Unconventional Resources 10 (2026) 100331

Contents lists available at ScienceDirect

Unconventional Resources

journal homepage: www.keaipublishing.com/en/journals/unconventional-resources/

Heterogeneous stacking strategy for modeling flowing bottom-hole pressure of oil wells

Deivid Campos [a], Bruno da Silva Macêdo [b], Oscar Ikechukwu Ogali [c], Matteo Bodini [d], Dmitriy A. Martyushev [e], Farouk Abduh Kamil Al-Fahaidy [f,g], Camila Martins Saporetti [h], Leonardo Goliatt [i,*]

[a] Computational Modeling Program, Engineering Faculty, Federal University of Juiz de Fora, Juiz de Fora, 36036-900, Brazil
[b] Systems and Automation Engineering Graduate Program, Federal University of Lavras, Lavras, 37200-000, MG, Brazil
[c] Department of Petroleum and Gas Engineering, University of Port Harcourt, Port Harcourt, Nigeria
[d] Dipartimento di Economia, Management e Metodi Quantitativi, Università degli Studi di Milano, Via Conservatorio 7, 20122 Milano, Italy
[e] Department of Oil and Gas Technologies, Perm National Research Polytechnic University, Perm, 614990, Russia
[f] Electrical Engineering Department, Ibb University, Yemen
[g] Information Technology Department, Emirates International University, Yemen
[h] Department of Computational Modeling, Polytechnic Institute, Rio de Janeiro State University, Nova, Friburgo, 28625-570, Brazil
[i] Department of Computational and Applied Mechanics, Federal University of Juiz de Fora, Juiz de Fora, 36036-900, Brazil

ARTICLE INFO

Keywords:
Flowing bottom hole pressure
Machine learning
Stacking model
Artificial intelligence
Production optimization

ABSTRACT

Accurately predicting Flowing Bottom-Hole Pressure (FBHP) is critical for optimizing oil and gas production. Existing predictive methods often rely on oversimplified or complex, yet computationally expensive, models that fail to capture the intrinsic nonlinearities of well dynamics, leading to inaccurate predictions and potential economic losses. This paper introduces a three-layer heterogeneous stacking ensemble model to address the latter challenge. In particular, the key novelty of the developed work is a hierarchical architecture that integrates five distinct Machine Learning (ML) base learners, two meta-learners, and a final super-learner, *i.e.*, an additional meta-model that combines the outputs of the meta-learners to capture complex, non-linear relationships in the data. When evaluated on a field dataset (total dataset samples $N = 795$; test set samples $N = 199$), the proposed Super Learner Stacking model (ST-S) demonstrated superior predictive performance on the independent test set, achieving R-squared (R^2) $= 0.857 \pm 0.006$ and Root Mean Squared Error (RMSE) $= 146.382 \pm 2.806$. In addition, the ST-S model outperformed all individual models and simpler stacking ensembles reported in the article. As a result, the developed ST-S model provides a robust, data-driven tool for FBHP prediction, achieving high predictive accuracy without resorting to computationally expensive methods, thereby supporting improved well management and production optimization.

1. Introduction

Over the last few decades, researchers in petroleum engineering have invested considerable effort in accurately predicting FBHP in oil and gas wells. Several researchers have attempted to develop relationships to model pressure drops in pipes, and researchers have reported promising findings [1]. Nonetheless, the more unknown variables are involved, the less effective these correlations become. Parameters of multiphase flow problems include the gas–oil ratio in the two-phase system and the pipe inclination angle, among others. Generally, proposed methods use small-scale wells in lab experiments to establish correlations for oil and gas production rates [2].

Most existing techniques for FBHP prediction, particularly empirical and mechanistic methods, make assumptions that can lead to errors and compromise their predictive accuracy. As a result, there is a need to explore the use of Machine Learning (ML) algorithms as alternatives to FBHP prediction. Uncertainties in conventional assumptions highlight their limitations in ensuring reliable FBHP prediction, in addition to

Peer review under the responsibility of KeAi Communications Co., Ltd.
* Corresponding author.
E-mail addresses: deivid.campos@estudante.ufjf.br (D. Campos), bruno.macedo2@estudante.ufla.br (B. da Silva Macêdo), oscar.ogali@uniport.edu.ng (O.I. Ogali), matteo.bodini@unimi.it (M. Bodini), martyushevd@inbox.ru (D.A. Martyushev), farouqakh@ibbuniv.edu.ye (F.A.K. Al-Fahaidy), camila.saporetti@iprj.uerj.br (C.M. Saporetti), leonardo.goliatt@ufjf.br (L. Goliatt).

https://doi.org/10.1016/j.uncres.2026.100331
Received 20 July 2025; Received in revised form 21 January 2026; Accepted 3 February 2026
Available online 5 February 2026

the fact that ML techniques have been successful in making accurate predictions within the oil and gas industry [3,4].

In petroleum engineering, determining the Bottom-Hole Pressure (BHP) in oil and gas wells is a challenging task. Multiphase flows are common phenomena in applications such as drilling, well clean-out, and the transportation of hydrocarbons comprising more than one phase, including oils, gases, and solids [5]. The presence of multiple phases makes it challenging to predict FBHP because they lack distinct flow patterns [6]. Although direct pressure measurements using bottom-hole gauges are straightforward, these methods require frequent calibration and maintenance, including costly well interventions that can occasionally disrupt production.

Researchers have developed correlations to estimate pressure drops in pipes. Still, the effectiveness of these models diminishes as the number of unknown variables increases [1], limiting their ability to fully capture the complex relationships among the various parameters involved in multiphase flow problems, such as gas-to-oil ratio and pipe inclination angle. Addressing this challenge remains an important research priority in the petroleum industry. Laboratory tests in small-scale models are also essential for uncovering relationships between oil and gas output levels and for inferring well operation. These experiments aim to determine the relationships governing production rates at small scales, which can then be applied to larger-scale operations [2].

Established FBHP forecasting approaches, including empirical and mechanistic models, often rest on presumptions that can lead to inaccuracies and compromise prediction reliability. Machine learning (ML) methods show promise as an alternative for FBHP prediction. The proven success of ML techniques in delivering precise forecasts within the oil and gas industry, along with the substantial uncertainties inherent in conventional assumptions, indicates their potential to enhance FBHP prediction [3,4].

Over the last few years, some studies have developed ML models to determine multiphase FBHP in well bores [7,8]. Jin and Emami-Meybodi [9] developed an integrated Artificial Neural Network (ANN) algorithm that integrates and classifies multiphase flow physics to improve FBHP predictions under various operating conditions. The results showed that the method can be applied to a wide range of data and production conditions. Nnaemeka and M. Chinyereugo [10] investigated the use of ML algorithms to predict FBHP production from a sandstone reservoir. They used Support Vector Regressor (SVR), Multilayer Perceptron (MLP), and Random Forest Regression (RFR) algorithms. The results showed that SVR estimated FBHP more accurately than other methods for sandstone formations.

Jin and Emami-Meybodi [11] used a Long Short-Term Memory Neural Network (LSTM-ANN) approach to estimate FBHP in unconventional wells during gas operations. The results showed that the LSTM-ANN approach performed well, constituting an alternative for estimating FBHP. Abdullahi and Ezeh [12] explored the use of Gated Recurrent Units (GRU) and LSTM to predict FBHP in oil and gas wells based on a lean and optimal set of variables. The results showed that GRU outperformed LSTM, with an average 30% reduction in error metrics. Agwu et al. [13] conducted a study on the interpretable multivariate adaptive regression spline method for estimating FBHP. The results show that the method performs well, with improved predictive accuracy. The model achieved the best result among the others.

Hybrid methods that integrate optimization and ML techniques have been proposed to predict FBHP, thereby improving performance and optimizing hyperparameter selection. Campos et al. [14] proposed a data-driven hybrid approach that combines a Radial Basis Function Neural Network (RBFNN) with the Particle Swarm Optimization (PSO) algorithm to build an automated ML model for predicting FBHP. The results showed that the hybrid approach outperformed traditional algorithms in the literature. A hybrid model using machine learning (ML) methods and Differential Evolution (DE) optimization for FBHP prediction was proposed by [15]. The hybridization of multi-adaptive regression splines and differential evolution achieved the best performance for the FBHP estimation.

Though hybrid methods have improved performance in FBHP estimation, they can be surprisingly complicated, requiring shared hyperparameters across distinct approaches. This task can be computationally intensive. In this context, simpler techniques can yield improved solutions for FBHP estimation.

Studies use ML to evaluate their databases and investigate intrinsic patterns. Although computational intelligence has made progress in developing solutions for FBHP prediction, much research has focused on ensemble methods, such as bagging, stacking, and voting. Stacking methods have been widely employed to solve engineering problems [16–20]. However, such models for FHBP prediction are scarce in the literature.

Nevertheless, such ensemble methods for FBHP estimation are rarely seen in the literature. Given the efforts to build ML models for FBHP from well data, only a limited amount of research has evaluated ensemble techniques comprising many weak methods.

Understanding and forecasting diverse scenarios for experts, academics, and decision-makers has become increasingly challenging due to the growing complexity of today's issues and the exponential growth in data from multiple processes. Models that can capture input/output nonlinear relationships and be resilient to data noise and uncertainty are necessary for solving complex machine learning and data science challenges.

The development and application of heterogeneous stacking ensembles for FBHP prediction address a critical need for accurate, reliable, and robust predictions to support informed decision-making, optimize production operations, and enhance reservoir management in the oil and gas industry.

Current research in the FBHP domain predominantly relies on standalone machine learning models [21–23] or conventional single-layer stacking [24]. However, these approaches often yield suboptimal precision and struggle to address the inherent uncertainty of multiphase flow data. This study addresses these limitations by advancing from flat ensemble structures to a hierarchical three-layer Super Learner (ST-S) framework. To the best of our knowledge, this is the first work to investigate deep super-learner architectures with heterogeneous ensembles specifically for FBHP prediction. By introducing an additional aggregation layer to synthesize the outputs of intermediate meta-learners, the proposed architecture significantly enhances both accuracy and robustness compared to traditional stacking models.

FBHP in oil wells is a significant concern in reservoir management, production optimization, and well performance evaluation [25]. This research presents a stacking ensemble model designed to enhance the accuracy of FBHP predictions within oil and gas wells. The developed model outperforms individual ML models in predictive performance. A grid search algorithm optimized performance and adapted model parameters to the analyzed dataset. The novelty of this approach lies in the use of multiple stacked ensembles with a superlearner architecture to predict FBHP. The use of an additional ensemble level, which combines outputs from several meta-learners, improves predictions and increases accuracy compared to typical stacking models. The proposed approach integrates several base learners with diverse learning methods, resulting in a enhanced, resilient and versatile model that overcomes the limitations of individual models.

The proposed stacking ensemble approach provides a valuable tool for petroleum engineers to enhance FBHP prediction accuracy and mitigate uncertainty. This has significant implications for reservoir management [26,27], production optimization [28] and well performance evaluation [27].

The stacking ensemble strategy used provides petroleum professionals with a valuable tool to enhance the precision of FBHP forecasting and reduce variability. This has substantial consequences for reservoir management [26,27], production optimization [28], and well performance assessment [27].

The format of this document is as follows. The dataset is presented and the computational methodology structure is explained in Section 2. The computational experiments and discussion are presented in Section 3. Section 4 contains the concluding ponderations.

Table 1
Description of the input features, their units, and physical relevance to FBHP prediction.

Feature	Description	Unit	Physical Meaning/Importance
WHP	Wellhead Pressure	psig	Represents the pressure at the surface; strongly influences the overall pressure gradient along the wellbore and is one of the strongest predictors of FBHP.
WFR	Water Flow Rate	bpd	Indicates water production volume; affects mixture density, frictional losses, and multiphase flow behavior in the tubing.
ID	Tubing Internal Diameter	in	Directly controls the available flow area; smaller diameters lead to higher frictional pressure losses and increased FBHP.
GFR	Gas Flow Rate	scf/d	Determines gas content in the produced stream; strongly affects gas–liquid flow regime, compressibility, and pressure losses.
OFR	Oil Flow Rate	bpd	Reflects the primary hydrocarbon production rate; influences mixture properties and frictional pressure behavior.
API	API Oil Gravity	°API	Measures oil density; lighter oils (higher API) reduce hydrostatic pressure and typically decrease FBHP.
WPD	Water Production Depth	ft	Represents the producing depth or reservoir penetration point; deeper production zones generally result in higher hydrostatic and flowing bottom-hole pressure.
WBHT	Well Bottom-Hole Temperature	°F	Affects fluid viscosity and phase behavior; higher temperatures reduce viscosity and can decrease frictional losses.
FBHP	Flowing Bottom-Hole Pressure	psig	Target output representing the pressure at the bottom of the well during flowing conditions.

Table 2
Statistics from training and test sets.

Dataset	Variable	mean	std	min	25%	50%	75%	max
Training (596 samples)	WHP	414.517	246.038	92.0	240.0	312.50	540.0	1550.0
	WFR	2239.124	230 .921	0.0	371.0	1483.50	3471.50	11395.0
	ID	3.962	0.592	1.995	3.958	3.958	4.0	6.276
	GFR	2741.696	2448.878	21.0	849.50	2165.50	3809.750	17859.0
	OFR	5451.792	3743.512	176.0	2479.750	4654.0	7459.0	17663.0
	API	33.837	3.092	25.40	33.40	33.40	37.30	47.50
	WPD	6326.179	507.840	4243.0	6093.750	6440.50	6645.250	8620.0
	WBHT	210.285	18.562	160.0	209.0	217.0	217.0	233.0
	FBHP	2476.574	387.141	1198.0	2192.0	2423.50	2771.750	3604.0
Test (199 samples)	WHP	451.663	274.310	170.0	255.0	360.0	530.0	1410.0
	WFR	2143.281	2280.622	0.0	335.50	1544.0	3225.50	10171.0
	ID	3.931	0.492	2.441	3.958	3.958	4.0	6.276
	GFR	2571.724	2118.727	9.0	865.50	2101.0	3552.0	12016.0
	OFR	5335.271	3665.280	348.0	2527.0	4583.0	7317.0	17040.0
	API	33.923	3.167	26.20	33.40	33.40	37.30	47.50
	WPD	6329.029	523.086	4650.0	6074.50	6440.0	6645.0	8478.0
	WBHT	210.136	17.373	160.0	20.0	217.0	217.0	233.0
	FBHP	2449.236	387.732	1705.0	2168.50	2359.0	2734.50	3698.0

2. Material and methods

2.1. Study area and dataset

For this work, we employed 596 training samples and 199 testing samples, providing data on the productivity of wells from different regions of the Middle East. The attributes present in the data are: Wellhead Pressure (WHP), Water Flow Rate (WFR), Internal Diameter (ID), Gas Flow Rate (GFR), Oil Flow Rate (OFR), American Petroleum Institute oil gravity (API), Water Production Depth (WPD), and Well Bottom-Hole Temperature (WBHT); the target output is the Flow Bottom-Hole Pressure (FBHP) [29].

Table 1 summarizes each feature, including its physical meaning and measurement units, offering a clear reference for interpreting the input variables. Table 2 presents the descriptive statistics of the features and FBHP for both the training and test datasets. The table reveals variations in values across features.

Fig. 1 provides a visualization of the sample distribution in the training set. It can be observed that the GFR and OFR characteristics are correlated, as the sample distribution lies on a straight line. This indicates that when the value of one characteristic increases, the value of the other tends to increase, and the WBHT and API characteristics display similar behavior. The OFR and WHP characteristics are unrelated, and there is no noticeable difference in the distribution of samples across them.

Fig. 2 illustrates the correlation values that support the observations made in the distribution chart. In the training set, the highest correlation, 0.80, occurs between the GFR and OFR features, followed by the WBHT and API features, which correlate 0.68. For FBHP, the WHP, WPD, and API characteristics are most strongly correlated. For the test set, the correlations are similar to those in the training set for most characteristic pairs: between GFR and OFR, 0.82; between WBHT and API, 0.60. Regarding FBHP, the most strongly correlated characteristics are WHP, API, and WPD.

2.2. Data preprocessing

All preprocessing procedures followed a controlled workflow to preserve data integrity and avoid any form of information leakage during model development. The dataset was first examined for structural inconsistencies, missing values and duplicated records. Samples

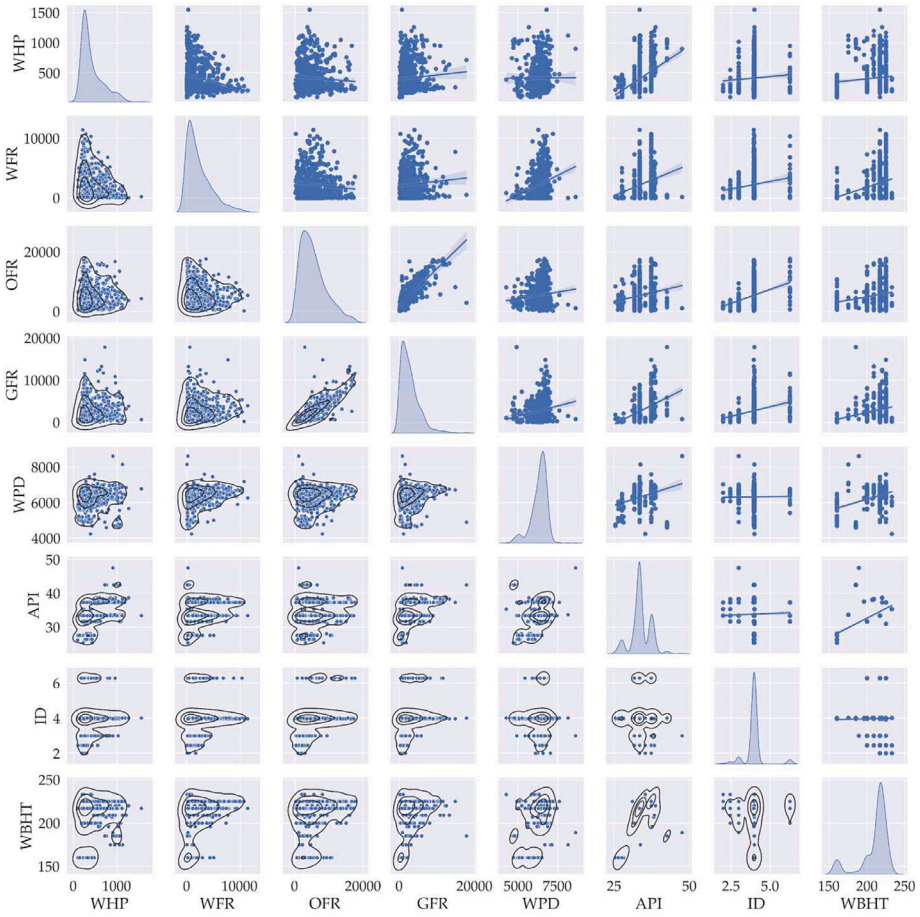

Fig. 1. Distribution of samples in the training set.

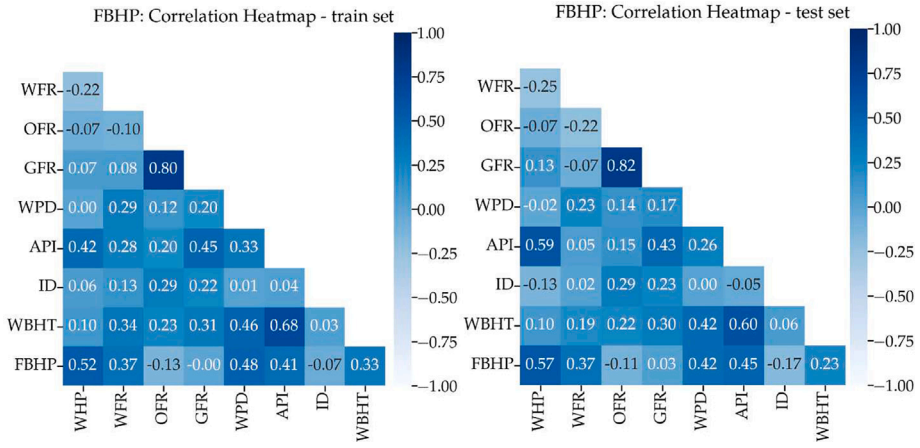

Fig. 2. Correlation of the variables in the training set (left) and test set (right).

with missing entries were removed to prevent bias, and no duplicated instances were detected after loading the data. Outlier removal was not applied because abrupt variations in pressure, temperature, or production rates may represent legitimate operational or reservoir behaviors and should remain in the analysis.

After this cleaning step, all numerical predictors were standardized using the z-score transformation implemented through the `StandardScaler` in scikit-learn. This transformation centers each feature by subtracting its mean and dividing by its standard deviation, which ensures that variables with different physical units contribute proportionally during the learning stage. To avoid data leakage, the scaler was fitted only on the training subset, and the resulting scaling parameters were applied unchanged to both the validation folds and the test set.

The normalization procedure was also embedded in the 5-fold cross-validation routine adopted in the stacking models. For every fold, scaling statistics were derived solely from the corresponding training partition. This guarantees that the validation portion of each fold remained completely unseen during the computation of preprocessing

parameters. The same workflow was applied during the final training stage, which maintained full consistency between the preprocessing and modeling steps.

2.3. Machine learning models

This section summarizes the ML models employed in this paper. We chose ML models because they consist of diverse approaches to FBHP prediction, each with strengths and limitations. For standalone approaches, the choice of the most suitable ML model depends on factors such as the complexity of the relationships between FBHP and input features, the availability of training data, and computational resources. The ML models presented in Table 3 serve as the basis for combining them using an ensemble or stacking approach, potentially achieving superior performance compared to individual models.

K-Nearest Neighbors (KNN) is a supervised learning method based on proximity. The training data consists of n-dimensional vectors, where each sample corresponds to a point in n-dimensional space. KNN uses the number of neighbors (K) as a hyperparameter to optimize predictive performance [30].

Linear Regression (LR) creates a model that represents the relationship between variables by fitting a linear function to the empirical data [31]. The model considers one or more independent variables and one dependent variable. The linear regression equation is $\mathbf{y} = \mathbf{aX} + \mathrm{b}$, where $\mathbf{y}$ is the dependent variable and $\mathbf{X}$ denotes the independent variable.

The Extreme Learning Machine (ELM) is a neural network model that has only one hidden layer and differs from other neural network architectures in its weight-adjustment strategy [32]. The weights in the hidden layer are initialized randomly, whereas those in the output layer are initialized analytically.

Random Forests (RF) is an ensemble technique that creates a collection of decision trees through k iterations, using random sampling from the [33] training set. Each iteration builds decision trees based on different random subsets of features and samples, and the ensemble of trees is used to improve overall prediction accuracy.

Support Vector Regression (SVR) [34] is often employed to map data into a higher-dimensional feature space. Considering a set $(x_1, z_1), \ldots, (x_l, z_l)$, where $x_i \in R^n$ is a characteristic vector and $z_i \in R^1$ is the target value.

2.4. Stacking models

Ensemble learning leverages the complementary strengths of multiple learning algorithms to mitigate individual models' weaknesses and improve predictive accuracy [35]. Among ensemble methods, stacking adopts a hierarchical architecture in which base learners, trained directly on the original input features, generate intermediate predictions. Such predictions are then passed to a higher-level meta-learner, which is trained to combine them effectively and generate the final output prediction [36].

There are three main strategies for integrating these models within a stacking framework: weighted averaging, stacking with regression models, and heterogeneous ensembles. Weighted averaging combines predictions from individual models by assigning weights based on their reliability. Weights are assigned based on each model's performance metrics. In the stacking approach, the first level of the stacking framework can incorporate any of the models presented in the table as base learners. The final meta-learner can then be another regression model. The stacking framework for heterogeneous ensembles can be designed to encompass a diverse set of learners. This may include KNN, LR, RF, and potentially other models. This approach yields a model that combines the strengths of various learning paradigms to achieve enhanced performance.

2.5. Stacking ensemble models framework

The proposed framework integrates multiple machine learning models with distinct training processes to enhance forecast accuracy, reduce uncertainty, and improve generalization. The approach consists of three primary stages: (1) developing base learners and training individual models, (2) constructing a stacking ensemble, and (3) implementing a super learner stacking model.

In the present work, the term "super learner" refers to an additional meta-learning layer that receives as input the predictions from multiple stacking ensembles (meta-learners) and learns an optimal combination of these predictions to produce the final FBHP estimate. For clarity, we denote this overall predictive architecture as the Super Learner Stacking (ST-S) model.

Five separate machine learning models are utilized as the foundation: KNN, LR, ELM, RF, and SVR. Each one is trained independently on the provided data, as depicted in Fig. 3(a). These models were chosen for their diverse capabilities, intended to encapsulate the multifaceted connection between the input features and the FBHP. Table 2 outlines the precise parameters and configurations used for each base learner.

The selection of KNN, LR, ELM, RF, and SVR as base learners was guided by the following three main criteria: (1) coverage of complementary learning paradigms, (2) suitability for the available dataset size, and (3) computational efficiency for repeated training runs. In particular, KNN provides a simple, nonparametric, instance-based method that captures local patterns in the FBHP–feature space; LR offers a linear, interpretable reference model that reflects approximate physical relationships and serves as a baseline; ELM is a fast single-hidden-layer neural network capable of approximating complex nonlinearities with low training cost; RF is a tree-based ensemble that captures nonlinear interactions and is robust to noisy or mildly collinear inputs; and SVR represents a margin-based kernel method with good generalization in medium-sized regression problems. The SVR model with an RBF kernel was chosen due to its ability to represent non-linear relationships between explanatory variables and the response variable, without the need to explicitly define the functional form. Furthermore, the RBF kernel exhibits good generalization capacity across various domains and allows for control of model complexity through the C and ε hyperparameters, which are adjusted through cross-validation [37–39].

Together, the five considered learners span linear and nonlinear, parametric and nonparametric, tree-based, kernel-based, instance-based, and neural-network paradigms, thereby increasing diversity in the stacking ensemble and improving its effectiveness. In contrast, more computationally demanding alternatives, such as deep neural networks, Gaussian process Regression, or large gradient-boosting ensembles, were not adopted as base learners to keep the training pipeline tractable for 100 independent runs with cross-validation and to reduce the risk of overfitting given the limited dataset size.

The base learners were configured with fixed hyperparameters, as summarized in Table 3, to reduce computational cost and ensure stability across multiple independent runs. Specifically, the SVR was implemented with an RBF kernel, regularization parameter $C = 1000$, and $\varepsilon = 0.1$. The Random Forest consisted of 30 decision trees, with a maximum depth of 5. KNN was configured with 5 neighbors, while Linear Regression was used without regularization with its default configuration. The ELM was defined with 50 neurons in the hidden layer and ReLU activation. No grid search was applied to the base learners to maintain methodological consistency and computational efficiency.

Grid search was applied exclusively to the meta-learners. For Ridge and Lasso regression models used in the higher stacking levels and the Super Learner, the regularization parameter α was explored over the interval $\alpha \in [10^{-6}, 10]$ with 500 equally spaced values, as also indicated in Table 3. The optimal α was automatically selected based on the mean error obtained via cross-validation.

Table 3
Parameterization of the models.

Acronym	Model name	Parameters	Meta-learner	Super-learner
KNN	K-Nearest Neighbor	Five neighbors	–	–
LR	Linear Regression	–	–	–
ELM	Extreme Learning	50 neurons, ReLU activation	–	–
RF	Random Forest	30 decision trees, depth limited to 5	–	–
SVR	Support Vector Regressor	$\epsilon = 0.1$, $C = 1000$, RBF kernel	–	–
ST-1	Stacking method	Stacked methods: ELM, RF, LR, KNN, SVR	Parameter-free LR	–
ST-2	Stacking method	Stacked methods: ELM, RF, LR, KNN, SVR	Ridge regression, exhaustive search on L_2 regularization	–
ST-3	Stacking method	Stacked methods: ELM, RF, LR, KNN, SVR	Lasso regression, exhaustive search on L_1 regularization	–
ST-S	Super-learner stacking method	Stacked models: ELM, RF, LR, KNN, SVR	Meta-learner 1 (ML-1): LR on (ELM, KNN, SVR); Meta-learner 2 (ML-2): LR on (RF, LR)	Ridge regression, exhaustive search on L_2 penalization with cross-validation

In the stacking ensemble construction step, three stacking ensemble models, *i.e.* ST-1, ST-2, and ST-3, are constructed using the base learners' predictions as input features for a meta-learner. Fig. 3(b) depicts this canonical stacking architecture. ST-1 employs Linear Regression as the meta-learner, whereas ST-2 and ST-3 use Ridge Regression and Lasso Regression, respectively, with L_2 and L_1 regularization to prevent overfitting and promote sparsity in the base-learner contributions.

To prevent the meta-learners from overfitting to the base learners' biases (data leakage), a 5-fold cross-validation scheme was uniformly applied across all stacking levels (ST-1, ST-2, ST-3, ML-1, ML-2, and ST-S). This method ensures that the inputs (meta-features) used to train any meta-learner are exclusively generated from predictions made by base models on out-of-fold samples. This reliance on independent, out-of-fold predictions guarantees the generalizability of the stacking ensemble.

Finally, the proposed ST-S model extends the stacking ensemble concept by incorporating an additional layer of ensemble learning, as illustrated in Fig. 3(c). Two meta-learners are employed: ML-1, which combines the predictions of ELM, SVR, and KNN, and ML-2, which combines the predictions of LR and RF. The outputs of ML-1 and ML-2 are used as input features for a final meta-learner that employs Ridge Regression with L2 regularization to produce a robust, accurate ensemble model.

To highlight the novelty of the ST-S model, it is essential to note that it introduces a dual-meta-learner structure, in which the base learners are divided into two functionally complementary groups. ML-1 combines ELM, KNN, and SVR to capture nonlinearities and local variability in the data, whereas ML-2 aggregates LR and RF to capture more global and stable patterns. The outputs of these two meta-learners are then combined in a final Ridge Regression layer, trained exclusively on out-of-fold predictions, thereby reducing bias and variance and enhancing generalization. By combining heterogeneous learners with two independent meta-learning pathways and a final regularized aggregation step, the ST-S framework distinguishes itself from previously reported super-learner architectures, such as those presented in [25,40]. This configuration enables the model to extract complementary modes of variability before the final aggregation, thereby improving robustness and predictive performance.

Fig. 4 depicts the overall methodology of the framework.

We implemented the framework relying on Python and well-known Python libraries, including pandas [41] and scikit-learn [42], and we executed 100 independent runs to assess the performance and stability of each model. The evaluation metrics include R, R^2, RMSE, MAE, and MAPE to quantify prediction accuracy and error rates. Additionally, we conducted an uncertainty analysis considering Mean Prediction Error (MPE), Width of Uncertainty Bandwidth (WUB), Mean Absolute Deviation (MAD), and Uncertainty to evaluate the reliability and precision of predictions. Furthermore, Shapley Additive exPlanations (SHAP) values are calculated to assess the relative importance of each input feature in predicting FBHP.

This approach mitigates potential biases and leverages the complementary strengths of each learner type. The number of learners that can be integrated within the ensemble framework is theoretically unbounded. However, practical considerations such as architectural complexity, memory constraints, and computational cost must be considered. Ideally, the resulting ensemble model should outperform the best individual learner in terms of prediction accuracy and robustness.

2.6. Performance metrics

The metrics used to assess the model's performance are: Person Coefficient (R), R-squared (R^2), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). Here, N denotes the dataset size, y_i represents the true FBHP for sample i, $\hat{y}_i$ represents its predicted FBHP value, $\bar{y}$ is the mean of the actual FBHP values, and $\bar{\hat{y}}$ is the mean of the estimated FBHP values. These metrics were chosen for their ability to capture the diverse behaviors demonstrated by the models under review.

- $\mathrm{R} = \dfrac{\sum_{i=1}^{N}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}\sqrt{\sum_{i=1}^{N}(\hat{y}_i - \bar{\hat{y}})^2}}$
- $R^2 = \sum_{i=1}^{N}(y_i - \hat{y}_i)^2 / \sum_{i=1}^{N}(y_i - \bar{y})$
- $\mathrm{MAE} = \dfrac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}|$
- $\mathrm{RMSE} = \sqrt{\dfrac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2}$
- $\mathrm{MAPE} = \dfrac{100}{N}\sum_{i=1}^{N}\left|\dfrac{y_i - \hat{y}}{y_i}\right|$

To provide the reader an immediate and concrete reference, for the best-performing ST-S model on the independent test set we obtained an average Pearson correlation coefficient $R = 0.930 \pm 0.003$, a coefficient of determination $R^2 = 0.857 \pm 0.006$, RMSE = 146.382 ± 2.806, MAE = 119.539 ± 2.081, and MAPE = $4.941 \pm 0.093\%$, computed across 100 independent runs on different train–test splits.

Error analysis was conducted to assess the models' performance in FBHP prediction. The error (e_j) for sample j is defined as $\bar{e} = \sum_{j=1}^{N} e_j$

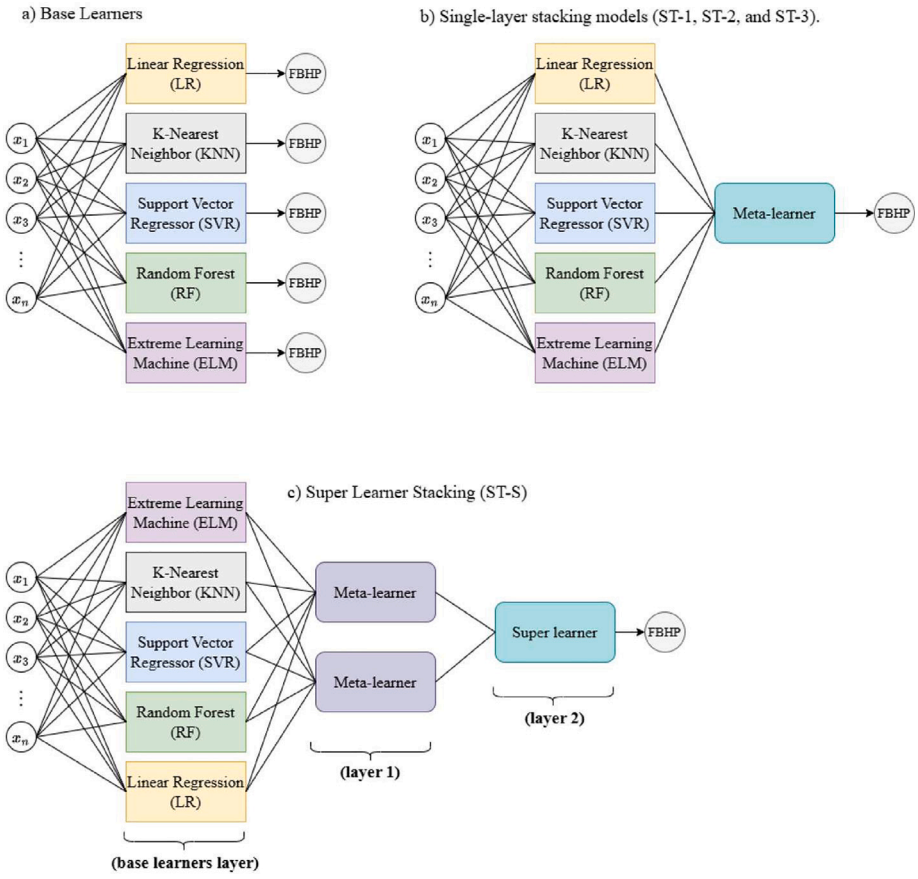

Fig. 3. Methodology chart illustrating the hierarchical stacking framework. The proposed approach is organized into three stages: (a) base learners, where individual machine learning models (LR, KNN, SVR, RF, and ELM) are independently trained on the original input features to generate preliminary FBHP predictions; (b) conventional stacking ensembles, consisting of single-layer stacking models (ST-1, ST-2, and ST-3) in which base-learner predictions are combined using a single meta-learner with different regularization strategies; and (c) the proposed Super Learner Stacking (ST-S), a multi-layer architecture that introduces two parallel meta-learners whose outputs are subsequently aggregated by a final super-learner to improve predictive accuracy, robustness, and generalization.

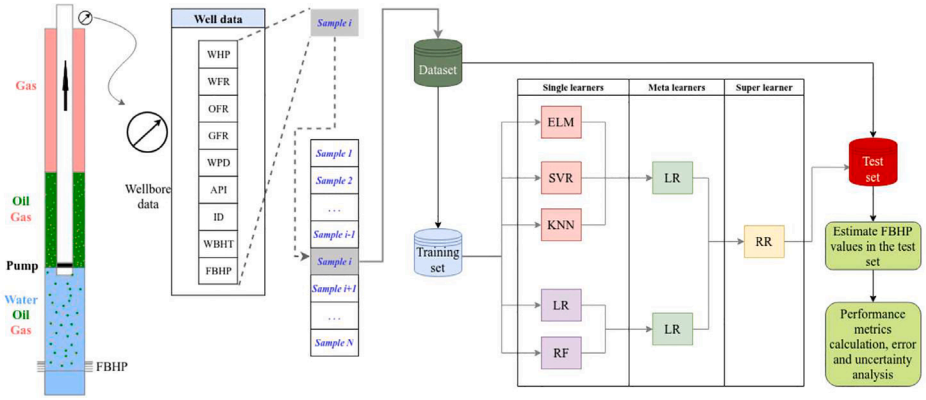

Fig. 4. Stacking framework.

and the confidence interval around at 95% level is $(\bar{e}–1.96S_e,\ \bar{e}+1.96S_e)$, where $S_e = \sqrt{\sum_{k=1}^{N}(\bar{e}-e_k)^2/(N-1)}$ represents the standard deviation.

The uncertainty analysis aims to quantify the extent to which output fluctuations respond to variations in input. Statistical measures such as the median, mean, and population quantiles are typically employed for this analysis [43]. Estimating uncertainty often involves propagation methods, and the corresponding confidence intervals are derived from these estimates, given the relatively limited sample size.

Within this study, input parameter variations were modeled using uniform distributions. MAD serves as an uncertainty metric, calculated as

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

where N_{MC} is the number of Monte Carlo simulations (set to 120000 in this study), and FBHP_i is the estimated FBHP for the ith sample. The uncertainty percentage is then calculated as:

$$\text{Uncertainty } \% = \frac{\mathrm{MAD}}{median(\mathrm{FBHP})} \times 100 \quad (2)$$

Table 4
Python (3.11 Version) libraries and versions used. The data were divided into 596 for training and 199 for the test set, and the data were run over 0 to 99; at each iteration, a different random seed was used to obtain different splits in a k-fold cross-validation, with k = 5 in the training set. Each run uses a deterministic seed.

Library	Version
scikit-learn	1.7.1
pandas	2.3.1
hydroeval	0.1.0
numpy	2.3.2
scipy	1.16.1
gmdhpy	1.0.3
xgboost	3.0.4

3. Results and discussion

The present section reports the numerical results obtained with the individual ML models and with the proposed ST-S ensembles, following the methodological framework described in Section 2. In particular, we first summarize the experimental setup and the evaluation protocol, and then compare the predictive performance and uncertainty characteristics of all models. Finally, we analyze feature-importance patterns and discuss the implications of the stacking strategy for FBHP prediction and reservoir management.

We conducted the experiments relying on Python, with the pandas [41] and scikit-learn [42] libraries for implementation. The computational environment consisted of an Intel(R) Core(TM) i7-9700F processor (8 cores at 3 GHz, 6MB cache), 32 GB of RAM, and the Linux Ubuntu 18 operating system. Table 4 shows the libraries used and their respective versions.

3.1. Performance comparison analysis

Table 5 presents the mean and standard deviation of the performance metrics, demonstrating the effectiveness of stacking ensembles in FBHP prediction. The results in Table 5 provide a comprehensive comparison across three distinct modeling paradigms: individual learners, homogeneous ensembles, and heterogeneous stacking. Individual learners such as LR and SVR showed limited predictive capability, likely due to their inability to handle the high dimensionality and nonlinearity of multiphase flow. Random Forest (RF), which represents a homogeneous ensemble approach, achieved a R^2 of 0.822; however, it was still surpassed by the ST-S model. Furthermore, ST-S exhibited the lowest Root Mean Squared Error (RMSE) (146.382), Mean Absolute Error (MAE) (119.539), and Mean Absolute Percentage Error (MAPE) (4.941). This indicates that the diversity of learning paradigms in a heterogeneous ensemble — combining tree-based, instance-based, and neural-network-based learners — provides a more robust approximation of FBHP than homogeneous methods.

To assess whether the observed differences in performance among all evaluated models, including individual learners (ELM, KNN, LR, SVR), the homogeneous ensemble (RF), and heterogeneous stacking models (ST-1, ST-2, ST-3, ST-S), are statistically significant, a non-parametric Kruskal–Wallis test was conducted on each evaluation metric. Table 6 presents the corresponding p-values. All metrics exhibit p-values below 0.001, indicating that the differences in performance across models are statistically significant. This confirms that the superior accuracy of the ST-S stacking model represents a meaningful improvement over both individual learners and the homogeneous ensemble, rather than being due to random variation.

The stacking approach combines the strengths of individual base learners (KNN, LR, ELM, RF, SVR) while mitigating their weaknesses. This ensemble approach leads to a more robust and accurate model than any single learner. Additionally, ST-S incorporates a diverse set of base learners, each with its own algorithms and learning mechanisms. This diversity enables the model to capture diverse aspects of the data and improve its generalizability. The additional Super Learner layer refines predictions by combining outputs from multiple meta-learners, enhancing accuracy and robustness.

On the other hand, SVR exhibited the lowest performance across all metrics The low R-value (0.743) and R-squared (0.104) indicate a weak correlation between SVR predictions and actual FBHP values, suggesting a poor fit of the model to the data. Additionally, significantly higher RMSE (366.144), MAE (295.693), and MAPE (12.072) values indicate large deviations between SVR predictions and actual FBHP values. A potential reason for poor SVR performance is hyperparameter sensitivity, as SVR performance is highly sensitive to hyperparameter selection. In this regard, the chosen hyperparameters (C = 1000, ε = 0.1) may not be optimal for this data and task, leading to subpar results. SVR can also be sensitive to feature scaling and distribution. If features have different scales or non-normal distributions, it can negatively impact performance. The choice of kernel function in SVR significantly influences its behavior. The study does not specify the chosen kernel; however, using an inappropriate one could lead to poor performance.

The superior performance of the ST-S model (R^2 = 0.857) over individual learners is rooted in the principle of methodological diversity. Heterogeneous stacking succeeds here because the chosen base learners represent fundamentally different ways of approximating the FBHP function: KNN captures local spatial similarities, ELM approximates complex non-linear maps, RF handles high-dimensional interactions, and LR maintains a stable linear baseline. The ensemble effectively votes across different mathematical perspectives, which decorrelates individual model errors and results in a more robust and generalized prediction than any single model could achieve. This is particularly critical in petroleum engineering, where noisy field data can easily lead a single algorithm into overfitting or missing specific flow-regime transitions.

3.2. Analysis of stacking coefficients

We performed an analysis of the final coefficients obtained on the independent test set to assess each base learner's contribution to the stacking models. Fig. 5 compares the final coefficients assigned to each base learner within the three stacking methods: ST-1, ST-2, and ST-3. These coefficients are obtained from the test set, derived from 100 independent runs, are presented as box plots, and offer important information about the relative influence of each base classifier on the overall predictions of the stacking sets. The boxplots show that the base learners affect the stacking models differently. Although the SVR coefficients are lower than those of the other models, the distributions and ranges of coefficients differ significantly across models and individual learners, indicating varying degrees of importance in their contributions to the final predictions. With a broader range of coefficients, ELM and SVR remain influential, whereas KNN contributes only moderately.

Fig. 6 presents the results of the ST-S approach. The boxplots on the left depict the distribution of coefficients obtained from predictions on the test set across 100 independent runs. The scatter plot (right) shows the relationship between the coefficients within the Super Learner ensemble for the same test evaluations. The boxplot for ML-1 exhibits a broader range and generally higher coefficient values than ML-2. This suggests that ML-1, which combines the predictions of ELM, SVR, and KNN, plays a more dominant role in the Super Learner model's final test predictions. While less dominant, ML-2 still exhibits a range of positive coefficient values, indicating its contribution to the overall test-based predictions. Its combination of LR and RF models seems to provide valuable information that complements the predictions from ML-1.

The architectural logic of the ST-S provides a further layer of refinement through functional decomposition. As shown in Fig. 6, the model

Table 5
Performance evaluation of FBHP prediction methods on the independent test dataset comparing individual base learners (ELM, KNN, LR, SVR), homogeneous ensembles (RF), and heterogeneous stacking models (ST-1, ST-2, ST-3, ST-S). The reported metrics (R, R^2, RMSE, MAE, MAPE) correspond to model performance on the test set. Mean and standard deviation (in parentheses) were calculated across 100 independent runs with different random seeds.

Model	R	R^2	RMSE	MAE	MAPE
ELM	0.916 (0.011)	0.820 (0.022)	163.676 (9.224)	132.249 (4.105)	5.432 (0.170)
KNN	0.895 (–)	0.801 (–)	172.682 (–)	136.585 (–)	5.580 (–)
LR	0.849 (–)	0.689 (–)	215.587 (–)	181.016 (–)	7.792 (–)
RF	0.909 (0.004)	0.822 (0.007)	163.199 (3.212)	132.134 (2.861)	5.508 (0.122)
SVR	0.743 (–)	0.104 (–)	366.144 (–)	295.693 (–)	12.072 (–)
ST-1	**0.930** (0.003)	0.854 (0.006)	147.853 (2.854)	120.637 (2.283)	5.008 (0.094)
ST-2	**0.930** (0.004)	0.853 (0.007)	148.376 (3.295)	121.054 (2.356)	5.030 (0.100)
ST-3	**0.930** (0.003)	0.853 (0.006)	148.330 (3.077)	120.737 (2.330)	5.012 (0.097)
ST-S	**0.930** (0.003)	**0.857** (0.006)	**146.382** (2.806)	**119.539** (2.081)	**4.941** (0.093)

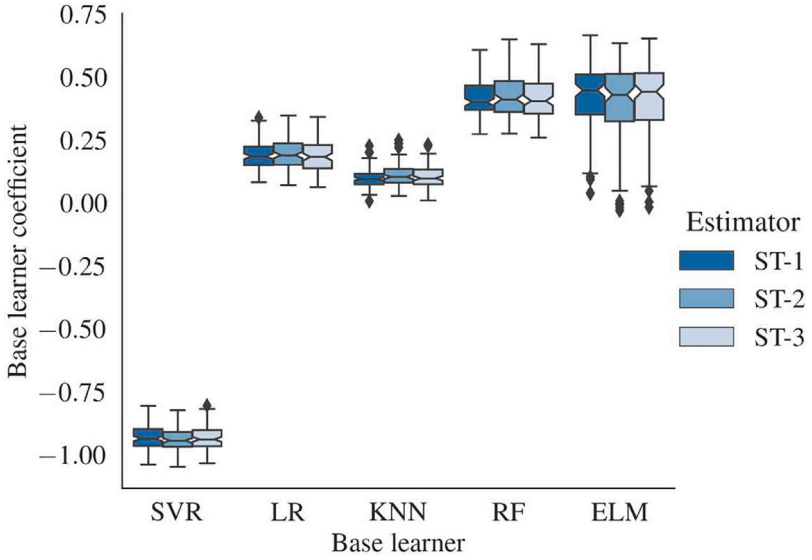

Fig. 5. Stacking models (ST-1, ST-2, ST-3) coefficients evaluated on the test set..

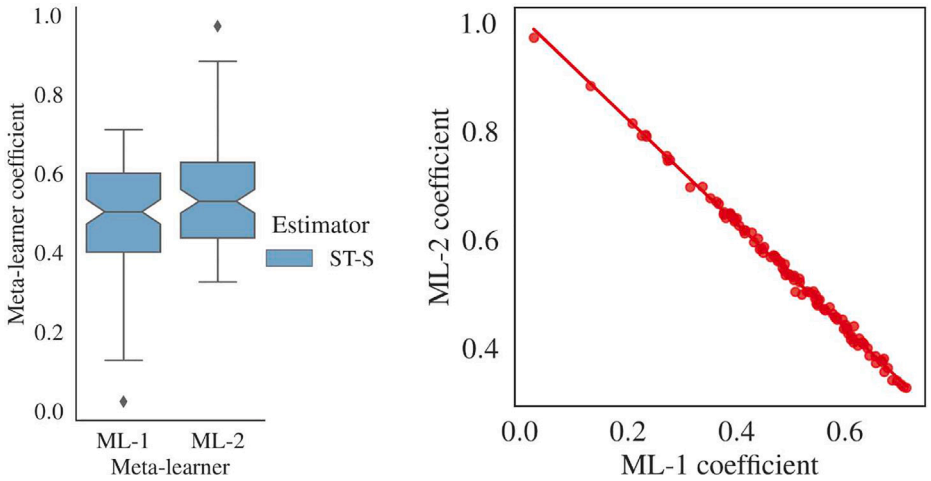

Fig. 6. Left: Distribution of coefficients obtained from the Super Learner ensemble across 100 independent test runs. Right: Visualization of the relationship between coefficients within the Super Learner ensemble across 100 independent test evaluations.

assigns a more dominant weight to ML-1, which aggregates learners (ELM, SVR, KNN) specifically suited for capturing local nonlinearities and high-frequency variability in the pressure data. Conversely, ML-2 (LR, RF) contributes by providing a stable global context. The strong linear correlation between ML-1 and ML-2 coefficients (Fig. 6, right) suggests that the final Super Learner (Ridge Regression) dynamically balances these two modes of information (local detail vs. global stability) to mitigate the high variance typically found in bottom-hole pressure signals.

The scatterplot on the right side of Fig. 6 exhibits a strong linear correlation between the coefficients of ML-1 and ML-2, suggesting that the contributions of the two meta-learners are directly dependent on

Table 6
Kruskal–Wallis test p-values for evaluation metrics across FBHP prediction models. Values below 0.05 indicate statistically significant differences.

Metric	R	R^2	RMSE	MAE	MAPE
p-value	<0.001	<0.001	<0.001	<0.001	<0.001

each other. However, they can vary independently across different test runs. The correlation indicates that the ST-S model can exhibit varying ensemble dynamics depending on the independent test evaluation and the performance of individual base learners within each meta-learner.

3.3. Uncertainty analysis of FBHP predictions

We present a comparative analysis of models' performance in predicting FBHP in Table 7. The latter table presents metrics for assessing model uncertainty: MPE, WUB, MAD, and overall uncertainty. The first column identifies the specific model. The second column reports the MPE, defined as the average deviation between the predicted and actual FBHP values. The WUB, displayed in the third column, quantifies the range of predicted FBHP values with 95% confidence level. The fourth column shows the median predicted FBHP for each model. The fifth column reports the MAD, which reflects the average magnitude of the prediction errors. Finally, the sixth column presents the uncertainty.

The results show the advantages of stacking ensembles, particularly the proposed approach, in achieving superior accuracy and reducing prediction uncertainty. As a consequence, the model can potentially be used as a complementary tool to enhance decision-making capabilities within petroleum engineering applications. In particular, coupling the ST-S model with WHP and WFR control allows operators to quantify the FBHP consequences of alternative operating scenarios (e.g., tightening or relaxing choke settings, accepting higher WFR within processing limits) and to select options that balance production targets, flow-assurance constraints, and intervention risk.

The MPE indicates potential bias in the models. Stacking models, especially ST-S, exhibit lower MPE than individual learners, indicating less overall bias in their predictions. WUB reflects the variability in predictions. SVR, the lowest-performing model in Table 5, demonstrates the largest WUB, indicating high variability. Conversely, stacking models consistently exhibit lower WUB scores, suggesting greater precision. MAD and Uncertainty, which measure the spread of predictions around the median, follow a similar trend. Stacking models, particularly ST-S, show lower MAD and Uncertainty, indicating better predictability and more consistent performance.

The observed performance differences can be explained by several key factors. First, stacking models reduces uncertainty through regularization, achieved by combining predictions from multiple base learners. Such a process inherently reduces both bias and variance, yielding more accurate and reliable predictions. Furthermore, the inclusion of diverse base learners with complementary strengths allows stacking models to effectively capture a wider range of data complexities, resulting in lower overall uncertainty. Finally, the proposed architecture further enhances predictions by incorporating outputs from multiple meta-learners, thus achieving additional reductions in bias and uncertainty compared to simpler stacking configurations. In contrast, individual models, such as SVR, may be limited by hyperparameter selection, data scaling, or kernel choice, leading to higher bias, variance, and uncertainty in their predictions.

The uncertainty analysis in Table 7 complements the performance evaluation in Table 5. Models that demonstrate high R and R^2 values, and low RMSE, MAE, and MAPE in Table 5, also exhibit lower MPE, WUB, MAD, and Uncertainty in Table 7. This connection reinforces the superior performance and reliability of the stacking models, particularly ST-S.

Accurate FBHP prediction is crucial for reservoir management, production optimization, and well performance evaluation [44]. Uncertainty in FBHP predictions can have substantial economic and operational consequences, impacting decisions regarding production strategies, well interventions, and reservoir development [45]. Stacking models is a promising approach to improve predictive accuracy and reduce uncertainty, offering valuable tools for petroleum engineers. Furthermore, understanding and quantifying the uncertainty associated with FBHP predictions is essential for risk assessment and informed decision-making.

Fig. 7 (left) presents a scatter plot illustrating the relationship between uncertainty and Root Mean Squared Error (RMSE) for all independent model runs. The x-axis represents Uncertainty, calculated as the MAD divided by the predicted FBHP median, expressed as a percentage. Higher uncertainty values indicate larger spread and lower confidence in the predictions. The y-axis depicts RMSE, which quantifies the average magnitude of prediction errors. Lower RMSE values correspond to superior prediction accuracy.

The figure reveals that the SVR model achieves the lowest uncertainty but exhibits the highest RMSE (lowest accuracy). The ELM model performs worst in terms of uncertainty, with median RMSE values. The RF model balances uncertainty and accuracy. Notably, the stacking models, particularly ST-S, achieve a favorable equilibrium between these two metrics. This outcome highlights the stacking approach's ability to effectively leverage the strengths of individual models.

As shown in the figure, there is a general inverse relationship between RMSE and Uncertainty. Models with superior accuracy (lower RMSE) tend to exhibit higher uncertainty, highlighting the inherent challenge of simultaneously achieving high accuracy and high confidence in predictions. Each data point represents the final solution from an independent model run, enabling a comparative analysis of their performance and uncertainty characteristics.

Fig. 7 (left) depicts the Pareto front, represented by the black dashed line. A Pareto front denotes the set of optimal (or near-optimal) solutions where improvement in one objective function (e.g., accuracy) necessarily comes at the expense of another (e.g., uncertainty). This visualization shows the inherent trade-off between accuracy and uncertainty in FBHP prediction models.

Fig. 7 (left) can help decision-makers construct informed choices based on specific project scenarios. If minimizing uncertainty is the primary objective, with some flexibility in accuracy, the SVR model is a viable alternative. Conversely, for scenarios prioritizing a balance between confidence and accuracy, the KNN and RF models are favorable options due to their suggested balance between these two aspects. However, if the predictions require highly accurate models with a degree of relative confidence, the ST-S stacking model represents the optimal choice.

3.4. Feature importance analysis

Feature importance scores were measured using the SHAP method [46] for each model during regression of the dataset. The average feature importance scores are presented in Table 8.

It should be noted that the scores presented in Table 8 represent the absolute mean contribution of each feature to the model's output magnitude. These values are not normalized, which explains the significant variation in scale between different learners. For example, the SVR model's importance scores are nearly an order of magnitude lower than those of the ST-S model; this is a direct consequence of the SVR's poor predictive fit ($R^2 = 0.104$), which resulted in minimal feature-to-output attribution.

Across all models, Wellhead Pressure (WHP) and Water Flow Rate (WFR) consistently exhibit the highest SHAP values, reflecting their significant influence on FBHP predictions. This result aligns with the established understanding of WHP and WFR as key factors governing pressure conditions within the wellbore [47,48]. Other features such

Table 7
Uncertainty and error analysis.

Model	MPE	WUB	Median	MAD	Uncertainty %
ELM	42.679 (5.771)	158.253 (10.121)	2935.134 (163.110)	614.870 (80.716)	20.993 (2.834)
KNN	−7.489 (–)	172.955 (–)	2783.792 (1.044)	223.801 (0.613)	8.039 (0.024)
LR	67.917 (–)	205.126 (–)	3059.848 (2.337)	485.386 (1.022)	15.863 (0.036)
RF	20.292 (3.346)	162.307 (3.180)	2790.372 (10.590)	312.298 (11.290)	11.192 (0.404)
SVR	−31.186 (–)	365.733 (–)	2427.982 (0.046)	11.054 (0.024)	0.455 (0.001)
ST-1	40.784 (3.181)	142.440 (2.875)	2895.283 (90.857)	442.265 (45.361)	15.283 (1.560)
ST-2	40.250 (3.475)	143.126 (3.489)	2887.649 (79.222)	432.147 (46.567)	14.973 (1.631)
ST-3	40.258 (3.540)	143.075 (3.269)	2908.879 (75.097)	440.415 (45.139)	15.143 (1.534)
ST-S	34.371 (3.952)	142.592 (2.883)	2873.618 (66.068)	404.195 (39.888)	14.075 (1.453)

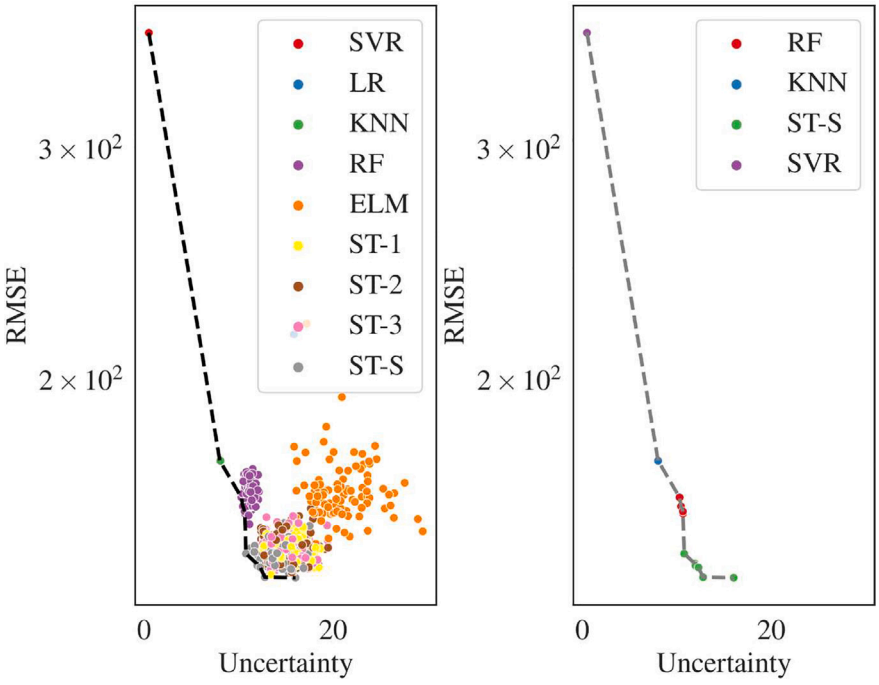

Fig. 7. Comparison between uncertainty and RMSE for all independent runs (left) and for those solutions in the nondominated front (right). The non-dominated front is represented by the dashed lines.

Table 8
Significance of input variables using SHAP analysis for test dataset analysis. Values represent the absolute mean impact on model output; higher magnitudes correlate with stronger predictive patterns captured by the model.

	WHP	WFR	OFR	GFR	WPD	API	ID	WBHT
ELM	1.13 (0.06)	0.51 (0.04)	0.28 (0.07)	0.29 (0.10)	0.32 (0.03)	0.10 (0.08)	0.06 (0.02)	0.13 (0.05)
KNN	0.70 (–)	0.33 (–)	0.09 (–)	0.02 (–)	0.19 (–)	0.06 (–)	0.02 (–)	0.06 (–)
LR	1.14 (–)	0.48 (–)	0.06 (–)	0.07 (–)	0.28 (–)	0.00 (–)	0.03 (–)	0.00 (–)
RF	0.68 (0.017)	0.29 (0.01)	0.04 (0.06)	0.08 (0.08)	0.27 (0.01)	0.10 (0.01)	0.00 (0.02)	0.00 (0.02)
SVR	0.04 (–)	0.03 (–)	0.01 (–)	0.01 (–)	0.01 (–)	0.02 (–)	0.01 (–)	0.02 (–)
ST-1	0.98 (0.04)	0.41 (0.02)	0.11 (0.03)	0.14 (0.04)	0.31 (0.01)	0.06 (0.01)	0.02 (0.09)	0.03 (0.01)
ST-2	0.97 (0.04)	0.41 (0.02)	0.11 (0.03)	0.13 (0.04)	0.30 (0.01)	0.06 (0.02)	0.02 (0.08)	0.03 (0.02)
ST-3	0.97 (0.04)	0.41 (0.02)	0.11 (0.04)	0.14 (0.04)	0.31 (0.01)	0.05 (0.02)	0.02 (0.09)	0.03 (0.02)
ST-S	0.94 (0.04)	0.40 (0.02)	0.09 (0.02)	0.12 (0.03)	0.29 (0.01)	0.05 (0.01)	0.02 (0.07)	0.03 (0.01)

as Oil Flow Rate (OFR), Gas Flow Rate (GFR), and Water Production Depth (WPD) show moderate importance, suggesting their contribution to FBHP but to a lesser extent. Well Bottom-Hole Temperature (WBHT), Internal Diameter (ID), and API Gravity (API) have relatively low SHAP values, indicating minimal impact. Notably, the relative importance remains consistent across individual learners and stacking ensembles, indicating that various modeling approaches effectively capture these underlying relationships.

This observed feature importance can be attributed to the underlying physical principles that govern it. WHP directly influences the pressure at the bottom hole, while WFR affects pressure losses due to friction within the wellbore [24]. OFR, GFR, and WPD relate to the properties and flow dynamics of the produced fluids, which can, in turn, indirectly influence pressure conditions. API Gravity, ID, and WBHT may have secondary effects through their influence on fluid properties or flow behavior, but their overall impact appears to be less significant.

Understanding the importance of features suggests practical insights for reservoir management and FBHP prediction. The dominance of WHP and WFR highlights the importance of accurate pressure and flow rate measurements. In this case, failures in the sensors associated with these variables can significantly affect well production. The moderate

importance of fluid properties suggests incorporating them into models to achieve greater accuracy. The relatively low importance of API Gravity, ID, and WBHT may indicate limited value for FBHP prediction in this context.

Building on these insights, from an operational standpoint, the strong influence of WHP and WFR indicates that the latter variables can be treated as primary control knobs to steer FBHP within target ranges without requiring immediate downhole interventions. In practical terms, integrating the ST-S predictions with real-time WHP and WFR measurements enables production engineers to evaluate alternative choke settings, adjust artificial-lift operating conditions, and anticipate the impact of increasing water production on FBHP before committing to workovers or water-handling upgrades.

Fig. 8 presents SHAP scores, providing a more precise visualization of feature importance across all estimators. According to the analysis, except for the SVR model, WHP, WFR, and WPD consistently have the most significant impact on FBHP predictions, followed by OFR and GFR. API, ID, and WBHT exhibit very low SHAP values and thus have minimal influence on the models. This result confirms that the ML models captured relationships that are physically consistent with the hydraulic and thermodynamic behavior of the production system.

Wellhead Pressure WHP exhibits the highest SHAP magnitude in nearly all models, reaffirming its dominant role as the primary driver of FBHP. Physically, WHP defines the boundary condition that initiates the pressure gradient along the wellbore. An increase in WHP provides greater hydraulic energy at the surface, which propagates through the fluid column, overcoming hydrostatic and frictional losses. The high SHAP values associated with WHP indicate that the models correctly captured this dependency. This finding aligns with petroleum-production theory, where surface pressure control governs fluid lifting efficiency, flow stability, and drawdown behavior, confirming that WHP is the primary determinant of downhole pressure transmission [13,49].

Water Flow Ratio WFR appears as the second most influential feature, emphasizing its impact on overall flow resistance and multiphase behavior. As the proportion of water in the mixture increases, the overall density and viscosity rise, intensifying frictional losses and modifying the flow regime within the tubing. These effects are documented in multiphase-flow theory, where higher water cuts are associated with greater pressure drops and reduced flow efficiency [50]. The intermediate SHAP importance of WFR reflects the model's ability to internalize these dynamics, capturing how variations in water fraction affect pressure-propagation efficiency and energy dissipation. In practical terms, this means that changes in fluid composition can significantly alter the well's pressure profile and production stability.

Oil and Gas Flow Rates OFR and GFR show moderate SHAP contributions across most models. These variables represent the volumetric dynamics of each production phase and are directly related to total flow rate, which governs kinetic energy and turbulence within the wellbore. Higher OFR and GFR values increase momentum and accelerate the flow, but also intensify wall shear and frictional losses. Their moderate importance suggests that, while they contribute to short-term FBHP variations, their effects are partially embedded within composite parameters such as WFR and WHP, which already capture the overall energy balance and multiphase behavior of the system [13].

Well Production Depth WPD exhibits relevant SHAP scores, reflecting its indirect influence through formation pressure and hydrostatic effects. Deeper wells generally experience higher formation pressures due to overburden stress and compaction, providing additional drive energy to sustain flow. However, increased depth also extends the hydrostatic column and frictional path length, partially offsetting this advantage. The resulting balance explains the moderate SHAP contribution of WPD, indicating that the models correctly captured the equilibrium between reservoir drive and wellbore pressure losses — an essential component of realistic FBHP estimation [51].

API Gravity API, Inside Diameter ID, and Wellbore Temperature WBHT show very low SHAP values in the models analyzed, indicating that the models correctly capture their minor influence on FBHP under the studied conditions. This is consistent with [52], which states that variations in the temperature of the condensate gas reservoir had only a minor effect on bottom-hole pressure calculations.

The SHAP analysis results (Table 8, Fig. 8) demonstrate that the ST-S model is not merely a black box but is physically consistent with established hydraulic and thermodynamic behaviors. While traditional physics-based mechanistic models (e.g., [1]) often struggle as the number of unknown variables increases, the ST-S model correctly identifies Wellhead Pressure (WHP) and Water Flow Rate (WFR) as the dominant physical drivers of pressure drop, effectively learning the physics that empirical correlations often oversimplify.

3.5. Model strengths and limitations

The findings highlight the advantages of stacking ensembles, especially the super learner approach, for achieving higher accuracy and reducing uncertainty, thereby improving decision-making in petroleum engineering applications. This study demonstrates the potential of these models to achieve greater accuracy and robustness than individual learners. Ensemble methods such as stacking can improve prediction accuracy and robustness compared to individual learners [54,55], especially for complex tasks in FBHP prediction [56].

Heterogeneous ensembles with diverse base learners are generally more effective than homogeneous ensembles [57]. The expectation is for enhanced performance due to methodological differences among inductive learning algorithms, which enable the production of diverse learning models that are more likely to complement one another. When the errors of heterogeneous individual models exhibit lower correlation, this can decrease the ensemble's overall error [58]. Super learner approaches with multiple stacking layers significantly improve model performance by combining predictions from diverse base models, effectively leveraging the strengths of different algorithms. This process helps capture complex relationships in the data and minimize errors by aggregating complementary predictions from various models [40].

Model performance is highly dependent on the choice of algorithm, hyperparameter tuning, and data preprocessing. Different algorithms have varying strengths and weaknesses: while some excel at handling large datasets with complex relationships, others are more suitable for small datasets with clear patterns [59]. Hyperparameter tuning involves optimizing the settings that govern how the algorithm learns from the data, often requiring experimentation and validation to find the optimal configuration that minimizes errors and maximizes predictive accuracy [60]. Data preprocessing is essential for preparing data in a format suitable for the chosen algorithm, ensuring effective learning and accurate predictions [61]. Understanding the strengths and weaknesses of different models is paramount for selecting suitable methods and constructing effective ensembles. By identifying which models best capture specific data patterns or relationships, practitioners can strategically combine them to leverage their complementary strengths, thereby enhancing predictive accuracy and mitigating individual model biases or limitations [62].

The ST-S configuration demonstrates superior performance compared to individual learners and simpler stacking ensembles. High R and R^2 values combined with low RMSE, MAE, and MAPE in ST-S indicate its ability to capture complex relationships between features and FBHP [63]. Uncertainty analysis reinforces this strength, with lower MPE, WUB, MAD, and Uncertainty values signifying more precise and reliable predictions. The use of diverse base learners in the stacking ensemble contributes to robustness and generalizability, which are crucial for real-world well conditions. Additionally, SHAP analysis reveals the model's effectiveness in utilizing relevant features (WHP, WFR) and fluid flow dynamics.

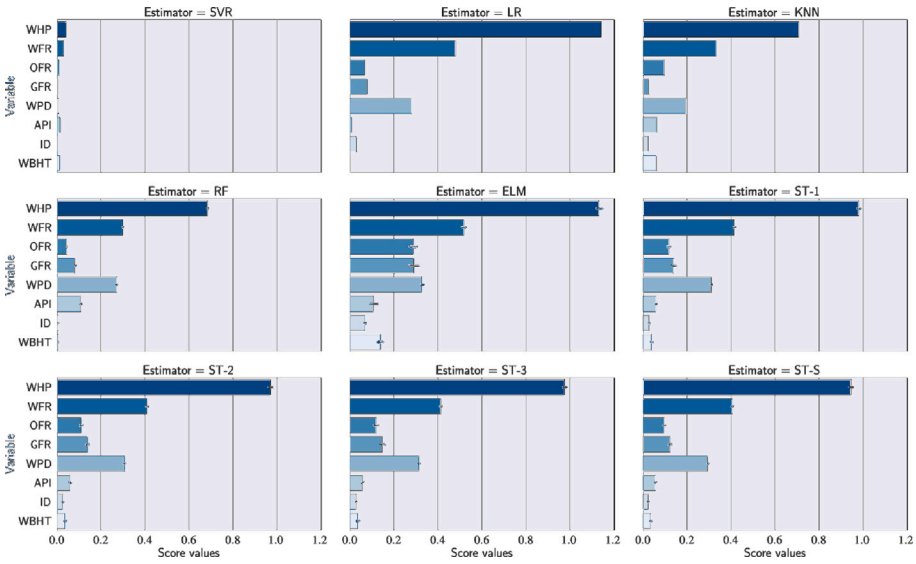

Fig. 8. SHAP [53] importance scores for the input features used in FBHP prediction across all evaluated models. Bars represent the average contribution of each feature to the model outputs, and markers indicate the variability across 100 independent runs.

The practical applicability of the ST-S model is supported by its balance between structural complexity and computational efficiency. Unlike hybrid models that rely on computationary expensive optimization procedures [64,65], the proposed framework achieves high predictive accuracy through a hierarchical arrangement of base learners that are inherently fast to train. Despite the increased number of layers in the model, the use of ELM (which avoids iterative optimization) and Ridge/Lasso meta-learners ensures that the training pipeline remains tractable, even when executing 100 independent simulations for uncertainty quantification.

Although the heterogeneous stacking configuration STS achieved the best overall predictive performance, the numerical improvements over simpler stacking ensembles ST1, ST2, and ST3 remain modest. However, these incremental gains were statistically consistent across 100 independent runs, accompanied by measurable reductions in prediction uncertainty. This behavior demonstrates the model's robustness and stability across different data splits. In practical petroleum engineering applications, where FBHP estimates support decisions on lift optimization, flow assurance, and intervention planning, even small improvements in accuracy and reliability can reduce operational risk and enable more confident decision-making. Nevertheless, adding a stacking layer increases training time and reduces interpretability, creating a trade-off that must be weighed against operational needs. Therefore, although the improvements achieved by the STS model are small in magnitude, they provide tangible operational value, particularly in scenarios where prediction stability is essential.

The proposed model has several important limitations. Computational complexity is high due to training multiple learners, and interpreting the reasoning behind the stacking ensemble's predictions remains challenging. In addition, the model's performance depends directly on the quality and representativeness of the training data. In this study, the model was developed and validated using a dataset from wells in the Middle East, which exhibit relatively homogeneous geological, fluid, and completion characteristics and operate within defined statistical ranges for variables such as WHP, WFR, and WPD. For this reason, the predictive accuracy observed may not generalize reliably to wells under substantially different conditions, such as reservoirs with high-viscosity oil, deepwater environments, or shale formations, where the underlying physical relationships and production regimes may differ significantly. In such cases, direct application of the model may require retraining or adaptation using local data, highlighting the importance of assessing model transferability before applying it to new basins or reservoir types.

Concrete examples from the independent test evaluations highlight the above discussed limitations. For instance, the SVR base learner yielded substantially degraded test performance ($R = 0.743$, $R^2 = 0.104$, RMSE = 366.144), and it also presented the largest width of uncertainty bandwidth (WUB = 365.733), thus reporting how sensitive some learners can be to fixed hyperparameter choices and to distributional characteristics of the data. Moreover, although the proposed ST-S achieved the best overall test accuracy ($R^2 = 0.857$ and RMSE = 146.382), the improvement over simpler stacking variants remained limited (e.g., ST-1: $R^2 = 0.854$ and RMSE = 147.853), therefore indicating diminishing returns when adding an extra stacking layer while still increasing training time and reducing interpretability. Last, but not least, it is important to notice that the latter test results were obtained from random splits of the same Middle-East dataset; thus, the model was not stress-tested on out-of-region wells (e.g., deepwater, shale, or high-viscosity oil settings), and transferability to substantially different operating envelopes remains an open issue that may require retraining with additional local data.

Future work should explore refined model configurations, enhance interpretability, and incorporate domain knowledge from petroleum engineering. Although the present work relied exclusively on the original input variables, future research could benefit from incorporating domain-informed features derived from physical principles. To specifically address the interpretability challenge associated with stacking models, future research should prioritize the utilization of advanced model-agnostic techniques. These include LIME (Local Interpretable Model-agnostic Explanations) [66], which provides local prediction explanations, thereby significantly enhancing model transparency.

In petroleum production systems, several physical relationships are well established, for example, the pressure drop along the wellbore is often proportional to the square of the flow rate [67], and interactions between fluid properties such as the water–oil ratio and API gravity strongly influence flow dynamics [68,69]. Therefore, introducing variables such as WHP^2, (WFR/API), or other combinations that capture these physical dependencies may help models more effectively capture the nonlinear relationships inherent to multiphase flow.

The incorporation of such domain-knowledge-based transformations represents a promising direction for advancing FBHP modeling, as it can enrich the feature space, improve generalization capability, and enhance model interpretability. Furthermore, understanding the impact of data quality on feature importance remains an essential line of investigation. These advancements could lead to even more accurate and practical models to support decision-making in reservoir management and production optimization.

3.6. Future works

Future investigations may explore extending the proposed heterogeneous stacking architecture to other problems in petroleum engineering and related data-driven applications. Recent studies have demonstrated the increasing relevance of advanced machine learning algorithms for modeling complex subsurface phenomena, such as EUR prediction in shale gas wells using GMDH, Lasso, and Bayesian Deep Learning [70] and hybrid intelligent systems for loss-circulation forecasting during drilling operations [71]. These studies indicate that ensemble-based and hierarchical learning approaches, similar in spirit to the Super Learner framework proposed here, may also improve predictive reliability in broader petroleum workflows.

Additionally, machine learning solutions have been increasingly applied to environmental and hydro-geological systems, such as groundwater-level forecasting under climate-change scenarios [72]. These examples reinforce the potential of stacking models for nonlinear, multivariate, and noisy data domains. Although these applications differ from FBHP estimation, they highlight the generality of data-driven modeling techniques and the potential for cross-domain methodological advances.

Future research may therefore focus on: (i) incorporating physical constraints or physically informed stacking architectures; (ii) expanding the meta-learner design to include uncertainty-aware or Bayesian super learners; and (iii) applying the proposed ST-S model to dynamic production variables, rate-transient analysis, and well integrity diagnostics.

4. Concluding remarks

The present study successfully developed and validated a heterogeneous stacking ensemble model for accurate FBHP prediction in oil wells. The key findings and their practical implications are summarized as follows:

- Superior predictive performance: The proposed ST-S model demonstrated superior accuracy and lower error rates compared to individual ML models and simpler stacking ensembles. Such improved predictive capability directly supports more reliable reservoir management and production optimization decisions by providing engineers with more reliable FBHP estimates.
- Robustness and generalizability: The integration of different base learners (*i.e.* ELM, KNN, LR, RF, SVR) within the stacking framework resulted in a model that is both robust and generalizable across multiple well conditions. This reliability across different operational scenarios reduces uncertainty in FBHP forecasting, thereby aiding the development of more effective production strategies.
- Reduced prediction uncertainty: The stacking ensemble approach significantly reduced prediction uncertainty, as evidenced by lower MPE, WUB, MAD, and uncertainty metrics. Such increased precision improves confidence in the model's outputs, thereby enabling better risk assessment and operational planning.
- Actionable feature insights: The SHAP analysis identified WHP and WFR as the most influential features on FBHP. This indicates that changes in the latter variables are likely to have the greatest impact on well performance, making them particularly valuable levers for optimization. Because such parameters are directly controllable at the surface, the model outputs can be embedded into day-to-day operational routines, such as setting WHP targets during production meetings, revising choke policies under changing water cuts, and screening candidate wells where small adjustments in WHP or WFR are predicted to yield the largest FBHP and production gains.
- Limitations and future work: Despite its strengths, the model's computational complexity and interpretability present challenges. Indeed, future research directions could focus on optimizing model efficiency, improving interpretability through additional techniques, and integrating physical knowledge for hybrid modeling.

This paper proposed a stacking ensemble model that provides a powerful, data-driven tool for FBHP prediction that not only advances the methodological state-of-the-art but also offers tangible benefits for practical reservoir management and production optimization, to help engineers make informed decisions.

Software and data availability

Software name: STS-FLOW — Stacked Two-Stage Flowing Bottom Hole Pressure Learning and Oilfield Wellbore Predictor, Developer: Leonardo Goliatt.

Contact information: leonardo.goliatt@ufjf.br.

First year available: October 01, 2025.

Program language: Python (version 3.11).

Source code: https://github.com/LGoliatt/sts-flow.

Data available: https://github.com/LGoliatt/sts-flow/tree/main/data/data_shammari.

Documentation: https://github.com/LGoliatt/sts-flow/blob/main/README.md.

CRediT authorship contribution statement

Deivid Campos: Writing – original draft, Software, Methodology, Data curation. **Bruno da Silva Macêdo:** Writing – original draft, Visualization, Software, Data curation. **Oscar Ikechukwu Ogali:** Writing – review & editing, Validation, Investigation, Formal analysis. **Matteo Bodini:** Writing – review & editing, Validation, Formal analysis. **Dmitriy A. Martyushev:** Writing – review & editing, Validation, Investigation, Formal analysis. **Farouk Abduh Kamil Al-Fahaidy:** Writing – review & editing, Validation, Investigation, Formal analysis. **Camila Martins Saporetti:** Writing – original draft, Software, Methodology, Data curation. **Leonardo Goliatt:** Writing – review & editing, Supervision, Project administration, 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.

Acknowledgments

The authors acknowledge the support of the funding agencies CNPq, Brazil (grants 307688/2022-4, 409433/2022-5, and 304646/2025-3), Fapemig, Brazil (grants APQ-02513-22, APQ-04458-23, and BPD-00083-22), Finep, Brazil (grant SOS Equipamentos 2021 AV02 0062/22), FAPERJ, Brazil (grant 10.432/2024-APQ1), and Capes, Brazil (Finance Code 001). This work has been supported by UFJF's High-Speed Integrated Research Network (RePesq) https://www.repesq.ufjf.br/.

References

[1] O. Adekomaya, A.A. Fadairo, O. Falode, Predictive tool for bottom-hole pressure in multiphase flowing wells, Pet. Coal 50 (3) (2008) 67–73.

[2] K. Aziz, G.W. Govier, Pressure drop in wells producing oil and gas, J. Can. Pet. Technol. 11 (1972) PETSOC–72–03–04, http://dx.doi.org/10.2118/72-03-04.

[3] E. Eltahan, R. Ganjdanesh, W. Yu, K. Sepehrnoori, R. Williams, J. Nohavitsa, Machine learning approach to improve calculated bottom-hole pressure, in: Proceedings of the 9th Unconventional Resources Technology Conference, in: URTeC 2021, American Association of Petroleum Geologists, 2021, http://dx.doi.org/10.15530/urtec-2021-5645.

[4] D. Molinari, S. Sankaran, Merging physics and data-driven methods for field-wide bottomhole pressure estimation in unconventional wells, in: Unconventional Resources Technology Conference, Houston, Texas, 26–28 July 2021, 2021, D031S074R004, http://dx.doi.org/10.15530/urtec-2021-5261.

[5] E. Kanin, A. Osiptsov, A.L. Vainshtein, E. Burnaev, A predictive model for steady-state multiphase pipe flow: Machine learning on lab data, J. Pet. Sci. Eng. 180 (2019) 727–746, http://dx.doi.org/10.1016/j.petrol.2019.05.055.

[6] A.E. Dukler, M.G. Hubbard, A model for gas-liquid slug flow in horizontal and near horizontal tubes, Ind. Eng. Chem. Fundam. 14 (4) (1975) 337–347.

[7] M.A. Fernandes, E. Gildin, M.A. Sampaio, Data-driven estimation of flowing bottom-hole pressure in petroleum wells using long short-term memory, in: 2024 International Conference on Machine Learning and Applications, ICMLA, 2024, pp. 1530–1537, http://dx.doi.org/10.1109/ICMLA61862.2024.00236.

[8] A.A. Al-Ghamdi, R.N. Gajbhiye, Development of an AI model to estimate flowing bottom-hole pressure in high-pressure high-temperature gas well, in: International Petroleum Technology Conference, in: 24IPTC, IPTC, 2024, http://dx.doi.org/10.2523/iptc-23664-ms.

[9] M. Jin, H. Emami-Meybodi, An integrated machine learning algorithm for unconventional flowing bottomhole pressure prediction during dynamic gas lift operation, SPE J. 30 (05) (2025) 2726–2737, http://dx.doi.org/10.2118/222222-PA.

[10] E. Nnaemeka, V. M. Chinyereugo, Predictive modelling of flowing bottom hole pressure in sandstone reservoir formations using machine learning, in: SPE Nigeria Annual International Conference and Exhibition, in: 25NAIC, SPE, 2025, http://dx.doi.org/10.2118/228717-ms.

[11] M. Jin, H. Emami-Meybodi, Prediction of flowing bottomhole pressure in gas-lifted wells using LSTM-ANN models, Energy Fuels 39 (31) (2025) 14992–15002.

[12] B. Abdullahi, M. Ezeh, Production optimization in oil and gas wells: A gated re-current unit approach to bottom hole flowing pressure prediction, in: SPE Nigeria Annual International Conference and Exhibition, SPE, 2024, D021S007R004.

[13] O.E. Agwu, S. Alatefi, A. Alkouh, R.R. Suppiah, Modelling the flowing bottom hole pressure of oil and gas wells using multivariate adaptive regression splines, J. Pet. Explor. Prod. Technol. 15 (2) (2025) 22.

[14] D. Campos, D.D.K. Wayo, R.B. De Santis, D.A. Martyushev, Z.M. Yaseen, U.I. Duru, C.M. Saporetti, L. Goliatt, Evolutionary automated radial basis function neural network for multiphase flowing bottom-hole pressure prediction, Fuel 377 (2024) 132666.

[15] L. Goliatt, R.S. Mohammad, S.I. Abba, Z.M. Yaseen, Development of hybrid computational data-intelligence model for flowing bottom-hole pressure of oil wells: New strategy for oil reservoir management and monitoring, Fuel 350 (2023) 128623, http://dx.doi.org/10.1016/j.fuel.2023.128623.

[16] A. Lasisi, N. Attoh-Okine, Machine learning ensembles and rail defects prediction: Multilayer stacking methodology, ASCE-ASME J. Risk Uncertain. Eng. Syst. Part A: Civ. Eng. 5 (4) (2019) 04019016, http://dx.doi.org/10.1061/AJRUA6.0001024.

[17] X. Wang, T. Han, Transformer fault diagnosis based on stacking ensemble learning, IEEJ Trans. Electr. Electron. Eng. 15 (12) (2020) 1734–1739, http://dx.doi.org/10.1002/tee.23247.

[18] S.A. El-Shorbagy, W.M. El-Gammal, W.M. Abdelmoez, Using SMOTE and heterogeneous stacking in ensemble learning for software defect prediction, in: Proceedings of the 7th International Conference on Software and Information Engineering, ICSIE '18, Association for Computing Machinery, New York, NY, USA, 2018, pp. 44–47, http://dx.doi.org/10.1145/3220267.3220286.

[19] N. Kardani, A. Zhou, M. Nazem, S.-L. Shen, Improved prediction of slope stability using a hybrid stacking ensemble method based on finite element analysis and field data, J. Rock Mech. Geotech. Eng. 13 (1) (2021) 188–201, http://dx.doi.org/10.1016/j.jrmge.2020.05.011.

[20] D.-N. Truong, J.-S. Chou, Fuzzy adaptive jellyfish search-optimized stacking machine learning for engineering planning and design, Autom. Constr. 143 (2022) 104579, http://dx.doi.org/10.1016/j.autcon.2022.104579.

[21] R. Huang, C. Wei, B. Wang, J. Yang, X. Xu, S. Wu, S. Huang, Well performance prediction based on long short-term memory (LSTM) neural network, J. Pet. Sci. Eng. 208 (2022) 109686, http://dx.doi.org/10.1016/j.petrol.2021.109686.

[22] N.M. Ibrahim, A.A. Alharbi, T.A. Alzahrani, A.M. Abdulkarim, I.A. Alessa, A.M. Hameed, A.S. Albabtain, D.A. Alqahtani, M.K. Alsawwaf, A.A. Almuqhim, Well performance classification and prediction: Deep learning and machine learning long term regression experiments on oil, gas, and water production, Sensors 22 (14) (2022) 5326, http://dx.doi.org/10.3390/s22145326.

[23] F.I. Syed, S. Alnaqbi, T. Muther, A.K. Dahaghi, S. Negahban, Smart shale gas production performance analysis using machine learning applications, Pet. Res. 7 (1) (2022) 21–31, http://dx.doi.org/10.1016/j.ptlrs.2021.06.003.

[24] C.C. Nwanwe, U.I. Duru, An adaptive neuro-fuzzy inference system white-box model for real-time multiphase flowing bottom-hole pressure prediction in wellbores, Petroleum 9 (4) (2023) 629–646, http://dx.doi.org/10.1016/j.petlm.2023.03.003.

[25] L. Goliatt, C. Saporetti, E. Pereira, Super learner approach to predict total organic carbon using stacking machine learning models based on well logs, Fuel 353 (2023) 128682, http://dx.doi.org/10.1016/j.fuel.2023.128682.

[26] M. Zolfagharroshan, E. Khamehchi, Accurate artificial intelligence-based methods in predicting bottom-hole pressure in multiphase flow wells, a comparison approach, Arab. J. Geosci. 14 (2021) 284, http://dx.doi.org/10.1007/s12517-021-06661-y.

[27] H. Zhang, Y. Ren, Y. Zhang, S. Zheng, Intelligent prediction method for fracture pressure based on stacking ensemble algorithm, Geomech. Geophys. Geo-Energy Geo-Resources 9 (1) (2023) 149, http://dx.doi.org/10.1007/s40948-023-00690-5.

[28] S.A. Marfo, S. Asante-Okyere, Y.Y. Ziggah, A new flowing bottom hole pressure prediction model using M5 prime decision tree approach, Model. Earth Syst. Environ. 8 (2) (2021) 2065–2073, http://dx.doi.org/10.1007/s40808-021-01211-7.

[29] A. Shammari, Prediction of Pressure Drop for Two-Phase Flow in Vertical Pipes Using Artificial Intelligence (Master's Thesis), King Fahd University of Petroleum and Minerals, 2011, Supervisor: Muhammad Al-Marhoun.

[30] T. Cover, P. Hart, Nearest neighbor pattern classification, IEEE Trans. Inform. Theory 13 (1) (1967) 21–27, http://dx.doi.org/10.1109/tit.1967.1053964.

[31] X. Su, X. Yan, C.-L. Tsai, Linear regression, Wiley Interdiscip. Rev.: Comput. Stat. 4 (3) (2012) 275–294, http://dx.doi.org/10.1002/wics.1198.

[32] Q.-Y. Zhu, A. Qin, P. Suganthan, G.-B. Huang, Evolutionary extreme learning machine, Pattern Recognit. 38 (10) (2005) 1759–1763, http://dx.doi.org/10.1016/j.patcog.2005.03.028.

[33] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32, http://dx.doi.org/10.1023/a:1010933404324.

[34] V. Vapnik, Estimation of Dependences Based on Empirical Data, Springer New York, 2006, http://dx.doi.org/10.1007/0-387-34239-7.

[35] I.D. Mienye, Y. Sun, A survey of ensemble learning: Concepts, algorithms, applications, and prospects, IEEE Access 10 (2022) 99129–99149, http://dx.doi.org/10.1109/ACCESS.2022.3207287.

[36] N. Liu, H. Gao, Z. Zhao, Y. Hu, L. Duan, A stacked generalization ensemble model for optimization and prediction of the gas well rate of penetration: a case study in Xinjiang, J. Pet. Explor. Prod. Technol. 12 (6) (2021) 1595–1608, http://dx.doi.org/10.1007/s13202-021-01402-z.

[37] K.-L. Du, B. Jiang, J. Lu, J. Hua, M. Swamy, Exploring kernel machines and support vector machines: Principles, techniques, and future directions, Mathematics 12 (24) (2024) 3935.

[38] R. Isnaeni, S. Sudarmin, Z. Rais, Analisis support vector regression (Svr) dengan kernel radial basis function (Rbf) untuk memprediksi laju inflasi di Indonesia, VARIANSI: J. Stat. Appl. Teach. Res. 4 (1) (2022) 30–38.

[39] R. Gupta, S.J. Nanda, Solving dynamic many-objective TSP using NSGA-III equipped with SVR-RBF kernel predictor, in: 2021 IEEE Congress on Evolutionary Computation, CEC, IEEE, 2021, pp. 95–102.

[40] S. Young, T. Abdou, A. Bener, Deep super learner: A deep ensemble for classification problems, in: E. Bagheri, J.C. Cheung (Eds.), Advances in Artificial Intelligence, Springer International Publishing, Cham, 2018, pp. 84–95, http://dx.doi.org/10.1007/978-3-319-89656-4_7.

[41] W. McKinney, et al., Pandas: a foundational Python library for data analysis and statistics, Python High Perform. Sci. Comput. 14 (9) (2011) 1–9.

[42] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, Scikit-learn: Machine learning in Python, J. Mach. Learn. Res. 12 (85) (2011) 2825–2830, URL: http://jmlr.org/papers/v12/pedregosa11a.html.

[43] L. Goliatt, Z.M. Yaseen, Development of a hybrid computational intelligent model for daily global solar radiation prediction, Expert Syst. Appl. 212 (2023) 118295, http://dx.doi.org/10.1016/j.eswa.2022.118295.

[44] R.M. El-Saghier, M. Abu El Ela, A. El-Banbi, A model for calculating bottom-hole pressure from simple surface data in pumped wells, J. Pet. Explor. Prod. Technol. 10 (5) (2020) 2069–2077, http://dx.doi.org/10.1007/s13202-020-00855-y.

[45] M. Jin, H. Emami-Meybodi, M. Ahmadi, Flowing bottomhole pressure during gas lift in unconventional oil wells, SPE J. 29 (05) (2024) 2432–2444, http://dx.doi.org/10.2118/214832-PA.

[46] S.M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, in: I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), in: Advances in Neural Information Processing Systems, vol. 30, Curran Associates, Inc., 2017, URL: https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf.

[47] B. Izgec, A.R. Hasan, D. Lin, C.S. Kabir, Flow-rate estimation from wellhead-pressure and temperature data, SPE Prod. Oper. 25 (01) (2009) 31–39, http://dx.doi.org/10.2118/115790-PA.

[48] S. Hari, S. Krishna, M. Patel, P. Bhatia, R.K. Vij, Influence of wellhead pressure and water cut in the optimization of oil production from gas lifted wells, Pet. Res. 7 (2) (2022) 253–262, http://dx.doi.org/10.1016/j.ptlrs.2021.09.008.

[49] Z. Tariq, et al., Real-time prognosis of flowing bottom-hole pressure in a vertical well for a multiphase flow using computational intelligence techniques, J. Pet. Explor. Prod. Technol. (2020) http://dx.doi.org/10.1007/s13202-019-0728-4.

[50] Z. Song, G. Han, Z. Ren, H. Su, S. Jia, T. Cheng, M. Li, J. Liang, Research on oil–water two-phase flow patterns in wellbore of heavy oil wells with medium-high water cut, Processes 12 (11) (2024) 2404, http://dx.doi.org/10.3390/pr12112404.

[51] X. Zhou, et al., Determination of reasonable bottom-hole pressure in unconventional tight oil reservoirs: A field case investigation, Lithosphere 2022 (12) (2022) 1–15, http://dx.doi.org/10.2113/2022/2007568.
[52] R. Chen, et al., Influence factors of the bottom hole flow pressure in a high-pressure and high-temperature condensate gas reservoir: Applicability analysis, Front. Energy Res. 10 (2022) 857354, http://dx.doi.org/10.3389/fenrg.2022.857354.
[53] S.M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, Adv. Neural Inf. Process. Syst. 30 (2017).
[54] S. Cui, Y. Yin, D. Wang, Z. Li, Y. Wang, A stacking-based ensemble learning method for earthquake casualty prediction, Appl. Soft Comput. 101 (2021) 107038, http://dx.doi.org/10.1016/j.asoc.2020.107038.
[55] J. Shi, C. Li, X. Yan, Artificial intelligence for load forecasting: A stacking learning approach based on ensemble diversity regularization, Energy 262 (2023) 125295, http://dx.doi.org/10.1016/j.energy.2022.125295.
[56] J. Figueroa, P. Baraldi, I. Chouybat, F. Ursini, E. Vignati, E. Zio, Estimation of real-time bottomhole parameters in CO2 injection wells during operations by means of an ensemble of neural networks, in: SPE Europe Energy Conference and Exhibition, in: 24EURO, SPE, 2024, http://dx.doi.org/10.2118/220048-ms.
[57] S. Bian, W. Wang, On diversity and accuracy of homogeneous and heterogeneous ensembles, Int. J. Hybrid Intell. Syst. 4 (2) (2007) 103–128.
[58] W. Wang, P. Jones, D. Partridge, Diversity between neural networks and decision trees for building multiple classifier systems, in: Multiple Classifier Systems, Springer Berlin Heidelberg, Berlin, Heidelberg, 2000, pp. 240–249, http://dx.doi.org/10.1007/3-540-45014-9_23.
[59] J.R. Rice, The algorithm selection problem, in: M. Rubinoff, M.C. Yovits (Eds.), in: Advances in Computers, vol. 15, Elsevier, 1976, pp. 65–118, http://dx.doi.org/10.1016/S0065-2458(08)60520-3.
[60] H.J.P. Weerts, A.C. Mueller, J. Vanschoren, Importance of tuning hyperparameters of machine learning algorithms, 2020, http://dx.doi.org/10.48550/ARXIV.2007.07588.
[61] S. García, J. Luengo, F. Herrera, Data Preprocessing in Data Mining, Springer International Publishing, 2015, http://dx.doi.org/10.1007/978-3-319-10247-4.
[62] G. Seni, J.F. Elder, Ensemble Methods in Data Mining: Improving Accuracy Through Combining Predictions, Springer International Publishing, 2010, http://dx.doi.org/10.1007/978-3-031-01899-2.
[63] E. Kasuya, On the use of r and r squared in correlation and regression, Ecol. Res. 34 (1) (2019) 235–236, http://dx.doi.org/10.1111/1440-1703.1011.
[64] N.O. Nikitin, I. Revin, A. Hvatov, P. Vychuzhanin, A.V. Kalyuzhnaya, Hybrid and automated machine learning approaches for oil fields development: The case study of volve field, north sea, Comput. Geosci. 161 (2022) 105061.
[65] N.O. Nikitin, M. Pinchuk, V. Pokrovskii, P. Shevchenko, A. Getmanov, Y. Aksenkin, I. Revin, A. Stebenkov, V. Latypov, E. Poslavskaya, et al., Integration of evolutionary automated machine learning with structural sensitivity analysis for composite pipelines, Knowl.-Based Syst. 302 (2024) 112363.
[66] M.T. Ribeiro, S. Singh, C. Guestrin, Why should I trust you? explaining the predictions of any classifier, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 1135–1144.
[67] Z. Xu, H. Liu, J. Sun, et al., A drilling wellbore pressure calculation model considering the effect of gas dissolution and suspension, Front. Earth Sci. (2022) http://dx.doi.org/10.3389/feart.2022.993876.
[68] X. Chen, Y. Zhang, H. Li, et al., Further study on oil/water relative permeability ratio model and waterflooding performance prediction model for high water-cut oilfields sustainable development, J. Pet. Explor. Prod. Technol. (2024) http://dx.doi.org/10.1007/s13202-024-01753-3.
[69] M.A.A. El-Moniem, A.H. El-Banbi, Effects of production, PVT and pipe roughness on multiphase flow correlations in gas wells, J. Pet. Explor. Prod. Technol. (2020) http://dx.doi.org/10.1007/s13202-020-00944-y.
[70] S. Beheshtian, S.K. Roodbari, H. Ghorbani, M. Azodinia, M. Mudabbir, A.R. Varkonyi-Koczy, Comparative evaluation of machine learning and Bayesian deep learning methods for estimating ultimate recovery in shale well reservoirs, in: 2024 IEEE 11th International Conference on Computational Cybernetics and Cyber-Medical Systems, ICCC, 2024, pp. 000017–000024, http://dx.doi.org/10.1109/ICCC62278.2024.10583113.
[71] S. Beheshtian, S.K. Roodbari, H. Ghorbani, M. Azodinia, M. Mudabbir, A.R. Varkonyi-Koczy, Advanced machine learning methods for accurate prediction of loss circulation in drilling well log, in: 2024 IEEE 11th International Conference on Computational Cybernetics and Cyber-Medical Systems, ICCC, 2024, pp. 000031–000036, http://dx.doi.org/10.1109/ICCC62278.2024.10582962.
[72] X. Lu, Z. Wang, M. Zhao, S. Peng, S. Geng, H. Ghorbani, Data-driven insights into climate change effects on groundwater levels using machine learning, Water Resour. Manag. 39 (7) (2025) 3521–3536, http://dx.doi.org/10.1007/s11269-025-04120-x.