Fuel 377 (2024) 132666

Contents lists available at ScienceDirect

Fuel

journal homepage: www.elsevier.com/locate/fuel

Full length article

Evolutionary automated radial basis function neural network for multiphase flowing bottom-hole pressure prediction

Deivid Campos [a], Dennis Delali Kwesi Wayo [b,c], Rodrigo Barbosa De Santis [d], Dmitriy A. Martyushev [e], Zaher Mundher Yaseen [f], Ugochukwu Ilozurike Duru [g], Camila M. Saporetti [h], Leonardo Goliatt [i,a,*]

[a] Computational Modeling Program, Engineering Faculty, Federal University of Juiz de Fora, Juiz de, Fora, 36036-900, Brazil
[b] Department of Petroleum Engineering, School of Mining and Geosciences, Nazarbayev University, Astana 010000, Kazakhstan
[c] Faculty of Chemical and Process Engineering Technology, Universiti Malaysia Pahang Al-Sultan Abdullah, Kuantan 26300, Malaysia
[d] Graduate Program in Industrial Engineering, Federal University of Minas Gerais, Belo, Horizonte, 31270-901, Brazil
[e] Department of Oil and Gas Technologies, Perm National Research Polytechnic University, Perm, 614990, Russia
[f] Civil and Environmental Engineering Department, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia
[g] Department of Petroleum Engineering, Federal University of Technology, Owerri, Imo State, PMB 1526, Nigeria
[h] Department of Computational Modeling, Polytechnic Institute, Rio de Janeiro State University, Nova, Friburgo, 22000-900, Brazil
[i] Department of Computational and Applied Mechanics, Federal University of Juiz de Fora, Juiz de, Fora, 36036-900, Brazil

ARTICLE INFO

Keywords:
Machine learning
Flowing bottom-hole pressure
Neural networks
Evolutionary optimization

ABSTRACT

Accurate multiphase flowing bottom-hole pressure prediction within wellbores is a critical requirement to improve tube design and production optimization. Existing models often struggle to achieve reliable accuracy across the full range of operational conditions encountered in oil and gas wells. This can lead to misallocating resources during well design, inefficient production strategies resulting in lost revenue, increased risk of wellbore damage, and poorly informed investment decisions. This research presents a data-driven hybrid approach that uses a Radial Basis Function Neural Network and a Particle Swarm Optimization algorithm to construct an automated hybrid machine learning model. The proposed model was compared with several well-established machine learning models in the literature using the same computational framework. The modeling results demonstrated the superiority of the hybrid approach. The model achieved superior performance with lower errors, as evidenced by a Relative Root Mean Squared Error (RRMSE) of 0.055. Furthermore, the model exhibited a low level of uncertainty throughout the analysis, indicating its high degree of reliability. These findings suggest the proposed data-driven approach offers a robust and practical solution for FBHP prediction in oil and gas wells.

1. Introduction

Multiphase flow in circular pipes is a common phenomenon across various engineering disciplines [1]. Within the petroleum industry, it is particularly relevant in wellbore environments such as drill-pipe/casing annuli, oil and gas production wells, and hydrocarbon transportation pipelines [2]. This type of flow involves the simultaneous movement of multiple components, forming mixtures like oil-water, gas-liquid, or even more complex combinations like oil-water-gas and gas-liquid–solid [3]. The presence of these distinct fluid phases and their varying properties makes multiphase flow significantly more intricate compared to single-phase (liquid or gas) flow. This complexity arises from the difficulty in establishing readily applicable flow standards, unlike single-phase scenarios [4]. Consequently, accurately estimating Flowing Bottom-Hole Pressure (FBHP) in multiphase flow presents a significant challenge.

Accurate prediction of flowing bottom-hole pressure (FBHP) is a critical aspect of optimizing oil and gas well operations, ensuring safe and efficient production, and maximizing profitability [5]. FBHP represents the pressure at the bottom of a wellbore, a crucial parameter that governs numerous aspects of well management [6]. Precise FBHP estimations are vital for well design, production optimization, and risk mitigation [7].

* Corresponding author at: Computational Modeling Program, Engineering Faculty, Federal University of Juiz de Fora, Juiz de, Fora, 36036-900, Brazil.
E-mail addresses: deivid.campos@engenharia.ufjf.br (D. Campos), dennis.wayo@nu.edu.kz (D.D.K. Wayo), rsantis@ufmg.br (R.B. De Santis), martyushevd@inbox.ru (D.A. Martyushev), z.yaseen@kfupm.edu.sa (Z.M. Yaseen), ugochukwu.duru@futo.edu.ng (U.I. Duru), camila.saporetti@iprj.uerj.br (C.M. Saporetti), leonardo.goliatt@ufjf.br (L. Goliatt).

https://doi.org/10.1016/j.fuel.2024.132666
Received 19 March 2024; Received in revised form 20 July 2024; Accepted 29 July 2024
Available online 14 August 2024

Table 1
Basic statistics for the training dataset (798 samples).

#	Production parameters	Variable	Unit	min	mean	std	max
1	Oil flow rate	OFR	stb/d	176.00	5422.62	3722.12	17 663.00
2	Gas flow rate	GFR	Mscf/d	9.00	2699.14	2370.40	17 859.00
3	Well perforation depth	WPD	feet	4243.00	6326.89	511.36	8620.00
4	Internal diameter of tubing	ID	inch	1.99	3.95	0.56	6.27
5	Water flow rate	WFR	stb/d	0.00	2215.13	2294.79	11 395.00
6	Oil API gravity	API	° API	25.40	33.85	3.10	47.50
7	Wellhead pressure	WHP	psia	92.00	423.81	253.74	1550.00
8	Well bottom-hole temperature	WBHT	° F	160.00	210.24	18.26	233.00
9	Flowing bottom-hole pressure	FBHP	psia	1198.00	2469.73	387.22	3698.00

Although there are efforts to replace fossil fuels, the use of these options, such as biofuel, is still small, and one of the reasons is due to low price competitiveness [8,9]. There is still a need to develop technologies in this sector so that there is the possibility of expanding their use. Therefore, the use of fossil fuels will be the majority for a long time, justifying the need to leverage studies of various factors that influence production. Enabling increased production while maintaining FBHP also attracts the attention of investors and can have an impact on shares and the market [10–12]. FBHP modeling can involve strongly nonlinear relationships between parameters measured directly in the well. The nonlinear relationships can be derived from the diagenetic processes and the distribution of heterogeneities [13]. These variations essentially introduce noise and randomness, ultimately leading to a decline in the model's ability to learn accurate patterns and predict effectively [14]. Consequently, some machine learning models may exhibit low precision depending on well operations, the sedimentary condition of the basin and the processes involved in the diagenetic evolution of the source rocks [15]. Different artificial intelligence techniques have been applied to solve complex real-world problems [16, 17]. In this way, ML models have received considerable attention in predicting FBHP in the oil and gas industries and have recently been used as data-driven alternatives to predict FBHP from data collected from wells [18–22].

Machine learning offers a promising alternative to traditional methods due to their capabilities to model complex non-linear relationships and adapt to changing wellbore conditions. Recent research has explored various ML models for FBHP prediction, including artificial neural networks [23–29], support vector machines [30], multiple linear regression [31], gradient boosting regression [7,32–35], tree-based models [36], and extreme learning machines [37].

Machine learning models may exhibit low precision depending on well operations, the sedimentary conditions of the basin, and the processes involved in the diagenetic evolution of the source rocks [15]. Combining ML models and optimization algorithms results in hybrid models with different characteristics to improve ML performance, allowing the creation of robust and accurate computational alternatives for FBHP modeling. Recent proposals combine Particle Swarm Optimization (PSO) with neural networks [38] and simulated annealing (SA) with support vector regression (SVR) [39] to estimate FBHP in a vertically oriented well with multiphase flow.

Although neural network models have been widely used to model FBHP, a specific type of network, the Radial Basis Function Neural Network (RBFNN), has not yet been used as an alternative for FBHP modeling. Existing models for FBHP prediction often rely on single neural network architectures or simpler methods such as linear regression. This research aims to address the limitations of existing ML models for FBHP prediction by introducing a data-driven hybrid approach that combines the strengths of a Radial Basis Function Neural Network (RBFNN) with a Particle Swarm Optimization (PSO) algorithm.

While traditional methods often utilize single neural networks or single ML techniques, this study explores the potential of Radial Basis Function Neural Networks (RBFNN) combined with Particle Swarm Optimization (PSO). Although neural network models have been widely used to model FBHP, a specific type of network, RBFNN networks have not yet been used as an alternative for FBHP modeling. RBFNN is generally trained faster than other neural network types, particularly multi-layered perceptron (MLP), due to the use of linear equations in the hidden-to-output layer weight calculations. Additionally, RBFNN demonstrates resilience to noise in data, a common challenge in real-world oil and gas production. Their radial basis functions effectively smooth noisy data, minimizing the influence of outliers on predictions. Furthermore, combining RBFNN with metaheuristics to perform optimized model selection can further enhance prediction accuracy. This approach results in a hybrid model that can capture the complex relationships between wellbore properties and FBHP, especially critical in multiphase flow scenarios where various fluid phases interact.

This research contributes to the field of FBHP prediction in oil and gas wells in the following aspects. Firstly, the research introduces a novel application of machine learning – the RBFNN-PSO model – as an automated approach for advanced FBHP analysis. This model addresses the limitations of existing automated methods by incorporating appropriate algorithms that leverage the available FBHP data. Secondly, the study complements the machine learning approach with a physics-driven mathematical model designed to replicate the multiphase flow behavior within both conventional and unconventional petroleum reservoirs.

This paper is organized as follows. Section 2 outlines the dataset and outlines the computational ensemble framework. In Section 3, the computational experiments are illustrated and discussed. Concluding remarks and the overall conclusion can be found in Section 4.

2. Methodology

2.1. Dataset

The input data consists of publicly available well data points [40, 41], including the features gas flow rate (GFR), oil flow rate (OFR), well perforation depth (WPD), internal diameter of tubing (ID), water flow rate (WFR), API oil gravity, wellhead pressure (WHP), well bottom-hole temperature (WBHT), and the target variable, FBHP [42]. Tables 1 and 2 display the fundamental statistics for the training and test datasets, respectively.

Fig. 1 shows the correlations among the features for the training and test datasets. It can be observed that GRF and OFR, WBHT, and WPD have a high correlation. Considering the FBHP as the target variable, the WPD, WBHT, and WFR characteristics have the highest correlations between 0.51 and 0.61.

2.2. Radial basis function neural network (RBFNN)

Consisting of three layers—the input layer, hidden, and the output layers— a Radial Basis Function Neural Network (RBFNN) is an Artificial Neural Network (ANN) type that utilizes radial basis functions as the activation function within its hidden layers [43]. A standard configuration of a multi-input, single-output RBFNN is illustrated in Fig. 2.

Table 2
Basic statistics for the training dataset (206 samples).

#	Production parameters	Variable	Unit	min	mean	std	max
1	Oil flow rate	OFR	stb/d	280.00	6321.51	4835.15	19 618.0
2	Gas flow rate	GFR	Mscf/d	33.60	3416.07	3068.43	13 562.2
3	Well perforation depth	WPD	feet	4550.00	6359.86	566.27	7100.0
4	Internal diameter of tubing	ID	inch	1.99	3.83	0.38	4.0
5	Water flow rate	WFR	stb/d	0.00	2700.00	2793.08	11 000.0
6	Oil API gravity	API	° API	30.00	33.77	2.31	37.0
7	Wellhead pressure	WHP	psia	80.00	321.07	153.56	960.0
8	Well bottom-hole temperature	WBHT	° F	157.00	203.64	16.95	215.0
9	Flowing bottom-hole pressure	FBHP	psia	1227.00	2489.03	302.16	3217.0

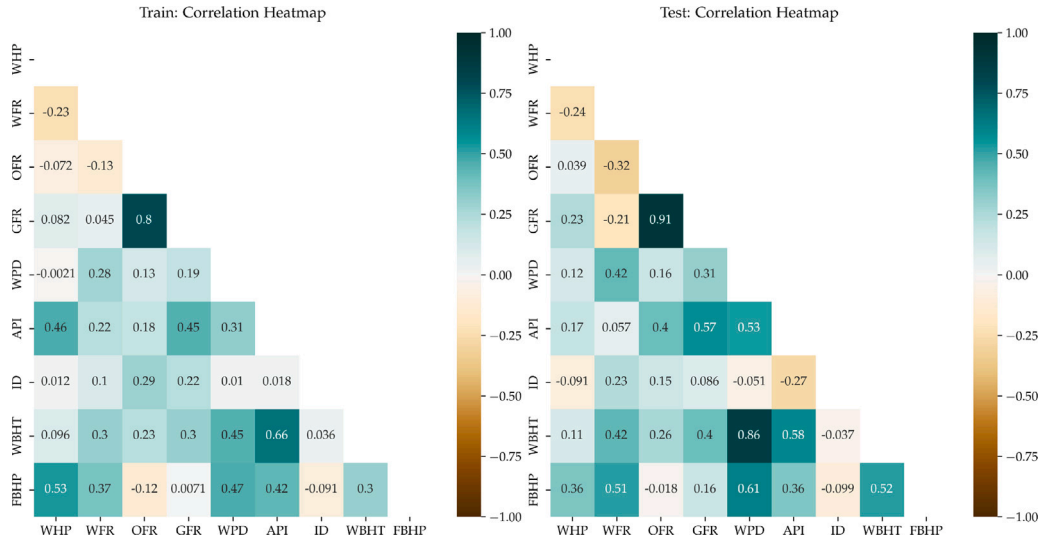

Fig. 1. Correlation coefficients for the training set (left) and test set (right). Each entry in the matrix represents the correlation coefficient between the variables, which appear on the rows and columns of the matrix.

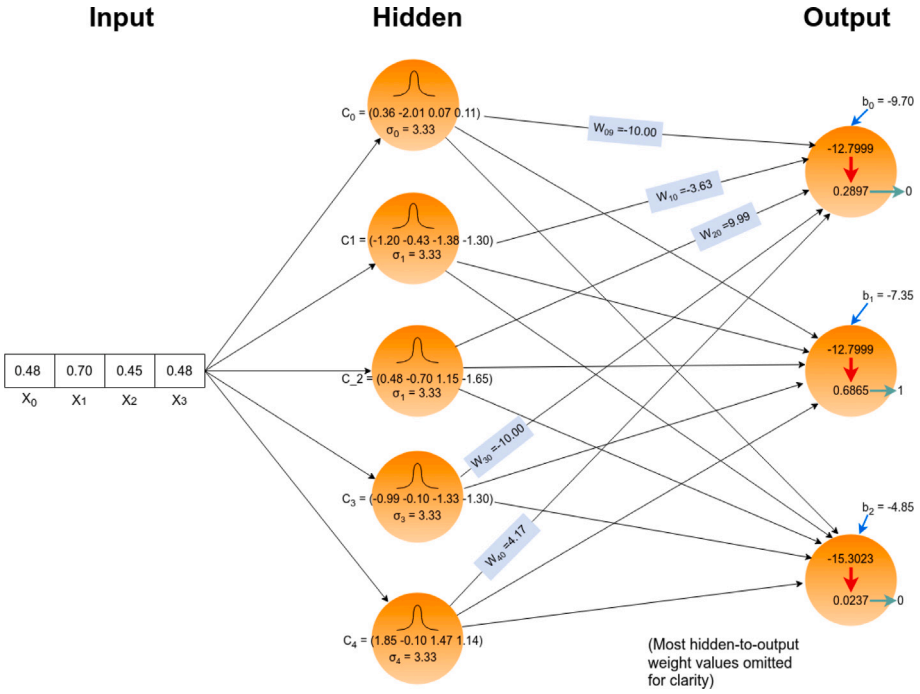

Fig. 2. A multi-input, multi-output RBFNN. This example showcases a Radial Basis Function Neural Network (RBFNN) with four inputs and three outputs. The hidden layer comprises five neurons, each equipped with activation functions. Parameters C_0 and σ_0 govern the radial basis functions, which in turn influence the network's output. The weights denoted by w_{ij} represent the strength of connections between the hidden and output layers. Finally, individual biases (b_i) are set for each output neuron, contributing to the overall network output.

The baseline RBFNN is a $k-m-1$ network, where each neuron in the RBFNN hidden layer is modeled as a Gaussian function. The mathematical expression for the output of the RBFNN is as follows:

$$\hat{y}(x) = \sum_{j=1}^{m} w_j(x)\theta_j(x) \tag{1}$$

where $w_j(x)$ represents the weight connecting the jth neuron in the hidden layer to the output layer, $\theta_j(t)$ corresponds to the output produced by the jth neuron in the hidden layer, and m is the number of hidden neurons. The output $\theta_j(x)$ is defined as:

$$\theta_j(x) = e^{-\frac{\|x-c_j\|^2}{2\sigma_j^2}} \tag{2}$$

where $x = [x_1, x_2, \ldots, x_n]^T$ represents the input vector, $c_j = [c_{1,j}, c_{2,j}(t), \ldots, c_{n,j}]^T$ is the center vector of the jth hidden neuron, n is the input vector's dimension, $\|x - c_j\|$ denotes the Euclidean distance between vectors x and c_j, and σ_j corresponds to the width of the jth hidden neuron.

2.3. Framework for machine learning model selection and optimization

The purpose of the PSO strategy is to adjust the predictor's internal parameters that leads to computing results consistent with the training data's actual results. PSO is a metaheuristic optimization approach based on the behavior of bird group [44]. This method has been adopted in studies due to its robustness, straightforward implementation, rapid convergence, and ability to address challenging optimization problems. The enhancement procedure of the PSO for the subsequent step is outlined as follows (Eqs. (3) and (4))

$$A_j^{i+1} = nA_j^i + e_1 v_1(p_{best}{}_j^i - X_j^i) + e_2 v_2(g_{best}^i - X_j^i) \tag{3}$$

$$X_j^{i+1} = X^K + A_j^{i+1} \tag{4}$$

where i is the actual step, n presents inactivity mass, $j = i, \ldots, R$, R denotes the number of particles in the group, v_1 and v_2 are randomly variables ([0,1]), e_1 and e_2 are excitation factors. X_j and A_j are position and acceleration of particle j, respectively. The individuals p_{best} is the local best point of particle j and g_{best} is the global best point of each particle, and K is the current iteration number.

Fig. 3 presents the framework of the proposed approach and the integration of PSO to build an automated RBFNN model. The parameters of the RBFNN method that were optimized using PSO, along with their corresponding search intervals, are described below. The RBFNN relies on three key parameters to achieve its approximation capabilities. These parameters are the smoothing parameter (w_1), the adjustable constant for activation functions (w_2), and the activation function (w_3). The smoothing parameter (w_1) that governs the smoothness of the function approximation by the RBFNN. A higher value of w_1 leads to a smoother but potentially less accurate approximation. The adjustable constant for activation functions (w_2) controls the behavior of the chosen activation function within the RBF node. The activation function (w_3) determines the output of each node based on its individual inputs and the radius of its influence.

The performance of an RBFNN is heavily dependent on the optimal selection of these three parameters. The Particle Swarm Optimization (PSO) algorithm searches for the most suitable combination of w_1, w_2, and w_3. This optimization process aims to identify the parameter set that yields the best performance on the training data. The PSO algorithm operates within a defined search space for each parameter. The parameters w_1 and w_2 are treated as continuous values and their search space ranges from 1 to 50 while w_3 is considered discrete taking values from the following functions: Multiquadric, Inverse, Gaussian, Linear, Cubic, Quintic, Thin plate, Sigmoid, ReLU, and Swish.

Table 3
Expressions for performance metrics.

Name	Expression
R	$\frac{\sum_{i=1}^{n}(y_i-\bar{y})(\hat{y}_i-\bar{\hat{y}}_i)}{\sqrt{\sum_{i=1}^{N}(y_i-\bar{y})^2}\sqrt{\sum_{i=1}^{N}(\hat{y}_i-\bar{\hat{y}})^2}}$
R^2	$\sum_{i=1}^{N}(y_i-\hat{y}_i)^2/\sum_{i=1}^{N}(y_i-\bar{y})$
MAE	$\frac{1}{N}\sum_{i=1}^{N}\lvert y_i-\hat{y}\rvert$
RRMSE	$\frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i-\hat{y})^2}}{\bar{y}}$
MAPE	$\frac{100}{N}\sum_{i=1}^{N}\left\lvert\frac{y_i-\hat{y}}{y_i}\right\rvert$

2.4. Performance metrics

The model's performance was assessed using the following metrics: R-squared (R^2), Pearson's Coefficient (R), Relative Root Mean Squared Error (RRMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). Detailed descriptions of these metrics can be found in Table 3. These metrics were selected for their ability to capture the diverse behaviors exhibited by the analyzed models comprehensively.

2.5. Error estimation, uncertainty analysis and feature importance

An error analysis was carried out to evaluate the models' capability to predict FBHP results. The error e_j for the sample j is calculated as the disparity between the measured FBHP value and the FBHP predicted by the machine learning models. The 95% confidence interval around the predicted FBHP values is determined by: $(\bar{e} - 1.96S_e,\ \bar{e} + 1.96S_e)$, where $\bar{e} = \sum_{j=1}^{N} e_j$ is the mean of the estimation and $S_e = \sqrt{\sum_{k=1}^{N}(\bar{e} - e_k)^2/(N-1)}$ represents the standard deviation.

The uncertainty analysis is a technique aimed at quantifying the extent to which output fluctuates in response to variations in input. Statistical measures like the median, mean, and population quantiles are typically employed to perform this analysis [45]. Estimating uncertainty often involves employing propagation methods, and the corresponding confidence intervals are derived based on these estimated values, given the relatively limited sample size.

In the experiments conducted within this study, the variation of input parameters was modeled using uniform distributions. The Mean Absolute Deviation (MAD)

$$\text{MAD} = \frac{1}{N_{MC}}\sum_{i=1}^{N_{MC}} |\text{FBHP}_i - median(\text{FBHP})| \tag{5}$$

is utilized as an assessment of the uncertainty associated with an output, which is computed as

$$\text{Uncertainty } \% = \frac{\text{MAD}}{median(\text{FBHP})} \times 100 \tag{6}$$

where $N_{MC} = 120000$ and FBHP_i is the flow bottom-hole pressure estimated for the ith sample.

In machine learning, feature importance techniques unveil the relative significance of individual features within a dataset for a specific prediction task. These techniques assign scores to each feature, indicating its contribution to the model's ability to make accurate predictions. Features with higher scores are deemed more influential in driving the model's output. Among several techniques, permutation feature importance is a robust and simple alternative. Permutation feature importance quantifies the change in the prediction error of a model following the permutation of a feature's values. This perturbation disrupts the connection between the feature and the actual outcome [46].

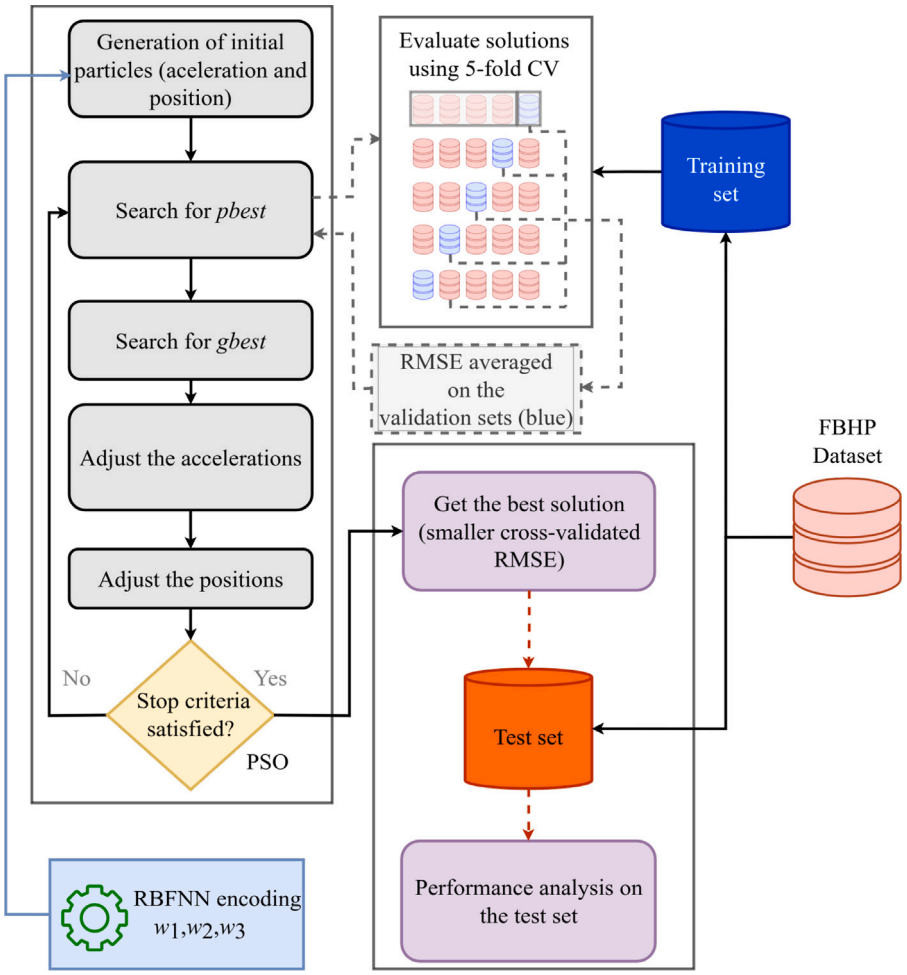

Fig. 3. Schematic workflow of the evolutionary automated RBFNN model. The bioinspired PSO algorithm initializes a population of candidate parameters (w_1, w_2, w_3) for the RBFNN model. This population iterates within an evolutionary cycle, where particle positions are adjusted based on performance evaluations. Cross-validation is employed to objectively compare the performance of each candidate solution. This iterative process continues until a predetermined stopping criterion is satisfied. After the evolutionary search, the final population is analyzed to identify the optimal candidate solution. This solution is then decoded into the corresponding RBFNN model architecture. Finally, the constructed model is evaluated on the test set, and performance metrics are calculated to quantify its effectiveness in predicting flowing bottom-hole pressure (FBHP).

3. Computational experiments, results and discussion

3.1. Comparative performance analysis

To comprehensively evaluate the performance of the proposed model against established machine learning techniques commonly employed in the literature, a series of experiments were conducted within a unified computational framework. The following well-recognized machine learning models were implemented and compared: Extreme Learning Machines (ELM), Elastic Net (EN), Kernel Ridge Regression (KRR), and Support Vector Regression (SVR)

Table 4 depicts the mean and standard deviation of the metrics discussed in Section 2.4. The outcomes were obtained from executing ELM, EN, KRR, and SVR models optimized with PSO in 100 separate iterations. Two recent research using the same dataset are also included. Notably, except for the MAE metric, the RBFNN exhibited superior performance compared to the other methods examined in this study. Furthermore, the RBFNN-PSO model demonstrated consistent results across the independent runs, as evident from the highest averaged values compared to the averages produced by the other models. The superior average metrics highlight the effectiveness of evolutionary model selection in enhancing the modeling process.

Fig. 4 shows a graphical comparison of the averaged models' performance. Each axis in the radar plot represents a different performance metric. The higher the value on a particular axis for a model, the better its performance on the corresponding metric. The figure shows that RBFNN outperforms the other models in all metrics, as observed in Table 4.

Table 5 displays the p-values resulting from the ANOVA test, which was used to compare all ML models. Notably, all values are below 0.05, indicating that, with 95% confidence, the variations between the method outcomes are statistically significant. This statistically significant result indicates that the methods produced demonstrably different outcomes for all metrics. This means that for all metrics, the difference between the results of the methods is statistically significant, and the value relatively smaller than 0.05 indicates that this difference was clearly detected by the methods and the null hypothesis (there is no difference in means) is strongly rejected.

3.2. Parameter distribution

Fig. 5 illustrates the parameter distribution of the RBFNN model. The RBFNN smoothing parameter (encoded as w_1) is well-distributed between 0 and 10, with a median value close to 3. Similarly, parameter ϵ (encoded as w_2) exhibits a well-distributed range between 0 and 500. Regarding the RBFNN activation functions (encoded as w_3), the Multiquadric function was chosen in 69 runs, followed by the Swish function, which was selected 29 times out of 100 runs.

Fig. 6 displays scatterplots showing the alignment between the measured and estimated flowing bottom-hole pressure (FBHP) for the

Table 4
The mean and standard deviation values for R, R^2, RRMSE, MAE, and MAPE are presented. Entries indicated with – are not available.

Estimator	R	R^2	RRMSE	MAE (psia)	MAPE (%)
ELM-PSO	0.869 (0.008)	0.732 (0.019)	0.063 (0.002)	110.118 (3.849)	4.587 (0.160)
EN-PSO	0.780 (0.001)	0.464 (0.002)	0.089 (0.000)	174.318 (0.289)	7.263 (0.012)
KRR-PSO	0.873 (0.013)	0.759 (0.029)	0.059 (0.003)	106.370 (5.592)	4.503 (0.216)
RBFNN-PSO	0.894 (0.004)	0.796 (0.011)	0.055 (0.001)	95.213 (2.918)	3.990 (0.113)
SVR-PSO	0.851 (0.003)	0.688 (0.099)	0.067 (0.008)	111.341 (19.074)	4.620 (0.798)
Ref. [28]	–	–	–	67.82	6.47
Ref. [42]	–	–	–	75.19	7.34

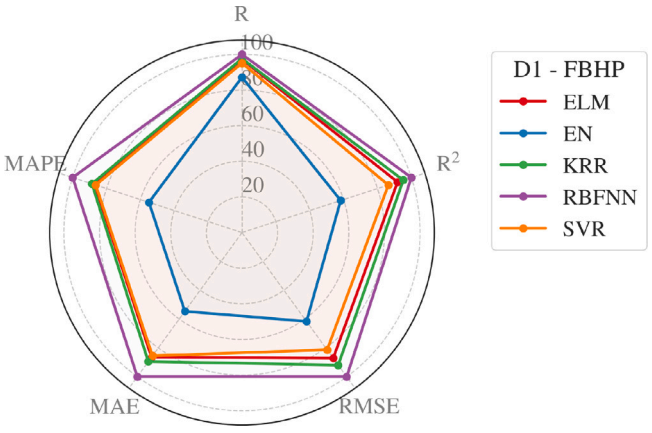

Fig. 4. Radar plot comparing the average performance metrics. Each axis represents a single metric used for comparison. The number of axes corresponds to the number of metrics. Greater distances from the center on an axis mean improved performance for the associated metric. Polygons with a larger area (circumscribed polygons) indicate superior overall performance.

Table 5
Obtained *p*-values in ANOVA test comparing all ML models.

Output	MAE	MAPE	R	R^2	RRMSE
FBHP	$\leq 10^{-7}$	$\leq 10^{-7}$	$\leq 10^{-7}$	$\leq 10^{-7}$	$\leq 10^{-7}$

top-performing models. The solid line represents the optimal fit, while the dots indicate the predictions made by the machine learning models based on their respective input variables. The estimations generated by RBFNN, KRR, and ELM methods exhibited substantial concordance with the samples, as indicated by R values surpassing 0.88.

3.3. Error and uncertainty analysis

The outcomes of the error analysis are presented in Table 6, encompassing mean prediction errors (MPE), width of uncertainty band (WUB), and 95% confidence interval estimation lower and upper bands (PEI95). The results indicate that most models, except for KRR, exhibited negative averages in their prediction errors. This suggests that the prediction models tended to underestimate the observed values. The computed uncertainty bands for both models spanned from ±133.792 to ±208.181, with the RBFNN model having a narrower band than the other models. Similarly, the RBFNN model featured a more compact 95% confidence prediction error interval, ranging from −252.181 to 272.282, in contrast to the other models. These findings underscore that the RBFNN variations demonstrated greater precision than the alternative approaches.

Table 7 presents the outcomes of the uncertainty analysis for FBHP using the proposed models. The analysis reveals that all models exhibited minimal uncertainties. Nevertheless, it is evident that the SVR and RBFNN models showcased lower uncertainty outcomes than the other models.

Fig. 7 shows the graphical comparison concerning RMSE and uncertainty. Coordinates (x, y) characterize a specific method (ELM, EN, KRR, RBFNN, and SVR), identified in different colors and markers. It can be observed that the RBFNN produced small uncertainty and RMSE compared with other models. In all cases explored, the RBFNN occasioned RMSE below 150 psia.

Table 6
Error analysis.

Model	MPE (psia)	WUB (psia)	PEI95 (psia)
ELM	−37.670	±144.254	−245.068 to 320.409
EN	−73.456	±208.181	−334.578 to 481.489
KRR	+2.910	±143.106	−283.397 to 277.578
RBFNN	−10.050	±133.792	−252.181 to 272.282
SVR	−26.625	±157.769	−282.601 to 335.852

MPE: mean prediction errors; WUB: width of uncertainty band; PEI95: 95% confidence interval estimation lower and upper bands.

Fig. 8 shows the importance coefficients associated with each variable. It can be observed that API, ID, and WBHT are less important in predicting FBHP in most methods. WFR had greater importance in predicting the EN, SVR, and RBFNN methods. The input features GFR, WHP, OFR, and WPD had similar degrees of importance for the RBFNN and SVR methods. For the ELM and KRR methods, the two features with the highest degrees of importance were GFR and OFR.

3.4. Discussion

An advantage is the population approach of the particle swarm algorithm since it allows the generation of multiple candidate solutions with similar performance levels [45]. Based on the objective function value, as determined by the performance metric, candidate solutions can be ranked (RMSE). The results presented in this study demonstrate the potential of the data-driven hybrid approach, however, it is crucial to acknowledge inherent limitations and consider practical applications in a systematic manner. The applicability of the proposed approach, as with other data-driven techniques, has a limitation that lies in the model's potential for generalization and extrapolation to unseen data. While the model performs well on the training dataset, it may not accurately predict flowing bottom-hole pressure (FBHP) in scenarios substantially different from those encountered during training. As shown in Fig. 6, there are a few data points that the model does not fit properly. This suggests that the model may overfit the training data and struggle to generalize to new, previously unseen conditions.

In the context of the workflow presented in Fig. 3, ensuring the appropriate acquisition and quality of data is essential for the successful implementation of the proposed RBFNN-PSO model. High-quality data is the foundation for these models, directly impacting their performance and prediction accuracy. The integration of the RBFNN-PSO model into existing workflows offers a range of benefits to engineers and specialists in the oil and gas industry, including optimized production processes, enhanced safety measures, and more informed decisions based on data-driven insights.

The performance of any machine learning model heavily relies on the quality and quantity of the training data. The model's accuracy may be limited if the training data does not encompass the full range of possible wellbore scenarios, such as extreme flow rates, variations in wellbore geometry, or different reservoir properties. Integrating

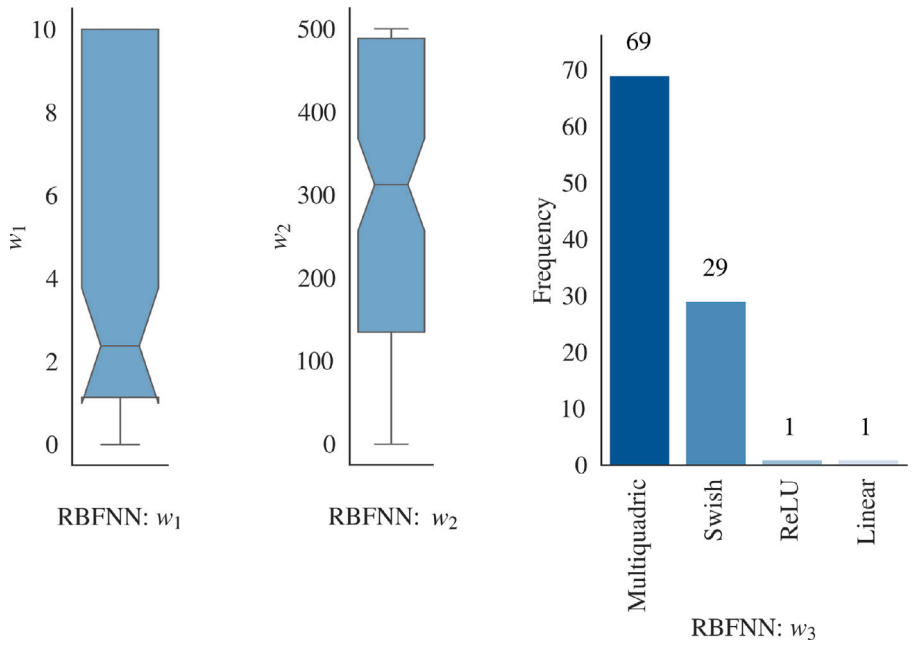

Fig. 5. RBFNN parameters distribution over 100 independent runs. The boxplot on the right shows the distribution of the smoothing parameter of the RBFNN network (w_1), and the boxplot in the center shows the distribution of the parameter ϵ (encoded as w_2) that controls the form of the radial basis functions and the figure on the left shows the barplot for activation functions (w_3) chosen for the final models in each independent run.

Table 7
Uncertainty analysis.

Model	Median (psia)	MAD (psia)	Uncertainty (%)	RMSE (psia)
ELM	2594.827 (93.633)	403.250 (25.456)	15.571 (1.313)	156.017 (5.272)
EN	2738.690 (5.871)	384.719 (1.340)	14.048 (0.036)	220.632 (0.369)
KRR	2367.839 (151.647)	476.249 (98.759)	20.173 (4.551)	147.893 (8.175)
RBFNN	2651.042 (34.449)	284.170 (11.166)	10.727 (0.633)	135.946 (3.260)
SVR	2487.214 (29.325)	261.040 (37.906)	10.497 (1.540)	167.185 (19.386)

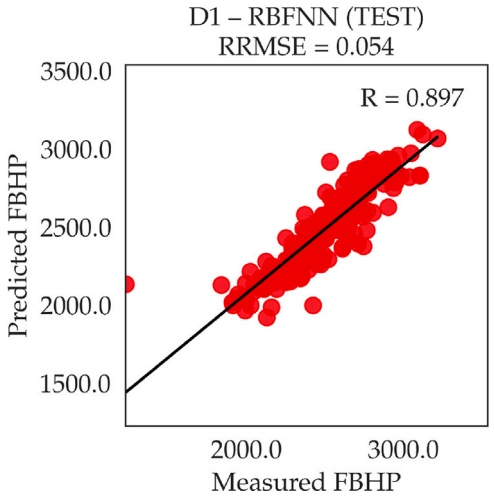

Fig. 6. Scatterplot for the best RBFNN model. The black line represents the best prediction, while the red dots indicate the predictions for the test set.

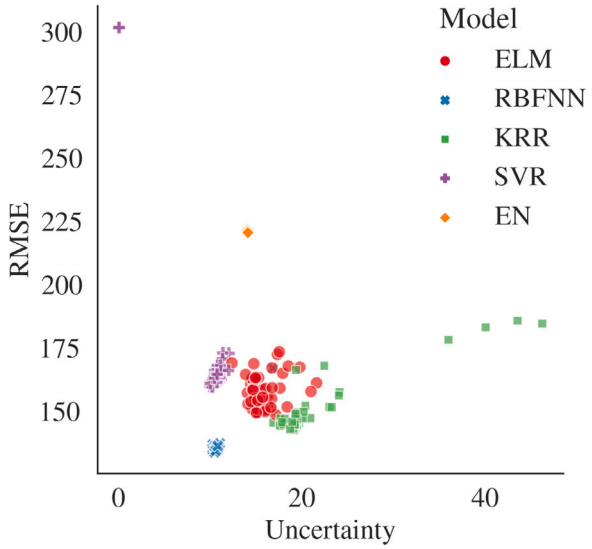

Fig. 7. Visual comparison depicting RMSE (psia) and Uncertainty (%) for all models. The data originates from Table 7. The level of uncertainty is displayed on the *x*-axis, while the RMSE values obtained by the model are presented on the *y*-axis.

machine learning with the metaheuristic search for internal parameters offers an alternative data-driven approach to modeling FBHP. As a result, an alternative model is built to adjust the parameters to compensate for the lack of input data, either due to sensor failure or insufficient data [47].

Although the feature importance analysis helped explore the relationships between model inputs and enhance interpretability, the proposed model primarily relies on data-driven learning without explicitly incorporating the fundamental physical principles governing multiphase flow. This limits its ability to accurately predict FBHP under conditions where the interplay of physical factors is dominant.

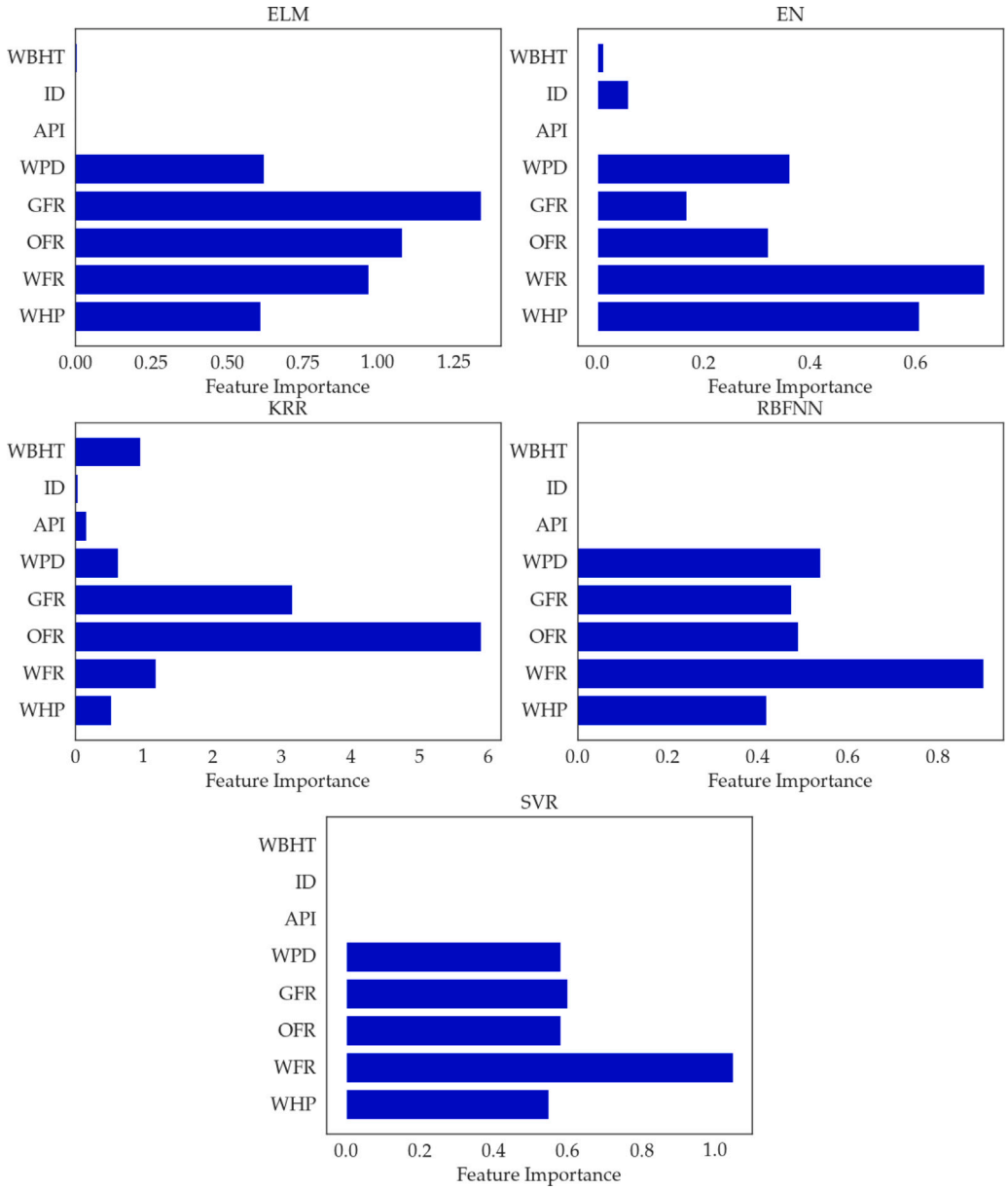

Fig. 8. Feature importance calculated by permutation (averaged over 100 independent runs). The x-axis represents the feature importance calculated by the random permutation method, and the input variables are shown in the y-axis.

The RBFNN model can be adapted for use in parallel computing environments and federated learning frameworks [48], which can be particularly valuable for oil and gas applications with distributed data. This adaptation would facilitate the aggregation of data from multiple wells or platforms without compromising data privacy, leading to more robust and generalizable models. Moreover, parallel computing can accelerate training and prediction times, particularly when dealing with large datasets [49]. The model employed in this study utilizes a PSO algorithm as its search mechanism, however, the hybrid model possesses flexibility and can incorporate various population-based algorithms [50] to achieve the optimization objective.

The successful implementation of accurate FBHP prediction tools, such as the proposed RBFNN-PSO model, can lead to potential economic and environmental benefits for the oil and gas industry [51]. Improved well design and optimized production strategies can help decision-makers and engineers reduce drilling and completion costs, increase hydrocarbon recovery, and minimize operational downtime. Furthermore, accurate FBHP predictions can enhance risk mitigation by preventing wellbore damage, blowouts, and other production-related failures, ultimately leading to cost savings and reduced environmental impact.

4. Conclusion

A novel hybrid approach is presented in this work, integrating machine learning models with optimization algorithms to achieve superior average predictive performance and facilitate automated model parameter selection. The key findings show that the hybrid RBFNN algorithm exhibited superior performance with minimal errors. The model demonstrated a low degree of uncertainty in the conducted analyses, and the RBFNN model achieved the most favorable averaged performance metrics ($R = 0.89$, $R^2 = 0.79$, RRMSE = 0.055, MAE = 95.21 psia, and MAPE = 3.99%). The proposed hybrid model demonstrates remarkable flexibility in constructing an adaptive model across varying conditions. Given that the metaheuristic algorithm generates incremental solutions, a process initiated in a specific pipeline of a well

or an exploration field can be continued in another pipeline, leveraging the prior solution as a starting point. As a result, the insights gained in one pipeline can be transferred to another. The problem formulation outlined in this paper and its management through the particle swarm optimization algorithm ensures the generation of robust and dependable incremental solutions, culminating in dependable models for flowing bottom-hole pressure (FBHP) modeling.

List of abbreviations and acronyms

ANN	Artificial Neural Network
FBHP	Flowing bottom-hole pressure
ELM	Extreme Learning Machine
EN	Elastic Net
KRR	Kernel Ridge Regression
MAD	Mean Absolute Deviation
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
ML	Machine Learning
MLP	Multi-layered perceptron
MPE	Mean Prediction Error
PEI95	95% confidence interval estimation lower and upper bands
PSO	Particle Swarm Optimization
R	Pearson Correlation Coefficient
R^2	Coefficient of Determination
RBFNN	Radial Basis Function Neural Network
RMSE	Root Mean Square Error
RRMSE	Relative Root Mean Square Error
SA	Simulated Annealing
SVR	Support Vector Regressor
WUB	Width of uncertainty band

CRediT authorship contribution statement

Deivid Campos: Writing – original draft, Software, Visualization, Data curation. **Dennis Delali Kwesi Wayo:** Writing – original draft, Visualization, Data curation,Validation. **Rodrigo Barbosa De Santis:** Writing – original draft, Software, Methodology, Data curation. **Dmitriy A. Martyushev:** Writing – review & editing, Investigation, Validation. **Zaher Mundher Yaseen:** Writing – review & editing, Investigation, Supervision, Validation. **Ugochukwu Ilozurike Duru:** Writing – review & editing, Writing – original draft, Investigation. **Camila M. Saporetti:** Software, Methodology, Data curation, Investigation. **Leonardo Goliatt:** Writing – review & editing, Supervision, Project administration, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work was supported by the Federal University of Juiz de Fora (UFJF), Brazil.

References

[1] Lin Z, Liu X, Lao L, Liu H. Prediction of two-phase flow patterns in upward inclined pipes via deep learning. Energy 2020;210:118541.
[2] Ponomareva IN, Galkin VI, Martyushev DA. Operational method for determining bottom hole pressure in mechanized oil producing wells, based on the application of multivariate regression analysis. Petrol Res 2021;6:351–60.
[3] Farokhipour A, Mansoori Z, Saffar-Avval M, Ahmadi G. 3D computational modeling of sand erosion in gas-liquid-particle multiphase annular flows in bends. Wear 2020;450–451:203241.
[4] Beggs D, Brill J. A study of two-phase flow in inclined pipes. J Pet Technol 1973;25:607–17.
[5] Vochozka M, Rowland Z, Suler P, Marousek J. The influence of the international price of oil on the value of the eur/usd exchange rate. J Compet 2020.
[6] Ahmadi MA, Chen Z. Machine learning models to predict bottom hole pressure in multi-phase flow in vertical oil production wells. Can J Chem Eng 2019;97:2928–40.
[7] Duru UI, Wayo DDK, Ogu R, Cyril C, Nnani H. Computational analysis for optimum multiphase flowing bottom-hole pressure prediction. Transylv Rev 2022;30.
[8] Maroušek J, Gavurová B, Struneckỳ O, Maroušková A, Sekar M, Marek V. Techno-economic identification of production factors threatening the competitiveness of algae biodiesel. Fuel 2023;344:128056.
[9] Bencoova B, Grosos R, Gomory M, Bacova K, Michalkova S. Use of biogas plants on a national and international scale. Acta Montan Slovaca 2021;26.
[10] Anthony E, Grauwde G, Domangue M. Increased rates, reserves, and revenues from heavy oil reservoir using esp technology: A suriname case study. In: SPE europec featured at EAGE conference and exhibition? SPE; 2009, pp. SPE–122006.
[11] Akbari M, Loganathan N, Tavakolian H, Mardani A, Štreimikiene D. The dynamic effect of micro-structural shocks on private investment behavior. Acta Montan Slovaca 2021;26:1–17.
[12] Pavolova H, Bakalár T, Kyšel'a K, Klimek M, Hajduova Z, Zawada M. The analysis of investment into industries based on portfolio managers. Acta Montan Slovaca 2021;26.
[13] Saporetti CM, Goliatt L, Pereira E. Neural network boosted with differential evolution for lithology identification based on well logs information. Earth Sci Inform 2021;14:133–40.
[14] Zhao C, Jiang Y, Wang L. Data-driven diagenetic facies classification and well-logging identification based on machine learning methods: A case study on xujiahe tight sandstone in sichuan basin. J Pet Sci Eng 2022;217:110798.
[15] Saporetti CM, d. Fonseca LG, Pereira E, d. Oliveira LC. Machine learning approaches for petrographic classification of carbonate-siliciclastic rocks using well logs and textural information. J Appl Geophys 2018;155:217–25.
[16] Kliestik T, Nica E, Durana P, Popescu GH. Artificial intelligence-based predictive maintenance, time-sensitive networking, and big data-driven algorithmic decision-making in the economics of industrial internet of things. Oeconomia Copernicana 2023;14:1097–138.
[17] Valaskova K, Nagy M, Grecu G. Digital twin simulation modeling, artificial intelligence-based internet of manufacturing things systems, and virtual machine and cognitive computing algorithms in the industry 4.0-based slovak labor market. Oeconomia Copernicana 2024.
[18] Dong P, Chen Z, Liao X, Yu W. Application of deep learning on well-test interpretation for identifying pressure behavior and characterizing reservoirs. J Pet Sci Eng 2022;208:109264.
[19] Huang R, Wei C, Wang B, Yang J, Xu X, Wu S, et al. Well performance prediction based on long short-term memory (lstm) neural network. J Pet Sci Eng 2022;208:109686.
[20] Okoro EE, Obomanu T, Sanni SE, Olatunji DI, Igbinedion P. Application of artificial intelligence in predicting the dynamics of bottom hole pressure for under-balanced drilling: extra tree compared with feed forward neural network model. Petroleum 2021.
[21] Ruiz-Diaz C, Gomez-Camperos J, Hernandez-Cely M. Flow pattern identification of liquid-liquid (oil and water) in vertical pipelines using machine learning techniques. J Phys Conf Ser 2022;2163:012001.
[22] d. O. Werneck R, Prates R, Moura R, Gonçalves MM, Castro M, Soriano-Vargas A, et al. Data-driven deep-learning forecasting for oil production and pressure. J Pet Sci Eng 2022;210:109937.
[23] Firouzi M, Rathnayake S. Prediction of the flowing bottom-hole pressure using advanced data analytics. In: Unconventional resources technology conference. 2019, http://dx.doi.org/10.15530/ap-urtec-2019-198240.
[24] Akinsete O, Adesiji BA. Bottom-hole pressure estimation from wellhead data using artificial neural network. In: SPE Nigeria annual international conference and exhibition, Lagos, Nigeria. 2019, http://dx.doi.org/10.2118/198762-MS.
[25] Masini SR, Goswami S, Kumar A, Chennakrishnan B, Baghele A. Artificial intelligence assisted production forecasting and well surveillance. In: Offshore technology conference. 2020, http://dx.doi.org/10.4043/30332-MS.
[26] Al Shehri FH, Gryzlov A, Tayyar TA, Arsalan M. Utilizing machine learning methods to estimate flowing bottom-hole pressure in unconventional gas condensate tight sand fractured wells in Saudi Arabia. Soc of Petrol Eng 2020. http://dx.doi.org/10.2118/SPE-201939-MS.

[27] Baki S, Dursun S. Flowing bottomhole pressure prediction with machine learning algorithms for horizontal wells, volume 2022-October. Society of Petroleum Engineers (SPE); 2022, http://dx.doi.org/10.2118/210235-MS.
[28] Nwanwe CC, Duru UI, Anyadiegwu C, Ekejuba AI. An artificial neural network visible mathematical model for real-time prediction of multiphase flowing bottom-hole pressure in wellbores. Petrol Res 2022.
[29] Zhang C-K, Zhang R, Zhu Z-P, Song X-Z, Su Y-A, Li G-S, et al. Bottom hole pressure prediction based on hybrid neural networks and bayesian optimization. Pet Sci 2023.
[30] Ahmadi MA, Galedarzadeh M, Shadizadeh SR. Low parameter model to monitor bottom hole pressure in vertical multiphase flow in oil production wells. Petroleum 2016;2:258–66.
[31] Rathnayake S, Rajora A, Firouzi M. A machine learning-based predictive model for real-time monitoring of flowing bottom-hole pressure of gas wells. Fuel 2022;317.
[32] Zakharov LA, Martyushev DA, Ponomareva IN. Predicting dynamic formation pressure using artificial intelligence methods. J Min Inst 2022;253:23–32.
[33] Ponomareva IN, Martyushev DA, Kumar Govindarajan S. A new approach to predict the formation pressure using multiple regression analysis: Case study from sukharev oil field reservoir – Russia. J King Saud Univ, Eng Sci 2022.
[34] Ashena R, Moghadasi J. Bottom hole pressure estimation using evolved neural networks by real coded ant colony optimization and genetic algorithm. J Pet Sci Eng 2011;77:375–85.
[35] Irani R, Nasimi R. Application of artificial bee colony-based neural network in bottom hole pressure prediction in underbalanced drilling. J Pet Sci Eng 2011;78:6–12.
[36] Marfo SA, Asante-Okyere S, Ziggah YY. A new flowing bottom hole pressure prediction model using m5 prime decision tree approach. Model Earth Syst Environ 2022;8:2065–73.
[37] Khamehchi E, Bemani A. Prediction of pressure in different two-phase flow conditions: machine learning applications. Measurement 2021;173:108665.
[38] Tariq Z, Mahmoud M, Abdulraheem A. Real-time prognosis of flowing bottom-hole pressure in a vertical well for a multiphase flow using computational intelligence techniques. J Petrol Explor Prod Technol 2020;10:1411–28.
[39] Liang H, Liu G, Zou J, Bai J, Jiang Y. Research on calculation model of bottom of the well pressure based on machine learning. Future Gener Comput Syst 2021;124:80–90.
[40] Ayoub MA. Development and testing of an artificial neural network model for predicting bottomhole pressure in vertical multiphase flow [Ph.D. thesis], King Fahd University of Petroleum and Minerals; 2005.
[41] Al-Shammari A. Prediction of pressure drop for two-phase flow in vertical pipes using artificial intelligence. King Fahd Univ Petrol Miner 2011.
[42] Nwanwe CC, Duru UI. An adaptive neuro-fuzzy inference system white-box model for real-time multiphase flowing bottom-hole pressure prediction in wellbores. Petroleum 2023.
[43] Yang Y, Wang P, Gao X. A novel radial basis function neural network with high generalization performance for nonlinear process modelling. Processes 2022;10.
[44] Eberhart R, Kennedy J. A new optimizer using particle swarm theory. In: Proceedings of the sixth international symposium on micro machine and human science. IEEE; 1995, p. 39–43.
[45] Goliatt L, Yaseen ZM. Development of a hybrid computational intelligent model for daily global solar radiation prediction. Expert Syst Appl 2023;212:118295.
[46] Breiman L. Random forests. Mach Learn 2001;45:5–32.
[47] Saporetti CM, da Fonseca LG, Pereira E. A lithology identification approach based on machine learning with evolutionary parameter tuning. IEEE Geosci Remote Sens Lett 2019;16:1819–23.
[48] Yaseen ZM. The next generation of soil and water bodies heavy metals prediction and detection: New expert system based edge cloud server and federated learning technology. Environ Pollut 2022;313:120081.
[49] Goliatt L, Mohammad RS, Abba SI, Yaseen ZM. Development of hybrid computational data-intelligence model for flowing bottom-hole pressure of oil wells: New strategy for oil reservoir management and monitoring. Fuel 2023;350:128623.
[50] Goliatt L, Saporetti C, Oliveira L, Pereira E. Performance of evolutionary optimized machine learning for modeling total organic carbon in core samples of shale gas fields. Petroleum 2024;10:150–64.
[51] Vochozka M, Horak J, Krulickỳ T, Pardal P. Predicting future brent oil price on global markets. Acta Montan Slovaca 2020;25.