JP&M Engineering
Journal of Petroleum and Mining Engineering

# Different Machine Learning Approaches to Predict Gas Deviation Factor (Z-factor)

Elsayed, M., A.*[1, 2] Alsabaa, A.[1],Salem, A. M.[2]
[1] Department of Petroleum Engineering,College of Petroleum Engineering & Geosciences King Fahd University of Petroleum & Minerals, Dhahran, Saudi Arabia.
[2] Department of Petroleum Engineering,Faculty of Petroleum and Mining Engineering Suez University, Suez, Egypt
*Corresponding author e-mail: 2018mattia@gmail.com

**Article Info**
Received 29 Nov. 2022
Revised 19 Jul. 2023
Accepted 6 Aug. 2023

**Keywords**
Z-factor; Artificial Intelligence; Support Vector Machine; Radial Basis Function; and Functional Network

## Abstract

The gas compressibility factor indicates the gas deviation from ideal gas behavior. Accurate values of gas compressibility factor affect the estimation of reservoir fluid properties, the initial gas in place, and the natural gas production and transportation process. Gas compressibility factor can be estimated in labs; however, this method is expensive and time-consuming. Due to these challenges, numerous studies created various empirical correlations depending on the results of the equation of state. The Standing and Katz chart is regarded as a standard for estimating gas compressibility factor. Many studies proposed approaches and correlations to fit this chart, however some did not cover the entire range of data, others provided implicit methods taking long time for calculation or faced high errors out of the data range. In this study, Support Vector Machine, Radial Basis Function, and Functional Network as machine learning approaches were implemented to predict the gas compressibility factor, based on 5490 data set of Standing and Katz chart. 70% of the data set was implemented in the training process and 30% in the testing process. The data set included pseudo-reduced pressure and pseudo-reduced temperature as inputs and Z-factor as an output. Different training functions were examined for each method for the best approach optimization. In addition, machine learning best approach was compared with other correlations. The best results in this work were obtained from Radial Basis Function with 0.14 average absolute percentage error and 0.99 correlation coefficient. The developed machine learning approach performed better than the examined correlations.

## Introduction

The oil and gas reserves that can be recovered can be estimated using numerical simulation and material balance. These techniques rely on the precision of various characteristics of the fluid to recognize the thermodynamic reservoirs' performance changes related to the gas composition, as well as the pressure and temperature of oil and gas reservoirs [1–3]. The gas deviation factor (Z-factor), commonly known as the gas compressibility factor, is one of the fundamental characteristics of fluids. Accurate estimation of compressibility factor is very essential, most especially when it comes to quick estimation of initial gas in place. The difference between the real and ideal gas at specific conditions of temperature and pressure is known as the gas compressibility factor [4]. It can be stated as the ratio of the actual volume ($V_a$) to ideal gas volume ($V_{id}$):

$$Z=\frac{V_a}{V_{id}}=\frac{\text{Actual volume of gas at specified pressure and temperature}}{\text{Ideal volume of gas at standard pressure and temperature}} \quad (1)$$

The gas masses typically tend to be insignificant in the ideal situation, when the pressure is relatively low, as seen in figure 1. This figure was taken from previous research [5]. As can be observed from figure 1, the optimum gas state is one where the gas compressibility factor will have a value of 1. This fact can be clarified by the fact that under ideal circumstances, molecules of gas are sufficiently separated from one another for attraction forces to be minimal. Real volume is bigger than what the ideal gas law predicts, and the ratio of actual gas volume to ideal volume is higher than 1. As the pressure value increases, the molecules of the gas get closer to one another, allowing interaction of repulsive type to take center stage. The real gas law, which can be expressed by equation 2, is used to calculate the gas compressibility factor in the lab.

$$pV=nZRT \quad (2)$$

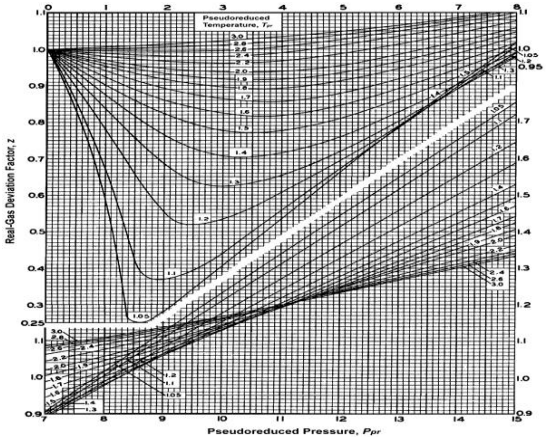


**Figure 1** Plots of gas compressibility factor.

Where "p" stands for gas pressure, "v" for gas volume, "n" for the number of moles, "R" for gas constant (commonly named universal constant), and "T" for temperature.

Equation 3 shows the law of real gases for a defined composition of the natural gas,

$$\frac{P_1V_1}{P_2V_2}=\frac{nZ_1RT_1}{nZ_2RT_2} \quad (3)$$

Equation 4 is showing the rearranged terms of equation 3 as per the previous assumption:

$$Z=\frac{P_1V_1T_2}{P_2V_2T_1} \quad (4)$$

The typical PVT cell is used in the lab to calculate the gas deviation factor. Z-factor lab measurements are costly and time-consuming to perform. Due to these challenges, numerous studies created various empirical relationships. These relationships were created using the equation of state's findings (EOS) to calculate the Z-factor. Standing and Katz [5] suggested a gas compressibility factor chart based on the law which asserts that at the same pseudo-reduced pressure ($P_{pr}$) and pseudo-reduced temperature ($T_{pr}$), different gas mixes vary to roughly the same degree (nearly similar Z-factor). Pseudo-reduced temperature and pressure can be estimated using equation 5 as defined by Dranchuk et al [6]. The Z-factor is represented as a function of these values for generalization purposes:

$$T_{pr}=\frac{T}{T_{pc}} \quad , \quad P_{pr}=\frac{P}{P_{pc}} \quad (5)$$

Where "$T_{pc}$" stands for pseudo-reduced temperature, "$P_{pc}$" stands for pseudo-reduced pressure

The molar abundance (mole fraction weighted) means the critical qualities of the components that make up natural gas is what is known as the pseudo-critical properties of the natural gas.

$$T_{pc}=\sum_{i=1}^{n} yiTci \quad , P_{pc}=\sum_{i=1}^{n} yiPci \quad , \quad (6)$$

Where "yi" stands for mole fraction of component i in the gas mixture, "Pci" stands for pseudo-critical pressure, "Tci" stands for pseudo-critical temperature.

As a function of specific gravity (air = 1.0), Sutton [7] provides equation 7 as follows:

$$T_{pc}=169.2+349.5\gamma_g-74.0\gamma_g^2 \quad , \quad (7)$$

$$P_{pc}=756.8-131.07\gamma_g-3.6\gamma_g^2$$

Where "$\gamma_g$" stands for Gas specific gravity

The purpose of this research is to add technical contributions to gas compressibility factor estimation by employing machine learning tools. The research presents three machine learning approaches, the Support Victor Machine (SVM), Radial Basis Function (RBF), and Functional Network (FN)) to calculate the Z-factor based on Standing and Katz chart data. These three approaches are trained and tested using Standing and Katz charts data and evaluated by monitoring two critical statistical metrics, average absolute percentage error (AAPE) and correlation coefficient (R) between the predicted values and actual measurements. In addition, the performance of the best approach among these three techniques is compared with other common empirical correlations from the literature.

## Literature Review

### Z-factor Implicit and Explicit Correlations

To determine the exact Z-factor value of a gas sample that is containing non-hydrocarbon components; laboratory studies should be carried out. However, correlations and the equation of state (EOS) have historically been more reliable in petroleum engineering [5–8], Standing & Katz, 1942 Z-factor chart is considered the most popular correlation in petroleum engineering to be used. There have been numerous attempts to produce these charts by creating implicit or explicit empirical correlations that can be utilized in place of the Standing and Katz charts method. The following three implicit correlations are examples of these Z-factor implicit correlations to be used for their accuracy, nearly unit regression coefficient correlation, and small maximum errors. Hall and Yarborough [9] developed an implicit Z-factor correlation with 1500 data points taken from Standing and Katz's original Z-factor chart and constants produced by regression, Hall and Yarborough's correlation modifies the hard sphere Carnahan-Starling equation of state. Dranchuk and Abou-Kassem (DAK) [6] correspondingly modified an implicit Z-factor correlation based on Benedict-Webb-Rubin equation of state using eleven constants based on regression analysis to determine these constants, based on 1500 data points that were taken from Katz's and standing charts. Dranchuk, Purvis, and Robinson's Correlation (DPR) calculates the Z-factor with minimal processing effort, because it only includes eight constants [10]. These correlations are useful, but when the systems' temperatures are close to the critical temperature, they fail to converge (or converge on incorrect pseudo-reduced density values). They also require expensive computations. These restrictions made the creation of the existing explicit linkages necessary [11].

Iterative processes are not necessary for explicit correlations. Therefore, unlike implicit correlations, they do not suffer the convergence problem. Beggs and Brills provided one of the best explicit correlations for evaluating the Z-factor [12]. Heidaryan [13], Azizi [14], and Sanjari and Lay [15], correlations are more recent examples. Brill and Beggs' compressibility factor [12] can be explained by equation 8 as follows:

$$Z=A+\frac{1-A}{e^{B}}+Cp_{pr}^{D} \quad (8)$$

$A=1.39\left(T_{pr}-0.92\right)^{0.5}-0.36T_{pr}-0.010$

$,B=(0.62-0.23T_{pr})pr+\left(\frac{0.066}{T_{pr}-0.86}-0.037\right)p_{pr}^{2}+\frac{0.32P_{pr}^{2}}{10^{E}}$

$C=0.132-0.32\log(T_{pr})$, $D=10^{F}$, $E=9(T_{pr}-1)$ and $F=0.3106-0.49T_{pr}+0.1824T_{Pr}^{2}$

Heidaryan [13] used regression analysis to create an explicit Z-factor correlation with a correlation coefficient of 0.99 and total of 22 constants based on the pseudo-reduced pressure range. Equation 9 expresses Heidaryan's explicit Z-factor correlation.

$$Z=\ln\left(\frac{A_1+A_3\ln\left(P_{pr}\right)+\frac{A_5}{T_{pr}}+A_7\left(\ln\left(P_{pr}\right)\right)^{2}+\frac{A_9}{T_{pr}^{2}}+\frac{A_{11}}{T_{pr}}\ln(}{1+A_2\ln\left(P_{pr}\right)+\frac{A_4}{T_{pr}}+A_6\left(\ln\left(P_{pr}\right)\right)^{2}+\frac{A_8}{T_{pr}^{2}}+\frac{A_{10}}{T_{pr}}\ln(}\right) \quad (9)$$

Where $A_1$ till $A_{11}$ are contestants and each constant has two different values depending on the data range of $P_{pr}$ greater or less than 3

Azizi modified an explicit Z-factor correlation with 20 constants within a pseudo-reduced temperature range of $1.1 \le T_{pr} \le 2$ and a pseudo-reduced pressure range of $0.2 \le P_{pr} \le 11$ [14]. Azizi explicit Z-factor correlation can be expressed by equation 10 below.

$$Z=A+\frac{B+c}{D+E} \quad (10)$$

$$A= aT_{pr}^{2.16}+bP_{pr}^{1.028}+cP_{pr}^{1.58}T_{pr}^{-2.1}+d\ln\left(T_{pr}^{-0.5}\right)$$

$$B= e+fT_{pr}^{2.4}+gP_{pr}^{1.56}+hP_{pr}^{0.124}T_{pr}^{3.033}$$

$$C= i\ln\left(T_{pr}^{-1.28}\right)+j\ln\left(T_{pr}^{1.37}\right)+k\ln\left(P_{pr}\right)+l\ln\left(P_{pr}^{2}\right)+m\ln\left(P_{pr}\right)\ln\left(T_{pr}\right)$$

$$D= 1+nT_{pr}^{5.55}+oP_{pr}^{0.68}T_{pr}^{0.33}$$

$$E= p\ln\left(T_{pr}^{1.18}\right)+q\ln\left(T_{pr}^{2.1}\right)+r\ln\left(P_{pr}\right)+s\ln\left(P_{pr}^{2}\right)+t\ln\left(P_{pr}\right)\ln\left(T_{pr}\right)$$

Where a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, and *t* are constants

Sanjari and Lay 2012 [15] introduced an explicit Z-factor correlation using 5844 data points. This correlation was developed using 16 constants overall depending on $P_{pr}$ values below and above 3 as expressed in equation 11.

$$Z=1+A_1P_{pr}+A_2P_{pr}^{2}+\frac{A_3P_{pr}^{A_4}}{T_{pr}^{A_5}}+\frac{A_6P_{pr}^{(A_4+1)}}{T_{pr}^{7}}+\frac{A_8P_{pr}^{(A_4+2)}}{T_{pr}^{(A_7+1)}} \quad (11)$$

Where $A_1$ till $A_8$ are constants.

Lateef developed an explicit z-factor correlation (Equation 12) as a multi-stage correlation based on Hall and Yarborough's implicit one within the range $1.15 \le T_{pr} \le 3$ and $6 \le P_{pr} \le 15$ with 19 constants using non-linear regression method [16].

$$Z=\frac{DP_{pr}\left(1+y+y^{2}-y^{3}\right)}{\left(DP_{pr}+Ey^{2}-Fy^{g}\right)(1-y)^{3}} \quad (12)$$

$$Y=\frac{DP_{pr}}{\frac{1+A^{2}}{C}-\frac{A^{2}B}{C^{3}}}$$

$$t=\frac{1}{T_{pr}}$$

$A=a_1te^{a_2(1-t)^{2}}P_{pr}$ ;

$B=a_3t+a_4t^{2}+a_5t^{6}P_{pr}^{6}$;

$C=a_9+a_8tP_{pr}+a_7t^{2}P_{pr}^{2}+a_6t^{3}Pr^{3}$;

$D=a_{10}te^{a_{11}(1-t)^{2}}$;

$E=a_{12}t+a_{13}t^{2}+a_{14}t^{3}$;

$F=a_{15}t+a_{16}t^{2}+a_{17}t^{3}$;

$G=a_{18}+a_{19}t$;

Where $a_1$ till $a_{19}$ are constants

The applications of machine learning methods for petroleum big data showed increasing growth over years for solving technical issues and providing optimum solutions for the operations for cost reduction for different industry segments such as drilling production and reservoir engineering [17–24].

## Machine Learning Methods

Support Vector Machine (SVM) is an efficient strong method for classification. It uses one or more vectors of features to predict labels after creating a decision boundary between two classes [25]. The decision boundary or the hyperplane, between the two classes, has an orientation that makes it the farthest from the closest data points of the two classes. The SVM approach is built by training on the given data set to determine weights and biases to build the hyperplane that separates the data with the maximum margin. In the early stage, SVM was just used to construct a linear classification [26]. The kernel method is a different style of using the SVM which, allows us to approach higher dimensional complex approachs [27]. A kernel function results in higher dimensional space for the non-linear problem by adding additional dimensions to the raw data. Calculations are done faster by Kernel function instead of doing computations in high dimensional space. The Radial basis function (RBF) structure is simple, but the application is similar to the Multilayer Perceptron (MLP) which is a static structure of the neural network which does not present feedback loops but the learning of which is supervised. The network structure for the RBF includes only three layers, which is simplifying the training process [28]. RBF networks is very efficient in dealing with extremely noisy data [29]. Non-linear transformation function is applied to the weighting vector of the hidden layer. This type of networks is used in the prediction of multi-variable continuous functions. The cost function immunization and oscillation control are used to determine the best solution [30]. There are three main components of functional

neural networks (FNs) those three components or three layers or the input and output layers containing neurons linked to the neurons existing on one or more multiple hidden layers [31,32]. They predicted the value would be the products of weights from the neurons, used to predict the ultimate value in the forward propagation. It takes the opposite direction with backpropagation as the output is used to determine the most correct wheats to get the optimum results [33–35]. Scalar outputs would be predicted by the functional neurons from the first hidden layer which is considered as a functional layer. Regular neural network layers are subsequent to the first functional hidden layer and for that, the forward propagation calculation would be straightforward [35–38].

## Methodology

This study followed a straight successful path utilizing three different artificial intelligence techniques to predict an important property for the petroleum industry. The chosen AI techniques were used before in other studies related to the petroleum industry and have proven to provide efficient results. Work started by reviewing and preparing data for the training process using artificial intelligence techniques.

### Data Description

In this study, 5490 measurements for the gas Z-factor, $T_{pr}$, and $P_{pr}$ were used to create the Z-factor (Generalized chart) for Standing and Katz [5]. Table 1 displays data statistics for every parameter. $T_{pr}$, as we can see, lies between 3 and 1.05, whereas $P_{pr}$ has a maximum of 15 and a minimum of 0.2. The Z-factor lies between 1.753 and 0.2992. The intricate link between the Z-factor, $P_{pr}$, and $T_{pr}$ will undoubtedly benefit from the use of AI approaches. Figure 2 displays the data distribution and repetition for the inputs ($P_{pr}$ and $T_{pr}$) and the Z-facto as predicted parameter.

**Table 1** Data statistical analysis

| Statistical Parameter | Pseudo-Reduced Pressure | Pseudo-Reduced Temperature | Z-factor |
|---|---|---|---|
| Mean | 7.6 | 1.74 | 1.05 |
| Median | 7.6 | 1.55 | 1.03 |
| Mode | 0.2 | 1.05 | 1 |
| Standard Deviation | 4.29 | 0.57 | 0.25 |
| Sample Variance | 18.38 | 0.33 | 0.06 |
| Kurtosis | -1.2 | -0.53 | -0.06 |
| Skewness | 2.5E-15 | 0.79 | -0.07 |
| Range | 14.8 | 1.95 | 1.45 |
| Minimum | 0.2 | 1.05 | 0.3 |
| Maximum | 15 | 3 | 1.75 |
| Sum | 45144 | 10320.75 | 6252.1 |
| Count | 5940 | 5940 | 5940 |

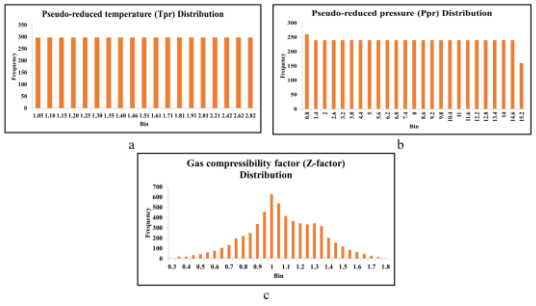


**Figure 2** Histogram plots for $T_{pr}$, $P_{pr}$, and Z-factor

### Model Development

This work is done by different machine learning tools that can be trained and tested to see the strength of these approaches to build an accurate approach which may determine the Z-factor from only two parameters which are, pseudo-reduced temperature and pseudo-reduced pressure included in standing and Katz charts data. The approach yielded trustworthy approachs based on artificial intelligence. The Support Victor Machine technique has been applied with the 5490 data sets randomized and separated into the training set and testing set considering that both sets should cover the full range. It was important to assure that the training and testing sets both must be covering the full range to assure the quality of the approachs. Even the validation set is determined following the same rule regarding the range of the data. The training testing ratio is determined to allow the maximum number of data enough to build the approach in the training and validation phases. At the same time, it was important to assure the quality of the approach and prove its generalization by enough percentage of testing data points. The 70% training set to 30% testing set was believed to give the best results and prove the validity of the approach. The data percentage was fixed for all the algorithms runs.

#### Support vector machine model development

Artificial intelligence techniques use a group of mathematical functions which can be defined as the kernel. The kernel is a way of computing the dot product of two vectors x and y in some (very high dimensional) feature space, which is why kernel functions are sometimes called "generalized dot product". The kernel function is to take data as input and transform it into the required form. The kernel functions return the inner product between two points in a suitable feature space. Thus, by defining a notion of similarity, with a little computational cost even in very high-dimensional spaces. Different SVM algorithms use different types of kernel functions. These functions can be different types. For example, linear, nonlinear, and polynomial. The code used for the SVM was run several times changing the kernel function and checking the accuracy of the results to get the optimum function. Other parameters were optimized by trying different inputs and evaluating the accuracy of the resulted Z-factor value against actual values. The main other values that can affect the SVM results are regularization strength (Lambda), the C parameter which is the penalty parameter of the error term, and the value

of epsilon (ϵ) which defines a margin of tolerance where no penalty is given to errors. The regularization strength (Lambda) was believed not to have a significant effect on the accuracy of the results in this study. Lambda was kept at the value of 0.0001 where the C parameter had a range of 0.045 to 4500 and the epsilon value was changed throughout several trials from 0.01 to 0.5.

**Radial basis function networks model development**

RBF mainly consists of two layers, the hidden layer, and the linear output layer. The design method of the RBF network can be one of two types which are newrb and newrbe. The transfer function of the RBF is radial basis transfer function (Radbas.). Euclidean distance weight function (dist) is used to calculate its weighted inputs and Product net input function (netprod) is used to calculate the net input. The linear transfer function (purelin) is used in the second layer and calculated. In the second layer, the weighted input is determined with dot prod, where sum net input function (netsum) is used to determine net inputs with. The first and second layers have biases. A two-layer network with 0% error on training vectors may be produced using newrbe, which also iteratively builds radial basis networks, one neuron at a time. Similar to newrbe, newrb uses a similar design approach. The distinction is that newrb develops neurons one by one. A radbas neuron is created at each iteration using the input vector that lowers the network error the best. The new network's error is examined, and if it's low enough, newrb is completed. If not, the next neuron is inserted. Until the error objective is attained, or the maximum number of neurons is achieved, this process is repeated. The number of neurons with newrb was set to change from 5 to 50 to determine the optimum architecture for the net. The spread constant is another value that affects the results of the RBF approach. In this study, different values ranging from 0.01 to 8 were set and results were evaluated to get the optimum spread constant value.

**Functional networks model development**

FNs are considered a unique generalization of a NN, which uses data to predict the functions of neurons and domain knowledge to design the network's structure. Functional networks' ability to handle functional restrictions based on functional characteristics that could be aware in the approach is a key property (e.g., associativity, distributivity, etc.). The components of a functional networks (FNs) are as follows: the input data is stored in one layer of input storage units. The output data is stored in one layer of output storage units. A set of input values from the preceding layer (of intermediate or input units) are evaluated by one or more layers of processing units, which then deliver a set of output values to the following layer (of intermediate or output units). Each neuron hence has a corresponding neuron function, which may be multivariate and contain an equal number of arguments and inputs. A functional cell is a unit (univariate) of a neuronal function. Oval shapes with the name of the associated function within are used to depict neurons. Nothing, a single layer, or many layers of units that hold the intermediate information generated by neuron units. The output of the processing units may be forced to coincide owing to these layers. a group of direct connections that join units in the input or intermediate layers with neurons, and neurons with intermediate or output units. When choosing a FN, there are two aspects to take into account: first, use the functions' family; second, choose the functions' items from the family [39]. Different functions 'family were used in this study such as functional network forward–backward (FNFBM), functional network backward-forward (FNBFM), FNFSM Functional network forward-selection method, FNESM Functional network exhaustive-search method, FNBEM Functional network backward-elimination method.

## Results

The machine learning approach for developing the approaches is mainly evaluated by determining the errors between the actual and predicted values for the Z factor using two statistical metrics named the correlation coefficient (R) and the average absolute percentage error (AAPE). The SVM approach code was run with different algorithms using different types of kernel functions such as Gaussian, Polynomial. The parameter C (penalty parameter of the error term), Lambda (regularization strength), and epsilon (defines a margin of tolerance where no penalty is given to errors) were changed for every run to reach the best result. The best result was achieved with Kernel function 'Gaussian" with C=450, Lambda= 0.000, and epsilon= 0.01. The training and testing data showed 0.58 and 0.6 for the AAPE respectively and 0.99 R for both as shown in figure 3.

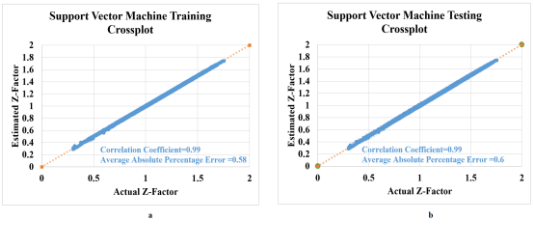


**Figure 3** Results cross plots for training (a) and testing processes (b) for Support Vector Machine approach.

The AAPE and R for the whole data were 0.59 and 0.99 respectively as mentioned in figure 4.

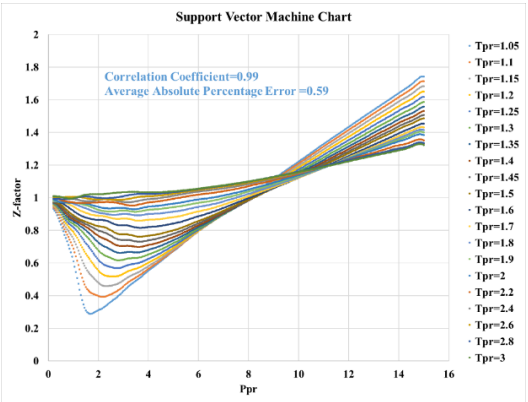


**Figure 4** Z- Factor chart for Support Vector Machine approach.

RBF approaches, newrbe and newrb codes were tested to predict the Z-factor using the same data set for the training and the testing data. Newrbe approach showed better results than newrb with 0.99 for R and 0.14 for AAPE. 0.13 and 0.16 AAPE were obtained for both the training and testing data respectively, while R is 0.99 for both as shown in figures 5 and 6.

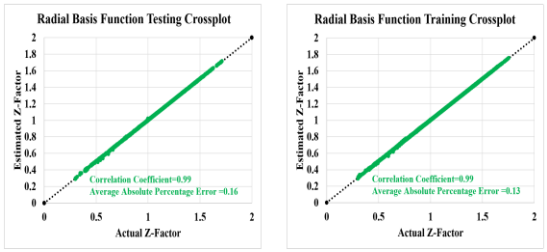


**Figure 5** Results cross plots for training (a) and testing processes (b) for Radial Basis Function approach.

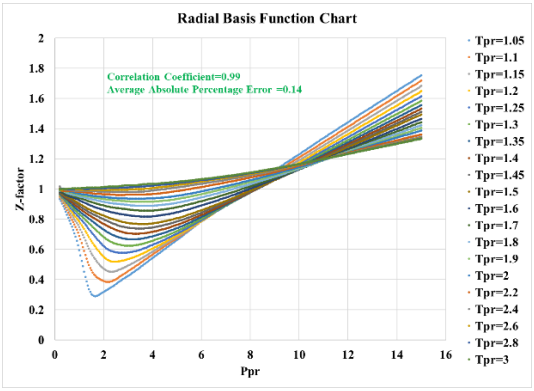


**Figure 6** Z- Factor chart for Radial Basis Function approach.

Table 2 and figure 7 summarize the best net main parameters of the best approach obtained from the RBF ML technique.

**Table 2** The developed Radial Basis Function

| Radial Basis Function Structure | Ranges |
|---|---|
| Input Features | 2 |
| Output | 1 (Z factor) |
| Hidden layer | 1 |
| Transfer function for hidden-layer | Radial Basis |
| Neurons Number | 4158 |
| Transfer function for outer layer | Purline |
| Training to Testing | 70 to 30 |

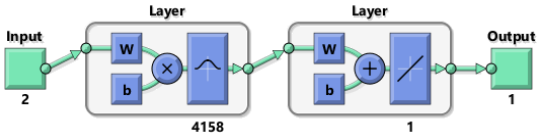


**Figure 7** Schematic of the structure of the best RBF net.

FNs different ML techniques were tested with the same data set giving high AAPE compared with SVM and RBF. It can be seen from figure 8 that regression between the estimated and actual data for both the training and testing have 2.5 and 2.8 AAPE respectively. The AAPE for the whole data set is 2.6 and shows 0.99 for R as shown in Figure 9.

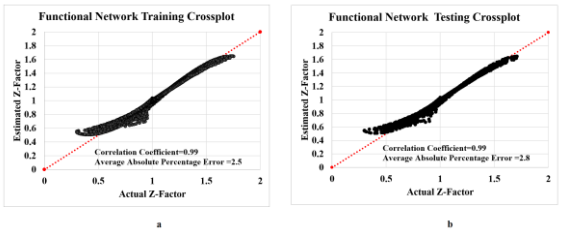


**Figure 8** Results cross plots for training (a) and testing processes (b) for Functional Network approach.

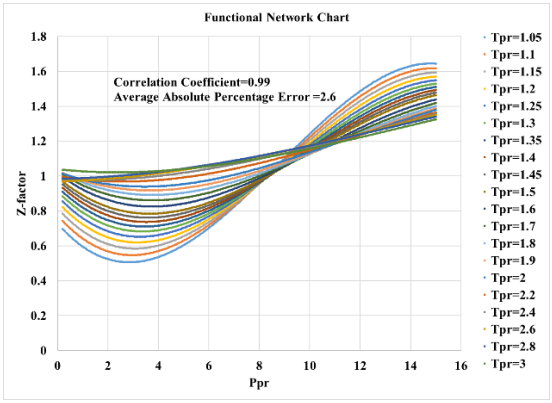


**Figure 9** Z- Factor chart for Functional Network approach.

After comparing the three approaches to predict Z-factor using Standing and Katz charts data, RBF showed better results than SVM and FN. In the following section, the results of the ML techniques will be compared with different published correlations from the literature to predict Z-factor using Standing and Katz charts data.

## Discussion

Scientists worked on this topic before as it is important to eliminate the need for the manual usage of Standing & Katz charts. This study is outperforming all the previous research in terms of accuracy and application range. In this section of the study, a detailed explanation and comparison with the most accurate Z-factor correlations from previous work, would emphasize the advantage of this research.

Lateef's explicit correlation was proposed to calculate the Z-factor [16]. Lateef's correlation was modified from the Hall and Yarborough's implicit correlation [9]. this correlation was proposed to be valid in the range of $1.15 \leq T_{pr} \leq 3$ and $0.2 \leq P_{pr} \leq 15$ with 0.44 AAPE and 0.99 R as shown in figure 10. Lateef's correlation compared with other explicit correlations such Sanjari and lay [15], Heidaryan et al [13], and Azizi [14] shows better performance to predict the Z-factor compared with the mentioned correlations.

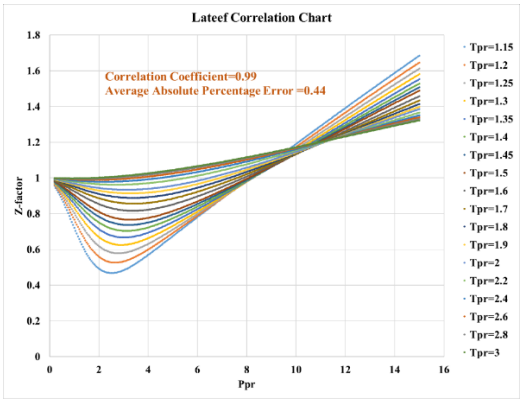


**Figure 10** the S-K chart predicted by Lateef correlation.

The best ML approach (RBF) from this research was compared to Lateef and other explicit correlations within the same data range (1.15 ≤Tpr ≤ 3 and 0.2 ≤ Ppr ≤ 15) to predict the Z-factor. Table 3 summarizes the results obtained by Lateef's correlation and what has been conducted in this research using RBF approach. RBF approach from this study predicting Z-factor is much better than the mentioned explicit correlations, resulting 0.11 AAPE and 0.99 R for the data range used in Lateef's correlation.

**Table 3** Comparison of the explicit correlations with the best ML approach (RBF)

| Correlation | AAPE (%) | R | MAXAE | MAXAAPE |
|---|---|---|---|---|
| Sanjari and Lay (2012) | 3.7 | 0.95 | 0.77 | 45.57 |
| Heidaryan et al. (2010) | 0.49 | 0.99 | 0.02 | 3.7 |
| Azizi et al. (2010) | 13.5 | 0.87 | 0.35 | 60 |
| Lateef correlation | 0.44 | 0.99 | 0.03 | 5.99 |
| The developed RBF model | 0.11 | 0.99 | 0.02 | 2 |

The above comparison proved the advantage of the current study over literature work in terms of accuracy. From another side, this study is outperforming the previous work in terms of range. The range of $T_{pr}$ used to develop Lateef's correlation is not the same range for Standing & Katz charts. All the AI approaches developed in this research used the full range same as the Standing & Katz charts' $P_{pr}$ and $T_{pr}$ ranges.

## Conclusions

Results from the different machine learning techniques (Support vector machine (SVM), Radial basis function (RBF), and Functional network (FN)) showed the charts data range (1.05 ≤$T_{pr}$ ≤ 3 and 0.2 ≤ $P_{pr}$ ≤ 15). RBF (newrbe) machine learning technique results performance is better than SVM and FN giving 0.99 for the correlation coefficient and 0.14 for the average absolute percentage error. The approach inputs are the pseudo-reduced temperature ($T_{pr}$) and pseudo-reduced pressure ($P_{pr}$) and Z-factor as the output. The developed RBF approach was compared with the other Z-factor explicit correlations and outperformed all previous work. The accuracy of the results of the machine learning technique (RBF) is 0.11 for the average absolute percentage error and 0.99 for the correlation coefficient. This performance accuracy exceeds the one by Lateef and the other empirical correlation within the range of (1.15 ≤$T_{pr}$ ≤ 3 and 0.2 ≤ $P_{pr}$ ≤ 15).

## Funding

There is no external fund for this research.

## Author Contributions

A.M. supervised the work and reviewed the manuscript writing. M.A., and A.A. performed the data collection, methodology work, results analysis, and manuscript writing.

## Conflicts of Interest

The author declares no conflict of interest.

## References

[1] R. Gharbi, Estimating the Isothermal Compressibility Coefficient of Undersaturated Middle East Crudes Using Neural Networks, Energy & Fuels. 11 (1997) 372–378. https://doi.org/10.1021/ef960123y.

[2] T. Kimura, M. Uematsu, K. Watanabe, Measurements of compressibility factor and vapor pressure for Refrigerant 502, J Chem Eng Data. 26 (1981) 109–112. https://doi.org/10.1021/je00024a001.

[3] A.M. Elsharkawy, Y.S.Kh.S. Hashem, A.A. Alikhan, Compressibility Factor for Gas Condensates, Energy & Fuels. 15 (2001) 807–816. https://doi.org/10.1021/ef000216m.

[4] W.D. McCain, The properties of petroleum fluids, 2nd ed., 1990. https://books.google.com.sa/books/about/The_Properties_of_Petroleum_Fluids.html?id=EWxUFzW61wkC&redir_esc=y.

[5] M.B. Standing, D.L. Katz, Density of Natural Gases, Transactions of the AIME. 146 (1942) 140–149. https://doi.org/10.2118/942140-G.

[6] P.M. Dranchuk, J.H. Abou-Kassem, CALCULATION OF Z FACTORS FOR NATURAL GASES USING EQUATIONS OF STATE., Journal of Canadian Petroleum Technology. 14 (1975) 34–36. https://doi.org/10.2118/75-03-03/2166813/PETSOC-75-03-03.PDF/1.

[7] R.P. Sutton, Compressibility Factors for High-Molecular-Weight Reservoir Gases, in: SPE Annual Technical Conference and Exhibition, Society of Petroleum Engineers, 1985. https://doi.org/10.2118/14265-MS.

[8] W.A. Fouad, M.I.L. Abutaqiya, K. Mogensen, Y.F. Yap, A. Goharzadeh, F.M. Vargas, L.F. Vega, Predictive Model for Pressure–Volume–Temperature Properties and Asphaltene Instability of Crude Oils under Gas Injection, Energy & Fuels. 32 (2018) 8318–8328. https://doi.org/10.1021/acs.energyfuels.8b01783.

[9] K.R. Hall, L. Yarborough, A new equation of state for Z-factor calculations, Oil Gas J. 71 (1973) 82–92.

[10] R.A. Purvis, D.B. Robinson, P.M. Dranchuk, A Reduced Equation of State Applied to Generalized Compressibility Factor Tablest, Journal of Canadian Petroleum Technology. 10 (1971). https://doi.org/10.2118/71-04-03.

[11] D. RA, P. DB, R. PM, Generalized compressibility factor tables, J Can Pet Technol. 10 (1971) 22–29.

[12] D.H. Beggs, J.P. Brill, A Study of Two-Phase Flow in Inclined Pipes, Journal of Petroleum Technology. 25 (1973) 607–617. https://doi.org/10.2118/4007-PA.

[13] E. Heidaryan, A. Salarabadi, J. Moghadasi, A novel correlation approach for prediction of natural gas compressibility factor, Journal of Natural Gas Chemistry. 19 (2010) 189–192. https://doi.org/10.1016/S1003-9953(09)60050-5.

[14] N. Azizi, R. Behbahani, M.A. Isazadeh, An efficient correlation for calculating compressibility factor of natural gases, Journal of Natural Gas Chemistry. 19 (2010) 642–645. https://doi.org/10.1016/S1003-9953(09)60081-5.

[15] E. Sanjari, E.N. Lay, An accurate empirical correlation for predicting natural gas compressibility factors, Journal of Natural Gas Chemistry. 21 (2012) 184–188. https://doi.org/10.1016/S1003-9953(11)60352-6.

[16] L.A. Kareem, T.M. Iwalewa, M. Al-Marhoun, New explicit correlation for the compressibility factor of natural gas: linearized z-factor isotherms, J Pet Explor Prod Technol. 6 (2016) 481–492. https://doi.org/10.1007/s13202-015-0209-3.

[17] S. Elkatatny, M. Mahmoud, Development of new correlations for the oil formation volume factor in oil reservoirs using artificial intelligent white box technique, Petroleum. 4 (2018) 178–186. https://doi.org/10.1016/j.petlm.2017.09.009.

[18] S. Elkatatny, Z. Tariq, M. Mahmoud, Real time prediction of drilling fluid rheological properties using Artificial Neural Networks visible mathematical model (white box), J Pet Sci Eng. 146 (2016) 1202–1210. https://doi.org/10.1016/j.petrol.2016.08.021.

[19] S. Mohaghegh, Virtual-Intelligence Applications in Petroleum Engineering: Part 3—Fuzzy Logic, Journal of Petroleum Technology. 52 (2000) 82–87. https://doi.org/10.2118/62415-JPT.

[20] S. Mohaghegh, R. Arefi, S. Ameri, M.H. Hefner, A Methodological Approach for Reservoir Heterogeneity Characterization Using Artificial Neural Networks, in: SPE Annual Technical Conference and Exhibition, Society of Petroleum Engineers, 1994. https://doi.org/10.2118/28394-MS.

[21] M. Mohamadi-Baghmolaei, R. Azin, S. Osfouri, R. Mohamadi-Baghmolaei, Z. Zarei, Prediction of gas compressibility factor using intelligent models, Natural Gas Industry B. 2 (2015) 283–294. https://doi.org/10.1016/j.ngib.2015.09.001.

[22] M. Al-Amri, M. Mahmoud, S. Elkatatny, H. Al-Yousef, T. Al-Ghamdi, Integrated petrophysical and reservoir characterization workflow to enhance permeability and water saturation prediction, Journal of African Earth Sciences. 131 (2017) 105–116. https://doi.org/10.1016/j.jafrearsci.2017.04.014.

[23] A.A.A. Mahmoud, S. Elkatatny, M. Mahmoud, M. Abouelresh, A. Abdulraheem, A. Ali, Determination of the total organic carbon (TOC) based on conventional well logs using artificial neural network, Int J Coal Geol. 179 (2017) 72–80. https://doi.org/10.1016/j.coal.2017.05.012.

[24] A. Choubineh, E. Khalafi, R. Kharrat, A. Bahreini, A.H. Hosseini, Forecasting gas density using artificial intelligence, Pet Sci Technol. 35 (2017) 903–909. https://doi.org/10.1080/10916466.2017.1303712.

[25] W.S. Noble, What is a support vector machine?, Nat Biotechnol. 24 (2006) 1565–1567. https://doi.org/10.1038/nbt1206-1565.

[26] L. Bottou, V. Vapnik, Local Learning Algorithms, Neural Comput. 4 (1992) 888–900. https://doi.org/10.1162/neco.1992.4.6.888.

[27] and L.I.R. M. A. Aizerman, E. M. Braverman, Theoretical foundations of the potential function method in pattern recognition learning. Automation and Remote Control, Ci.Nii.Ac.Jp. 25:821-837 (1964). https://ci.nii.ac.jp/naid/10021200712/ (accessed March 27, 2022).

[28] A. Tatar, A. Shokrollahi, M. Mesbah, S. Rashid, M. Arabloo, A. Bahadori, Implementing Radial Basis Function Networks for modeling CO2-reservoir oil minimum miscibility pressure, J Nat Gas Sci Eng. 15 (2013) 82–92. https://doi.org/10.1016/j.jngse.2013.09.008.

[29] A. Rostami, M. Kalantari-Meybodi, M. Karimi, A. Tatar, A.H. Mohammadi, Efficient estimation of hydrolyzed polyacrylamide (HPAM) solution

viscosity for enhanced oil recovery process by polymer flooding, Oil and Gas Science and Technology. 73 (2018). https://doi.org/10.2516/ogst/2018006.

[30] B.M. Wilamowski, R.C. Jaeger, Implementation of RBF type networks by MLP networks, in: IEEE International Conference on Neural Networks - Conference Proceedings, 1996: pp. 1670–1675. https://doi.org/10.1109/icnn.1996.549151.

[31] A.A. Mahmoud, S. Elkatatny, A. Al-AbdulJabbar, T. Moussa, H. Gamal, D. al Shehri, Artificial neural networks model for prediction of the rate of penetration while horizontally drilling carbonate formations, in: 54th U.S. Rock Mechanics/Geomechanics Symposium, 2020. https://www.researchgate.net/publication/344594897_Artificial_Neural_Networks_Model_for_Prediction_of_the_Rate_of_Penetration_While_Horizontally_Drilling_Carbonate_Formations (accessed January 18, 2021).

[32] Alsabaa, A., Gamal, H.A., S.M. Elkatatny, A. and Abdulraheem, Real-Time Prediction of Rheological Properties of All-Oil Mud Using Artificial Intelligence, American Rock Mechanics Association. (2020). https://doi.org/https://www.onepetro.org/conference-paper/ARMA-2020-1645.

[33] A. Alsabaa, H. Gamal, S. Elkatatny, A. Abdulraheem, New correlations for better monitoring the all-oil mud rheology by employing artificial neural networks, Flow Measurement and Instrumentation. (2021) 101914. https://doi.org/https://doi.org/10.1016/j.flowmeasinst.2021.101914.

[34] M. A.A., S. Elkatatny, A. Alsabaa, D. Shehri, Functional neural networks-based model for prediction of the static young's modulus for sandstone formations, in: 54th U.S. Rock Mechanics/Geomechanics Symposium, 2020.

[35] A. Alsabaa, H. Gamal, S. Elkatatny, Y. Abdelraouf, Machine Learning Model for Monitoring Rheological Properties of Synthetic Oil-Based Mud, ACS Omega. (2022). https://doi.org/10.1021/ACSOMEGA.2C00404.

[36] F. Rossi, B. Conan-Guez, F. Fleuret, Functional data analysis with multi layer perceptrons, in: Proceedings of the International Joint Conference on Neural Networks, 2002: pp. 2843–2848. https://doi.org/10.1109/ijcnn.2002.1007599.

[37] A. Alsabaa, S. Elkatatny, Improved Tracking of the Rheological Properties of Max-Bridge Oil-Based Mud Using Artificial Neural Networks, ACS Omega. (2021) acsomega.1c01230. https://doi.org/10.1021/acsomega.1c01230.

[38] A.M. Salem, M. Attia, A. Alsabaa, A. Abdelaal, Z. Tariq, Machine Learning Approaches for Compressibility Factor Prediction at High- and Low-Pressure Ranges, Arab J Sci Eng. (2022). https://doi.org/10.1007/S13369-022-06905-3.

[39] G. Zhou, Y. Zhou, H. Huang, Z. Tang, Functional networks and applications: A survey, Neurocomputing. 335 (2019) 384–399. https://doi.org/10.1016/j.neucom.2018.04.085.