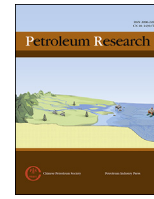
Petroleum Research 6 (2021) 417–422

Contents lists available at ScienceDirect

Petroleum Research

journal homepage: http://www.keaipublishing.com/en/journals/petroleum-research/



Forecasting multiphase flowing bottom-hole pressure of vertical oil wells using three machine learning techniques

Nagham Amer Sami [a], Dhorgham Skban Ibrahim [a, b, *]

[a] Petroleum and Natural Gas Institute, Faculty of Earth Science and Engineering, University of Miskolc, 3515, Egyetemváros, Hungary
[b] Department of Petroleum Engineering, College of Engineering, University of Baghdad, Aljadrea, 10071, Baghdad, Iraq

A R T I C L E I N F O

Article history:
Received 17 April 2021
Received in revised form
10 May 2021
Accepted 10 May 2021
Available online 14 May 2021

Keywords:
Machine learning
Artificial intelligence
Bottom hole pressure
Artificial neural network
Random forest
K-nearest neighbors

A B S T R A C T

Flowing bottom-hole pressure (FBHP) is a key metric parameter in the evaluation of performances of oil and gas production wells. An accurate prediction of FBHP is highly required in the petroleum industry for many applications, such the hydrocarbon production optimization, oil lifting cost, and assessment of workover operations. Production and reservoir engineers rely on empirical correlations and mechanistic models exist in open resources to estimate the FBHP. Several empirical models have been developed based on simulation and laboratory results that involved many assumptions that reduce the model's accuracy when they are applied for the field applications. The technologies of machine learning (ML) are one discipline of Artificial Intelligence (AI) techniques provide promising tools that help solving human's complex problems.

This study develops machine-learning based models to predict the multiphase FBHP using three machine learning techniques that are Random forest, K-Nearest Neighbors (KNN), and artificial neural network (ANN). Results showed that using an artificial neural network model give error of 2.5% to estimate the FBHP which is less than the random forest and K-nearest neighbor models with error of 3.6% and 4% respectively. The ML models were developed based on a surface production data, which makes the FBHP is predicted using actual field data. The accuracy of the proposed models from ML was evaluated by comparing the results with the actual dataset values to ensure the effectiveness of the work.

The results of this study show the potential of artificial intelligence in predicting the most complex parameter in the multiphase petroleum production process.

1. Introduction

Flowing bottom-hole pressure (FBHP) is one of the critical parameters in the evaluation of performance of oil and gas wells. The value of FBHP varies during the well's life so, accurate determination and prediction of it are continuously needed for different flow conditions. FBHP is the most effective variable in the well design facilities such as: tubing size, well head completion, predicting a suitable time for artificial lift, artificial lift installations and to find production analysis, maximum flow rate, well testing and productivity indexes of wells (Ahmed et al., 2016). Installation of down-hole gauges in oil wells to determine FBHP is a dominant process. However, it is not practical, economic, and/or feasible to deploy a pressure gauge in the well to measure the FBHP (Medhat and Hassan, 2016). These gauges require continuous maintenance and calibration to avoid erroneous readings also, by intervening a well from time to time to measure FBHP is an expensive task, associated with production risk and interruption (Tariq et al., 2020).

The influx of the formation fluid to the production tube occurs as soon as the well is open to flow at the surface. In this stage, the hydrocarbon fluid with different physiochemical properties flow to the well from the perforated formation zone. The wellbore configuration, the surface choke and properties of fluids lead to changes in the distribution of pressure, volume fractions, and velocities in the wellbore with respect to space and time. As the formation fluids flow to the surface, the mixture density, liquid hold up, fluid phase and flow regime pattern change and these will define local pressure drops and, consequently, the BHP (Spesivtsev et al., 2018). Flow up the tubing string in oil and gas wells is usually

* Corresponding author. Petroleum and Natural Gas Institute, Faculty of Earth Science and Engineering, University of Miskolc, 3515, Egyetemváros, Hungary
E-mail address: oljdhorg@uni-miskolc.hu (D.S. Ibrahim).

https://doi.org/10.1016/j.ptlrs.2021.05.004

Table 1
Sample of the dataset from open resources.

Sample No.	Input Parameters									Output parameters
	Qo (bbl/day)	Qg (MScf/day)	Qw (bbl/day)	ID (inches)	Depth (ft)	API	ST (F)	BT (F)	Pwh (psia)	Pwf (psia)
0	4600	2693.37	11000	4	6621	32.6	90	212	175	2804
1	700	411.69	1300	2.441	6271	32.6	90	212	230	2368
2	8616	4230.46	2500	3.813	6294	36.5	156	208	285	2343
3	2983	1023.09	905	3.958	6345	32.8	90	180	220	2161
4	3792	1749.74	3796	4	6340	32.6	90	212	210	2289
.										
.										
.										
207	13629	5162.91	458	3.958	6285	32.6	90	212	280	2349

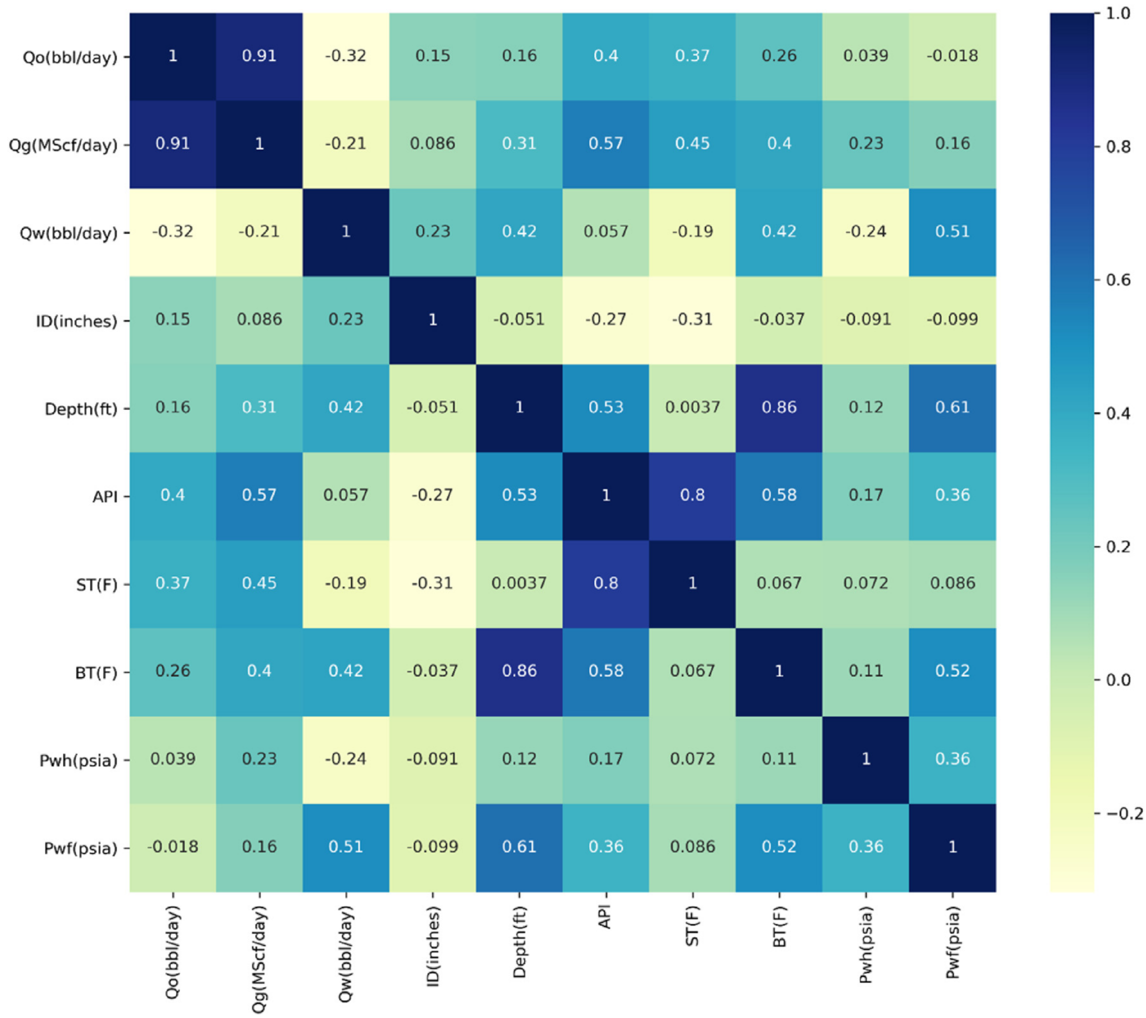


Fig. 1. The correlation plot of all attributes in the dataset.

multiphase flow.

Calculation of pressure drops in upward multiphase flow is not simple, due to the slippage of gas past liquid, along with the changing temperature and pressure conditions. The multiphase flow in oil and gas wells consists of oil, gas and water and the deformable nature of the interface in gas/liquid and liquid/liquid flows increase the complexity of the study (Pucknell et al., 1993). Although many correlations and models have been proposed in the literature to estimate pressure drop in vertical wells, the debate about their accuracy, applicability, generalization, and easy access to their parameters, persists (Orkiszewski, 1967; Aziz et al., 1972; Govier and Fogarasi, 1975; Ashiem, 1986; Ansari et al., 1990, 1994; Pucknell et al., 1993; Gomez et al., 2000).]. Petroleum industry relays on the multiphase correlations or mechanistic modelling to calculate the FBHP from the known parameters and surface measurements. These models involve many assumptions and require numerals parameters that requires much information, some are difficult to obtain, and therefore reduce the accuracy of the model and limit its applicability (Ahmadi et al., 2015). Fluid flow through a vertical production tube, needs to have a reliable technique for

Table 2
Statistical description of training and testing dataset.

Training dataset	Independent variables						Dependent variables
	Qo(bbl/day)	Qw(bbl/day)	ID (inches)	Depth(ft)	API	Pwh(psia)	Pwf(psia)
count	165	165	165	165	165	165	165
mean	6295.055	2623.848	3.812291	6374.406	33.85333	321.2182	2497
Std	4994.9	2825.31	0.42895	566.249	2.348184	156.7247	306.6559
min	280	0	1.995	4550	30	80	1227
25%	2300	0	3.813	6317	32.6	210	2293
50%	4700	1578	3.958	6533	32.6	270	2504
75%	9720	5078	3.958	6727	36.5	400	2710
max	19618	11000	4	7100	37	960	3217
Skew	0.843	0.901	−3.51	−1.634	0.008	1.311	−0.358
count	41	41	41	41	41	41	41
mean	6428	3006.512	3.919659	6301.366	33.44634	320.5122	2456.976
Std	4184.578	2670.903	0.070548	569.6029	2.189075	141.9467	284.7153
min	840	0	3.813	4550	30	180	1906
max	16437	10500	4	6933	37	750	2984
Skewness	0.9206	0.731	−0.829	−1.6155	0.3611	1.538	−0.0017

forecasting and evaluating the multiphase flow FBHP.

Artificial Intelligence is a potential technology that recently provided a reliable result for big data analysis. A machine learning algorithm is a promising tool of artificial intelligence that help human's understanding complex problems and used potentially in predicting and capturing nonlinear complex behaviors (Gene et al., 2019). The ability of machine learning prediction improves by learning from more database over time. Recently, machine learning (ML) techniques established as an alternative potential solution in many applications of oil and gas industry. (ML) is found to be trained on a large amount of previous and historical dataset for giving the problem and analytics a reasonably good approximation result. Use of data analytics and machine learning to study the processes taking place in the wellbore is a relatively new area.

Many researchers have paid attention to use machine learning technique to predict flowing bottom hole pressure in multiphase flow wells. Jahanandish et al. (2011) developed an artificial neural network (ANN) model to predict the bottomhole flowing pressure and the pressure drop in vertical multiphase flowing wells, a total of 413 wells field data used to fit and validate the ANN model. Li et al. (2014) modified BHP calculation procedure that combines the multiphase flow correlations and Back-propagation (BP) neural network models to achieve higher accuracy. The combined procedure gives an average absolute percent error of 23.0% and a standard deviation of 0.176. Ahmadi et al. (2016) utilized least square support vector machine (LSSVM) method to predict flowing bottom hole pressure with error less than 6%. Chen et al. (2017) investigated method based on support vector machine (SVM) and a support vector regression model with ϵ-insensitive loss function (ϵ-SVR) to predict the FBHP with error of 2.62%. Medhat and Hassan (2016) developed neural network models to predict the flowing bottom-hole pressure in vertical oil wells using the historical dataset measured from actual different oil fields. Spesivtsev et al. (2018) developed a model to predict multiphase bottom-hole pressure using the artificial neural network machine learning approach. They used a dataset consist of 3500 samples were generated using the simulator.

In this work, three different machine learning algorithms are used in order to predict the FBHP for multiphase flow in vertical wells. Real field data from open literature resource is used to build and evaluate the model. The accuracy of the proposed models was evaluated by a group of the datasets to validate the approach accurately of the BHPs obtained from ML models to ensure the effectiveness of the work.

2. Methodology

2.1. Data acquisition and description

A total of 206 multiphase flow data points; collected from Middle East fields; for vertical wells was obtained from open resources (Ayoub 2004). These wells are flowing naturally without any artificial lift process. During the measurements, the well bottom-hole flowing pressure is recorded using the down-hole pressure gauges inside the well just above the perforations. The dataset includes 9 production related variables that were used to predict the bottom hole flowing pressure (FBHP (psia)), flowing oil rate (Qo (bbl/day)), flowing gas rate (Qg (Mscf/day)), flowing water rate (Qw (bbl/day)), production tubing internal diameter (ID (inches)), well perforation depth (Depth (ft)), oil gravity (API), surface temperature (ST (F)), well bottom-hole temperature (BT (F)), and wellhead pressure (Pwh (psia)). The output is the measured flowing bottom hole pressure (FBHP (psia)). Table 1 lists some of the data points used for FBHP modelling.

2.2. Feature selection and scaling

Data analysis and preprocessing performed carefully in this study since, the prediction performance of any artificial intelligent (AI) models is highly depending on the quality of the data. The reduction of the data points was done to reduce the number of input parameters in order to make the model run more efficiently. Feature selection method is allowed using such data to train the ML models in least misleading way. Since there are nine independent variables (input attributes) in a dataset, it is important to select the right amount of it because too many attributes for training the model resulted in time consuming and overfitting. The first most imperative step taken in the process of data analysis of this study was to plot the attributes correlation matrix as shown in Fig. 1. It can be seen from this data exploration that some input parameters have a strong positive or negative correlation with each other and with the bottom hole flowing pressure. Anaconda open-source platform was used in this study to run Python codes in Jupyter notebook.

The existence of such strong correlations between independent variables causes the problems of collinearity and multicollinearity in linear machine learning model that reduce the predictive powers of the developed models (Munqith et al., 2017). Although, the selected machine learning algorithms are not affected by collinearity, feature selection is performed in this study to reduce the

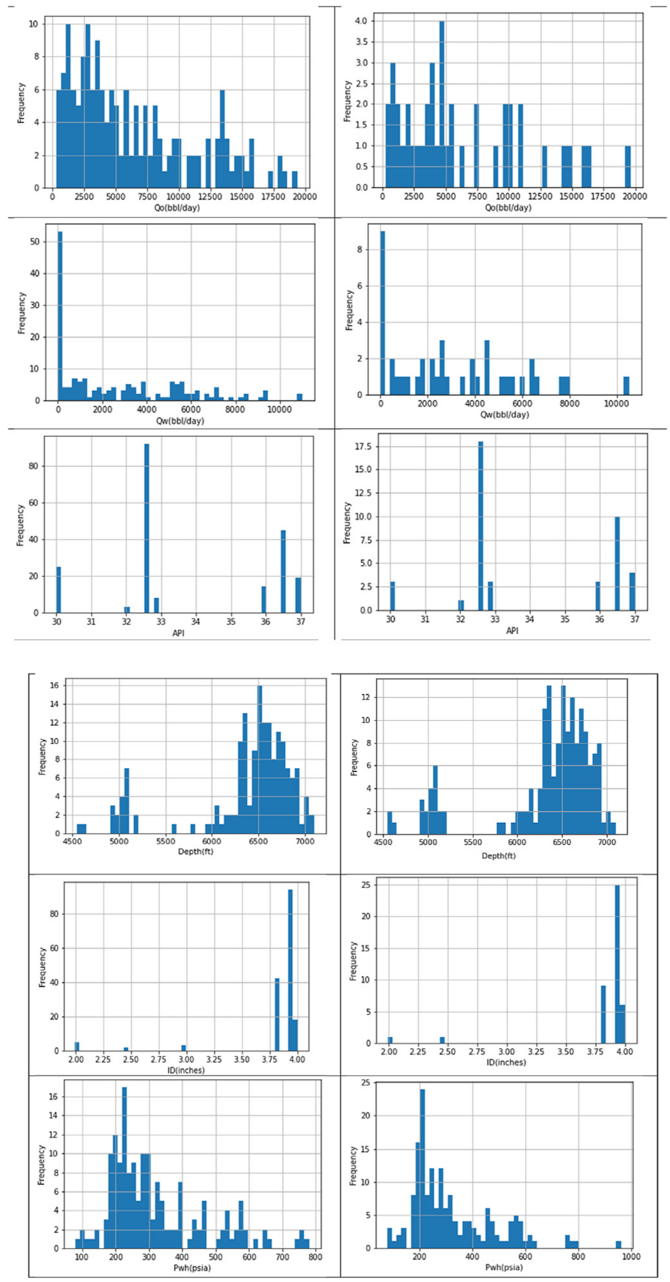


Fig. 2. Frequency histograms of a) training dataset and b) testing dataset.

number of independent variables required for further flowing bottom hole pressure prediction.

The filter method is selected in this study to reduce the number of the independent variables. Correlation filter method is generally used as a feature selection method independently of any machine learning algorithms. Instead, features are selected on the basis of their characteristics using various statistical tests for their correlation with the output variable.

A high correlation between the variables is often a useful property, if two variables are highly correlated, we can predict one from the other. The highly correlated variables provide redundant information in predict the target output. In this case, the second variable does not add additional information in the model, so removing it can help to reduce the dimensionality and essentially, the target can be predicted more accurately with just one of these redundant variables. The correlation between the variables was calculated using Pearson's correlation coefficient. Pearson correlation coefficient is a popular method to measure the linear relationship between two data variables, which can vary between 1 and -1. The correlation filter process using threshold of 0.8 reduces the number of elements from 2060 to 1442. Accordingly, the six following independent variables were finalized to be used as input variables:

Flowing oil rate (Qo (bbl/day)).
Flowing water rate (Qw (bbl/day)).
Production tubing internal diameter (ID (inches)).
Well perforation depth (Depth (ft)).
Oil gravity (API), And.
Wellhead pressure (Pwh (psia)).

Feature scaling is performed in this study during the data preprocessing to handle the highly varying units and values of the data. This technique is used to standardize the independent input variables of the dataset in a fixed range. If feature scaling is not done, machine learning algorithm tends to consider the weight of the great values, higher than the smaller values regardless of the variable's unit.

2.3. Machine learning model setup

A total of 206 tests was collected from different wells and divided into two sets to train and validate the model. The training dataset of 80% (165 data points) was used to prepare and training the models, and the remaining dataset of 20% (41 data points) as a testing set was used to validate the trained model and test the prediction capabilities of the developed models.

In this study, three machine learning algorithms: Random Forest Regressor (RF), K-Nearest Neighbors Regressor (KNN) and an Artificial Neural Network Algorithm Regressor (ANN) are applied to flowing bottom hole pressure prediction (Breiman, 2001; Cunningham and Delany, 2007; Yang et al., 2020). A random forest (RF) method randomly selects rows/observations and specific variables to build multiple decision trees form and vote the best solution of the predicted output. K-Nearest-Neighbors regression (KNN) method is a non-parametric algorithm that, approximates the association between independent input parameters and the continuous dependent outcome by averaging the observations in the same neighborhood. It uses feature similarity to predict the output values of any new input data points. This means that the new data is assigned a value based on how closely it resembles the points in the training set. The size of the neighborhood (K) needs to be set in order to select the size that minimizes the mean-squared error.

In this study the optimum K value and the optimal weight function that gives minimum error were found using the grid search. K value is found to be six and the weight function of 'distance' is used in which the points weight by the inverse of their distance. The ANN model based on three structural layers, namely an input layer, hidden layers, and an output layer. Input layer consists of six input attributes which are well depth, flowing oil rate, flowing water rate (Qw), production tubing internal diameter (ID), oil gravity (API), and wellhead pressure (Pwh).

Preparing and training of the ANN model involve forwarding input data from the input layer to the hidden layers and from the hidden layers to the output layer to generate the required output. Each layer contains multiple arranged neurons (nodes) and connects with the subsequent layer by connections or edges between them termed as weights and biases. At the output layer, an error function was used to measure the difference between the predicted and actual FBHP values. The resulting error going back to the layers, then the weights and biases tune up again to improve the

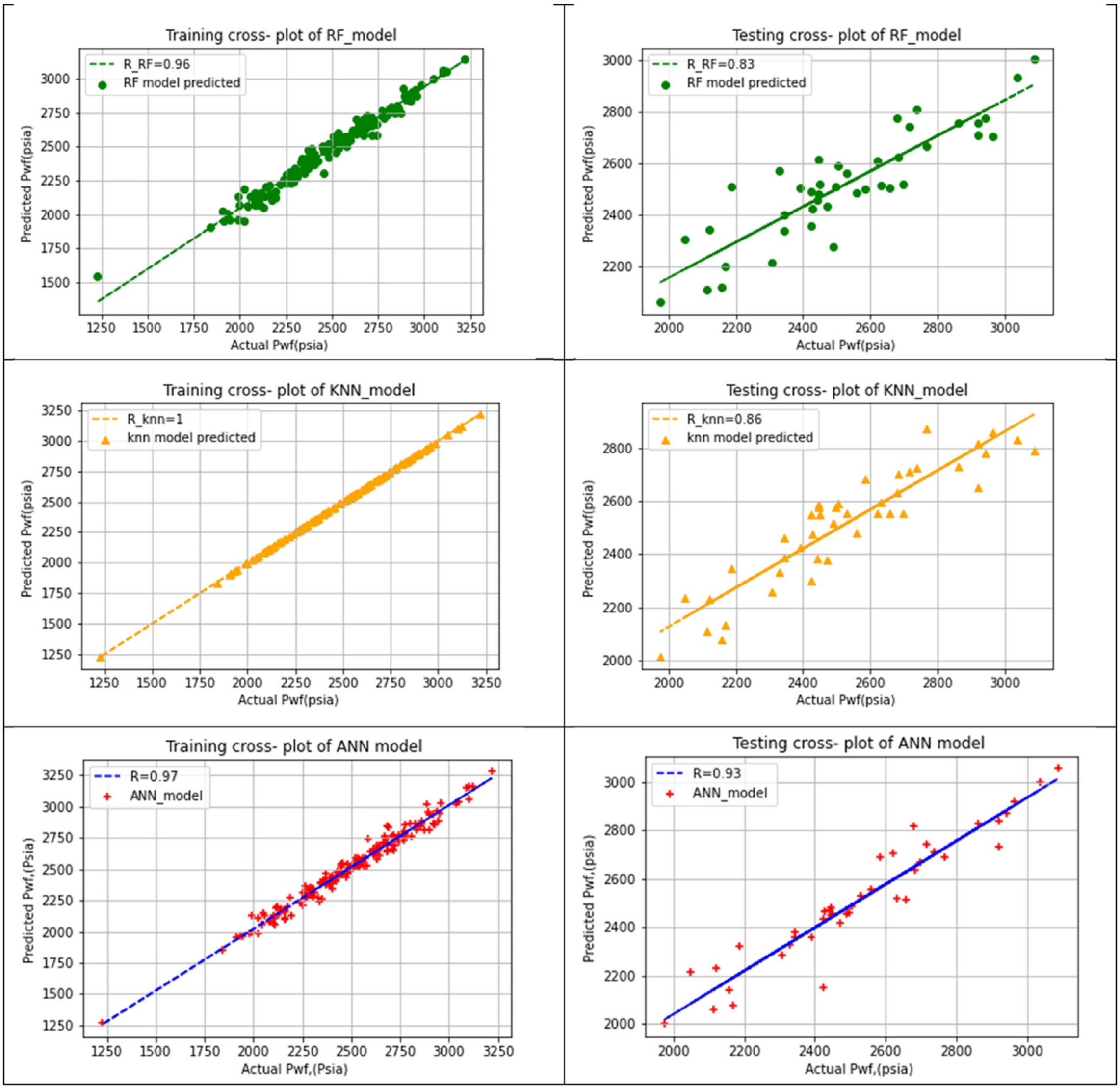


Fig. 3. Actual and predicted FBHP for the training and testing data using different machine learning model.

Table 3
The MSE and the R^2 values of the three models for both training and testing dataset.

Model	Training set		Testing set	
	R^2	MSE	R^2	MSE
Artificial Neural Network (ANN)	0.97	1.4%	0.93	2.5%
K-Nearest Neighbors Regression (KNN)	1.0	0%	0.86	3.6%
Random Forest Regression (RF)	0.96	1.7%	0.83	4%

predictive capability of the system. This iteration process is called as the epoch. With progressive training runs, ANN then self-corrects iteratively, until the model error was reduced to the minimum value. At the end of the training phase, the final model based on a representation of the transfer function between the inputs and the predicted output.

Sensitivity study was performed by varying the neuron number between 5 and 50 and the number of hidden layers between 1 and 5. The optimum number of neurons was found to be 50, and the maximum number of hidden layers found to be 5, since this combination ended up in highest R2 and lowest MAE in the training/testing set of the modelling. Sensitivity study for selecting optimum transfer function between input layer and the hidden layers was also executed between Sigmoid, Tanh, and Rectified linear units Relue. Relue function is selected in this study as the transfer function that result in less prediction error. A statistical calculation of the final selected data used in the training and testing the model is given in Table 2.

3. Results and discussion

Three machine learning algorithms: K-Nearest Neighbor, Random Forest regressor, and Artificial Neural Network apply to the selected input variables shown earlier. The histogram distribution plots for the datasets are given in Fig. 2.

After the data preprocessing steps, the selected machine learning models were built on the training dataset. The evaluation of each algorithm was essential to ensure the quality of the employed model. The models are fit to the training dataset and then used to predict the flowing bottom hole pressure for the training and testing dataset. A graphical description of the results is presented in Fig. 3, which show a comparison between the actual FBHP values and ML model predicted for both the training and testing data and also the figure displays the correlation of determination (R^2) for each model.

The comparison between all the ML methods which were used in this study is based on mean square error (MSE) and highest coefficient of determination. Table 3 shows the MSE and the R^2 values of the three models for both training and testing dataset.

This comparison clearly shows the power of the ML models in predicting the flowing bottom hole pressure. The developed RF and KNN models can highly predict the FBHP in the training dataset, but its show lower prediction accuracy of the testing dataset. On the other hand, ANN model can predict FBHP at a high accuracy for both the training and testing dataset.

4. Conclusions

Modelling and calculating of the flowing bottom hole pressure in multiphase flow of oil well associated with many assumptions that reduce the accuracy of the proposed correlations. Bottom hole pressure is the most important parameters to be calculated for both reservoir and production engineering. The oil industry relies on the use of empirical and/or mathematical correlations to calculate the FBHP. The emerging technologies of ML are one discipline of Artificial Intelligence, which provides the promising results, it is a breakthrough tool that helps solving human's complex problems. In this study machine learning techniques can accurately predict bottom hole pressure during the multiphase well production. Among the three developed machine-learning models, ANN stood the most optimum model. The results of this study show the potential of artificial intelligence in predicting the most complex parameter in oil and gas industry. This effort proves the role of data-driven computational models in the petroleum industry for the production scheme.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ptlrs.2021.05.004.

References

Ahmadi, M.A., Galedarzadeh, M., Shadizadeh, S.R., 2016. Low parameter model to monitor bottom hole pressure in vertical multiphase flow in oil production wells. Petroleum 2 (3), 258–266. https://doi.org/10.1016/j.petlm.2015.08.001.

Ahmadi, M.A., Soleimani, R., Lee, M., Kashiwao, T., Bahadori, A., 2015. Determination of oil well production performance using artificial neural network (ANN) linked to the particle swarm optimization (PSO) tool. Petroleum 1, 118–132. https://doi.org/10.1016/j.petlm.2015.06.004.

Ansari, A.M., Sylvester, N.D., Sarica, C., Shoham, O., Brill, J.P., 1994. A Comprehensive mechanistic model for upward two-phase flow in wellbores. SPE-20630-PA. https://doi.org/10.2118/20630-PA.

Ansari, A.M., Sylvester, N.D., Shoham, O., Brill, J.P., 1990. A comprehensive mechanistic model for upward two-phase flow in wellbores. SPE-20630-MS. https://doi.org/10.2118/20630-MS.

Asheim, H., 1986. MONA, an accurate two-phase well flow model based on phase slippage. SPE-12989-PA. https://doi.org/10.2118/12989-PA.

Ayoub, M.A., 2004. Development and Testing of an Artificial Neural Network Model for Predicting Bottomhole Pressure in Vertical Multiphase Flow. M.S. thesis. King Fahd University of Petroleum and Minerals.

Aziz, K., Govier, G.W., Fogarasi, M., 1972. Pressure drop in wells producing oil and gas. J. Can. Pet. Technol. 11, 38–48. https://doi.org/10.2118/72-03-04.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Chen, W., Di, Q., Ye, F., Zhang, J., Wenchang, W., 2017. Flowing bottomhole pressure prediction for gas wells based on support vector machine and random samples selection. Int. J. Hydrogen Energy 42 (29), 18333–18342. https://doi.org/10.1016/j.ijhydene.2017.04.134.

Cunningham, P., Delany, S.J., 2007. K-nearest Neighbour Classifiers: 2nd edition (with python examples). Technical Report. https://arxiv.org/abs/2004.04523.

Gene, M., Xingru, W., Kegang, L., 2019. An improved model for gas-liquid flow pattern prediction based on machine learning. J. Petrol. Sci. Eng. 183 https://doi.org/10.1016/j.petrol.2019.106370.

Gomez, L.E., Shoham, O., Schmidt, Z., Chokshi, R.N., Northug, T., 2000. Unified mechanistic model for steady-state two-phase flow: horizontal to vertical upward flow. SPE 65705 (PA. 5), 339–350. https://doi.org/10.2118/65705-PA.

Govier, G.W., Fogarasi, M., 1975. Pressure drop in wells producing gas and condensate. J. Can. Pet. Technol. 14, 28–41. https://doi.org/10.2118/75-04-03.

Jahanandish, I., Salimifard, B., Jalalifar, H., 2011. Predicting bottomhole pressure in vertical multiphase flowing wells using artificial neural networks. J. Petrol. Sci. Eng. 75, 336–342. https://doi.org/10.1016/j.petrol.2010.11.019.

Li, X., Miskimins, J.L., Hoffman, B.T., 2014. A combined bottom-hole pressure calculation procedure using multiphase correlations and artificial neural network models. SPE-170683-MS. https://doi.org/10.2118/170683-MS.

Medhat, A., Hassan, Y., 2016. Neural networks for flow bottom hole pressure prediction. Int. J. Energy a Clean Environ. (IJECE) 6 (4), 1839–1856. https://doi.org/10.11591/ijece.v6i4.

Munqith, A., Mingzhen, W., Baojun, B., Mortadha, A., 2017. Development of machine learning methodology for polymer gels screening for injection wells. J. Petrol. Sci. Eng. 151, 77–93. https://doi.org/10.1016/j.petrol.2016.12.038.

Orkiszewski, J., 1967. Predicting Two-phase Pressure Drops in Vertical Pipe. SPE. https://doi.org/10.2118/1546-PA.

Pucknell, J.K., Mason, J.N.E., Vervest, E.G., 1993. An evaluation of recent "mechanistic" models of multiphase flow for predicting pressure drops in oil and gas wells. SPE-1546-PA. https://doi.org/10.2118/26682-MS.

Spesivtsev, P., Sinkov, K., Sofronov, I., Zimina, A., Umnov, A., Yarullin, R., Vetrov, D., 2018. Predictive model for bottomhole pressure based on machine learning. J. Petrol. Sci. Eng. 166, 825–841. https://doi.org/10.1016/j.petrol.2018.03.046.

Tariq, Z., Mahmoud, M., Abdulraheem, A., 2020. Real-time prognosis of flowing bottom-hole pressure in a vertical well for a multiphase flow using computational intelligence techniques. J Pet Expl Prod Tech 10, 1411–1428. https://doi.org/10.1007/s13202-019-0728-4.

Yang, X., Dindoruk, B., Lu, L., 2020. A comparative analysis of bubble point pressure prediction using advanced machine learning algorithms and classical correlations. J. Petrol. Sci. Eng. 185, 106598. https://doi.org/10.1016/j.petrol.2019.106598.