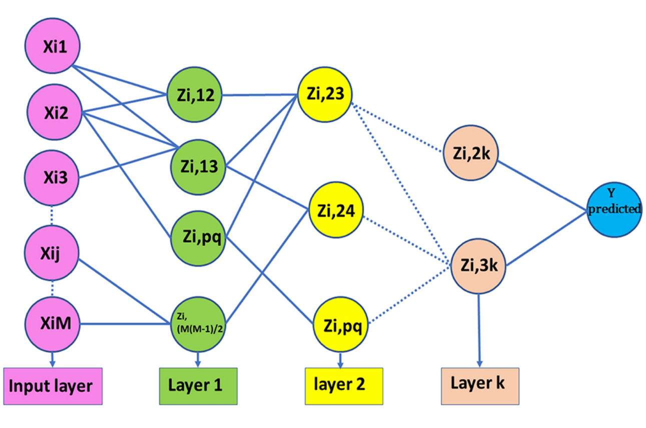


Figure 2. Self-organizing GMDH algorithm.[51]

In the GMDH algorithm, assuming that the dependent variable Y is a function of independent variables x_i, it can be represented by

$$Y = f(x_1, x_2, x_3, ..., x_n) \tag{1}$$

where X is the vector $x = (x_1, x_2, x_3, ..., x_n)$ where it acts as the input data needed to predict the closest output Y to the actual input Y. GMDH can remove the inaccurate, small, and noisy data to improve the accuracy of the model and provide a more straightforward structure than the usual physical model.[40]

GMDH can be a combinatorial multi-layer algorithm wherein a network of layers and nodes is obtained, applying the input parameters found from the data stream.[52] The network topology of GMDH has been initially obtained utilizing a layer-by-layer iterative process depending on a pre-selected approach of what constitutes the best nodes at each level[47,51]

$$y = a_0 + \sum_{i=1}^{m} a_i x_i + \sum_{i=1}^{m}\sum_{j=1}^{m} a_{ij}x_i x_j + \sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m} a_{ijk}x_i x_j x_k ... \tag{2}$$

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & \cdots & x_{1M} \\ x_{21} & x_{22} & \cdots & \cdots & x_{2M} \\ \cdots & \cdots & \cdots & x_{ij} & x_{iM} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{N1} & x_{N2} & \cdots & \cdots & x_{NM} \end{bmatrix} \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ \cdots \\ y_N \end{bmatrix} \tag{3}$$

As shown below, eqs 2 and 3 can be replaced with a partial polynomial

$$y = G(X_i X_j) \tag{4}$$

$$y = a_0 + a_1 x_i + a_2 x_j + a_3 x_{i2} + a_4 x_{j2} + a_5 x_i x_j \tag{5}$$

where $i,j = 1, 2, ..., M$, $i \neq j$.

3. RESULTS AND DISCUSSION

3.1. P_b GMDH Model. The GMDH model contained two hidden layers and one output layer. Figure 3 shows a schematic

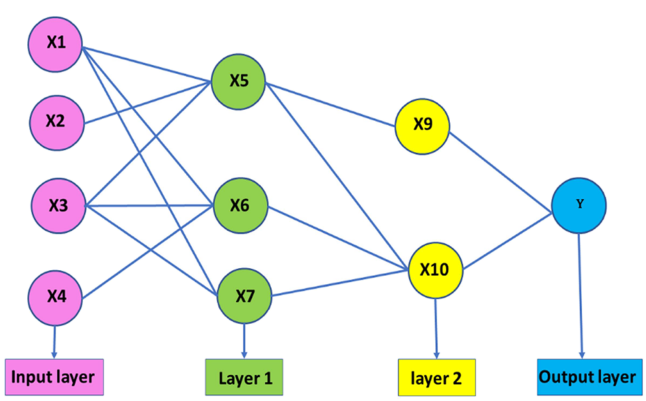


Figure 3. Schematic of the P_b GMDH Model.

of the developed GMDH model. After the model runs, a total of four input parameters was identified to impact the P_b predictions (Y). These parameters are R_s (X_1), γ_g (X_2), API (X_3), and T_f (X_4). The input parameters were chosen to depend on the summary of parameters applied in the existing correlations, as described in the literature review (Table 1).

https://doi.org/10.1021/acsomega.2c00651
ACS Omega 2022, 7, 13196−13209

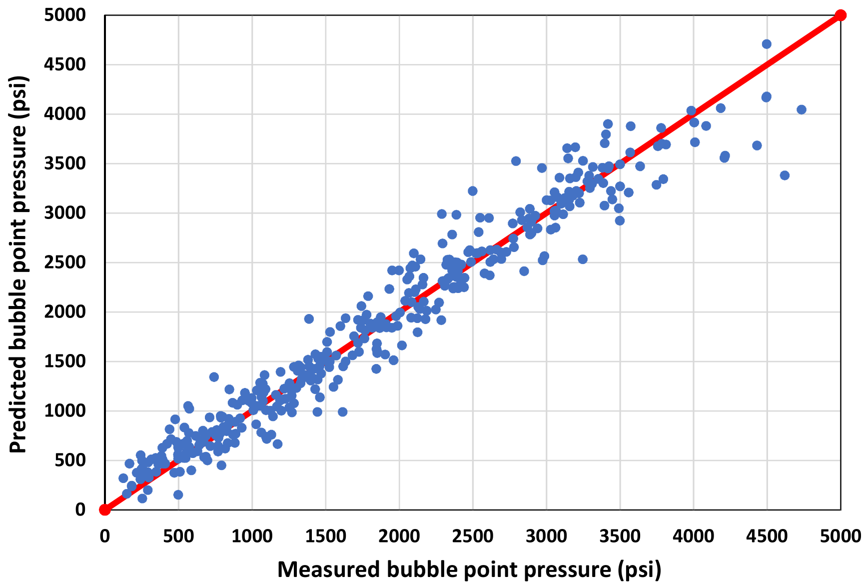


Figure 4. Cross plot of the training GMDH model.

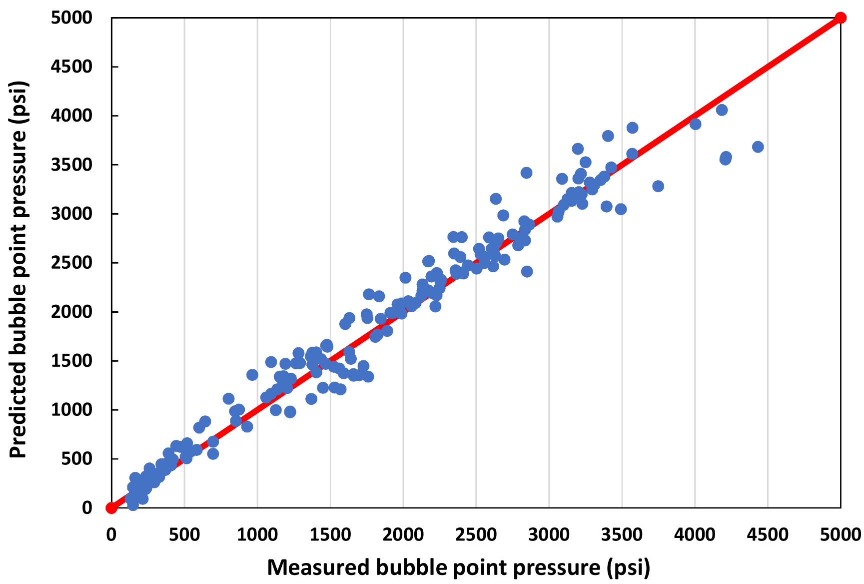


Figure 5. Cross plot of the validating GMDH model.

The GMDH model can use the parameters with the most impact on P_b.

The P_b model developed by the GMDH approach has three layers, as shown below:

Layer number 1:

Number of neurons: 3

$$X_5 = 4.36 \times 10^{+03} - 88.1X_3 - 5.1 \times 10^{+03}X_2 + 8.37X_1 + 71.5X_2X_3 - 0.0857X_1X_3 - 1.74X_1X_2 + 0.359{X_3}^2 + 969{X_2}^2 - 0.000533{X_1}^2 \quad (6)$$

$$X_6 = 933 + 1.2X_4 - 43.4X_3 + 5.83X_1 + 0.113X_3X_4 + 0.00387X_1X_4 - 0.0706X_1X_3 - 0.0112{X_4}^2 + 0.207{X_3}^2 - 0.000792{X_1}^2 \quad (7)$$

$$X_7 = 929 - 34.6X_3 + 6.55X_1 - 0.0719X_1X_3 + 0.371{X_3}^2 - 0.000648{X_1}^2 \quad (8)$$

Layer number 2:

Number of neurons: 2

$$X_9 = -385 + 0.581X_5 + 5.21X_4 - 0.0217X_1 - 0.00132X_4X_5 - 0.00173X_1X_5 + 0.00848X_1X_4 + 0.000405{X_5}^2 - 0.00989{X_4}^2 + 0.0016{X_1}^2 \quad (9)$$

$$X_{10} = -51.4 - 0.253X_7 + 0.773X_6 + 0.526X_5 + 0.000949X_6X_7 - 9.93 \times 10^{-05}X_5X_7 - 0.000641X_5X_6 - 0.000625{X_7}^2 - 8.53 \times 10^{-05}{X_6}^2 + 0.000489{X_5}^2 \quad (10)$$

https://doi.org/10.1021/acsomega.2c00651
ACS Omega 2022, 7, 13196−13209

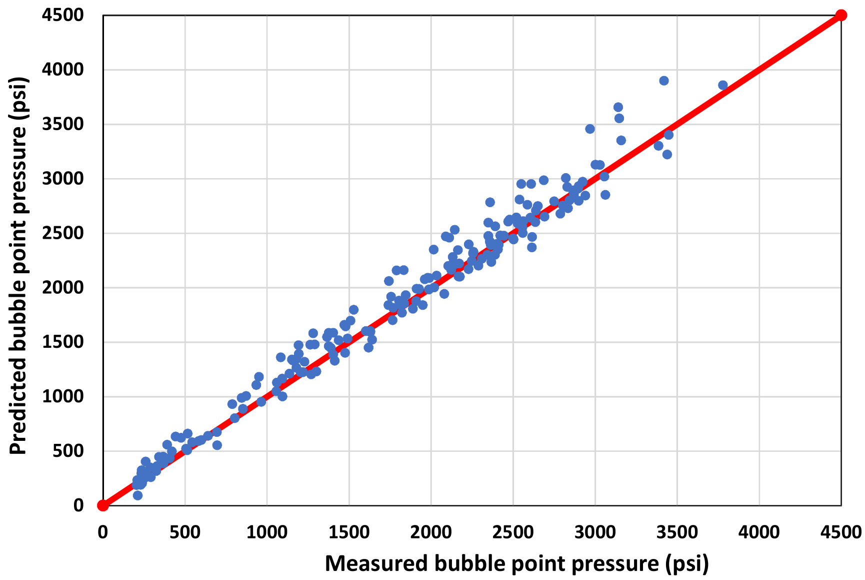


Figure 6. Cross plot of the testing GMDH model.

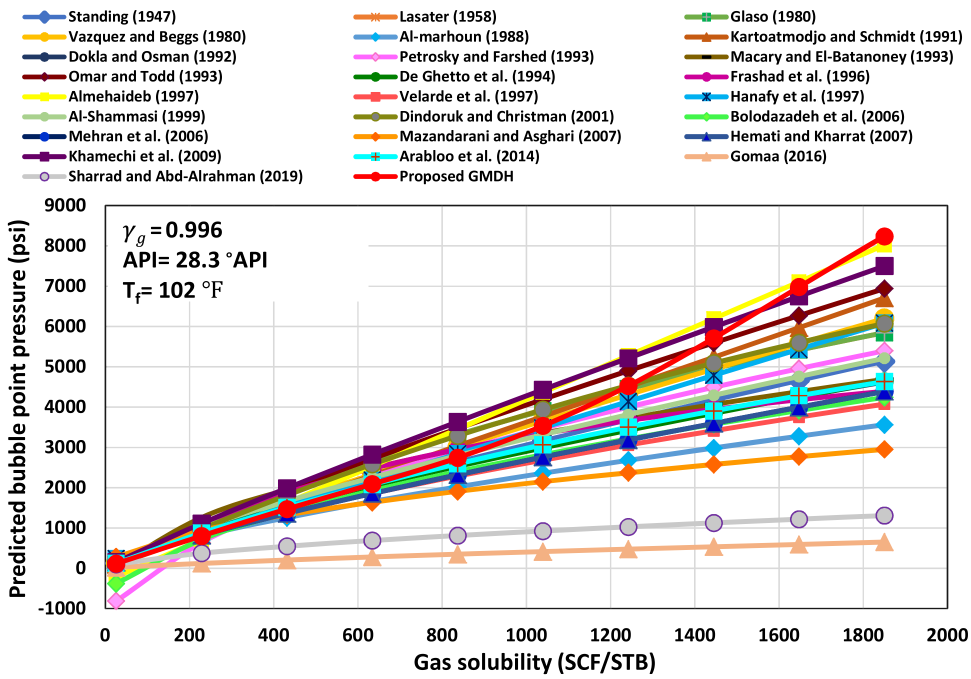


Figure 7. TA of R_s for the GMDH and all studied models.

Layer number 3:
Number of neurons: 1

$$Y = -66.1 + 0.183X_{10} + 0.82X_9 + 0.721X_4 - 0.00287X_9X_{10} - 0.00307X_4X_{10} + 0.00334X_4X_9 + 0.00157X_{10}{}^2 + 0.00129X_9{}^2 - 0.00326{*}X_4{}^2 \quad (11)$$

where X_1 is the gas solubility (R_s, SCF/STB), X_2 is the gas specific gravity (γ_g, fraction), X_3 is the oil specific gravity (API, °API), X_4 is the reservoir temperature (T_f, °F), and Y is the bubble point pressure (P_b, psi).

3.2. Cross Plot Analysis. The cross plots obtained from the GMDH model training, validating, and testing datasets are demonstrated in Figures 4, 5, and 6. Figure 4 shows that most of the training datasets are closer to the line 45°,indicating accurate predictions of the GMDH model. Similarly, most of the validating and testing datasets are also closer to the line 45°, suggesting the accuracy of the GMDH model predictions for P_b (see Figures 5 and 6).

3.3. Trend Analysis. Trend analysis (TA) can be an essential method to study the relationships between the features and targets to represent the actual physical behavior. In addition, the TA can show modeling errors to indicate unexpected relationships between the features and targets to prove the reliability of the studied model.[35,36] In addition, the TA is conducted to make the GMDH model sample by fixing model features that affect the targets or removing redundant parameters that have less or no effect on the studied model.[53] Trend analysis can also identify significant connections between input and outcomes, guiding the improvement of

https://doi.org/10.1021/acsomega.2c00651
ACS Omega 2022, 7, 13196−13209

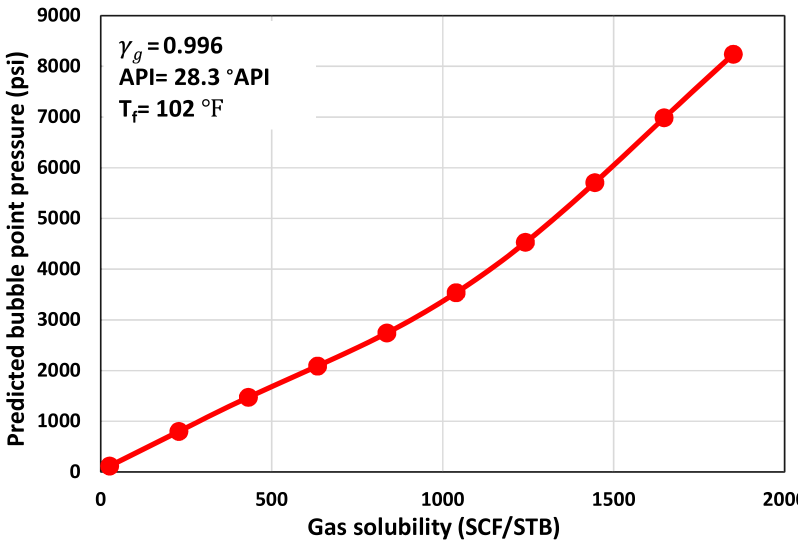


Figure 8. TA of R_s for the GMDH model.

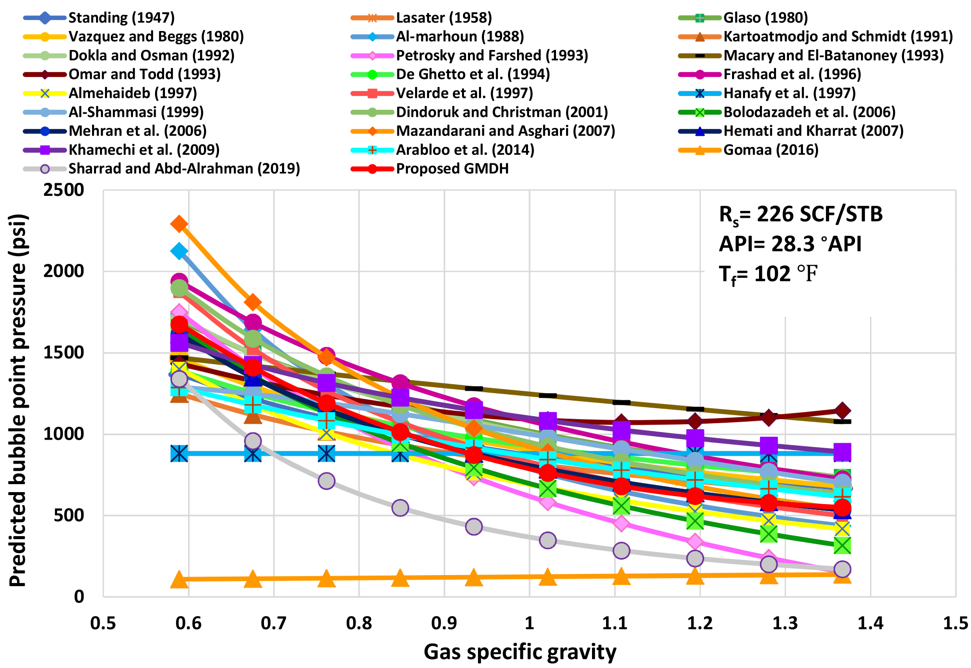


Figure 9. TA of γ_g for the GMDH and all studied models.

robust models.[54] Consequently, the trend analysis is an essential part of this research. In this study, four input parameters R_s, γ_g, API, and T_f have been chosen for the TA. The B_{ob} has not been used as a feature because it was only utilized by Almehaideb,[12] Hemmati and Kharrat,[24] and Omar and Todd.[10]

Figure 7 shows the trend of gas solubility (R_s). Petrosky and Farshad[17] built correlations for the gas solubility in 217−1406 SCF/STB. Therefore, Petrosky and Farshad's[17] correlation shows that the P_b was negative 812.6 at gas solubility 26 SCF/STB (Figure 7). Almehaideb[12] developed a correlation for the gas solubility range of 128−3871 SCF/STB. Thus, Almehaideb's[12] correlation indicates that the P_b was negative 207.5 at a gas solubility of 26 (Figure 7). The trend shown by the current GMDH model shows the proper relationship between the studied features and the target to represent the actual physical behavior for the gas solubility (Figure 8). Li et al.'s[55] experimental study indicates that increasing gas solubility increases the P_b.

Figure 9 demonstrates the trend of gas specific gravity (γ_g). Increasing the γ_g decreases the P_b, as displayed by all studied models. Nevertheless, Hanafy et al.'s[20] correlation suggests that P_b does not change with changing the γ_g. Gomaa's[28] correlation shows that increasing the γ_g slightly increases the P_b as shown in Figure 9, so his model contradicted the prior correlation behavior. Omar and Todd's[10] correlation represented that increasing the γ_g decreases the P_b and then increases it to display different trends for the γ_g. Therefore, some of the previous models' predictions of Omar and Todd,[10] Hanafy et al.,[20] and Gomaa[28] do not show the correct relationship between the γ_g and the P_b. Figure 10 shows the proper gas specific gravity trend for the proposed model. As we know, the bubble point pressure is known as the pressure at which the first bubble of gas comes out from the liquid at a

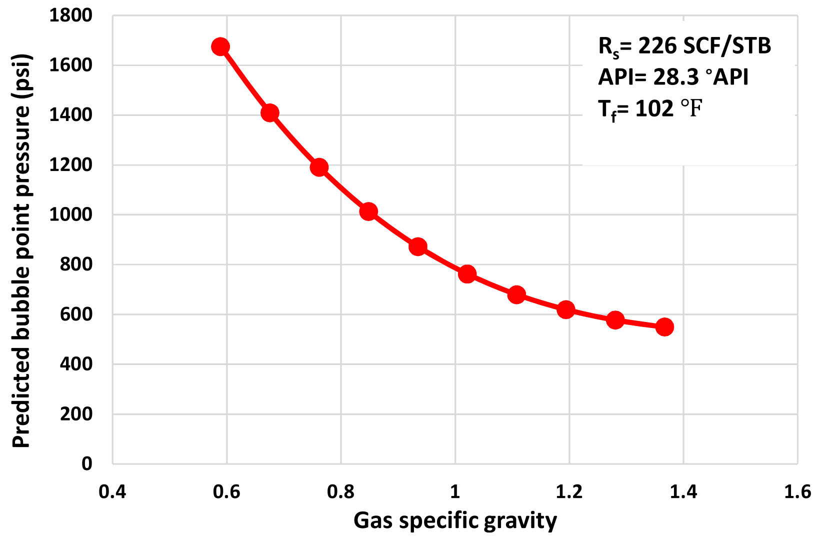


Figure 10. TA of γ_g for the GMDH model.

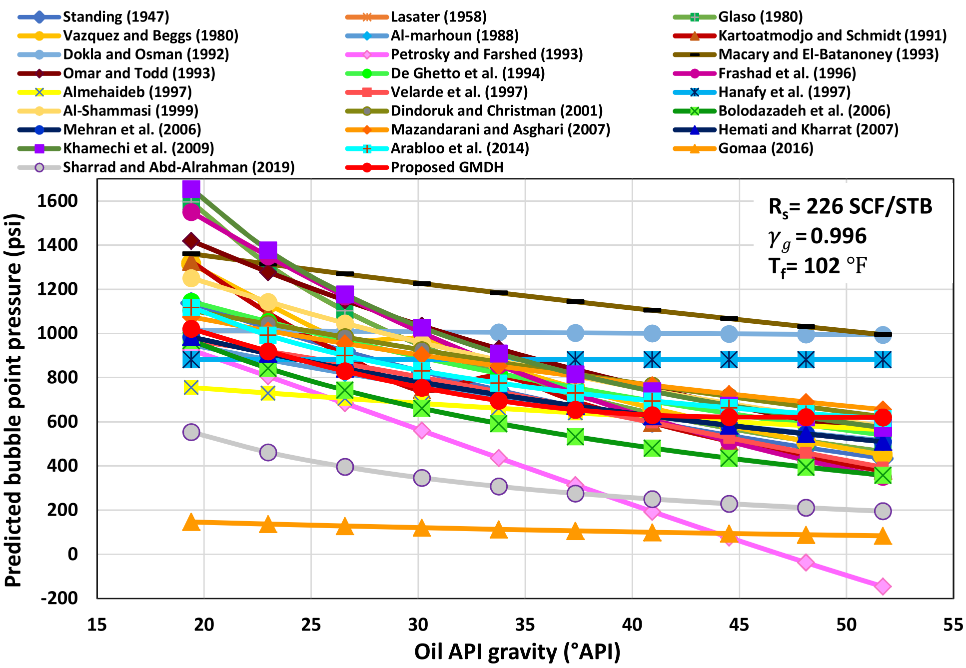


Figure 11. TA of API of the GMDH and all studied models.

given temperature. The gas that has less density can quickly move from the liquid. Therefore, increasing the gas density or gas specific gravity decreases the bubble point pressure. Al-Shammasi[13] studied the TA of the γ_g and showed that increasing the γ_g decreases the P_b.

Figure 11 shows the API's trend, and P_b decreases with increasing API. Other correlations did not consider API; therefore, the proposed model suggests that P_b is independent of API. Hanafy et al.'s,[20] Gomaa's,[28] and Dokla and Osman's[16] correlations showed that increasing the API slightly decreases the P_b to indicate that their corrections followed the correct trend. Therefore, Hanafy et al.[20] do not follow a proper physical behavior. Petrosky and Farshad's[17] equation presents that the P_b is negative 37.35, 145.91 psi at their API range 48.11, 51.7 °API because they built their model on the (16.3–45 °API) range (Figure 11). The GMDH model trend represents the correct relationships between the API and P_b to define the proper physical behavior (Figure 12). Increasing the API decreases the oil density. Light oil can easier expel the gas compared to heavy oil. As a result, increasing the API reduces the P_b. The TA of API was proposed by Al-Shammasi[13] and indicated that increasing the API declines the P_b.

Reservoir temperature (T_f) trend analysis is shown in Figure 13. Increasing the reservoir temperature increases the P_b, as shown in Figure 13. Nevertheless, Hanafy et al.'s[20] correlation indicates that changing the T_f does not change the P_b to show a horizontal line because they do not consider the T_f in their correlation; therefore, they failed to indicate the correct relationships between the T_f and the P_b. Dindoruk and Christman[21] and Arabloo et al.[26] consider the T_f in their correlation; nevertheless, the P_b is almost constant. Dokla and Osman's[16] equation indicates that the increase in reservoir temperature decreases the P_b. Therefore, Dokla and Osman's[16] correlation does not present the correct physical behavior. The

https://doi.org/10.1021/acsomega.2c00651
ACS Omega 2022, 7, 13196−13209

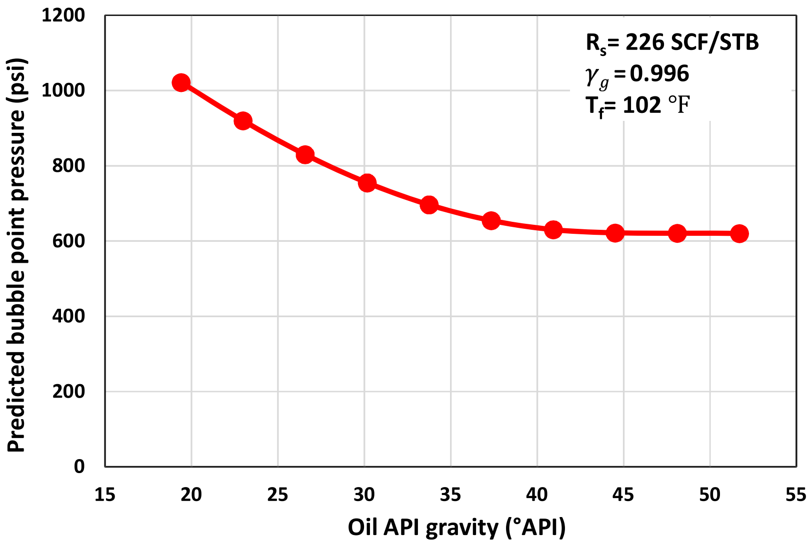


Figure 12. TA of API of the GMDH model.

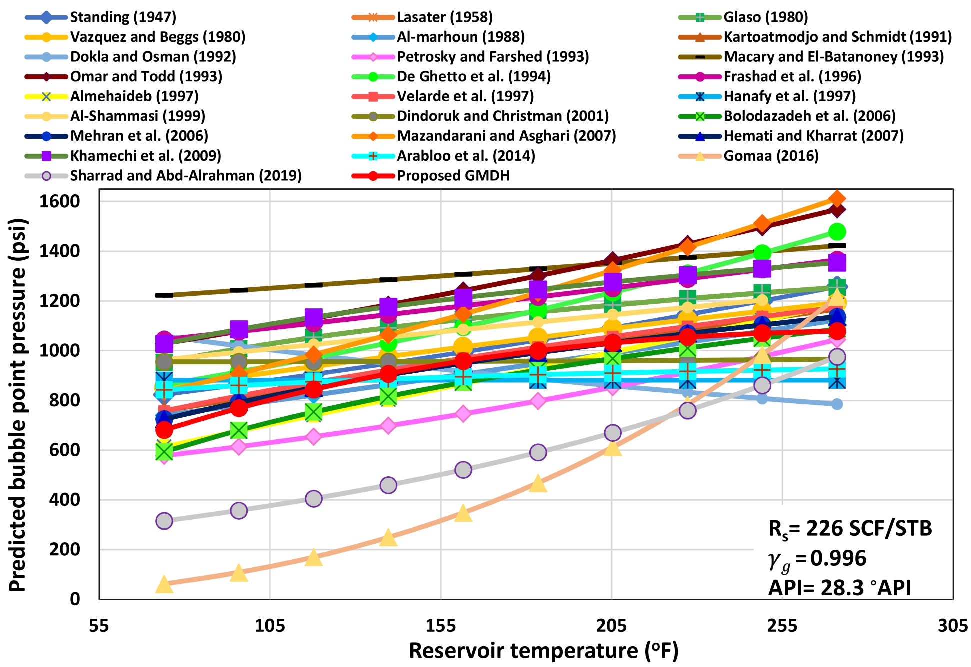


Figure 13. TA of the GMDH and all studied models.

GMDH model indicates the proper relationships between the T_f and the P_b to prove the proper physical behavior (Figure 14). The correct trend between the bubble point pressure and temperature should increase the bubble point pressure when the temperature increases. A higher temperature decreases gas specific gravity, and hence, based on Figure 10, the bubble point pressure increases.

The GMDH model represents the correct relationships between the T_f and P_b to indicate the proper physical laws. As a result, the TA showed the reliability of the GMDH model. The results of the TA indicate that the four studied features R_s, γ_g, API, and T_f follow the correct trends.

3.4. GMDH Model and Published Models' Comparison. Statistical analysis was conducted with trend analysis and cross-plotting analysis to validate and describe the efficiency of the proposed GMDH and all studied models. The statistical error analysis applied in this research are the following: R, AAPRE, E_{max}, E_{min}, SD, and RMSE (see the Supporting Information). The main indicators that are applied in this research are AAPRE and R.

Figure 15 demonstrates the AAPRE and R of GMDH and all studied models to represent the accuracy of the models in the clear picture. As shown in Figure 15, the GMDH model is the most accurate (AAPRE of 8.51% and R of 0.9883) model compared to all studied models. After the GMDH model, Velarde et al.'s[5] and Mehran et al.'s[23] models are the second and third rank models with R values of 0.972 and 0.969. Petrosky and Farshad's[17] correlation with 76.59% (AAPRE) became the least accurate among the studied models to predict the P_b. Sharrad and Abd-Alrahman's[30] model shows the lowest R of 0.8929.

This proposed GMDH and all studied models' prediction performance were conducted and compared using statistical error analysis (Figures 16 and 17). The R and AAPRE are used

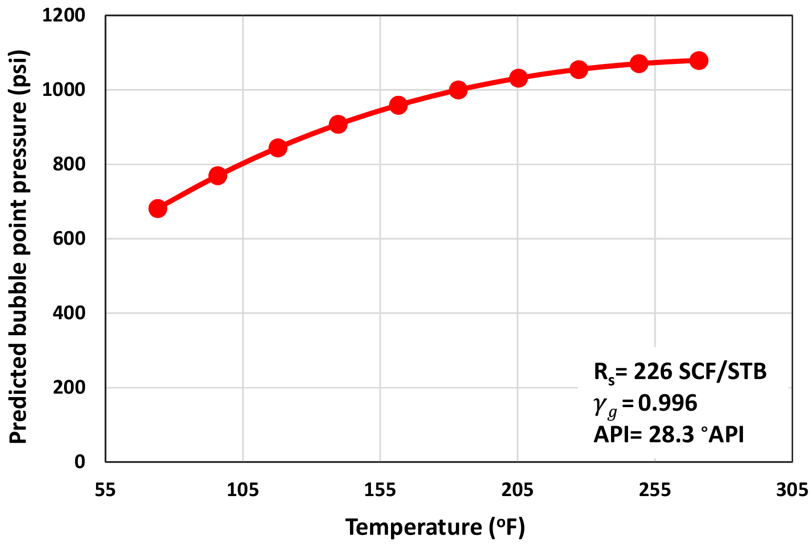


Figure 14. TA of T_f of the GMDH model.

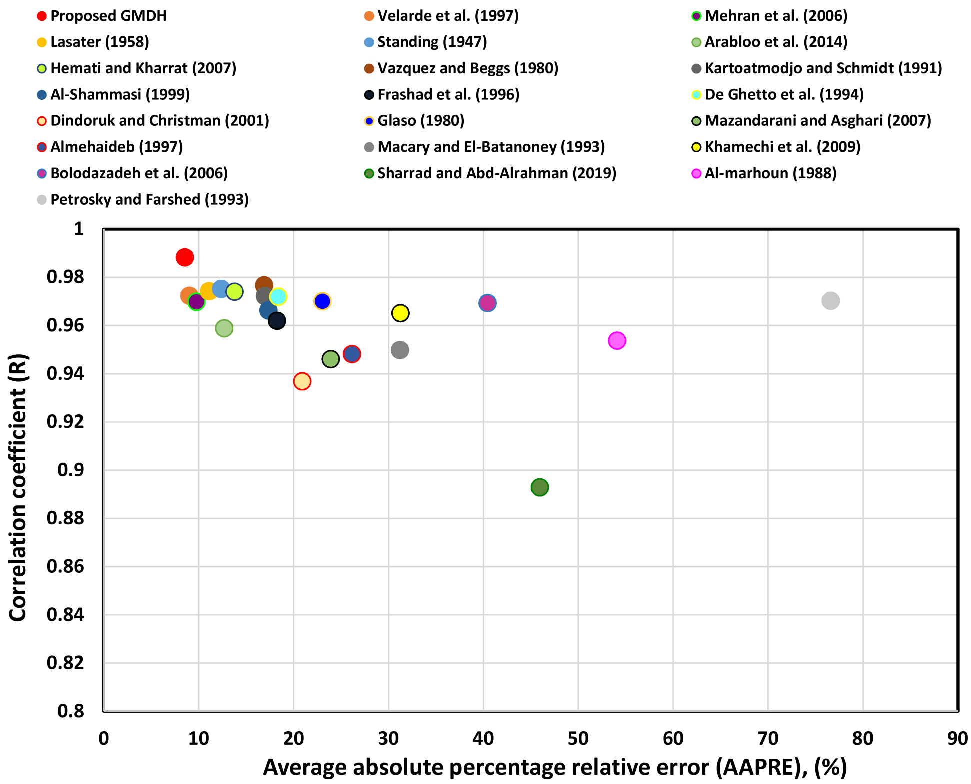


Figure 15. R and AAPRE (%) comparison of the proposed GMDH and all studied models.

as the leading indicators to show the accuracy for all models. The GMDH and all studied models are ranked based on the R and AAPRE. Figure 16 shows the AAPRE (%), $E_{max.}$ (%), and RMSE for the GMDH and all studied models. As shown in Figure 16 and Figure 17, this proposed GMHD model has the lowest AAPRE of 8.51%, $E_{max.}$ of 57.26%, $E_{min.}$ of 0.010%, SD of 0.09, R of 0.9883, and RMSE of 12.70. The second rank model is Velarde et al.'s[5] model with an AAPRE of 9%, $E_{max.}$ of 62.47%, $E_{min.}$ of 0.039%, RMSE of 13.04, SD of 0.095, and R of 0.9724 (Figure 16 and Figure 17). The third rank model is Mehran et al.'s[23] model with an AAPRE of 9.75%, $E_{max.}$ of 63.86%, $E_{min.}$ of 0.035%, RMSE of 13.60, SD of 0.095, and R of 0.9699 (Figure 16 and Figure 17). The fourth rank model is Lasater's[7] model with an AAPRE of 11.07%, $E_{max.}$ of 66.08%, $E_{min.}$ of 0.016%, RMSE of 15.31, SD of 0.106, and R of 0.9742 (Figure 16 and Figure 17).

Table 3 shows the models' used methods, AAPRE, and R to indicate the model performance. The linear regression method (Velarde et al.'s[5] model) becomes after the proposed GMDH model. However, the linear regression method (Almehaideb's[12] model) has an AAPRE of 54% and R of 0.9538 (Table 3). The linear regression methods assume that the relationships between the inputs and outputs are linear, but in most cases, are not linear. In addition, linear regression is sensitive to outliers. Nonlinear methods, i.e., Frashad et al.'s[11] and Al-Shammasi's[13] models, have AAPRE values of 18.23 and

https://doi.org/10.1021/acsomega.2c00651
ACS Omega 2022, 7, 13196−13209

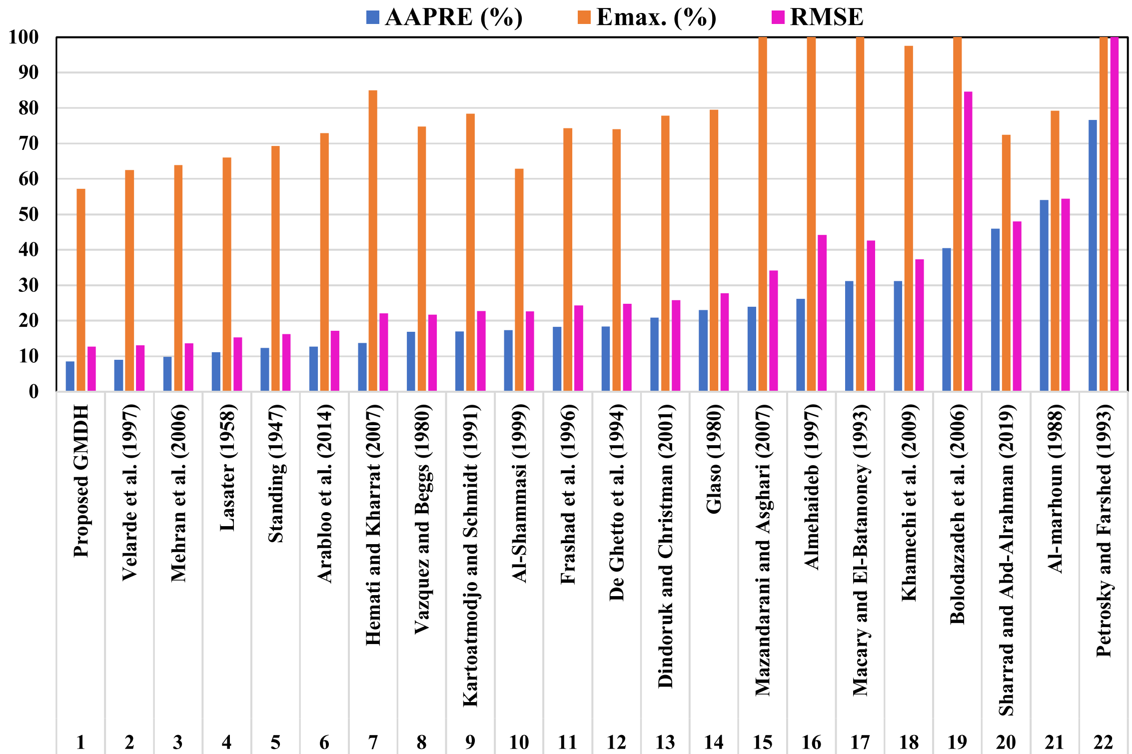


Figure 16. Proposed GMDH and all studied models' AAPRE (%), $E_{max.}$ (%), and RMSE comparison.

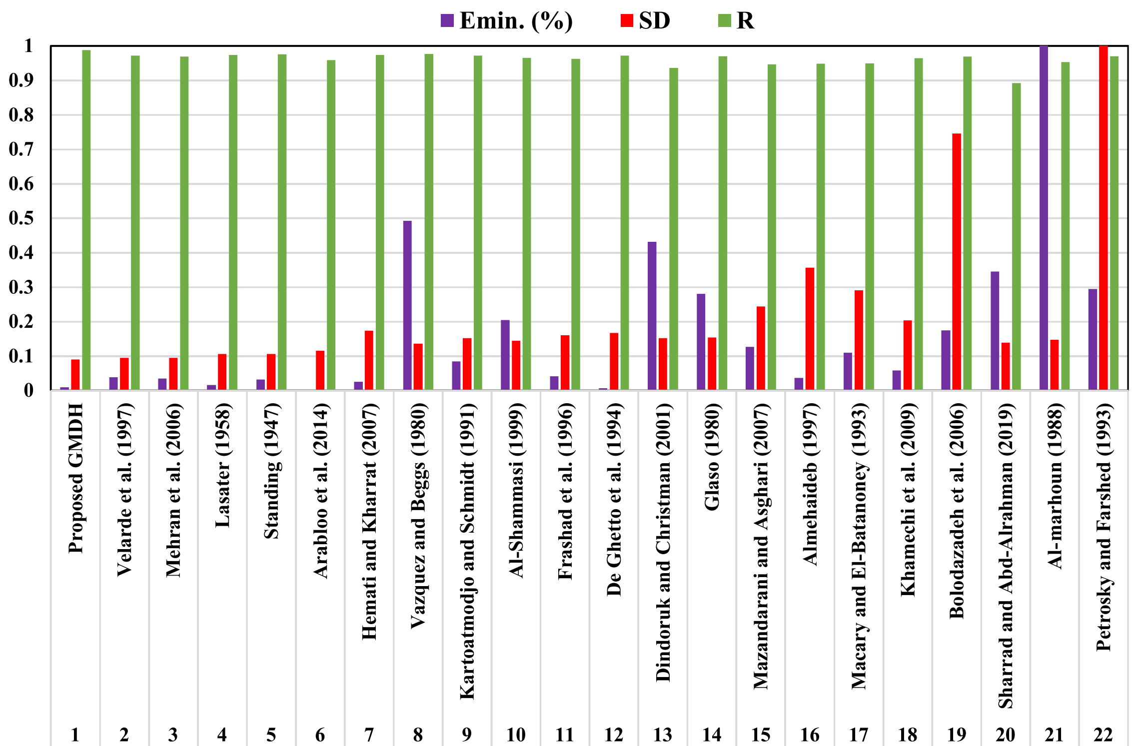


Figure 17. Proposed GMDH and all studied models' $E_{min.}$ (%), SD, and R comparison.

17.33% and R values of 0.9621 and 0.9663, respectively (Table 3). The multiple regression method has different AAPRE and R values for different models, as shown in Table 3. Graphical methods (Standing's[6] and Lasater's[7] models) have AAPRE values of 12.35 and 11.07% and R values of 0.9753 and 0.9742, respectively (Table 3). The Microsoft excel method (Dindoruk and Christman's[21] model) has an AAPRE of 20.89% and R of 0.9369.

4. CONCLUSIONS

The GMDH model uses four input parameters R_s, γ_g, API, and T_f. It provides the best calculation of the bubble point pressure

https://doi.org/10.1021/acsomega.2c00651
ACS Omega 2022, 7, 13196−13209

Table 3. Models' Methods, AAPRE, and *R*

no.	model	method	AAPRE (%)	*R*
1	Standing (1947)[6]	graphical methods	12.35	0.9753
2	Lasater (1958)[7]	graphical methods	11.07	0.9742
3	Almehaideb (1997)[12]	linear regression	54.06	0.9538
4	Velarde et al. (1997)[5]	linear regression	9.00	0.9724
5	Frashad et al. (1996)[11]	nonlinear regression	18.23	0.9621
6	Al-Shammasi (1999)[13]	nonlinear regression	17.33	0.9663
7	Glaso (1980)[9]	linear and nonlinear regressions	23.02	0.9701
8	Vasquez and Beggs (1980)[8]	multiple regression	16.88	0.9767
9	Al-marhoun (1988)[14]	multiple regression	54.06	0.9538
10	Kartoatmodjo and Schmidt (1991)[15]	multiple regression	16.94	0.9722
11	Petrosky and Farshed (1993)[17]	multiple regression	76.59	0.9703
12	Macary and El-Batanoney (1993)[18]	multiple regression	31.20	0.9499
13	De Ghetto et al. (1994)[19]	multiple regression	18.37	0.9720
14	Mehran et al. (2006)[23]	multiple regression	9.75	0.9699
15	Bolodazadeh et al. (2006)[22]	multiple regression	40.42	0.9694
16	Hemmati and Kharrat (2007)[24]	multiple regression	13.76	0.9741
17	Mazandarani and Asghari (2007)[25]	multiple regression	23.91	0.9462
18	Sharrad and Abd-Alrahman (2019)[30]	multiple regression	45.93	0.8929
19	Dindoruk and Christman (2001)[21]	Microsoft Excel	20.89	0.9369
20	Khamechi et al. (2009)[46]	genetic algorithm	31.24	0.9652
21	Arabloo et al. (2014)[26]	LINGO	12.66	0.9589
22	proposed GMDH	Group Method of Data Handling (GMDH)	8.51	0.9883

(P_b) that is substantial for performing petroleum engineering calculations related to reservoir characterization and simulation and well performance analysis.

In this study, the GMDH approach used 760 data points to build, validate, and compare the model's performance with the current models available in the literature. The GMDH model is validated by applying trend analysis, where it follows the physical behavior of input parameters. The trend analysis shows the reliability GMDH model.

AAPRE and *R* were selected as the primary indicators to show the accuracy of the GMDH and all studied models. The GMDH model presented 8.51%, the lowest AAPRE that outperforms the existing correlations. The model also showed an *R* of 0.9883, the highest of the other investigated correlations, which provided the GMDH model's robustness. The GMDH model showed the lowest RMSE, SD, $E_{max.}$, and $E_{min.}$ of 12.70, 0.09, 57.26%, and 0.010%, respectively, compared to existing models. The GMDH model can predict the P_b directly without using computer software like other machine learning and deep learning (LSTM) approaches and surpass all studied correlations. Future studies can be achieved utilizing the GMDH model for other petroleum industry problems, namely, predicting permeability, porosity and sand production, reservoir characterization, and rock mechanics properties. Moreover, the GMDH model provides a straightforward correlation that can be applied directly by many petroleum applications.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.2c00651.

Statistical error analysis equations (PDF)

AUTHOR INFORMATION

Corresponding Authors

Fahd Saeed Alakbari − *Petroleum Engineering Department, Universiti Teknologi PETRONAS, Bandar Seri Iskandar, Perak 32610, Malaysia*; orcid.org/0000-0002-3227-698X; Email: fahd_19001032@utp.edu.my, alakbarifahd@gmail.com

Mysara Eissa Mohyaldinn − *Petroleum Engineering Department, Universiti Teknologi PETRONAS, Bandar Seri Iskandar, Perak 32610, Malaysia*; Email: mysara.eissa@utp.edu.my

Authors

Mohammed Abdalla Ayoub − *Petroleum Engineering Department, Universiti Teknologi PETRONAS, Bandar Seri Iskandar, Perak 32610, Malaysia*

Ali Samer Muhsan − *Mechanical Engineering Department, Universiti Teknologi PETRONAS, Bandar Seri Iskandar, Perak 32610, Malaysia*

Ibnelwaleed A. Hussein − *Gas Processing Center, College of Engineering and Chemical Engineering Department, College of Engineering, Qatar University, P.O. Box 2713 Doha, Qatar*; orcid.org/0000-0002-6672-8649

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.2c00651

Author Contributions

F.S.A. implemented the original draft preparation, methodology, software, review and editing, and data collection. M.E.M. implemented the problem conceptualization, research supervision and funding, and drafting. M.A.A. implemented the model development, result verification, and review and editing. A.S.M. implemented the review and editing. I.A.H. implemented the visualization and the review and editing.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors would like to give their sincerest thanks to the Universiti Teknologi PETRONAS for supporting this study under YUTP-Grant cost center 015LC0-105.

NOMENCLATURE

Latin Synonyms

P_b bubble point pressure, psi
R correlation coefficient
$E_{max.}$ maximum absolute percent relative error

$E_{min.}$ minimum absolute percent relative error
API oil API gravity, °API
R_s gas solubility, SCF/STB
T_f reservoir temperature, °F
B_{ob} oil formation volume factor at the bubble point pressure, bbl/STB
Y dependent variable (bubble point pressure (P_b), psi)
x_i independent variables
X vector: $X = (x_1, x_2, x_3, ..., x_n)$
X_1 gas solubility (R_s), SCF/STB
X_2 gas specific gravity (γ_g), fraction
X_3 oil specific gravity (API), °API
X_4 reservoir temperature (T_f), °F

Greek Synonyms

γ_g gas specific gravity
γ_o oil specific gravity

Abbreviations

PVT pressure−volume−temperature
GMDH Group Method of Data Handling
AAE average absolute error
APRE average percent relative error
AAPRE average absolute percent relative error
RMSE root mean square error
SD standard deviation
USA United States of America
MS-Excel Microsoft Excel
LSTM long short-term memory
TA trend analysis

REFERENCES

(1) McCain, Jr, W. D. *The Properties of Petroleum Fluids*, 2nd ed.; PennWell Publishing company: Tulsa, Oklahoma, 1990.
(2) Bandyopadhyay, P.; Sharma, A. Development of a New Semi Analytical Model for Prediction of Bubble Point Pressure of Crude Oils. *J. Pet. Sci. Eng.* **2011**, *78*, 719−731.
(3) Farasat, A.; Shokrollahi, A.; Arabloo, M.; Gharagheizi, F.; Mohammadi, A. H. Toward an Intelligent Approach for Determination of Saturation Pressure of Crude Oil. *Fuel Process. Technol.* **2013**, *115*, 201−214.
(4) Valkó, P. P.; McCain, W. D. Reservoir Oil Bubblepoint Pressures Revisited; Solution Gas−Oil Ratios and Surface Gas Specific Gravities. *J. Pet. Sci. Eng.* **2003**, *37*, 153−169.
(5) Velarde, J.; Blasingame, T. A.; McCain, W. D. Correlation of Black Oil Properties At Pressures Below Bubble Point Pressure - A New Approach. In *Annual Technical Meeting* , Calgary, Alberta, June 1997; Petroleum Society of Canada, *1997*, DOI: 10.2118/97-93.
(6) Standing, M. B. A Pressure-Volume-Temperature Correlation for Mixtures of California Oils and Gases. In *Drilling and Production Practice*; American Petroleum Institute, 1947.
(7) Lasater, J. A. Bubble Point Pressure Correlation. *J. Pet. Technol.* **1958**, *10*, 65−67.
(8) Vasquez, M.; Beggs, H. D. Correlations for Fluid Physical Property Prediction. *J. Pet. Technol.* **1980**, *32*, 968−970.
(9) Glaso, O. Generalized Pressure-Volume-Temperature Correlations. *J. Pet. Technol.* **1980**, *32*, 785−795.
(10) Omar, M. I.; Todd, A. C. Development of New Modified Black Oil Correlations for Malaysian Crudes. In *SPE Asia Pacific Oil and Gas Conference,* Singapore, February 1993; Society of Petroleum Engineers, *1993*, DOI: 10.2118/25338-MS.
(11) Frashad, F.; LeBlanc, J. L.; Garber, J. D.; Osorio, J. G. Empirical PVT Correlations For Colombian Crude Oils. In *SPE Latin America/Caribbean petroleum engineering conference,* Port-of-Spain, Trinidad, April 1996; Society of Petroleum Engineers, **1996**, DOI: 10.2118/36105-MS.
(12) Almehaideb, R. A. Improved PVT Correlations For UAE Crude Oils. In *Middle east oil show and conference,* Bahrain, March 1997; Society of Petroleum Engineers, **1997**, DOI: 10.2118/37691-MS.
(13) Al-Shammasi, A. A. Bubble Point Pressure and Oil Formation Volume Factor Correlations. In *Middle East Oil Show and Conference*, Bahrain, February 1999. Society of Petroleum Engineers: Bahrain **1999**, p 17, DOI: 10.2118/53185-MS.
(14) Al-Marhoun, M. A. PVT Correlations for Middle East Crude Oils. *J. Pet. Technol.* **1988**, *40*, 650−666.
(15) Kartoatmodjo, T. R. S.; Schmidt, Z. *New Correlations for Crude Oil Physical Properties*. 1991.
(16) Dokla, M.; Osman, M. Correlation of PVT Properties for UAE Crudes (Includes Associated Papers 26135 and 26316). *SPE Form. Eval.* **1992**, *7*, 41−46.
(17) Petrosky, G. E.; Farshad, F. F. Pressure-Volume-Temperature Correlations for Gulf of Mexico Crude Oils. In *SPE Annual Technical Conference and Exhibition* , Houston, Texas , October 1993; Society of Petroleum Engineers, *1993*, DOI: 10.2118/26644-MS.
(18) Macary, S. M.; El-Batanoney, M. H. Derivation of PVT Correlations for the Gulf of Suez Crude Oils. *Sekiyu Gakkai Shi* **1993**, *36*, 472−478.
(19) De Ghetto, G.; Paone, F.; Villa, M. Reliability Analysis on PVT Correlations. In *the European Petroleum Conference* Held in London, UK, 25−27. October 1994; Society of Petroleum Engineers, **1994**.
(20) Hanafy, H. H.; Macary, S. M.; ElNady, Y. M.; Bayomi, A. A.; El Batanony, M. H. Empirical PVT Correlations Applied to Egyptian Crude Oils Exemplify Significance of Using Regional Correlations. In *International Symposium on Oilfield Chemistry*, Houston, Texas, February 1997; Society of Petroleum Engineers, *1997*, DOI: 10.2118/37295-MS.
(21) Dindoruk, B.; Christman, P. G. PVT Properties and Viscosity Correlations for Gulf of Mexico Oils. *SPE Annual Technical Conference and Exhibition,* New Orleans, Louisiana, September 2001; Society of Petroleum Engineers, **2001**, p 14, DOI: 10.2118/71633-MS.
(22) Bolondarzadeh, A.; Hashemi, S.; Solgani, B. The New PVT Generated Correlations of Iranian Oil Properties. In *4th Iranian Petroleum Engineering Student conference*; 2006.
(23) Mehran, F.; Movagharnejad, K.; Didanloo, A. New Correlation for Estimation of Formation Vilume Factor and Bubblepoint Pressure for Iranian Oil Fields. In *1st Iranian Pet. Eng. Conference*; 2006.
(24) Hemmati, M. N.; Kharrat, R. A Correlation Approach for Prediction of Crude Oil PVT Properties. In *SPE Middle East Oil and Gas Show and Conference,* Manama, Bahrain, March 2007; Society of Petroleum Engineers, *2007*, DOI: 10.2118/104543-MS.
(25) Mazandarani, M. T.; Asghari, S. M. Correlations for Predicting Solution Gas-Oil Ratio, Bubblepoint Pressure and Oil Formation Volume Factor at Bubblepoint of Iran Crude Oils. In *European Congress of Chemical Engineering,* Copenhagen; 2007.
(26) Arabloo, M.; Amooie, M.-A.; Hemmati-Sarapardeh, A.; Ghazanfari, M.-H.; Mohammadi, A. H. Application of Constrained Multi-Variable Search Methods for Prediction of PVT Properties of Crude Oil Systems. *Fluid Phase Equilib.* **2014**, *363*, 121−130.
(27) Ahmadi, M. A.; Zendehboudi, S.; James, L. A.; Elkamel, A.; Dusseault, M.; Chatzis, I.; Lohi, A. New Tools to Determine Bubble Point Pressure of Crude Oils: Experimental and Modeling Study. *J. Pet. Sci. Eng.* **2014**, *123*, 207−216.
(28) Gomaa, S. New Bubble Point Pressure Correlation for Middle East Crude Oils. *Int. Adv. Res. J. Sci. Eng. Tech* **2016**, *3*, 1−9.
(29) Alakbari, F. S.; Elkatatny, S.; Baarimah, S. O. Prediction of Bubble Point Pressure Using Artificial Intelligence AI Techniques. In *SPE Middle East Artificial Lift Conference and Exhibition*, Manama, Kingdom of Bahrain, November 2016; Society of Petroleum Engineers, DOI: 10.2118/184208-MS.
(30) Sharrad, M.; Abd-Alrahman, H. H. New derived correlations for libyan crude oil to estimate bubble-point pressure. *Sci. J. Appl. Sci. Sabratha Univ.* **2019**, *2*, 1−13.
(31) Elkatatny, S.; Mahmoud, M. Development of New Correlations for the Oil Formation Volume Factor in Oil Reservoirs Using

Artificial Intelligent White Box Technique. *Petroleum* **2018**, *4*, 178−186.
(32) Elkatatny, S.; Mahmoud, M. Development of a New Correlation for Bubble Point Pressure in Oil Reservoirs Using Artificial Intelligent Technique. *Arab. J. Sci. Eng.* **2018**, *43*, 2491−2500.
(33) Yang, X.; Dindoruk, B.; Lu, L. A Comparative Analysis of Bubble Point Pressure Prediction Using Advanced Machine Learning Algorithms and Classical Correlations. *J. Pet. Sci. Eng.* **2020**, *185*, 106598.
(34) Tariq, Z.; Mahmoud, M.; Abdulraheem, A. Machine Learning-Based Improved Pressure−Volume−Temperature Correlations for Black Oil Reservoirs. *J. Energy Resour. Technol.* **2021**, *143*, 113003.
(35) Alakbari, F. S.; Mohyaldinn, M. E.; Ayoub, M. A.; Muhsan, A. S. Deep Learning Approach for Robust Prediction of Reservoir Bubble Point Pressure. *ACS Omega* **2021**, *6*, 21499−21513.
(36) Ayoub, M. A.; Elhadi, A.; Fatherlhman, D.; Saleh, M. O.; Alakbari, F. S.; Mohyaldinn, M. E. A New Correlation for Accurate Prediction of Oil Formation Volume Factor at the Bubble Point Pressure Using Group Method of Data Handling Approach. *J. Pet. Sci. Eng.* **2022**, *208*, 109410.
(37) Ayoub, M. A.; Mohamed, A. A. Estimating the Lengthy Missing Log Interval Using Group Method of Data Handling (GMDH) Technique. *Appl. Mech. Mater.* **2014**, *695*, 850−853.
(38) Menad, N. A.; Noureddine, Z.; Hemmati-Sarapardeh, A.; Shamshirband, S.; Mosavi, A.; Chau, K. Modeling Temperature Dependency of Oil - Water Relative Permeability in Thermal Enhanced Oil Recovery Processes Using Group Method of Data Handling and Gene Expression Programming. *Eng. Appl. Comput. Fluid Mech.* **2019**, *13*, 724−743.
(39) Shen, C.; Asante-Okyere, S.; Yevenyo Ziggah, Y.; Wang, L.; Zhu, X. Group Method of Data Handling (GMDH) Lithology Identification Based on Wavelet Analysis and Dimensionality Reduction as Well Log Data Pre-Processing Techniques. *Energies* **2019**, *12*, 1509.
(40) Ayoub, M.; Almansour, A.; Hassan, A. A Novel Formula for Estimating Oil Compressibility Below Bubble Point Pressure Using Group Method of Data Handling: A Comparative Approach. In *SPE/IATMI Asia Pacific Oil & Gas Conference and Exhibition*, Bali, Indonesia, October 2019; Society of Petroleum Engineers, *2019*, DOI: 10.2118/196446-MS.
(41) Mathew Nkurlu, B.; Shen, C.; Asante-Okyere, S.; Mulashani, A. K.; Chungu, J.; Wang, L. Prediction of Permeability Using Group Method of Data Handling (GMDH) Neural Network from Well Log Data. *Energies* **2020**, *13*, 551.
(42) Hemmati-Sarapardeh, A.; Dabir, B.; Ahmadi, M.; Mohammadi, A. H.; Husein, M. M. Modelling Asphaltene Precipitation Titration Data: A Committee of Machines and a Group Method of Data Handling. *Can. J. Chem. Eng.* **2019**, *97*, 431−441.
(43) Dargahi-Zarandi, A.; Hemmati-Sarapardeh, A.; Shateri, M.; Menad, N. A.; Ahmadi, M. Modeling Minimum Miscibility Pressure of Pure/Impure CO2-Crude Oil Systems Using Adaptive Boosting Support Vector Regression: Application to Gas Injection Processes. *J. Pet. Sci. Eng.* **2020**, *184*, 106499.
(44) Amar, M. N.; Ghriga, M. A.; Ouaer, H.; El Amine Ben Seghier, M.; Pham, B. T.; Andersen, P. Ø. Modeling Viscosity of CO2 at High Temperature and Pressure Conditions. *J. Nat. Gas Sci. Eng.* **2020**, *77*, 103271.
(45) Rezaei, F.; Rezaei, A.; Jafari, S.; Hemmati-Sarapardeh, A.; Mohammadi, A. H.; Zendehboudi, S. On the Evaluation of Interfacial Tension (IFT) of CO2−Paraffin System for Enhanced Oil Recovery Process: Comparison of Empirical Correlations, Soft Computing Approaches, and Parachor Model. *Energies* **2021**, *14*, 3045.
(46) Khamehchi, E.; Rashidi, F.; Rasouli, H.; Ebrahimian, A. Novel Empirical Correlations for Estimation of Bubble Point Pressure, Saturated Viscosity and Gas Solubility of Crude Oils. *Pet. Sci.* **2009**, *6*, 86−90.
(47) Ivakhnenko, A. G. Polynomial Theory of Complex Systems. *IEEE Trans. Syst. Man. Cybern.* **1971**, *SMC-1*, 364−378.
(48) Oh, S.-K.; Pedrycz, W. The Design of Self-Organizing Polynomial Neural Networks. *Inf. Sci. (Ny).* **2002**, *141*, 237−258.
(49) Ma, S.-W.; Kou, C.-H.; Chen, L.; Wang, A.-P. Application of Group Method of Data Handling to Strean-Way Transition. In 2009 International Joint Conference on Artificial Intelligence, Hainan, China, 25−26 April 2009; IEEE, 2009; pp. 301−304, DOI: 10.1109/JCAI.2009.128.
(50) Onwubolu, G. C. Hybrid Self-Organizing Modeling Systems; Onwubolu, G. C., Ed.; *Studies in Computational Intelligence*; Springer Berlin Heidelberg: Berlin, Heidelberg, 2009; Vol. *211*, DOI: 10.1007/978-3-642-01530-4.
(51) Mohammed, M. A. A. GMDH multilayered algorithm in matlab. In *GMDH-Methodology and Implementation in MATLAB*; Onwubolu, G., Ed.; Imperial College Press: Covent Garden, London, UK, 2016; pp 75−124.
(52) Madala, H. R; Ivakhnenko, A. G. *Inductive Learning Algorithms for Complex Systems Modeling*; CRC press: Boca Raton, Florida, United States, 2019, DOI: 10.1201/9781351073493.
(53) Pannell, D. J. Sensitivity Analysis of Normative Economic Models: Theoretical Framework and Practical Strategies. *Agric. Econ.* **1997**, *16*, 139−152.
(54) Hill, M. C.; Tiedeman, C. R. *Effective Groundwater Model Calibration: With Analysis of Data, Sensitivities, Predictions, and Uncertainty*; John Wiley & Sons: Hoboken, New Jersey, United States, 2006.
(55) Li, S.; Li, Z.; Wang, Z. Experimental Study on the Performance of Foamy Oil Flow under Different Solution Gas−Oil Ratios. *RSC Adv.* **2015**, *5*, 66797−66806.

https://doi.org/10.1021/acsomega.2c00651
ACS Omega 2022, 7, 13196−13209