Energy 264 (2023) 126122

Contents lists available at ScienceDirect

Energy

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

Evaluation of proppant injection based on a data-driven approach integrating numerical and ensemble learning models

Lei Hou [a,b], Derek Elsworth [c,*], Fengshou Zhang [d], Zhiyuan Wang [e], Jianbo Zhang [e]

[a] China-UK Low Carbon College, Shanghai Jiao Tong University, Shanghai 201306, China
[b] School of Engineering, The University of Warwick, Coventry, CV4 7AL, UK
[c] Energy and Mineral Engineering & Geosciences, EMS Energy Institute and G3 Center, Pennsylvania State University, University Park, 16802, USA
[d] Department of Geotechnical Engineering, College of Civil Engineering, Tongji University, Shanghai, 200092, China
[e] School of Petroleum Engineering, China University of Petroleum (East China), Qingdao, 266580, China

ARTICLE INFO

Keywords:
Hydraulic fracturing
Proppant injection
Fracture evolution
Machine learning
Field application

ABSTRACT

Injecting proppant to prop open fluid-driven fractures in subsurface reservoirs is one of the key missions of hydraulic fracturing. However, quantitative evaluation of the distribution of successfully propped fractures is limited due to the infeasibility of direct measurement. This work defines an indexing parameter for field practice to estimate the proportion of proppant-filled fractures in the reservoir – the proppant filling index (PFI). A new data-driven workflow, combining numerical models and an ensemble learning algorithm, is proposed and trained on field records of both screen-out and near screen-out cases and is then applied to predict PFIs for regular cases. The algorithm performance is promoted *via* variable importance measure (VIM) analyses and a backward elimination strategy. Four screen-out and twelve regular cases are presented to demonstrate the predicted PFI and highlight its potential utilizations. The new PFI and workflow evaluate the proppant injection quantitatively and reveal any mismatch between proppant injection and underground fractures, which may be essential for post-fracturing analyses and reservoir characterization to improve both oil & gas recovery, the sequestration of CO_2, storage then recovery of H_2 and the recovery of deep geothermal fluids as important components in enabling the energy transition.

1. Introduction

Proppant injection is one of the key objectives of hydraulic fracturing – to establish sufficient permeability in artificially-created fracture networks, thus providing fluid access to the stimulated reservoir and enhancing production [1–3]. This is critical for unconventional reservoirs that lack initial permeability to fluids and is where low-viscosity proppant-laden fracturing fluids are applied. The slurries carry proppant far into the created fracture network as fluid overpressures drive the propagation of new artificial fractures [4,5]. The proppant settles within the fractures and is typically unevenly distributed and indeed may cause screenout – a situation where the proppant bridges the fracture and blocks flow to the advancing tip – thereby stalling the advancing fracture and potentially creating hazardous fluid overpressures [6–8]. Therefore, the evaluation of proppant injection is crucial for post-fracturing analysis, optimization of pumping schedules, maximization of proppant injection, and production enhancement. Dynamic numerical models are built to couple the proppant distribution and fracture propagation in a single fracture during the fracturing operation [9–11]. However, proppant injection is still difficult to quantify at field scales due to the difficulties in the direct detection of millimeter-scale fractures and micrometer-scale proppant from thousands of meters above the underground reservoir [12,13]. The field engineers usually use the injected proppant volume per stage length (also known as the proppant intensity) to compare the proppant injection among different stages from the same well under similar geological and operating conditions [14,15]. Other qualitative post-fracturing analysis approaches include the G-function-based/before-closure analyses that evaluate the injection indirectly based on the parameters of ISIP (instant shut-in pressure), closure pressure, fluid efficiency, and other parameters [16–19]. Core drilling through hydraulically fractured shale reservoirs has been used to spot the proppant packing, distribution, and embedment but this is a non-routine and exotic method of specialist post-mortem analysis [20,21], which is difficult to justify due

* Corresponding author.
E-mail addresses: leihou.upc@gmail.com (L. Hou), elsworth@psu.edu (D. Elsworth).

https://doi.org/10.1016/j.energy.2022.126122
Received 13 July 2022; Received in revised form 1 November 2022; Accepted 15 November 2022
Available online 16 November 2022
0360-5442/© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

to the high cost. Regardless, all of these methods have difficulty in evaluating how the injected proppant matches with the fractures in the reservoir.

Current data science analyses have extracted valuable information about proppant injection from the sand screen-out case [22,23] – an extreme injection situation when a sharp increase of pressure is encountered and the fracking operation is suspended. Most of the screen-out cases, occurring during the late period of fracturing, are induced by the injected proppant jamming across the fractures and exceeding the downstream carrying capacity of the evolving fractures [24,25]. Although many machine learning models have been built to interpret field records and to make predictions of screen-out [26–28], few of these experiences can be directly extended to evaluate the regular fracturing case. For one reason, the prediction of screen-out is usually treated as a classification of isolated incidents, while proppant transport evaluation requires a continuous quantity output [29]. Moreover, the common nature of most machine learning algorithms (the consistency between the categories of training data and predicted data) determines that their performances are highly reliant on the training datasets – namely an algorithm trained with screen-out data should mainly be restricted to predicting screen-out, and thus be inappropriate for regular cases. Proppant injection, however, is currently difficult to quantify for regular cases, forming a chicken-or-egg causality dilemma when we are trying to use data-driven methods for quantification. Therefore, a universal parameter determined to represent both screen-out and regular cases may extend the training datasets, thus boosting the application of trained algorithms for regular cases.

This work establishes a data-driven workflow and produces a real-time evaluation of proppant injection, towards improvement in the efficiency of transporting proppant into fractures, propping fractures, and the characterization of artificially-fractured reservoirs. Machine learning algorithms trained on screen-out cases are applied to evaluate proppant injection for regular cases. To accomplish this, a new comprehensive variable – the proppant filling index (PFI) – is defined based on the determining-element mechanism [29], which is adequate to estimate the proportion of proppant-filled fractures for both screen-out and regular cases. The PFI is used to label the training datasets containing screen-out and near screen-out (defined as the condition when rapid pressure rising is encountered and the pump rate has to be reduced to complete the remaining operation) cases. An ensemble-machine-learning workflow is built to extract and learn experiences from the training cases and then predict the PFI for regular cases. Numerical models, including fluid efficiency, stratified proppant flow, and bottom pressure, together with other parameters are integrated into the workflow for data pre-processing. Variable importance measure (VIM) analysis is applied to optimize input features and enhance the performance of the workflow, which also mitigates the "black-box" effect of deep learning. The new PFI and workflow produce a continuous and quantitative evaluation of proppant injection for regular fracturing operations compared with previous efforts focusing on discrete predictions of screen-out events, which can be essential for optimizing fracturing schedules, enhancing ESRV (the effective stimulated reservoir volume) and defining anticipated production. Moreover, the interpretation of PFI curves (evaluating the mismatch between proppant injection and underground fractures) may also improve the characterization of artificially fractured reservoirs – a core and generic technology for the geological sequestration of CO_2 [30,31], storage of hydrogen and the recovery of deep geothermal fluids with clear application to the energy transition [32].

2. Methodology

A new workflow is proposed to evaluate proppant injection for field practice (Fig. 1), which integrates a newly defined proppant filling index (PFI), numerical modeling, and an ensemble machine learning model. Importantly, the PFI is universal to screen-out and regular cases. The numerical models extract essential features from the original data, which significantly improves the performance. The Gated Recurrent Unit (GRU) and Random Forest (RF) models are assembled using the average strategy for training and predicting. The root-mean-square error (RMSE) and coefficient of determination (R^2) are used as the dual criteria for evaluating the performance of the workflow.

2.1. Data preparation

This work collects 63 stages of shale gas fracturing records (summarized in Table 1) from the Sichuan basin, China, in which 27 stages

Table 1
Summary of the training, testing, and application datasets.

	Training/ stage	Testing/ stage	Application/ stage	Notes
Near Screen-out cases	25	2 (Wells A and C)	/	/
Screen-out cases	22	2 (Wells B and D)	/	/
Regular cases	/	/	5 (Well E) 5 (Well F)	For comparisons with proppant intensity
			1 (Well G)	Pressure descending
			1 (Well H)	Pressure ascending

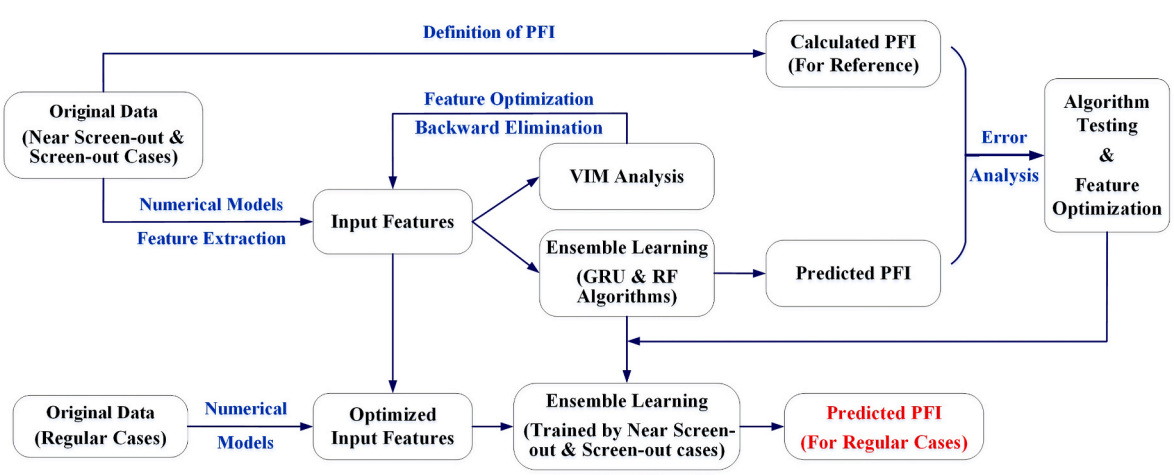

Fig. 1. The data-driven workflow for the evaluation of proppant injection.

are near screen-out cases, 24 stages are screen-out cases and 12 stages are regular cases. The testing set, for evaluating the performance of the workflow, consists of two screen-out and two near-screen-out cases. Two groups of regular cases are collected from Wells E and F, with each group containing five randomly selected stages. The stages from the same well are designed to control the geological and engineering uncertainty, in order to compare PFI with the proppant intensity. The remaining application cases exert characteristics of pressure ascending and descending, respectively.

The collected parameters include the geological features (well depth, vertical depth, minimum horizontal stress, pore pressure) and treatment/fracturing records (stage number, pump rate, fluid and proppant type, wellhead pressure, proppant concentration, and stage length). The original fluid and proppant types are non-numeric parameters and are replaced by the fluid viscosity and proppant diameter, as representative parameters of performance. The scale of fracturing of the collected cases is ~2000 m^3 on average for fracturing fluid (mainly slickwater) and 100 m^3 on average for proppant (mainly 40/70 size). Notably, for the screen-out cases, the incidents occur principally in the middle or at the end of fracturing operations when a certain amount of proppant has been injected. Therefore, the proppant accumulation in fractures becomes the main reason for screen-out.

2.2. Definition and derivation of PFI

The PFI is defined as the volume proportion of proppant-filled fracture of the total proppant-accessible fractures under certain hydraulic injecting conditions. It is a function of pump rate, proppant accumulation, and fracture volume/capacity according to the determining element of screen-out [29].

$$PFI = f\left(Q, V_p, V_{eff}, C_x\right) \tag{1}$$

where Q is pump rate; V_p is the injected proppant volume; V_{eff} is the effective fracture volume that allows the proppant flow under the pump rate Q; and C_x is the complexity of fracture networks.

Assuming the fracture networks are constant during the proppant injection, the PFI will have a linear correlation with V_p/V_{eff}. The general expression of PFI for both screen-out and regular cases is given by

$$PFI = \alpha \frac{V_p}{V_{eff}} C_x \tag{2}$$

where α is the coefficient and is a function of the pump rate

$$\alpha = \frac{1}{2}\left(1 + \frac{\Delta Q}{Q_i}\right) \times 100 \tag{3}$$

where ΔQ is the pump rate reduction when screen-out or near screen-out occurs. Q_i is the initial pump rate before adjustments. For screen-out cases, the pump usually is urgently shut-in to control the pressure and ensure the safety of people and equipment on the surface, where ΔQ equals Q_i.

The presence and location of underground fractures are currently difficult to detect, measure, and classify [33]. Therefore, the complexity of the fracture network (C_x) is simplified in this study and assigned a value of unity. For screen-out and near-screen-out cases, the effective fractures are near-completely filled by the injected proppant, leaving a limited channel for flow in the fractures, thus causing pressure to climb sharply. The effective fracture volume (V_{eff}) is approximately equal to the maximum injection volume of proppant (V_{pmax}) under the corresponding injecting conditions. Therefore, the PFI for screen-out and near-screen-out cases is calculated as

$$PFI = \frac{1}{2}\left(1 + \frac{\Delta Q}{Q_i}\right) \frac{V_p}{V_{p\,\max}} \times 100 \tag{4}$$

Remarkably, Eq. (4) is not suitable for regular cases where the injection volume of proppant does not approach the volume of the effective fractures (V_{eff} can not be replaced by V_{pmax}). The PFI of the regular case is predicted by the machine learning algorithm that is trained by the PFIs of screen-out and near-screen-out cases. According to the expression, the value of PFI mainly varies between 0 and 100. The higher the PFI value, the more fully the fractures are filled by the proppant. However, higher PFI also indicates a higher risk of sand screen-out.

The PFI estimates the pumpability of the proppant before screenout under given geological and engineering conditions. It evaluates the mismatch between proppant injection and underground fractures and thus can also be used for characterizing the fractured reservoir. The PFI reflects the proportion of effective fractures remaining open after injection, which can be significant for the evaluation of the ESRV, simulation of cracked reservoirs, optimization of production schedules, and design of re-fracturing.

2.3. Feature extraction by numerical models

Proppant-flowing related and fracture-volume related models, referring to the determining-element mechanism, are used for feature extractions during the data pre-processing. The ratio of injected proppant and fluid volumes (V_p/V_f) and the height of the slurry flowing layer (H_1) are the resulting parameters of the proppant-flowing models. The H_1 is calculated by the Bi-power correlations [34,35], as shown in Appendix A. The fracture volume is estimated indirectly by pressure interpretations and fluid efficiency. The slope of the wellhead pressure (ΔP_s) and conversion of the bottom pressure (P_b, in Appendix A) are calculated based on previous characterizations [36–38]. The fluid efficiency (η, in Appendix A) is obtained by the G-function analysis of the pressure drop after pump shut-in Ref. [18].

There are 11 original parameters and 5 extracted features used as inputs for training the machine learning algorithms, as summarized in Table 2. Eq. (4) and other equations in Appendix A are used to calculate PFI and extracted features. The calculations are carried out using the second-level field records from 63 fracturing stages, which involve more than 610,000 groups of data.

2.4. Ensemble machine learning model

The Gated Recurrent Unit (GRU) and Random Forest (RF) models are used for data processing based on previous experience [7,39,40]. The GRU model is optimized by grid search and walk-forward validation techniques [41,42], based on which a three-layer model (including the output layer) is established with 100 neural units on the first and hidden layers (Appendix B). A drop-out (drop rate is 0.2) layer is set after the first and hidden layers to avoid overfitting. The optimization of hyper-parameters is presented in Appendix B (Table B1), including the batch size (200), epoch (30), activation function ('ReLu'), and optimizer ('Adam') [43,44]. Besides, a Random Forest regression model with 50 estimators is also built to improve the prediction [45]. The GRU and RF models are assembled based on the average strategy [46], in which the final prediction is the average result of the GRU and RF productions

Table 2
Summary of input and output features of the data processing workflow.

Inputs		Output
Original Features	Extracted Features	
well depth; vertical depth	V_p/V_f	PFI
minimum horizontal stress; pore pressure	H_1	
stage number; stage length	P_b	
pump rate; wellhead pressure	ΔP_s	
fluid viscosity; proppant diameter and concentration	η	

trained by the entire training dataset respectively.

The new data-driven workflow is then established based on the ensemble learning model and the definition of PFI, consisting of data collection, feature extraction, and optimization, algorithm training, testing, and application, as shown in Fig. 1. The RMSE and R^2 are used as the dual criteria for error analyses. The contribution of each feature to the prediction is estimated by the variable importance measure (VIM) [47], based on which the extracted features are optimized using a backward elimination strategy. The optimized features and trained algorithm, based on the near screen-out and screen-out cases, are then applied to predict the PFI for regular cases (Fig. 1). Notably, the application of this workflow may be restricted by the data consistency. The workflow trained by data from a specific field will exhibit the best performance in the same region as the data source. For the application in a different region, the differences in formations, geological conditions, and operators could bring uncertainties in the predictions.

3. Results

By averaging the outputs of the GRU and RF algorithms, the predicted PFIs for testing cases are significantly improved and are consistent with the reference values. The extracted features are optimized based on the VIM analyses and backward elimination strategy, which boosts the computational efficiency and mitigates the "black-box" effect of GRU. The workflow is then deployed for twelve regular cases from four different wells to evaluate the proppant injection. Valuable evaluations of proppant injection are obtained by interpreting the evolution of PFIs.

3.1. Performance of the algorithms

The testing cases are collected from Wells A-D, containing two screen-out and two near-screen-out cases, summarized in Table 3. According to the averaged RMSE and R^2, the ensemble algorithm reduces the RMSE by 35.7%–45.2% and increases the R^2 by 70.8%–92.9% compared with the predictions produced by a single GRU or RF algorithm.

3.2. Feature analyses and optimization

The contribution of each feature (Table 2) to the prediction is estimated based on the VIM analyses using the variance importance and the permutation importance of the RF model. The variance importance computes how much each feature contributes to decreasing the weighted impurity, which is the variance for the regression trees [48]. The permutation importance measures feature importance by observing how the random re-shuffling of each predictor influences model performance [49]. The results are presented in Fig. 2, in which the sand-fluid ratio (V_p/V_f), fluid efficiency (η), and pump rate (Q) are at the top of the rankings. The outstanding importance of the V_p/V_f may be a consequence of its similarity to the PFI in the data characteristics (PFI is defined as the ratio of injected proppant volume and the maximum injection volume of proppant in Eq. (4)). The high importance of η proves the accurate interpretation of the fluid in fractures, which may characterize the underground fracture volume. The pump rate, one of the elements in the determining mechanism [29], exerts similar importance to fluid efficiency. Among the extracted features, importances of the ΔP_s and bottom pressure (P_b) are ranked way down the list.

Only the extracted features are optimized because they are able to be promoted by upgrading the corresponding calculations. The backward elimination strategy is performed by referring to the importance rankings in Fig. 2. The extracted feature is deleted in sequence from the model inputs in the lowest-importance order. Based on the same dataset, the training and testing are repeated when the input is updated to compare the averaged RMSE and R^2 with the reference group (based on all features). The elimination of the feature will proceed if the errors are improving, and will terminate if the errors increase. The elimination process is summarized in Table 4. Both the RMSE and R^2 are improved by eliminating ΔP_s and P_b. The process is terminated when the increase of RMSE and decrease of R^2 are observed. Therefore, ΔP_s and P_b are removed from the model inputs and H_1 is reserved for predictions.

Table 3
Performances of algorithms on the testing cases evaluated by the RMSE and R^2.

	GRU		RF		Ensemble (GRU + RF)		Notes
	RMSE	R^2	RMSE	R^2	RMSE	R^2	
Well A	22.31	−0.19	18.32	0.20	13.16	0.59	Near screen-out case
Well B	5.27	0.98	14.53	0.84	9.40	0.94	Screen-out case
Well C	17.67	0.51	25.29	−0.01	4.53	0.97	Near screen-out case
Well D	19.33	0.62	17.57	0.69	14.45	0.79	Screen-out case
Averaged	**16.14**	**0.48**	**18.93**	**0.43**	**10.38**	**0.82**	

3.3. Prediction of PFI for testing case

The testing results of PFI and original measurements are plotted in Fig. 3. The reference PFI (solid orange line) is the calculation using Eq. (4). The predicted PFI (dashed orange line) is the output of the ensemble learning workflow using the inputs in Table 2. Generally, the predicted PFI varies closely around the reference curve and shows similar trends to the calculated PFI, indicating that the ensemble learning model is well-trained and the performance of the workflow is reliable. The model provides more accurate predictions for Wells B and C. The predictions for Wells A and D may be more difficult regarding the near-flat pressure evolution during proppant injections, which are difficult to interpret. The underground conditions in Well C may be coincident with the deriving assumptions of Eq. (4), thus exerting the best-fitting results. The pressure is sensitive to proppant injection in Well B. The opened fractures may be narrow and complex, which are difficult to enter and fill. The lowest proppant concentration and shortest length of the proppant slug are observed in Fig. 3 (b), which results in lower predicted PFIs than the reference values.

The PFI calculation, Eq. (4), is derived based on the simplification of fracture complexity and the linear-correlation assumption (ignoring the fracture propagation during proppant injection). Therefore, the manually calculated PFI is a smooth step-rising curve related to the proppant injection (the solid orange curves in Fig. 3). It is worth noting that the simplification and assumption are only used for manually calculating the PFI. The PFI predicted by the machine learning workflow is produced without such presuppositions. Moreover, the geological features (stress, depth, etc.) and fluid efficiency (the ratio of fluid remaining in the fracture and total injected fluid – indirectly reflecting fractures) are used as inputs for predictions. Therefore, the predicted PFI can fluctuate by the evolution of fractures during proppant injection (the dashed orange curves in Fig. 3).

Consequently, interpreting the deviations between predicted PFI and references (as shown in Fig. 3) may be diagnostic of the underground evolutions of proppant flow and fractures. In Fig. 3 (a), the proppant injection for Well A starts from a long slug, which is considered to be radical for shale gas fracturing. The initially opened fractures are usually undeveloped and easily filled by the rapidly injected proppant. The predicted PFI rises and remains high thereafter, as shown in Fig. 3 (a). The net pressure in the proppant-packed fractures increases with continuous injection, which boosts the propagation of the initial fractures and then mitigates the increase in PFI. The sharp fluctuation at the end of fracturing may be caused by the 30/50 mesh proppant (the large

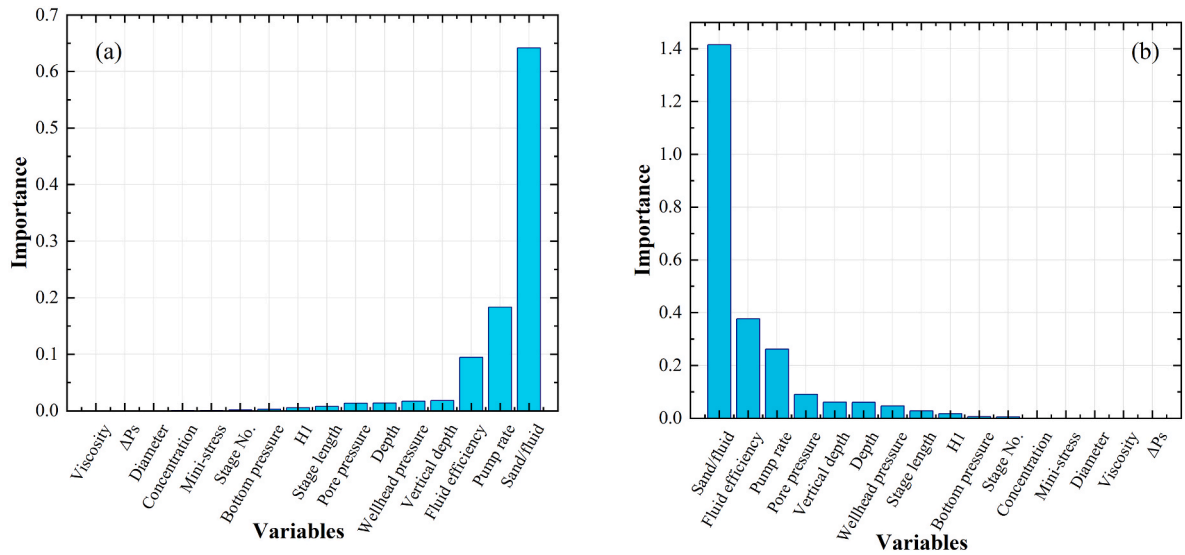

Fig. 2. The VIM analyses of the input variables based on (a) the variance importance and (b) the permutation importance.

Table 4
Optimization of extracted features using backward elimination.

	Reference (All features)		Step 1 (Eliminate ΔP_s)		Step 2 (Eliminate ΔP_s and P_b)		End of elimination (Eliminate ΔP_s, P_b, and H_1)	
	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2
Well A	13.16	0.59	14.41	0.50	12.02	0.65	13.55	0.56
Well B	9.40	0.94	9.28	0.94	11.55	0.90	10.07	0.93
Well C	4.53	0.97	3.46	0.98	4.06	0.97	5.36	0.95
Well D	14.45	0.79	13.59	0.81	11.97	0.86	13.79	0.81
Averaged	**10.38**	**0.82**	**10.18**	**0.81**	**9.90**	**0.85**	**10.69**	**0.81**

size of proppant that is not commonly used due to the narrow fracture width in shale reservoirs) injected in the last slug. A similar condition is observed in Well D. A large volume of fine (100 mesh) proppant in the first three slugs may fill and jam the minor fractures and boost the propagation of the main fractures, thus causing the fluctuation in the PFI curve at ~6500s, as shown in Fig. 3 (d). The sharp variation in PFI at the end of treatment may be induced by new fractures opening and filling. To summarize, the PFI prediction provides quantitative evaluations of proppant injection and qualitative perception of fracture evolution, which is significant for both post-fracture analysis and the characterization of fractured reservoirs.

3.4. Comparison between PFI and proppant intensity

Ten regular fracturing cases (Wells E−1 to E−5 and Wells F-1 to F-5) from two fracturing wells are evaluated for the PFI based on the new workflow. The averaged PFI after the last injection of the proppant slug and before pump-off is used as the final result and is compared with the proppant intensity. The proppant intensity assumes that all clusters are equally opened for the injection of slurry, which is not realistic. Therefore, fracturing stages from the same well are selected to control the geological and engineering uncertainties that cause the variation in the staged opening of clusters. A qualitative comparison between PFI and proppant intensity is presented in Fig. 4. The proppant intensity (the injected proppant volume per stage length) is often used in field operations to evaluate the efficiency of proppant injection among different stages of the same well or different wells in the same region [14,15]. The PFIs produced by the workflow show similar trends to the proppant intensity, thus demonstrating the reliability of the PFI. Moreover, the lower value of proppant intensity and higher value of PFI are observed for Well F. The proppant injection in Well F is still considered to be effective because the fracturing pressure is sensitive to proppant size and concentration, and 37% of the total proppant is 100 mesh in order to enhance the proppant intensity. The opened fractures in Well F may be narrower and more complex than those in Well E where 30/50 proppant is injected.

3.5. Evaluation of PFI for regular cases

Both pressure descending (Well G) and ascending (Well H) cases are optimized as additional application examples. The proppant injection in Well G is near-continuous, as shown in Fig. 5 (a). Although the pressure trend is declining, the PFI at the end of the operation exceeds a relatively high value of 80. It is considered to be a successful case because the PFI grows smoothly upwards to a high value while the pressure remains within a safe range, indicating that the fractures are efficiently filled with proppant. For Well H, the pressure is sensitive to the proppant injection, as shown in Fig. 5 (b). The PFI approaches 100 by the end of proppant injection, which is risky and may be near screen-out. The PFI curve remains low and grows smoothly before 7000s, then fluctuates. There is a jump in PFI between 6500s and 7000s induced by a slight increase in proppant concentration. Therefore, the lower proppant concentration and longer proppant injection slug may be an efficient strategy to control wellhead pressure and enhance the proppant intensity.

4. Discussion

4.1. Significance of the numerical models

Feature extraction using the numerical model plays an important role in the entire workflow, which is an essential data amplification method. The predictions based on both original features (Table 2) and all features are compared in Table 5. By introducing the extracted

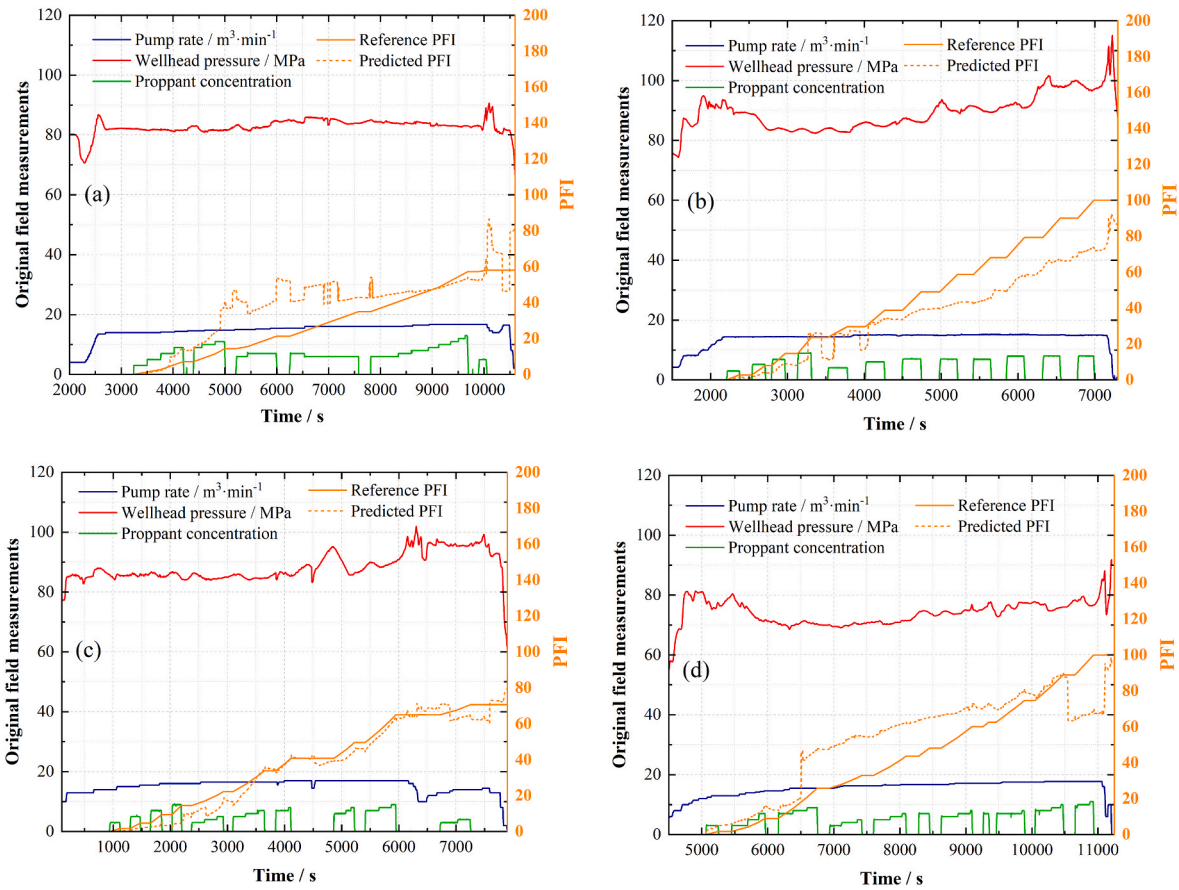

Fig. 3. Comparisons between the labeled PFIs and predictions based on testing cases of (a) Well A; (b) Well B; (c) Well C and (d) Well D. The original field measurements (pump rate – blue solid line, wellhead pressure – red solid line and proppant concentration – green solid line) are presented to show the fracturing process. The solid and dashed orange lines represent the reference (calculation result using Eq. (4)) and predicted (output of the ensemble learning workflow) PFIs, respectively.

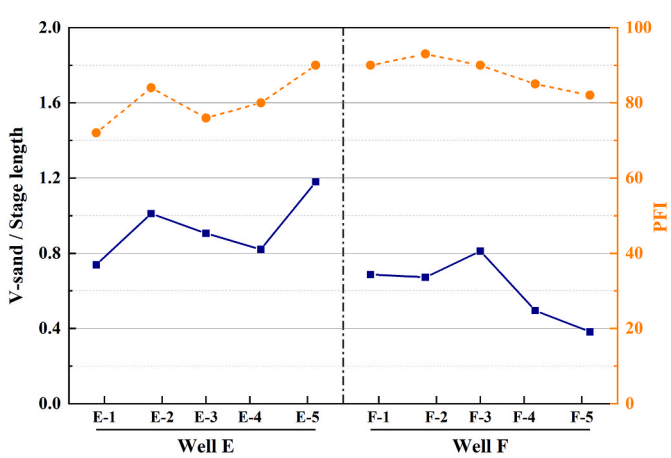

Fig. 4. Comparison between proppant intensity and PFI based on ten regular fracturing cases in Wells E and F. The bottom blue solid curve is the proppant intensity. The upper orange dashed curve is the PFI.

features, the averaged RMSE is reduced by 54.1%, and the averaged R^2 increases more than tenfold. These improvements are more significant than the gains made by algorithms in Table 3. Therefore, fundamental research is crucial, especially when a strong learning algorithm has been applied (e.g. the ensemble model in the current workflow). According to the VIM analyses and feature optimization, the contributions of the pressure-related calculations are deficient. This may be due to the flat pattern of the pressure variations, which is difficult to interpret (Figs. 2 and 3), thus requiring improvements to extract more valuable information.

4.2. Interpretation of the PFI curve

The interpretation of the PFI curve based on massive application cases could be crucial for both post-fracturing analysis and real-time adjustments. The evolution of the PFI curve reveals more clues about the downhole proppant transport during the entire process of the fracturing operation, including the effects of slug length, proppant concentration, and proppant size. The PFI combined with the total fracture volume measured by a micro-seismic detection may estimate the effective stimulated reservoir volume (ESRV). Besides, the slope of the PFI curve may be also an important indicator that aids the operator in real-time adjustments to enhance the proppant intensity and control the screen-out risk. For instance, the PFI slopes in Fig. 5 are approximately 0.0086 for Well F (considered as a successful case) and 0.0118 for Well G (considered as a risky case), respectively.

4.3. Remaining errors

The remaining errors in Table 4 may be reduced by improving the calculation of PFI and performance of the workflow in aspects of the estimation of fracture complexity and real-time propagation, the effect of proppant diameters, the size of training datasets, advanced numerical models, and new algorithms [50]. Noteworthy, collecting more data for algorithm training is a unique and efficient way to promote a

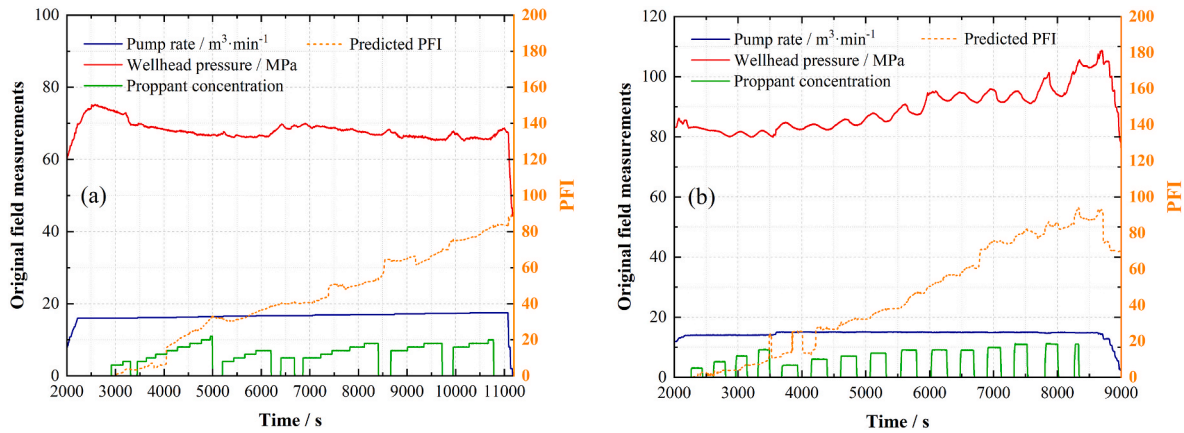

Fig. 5. PFI evaluation using the ensemble-learning workflow based on (a) Well G and (b) Well H. The original field measurements (pump rate – blue solid line, wellhead pressure – red solid line, and proppant concentration – green solid line) are presented to show the fracturing process. The dashed orange line represents the predicted PFI.

Table 5
Model performance based on original features and extracted features.

	Original Features		Original and Extracted Features	
	RMSE	R^2	RMSE	R^2
Well A	27.46	−0.80	13.16	0.59
Well B	14.61	0.84	9.40	0.94
Well C	30.75	−0.49	4.53	0.97
Well D	17.68	0.69	14.45	0.79
Averaged	**22.63**	**0.06**	**10.38**	**0.82**

data-driven workflow [51]. Due to the paucity of data, alternative modes of underground detection of proppant transport (core drilling, fiber optic sensors, etc.) are unavailable in this study. However, daily fracturing operations continually produce new data for the field operators. It is possible to upgrade the workflow by simply feeding new data [52]. Moreover, it is useful to try new input features based on experience and observations of the targeting field – namely the inputs of the workflow are customizable. The feature analyses and optimization methods presented in section 3.2 are helpful to further evaluate the new inputs and boost the performance of the data-driven workflow.

4.4. Limitations and implications

The limitations of this study are mainly in the necessary simplification of the problem and the resulting assumptions applied during the PFI derivation – each necessary due to the complexity of the influence factors. Fracture propagation during proppant injection and the complexity of the fracture networks are ignored when we manually calculate the PFI. This may be improved by accurate calculation of real-time fracture propagation and a strict description of the randomly generated fracture networks [53], which are beyond the scope of this work. Additionally, the PFI, according to its definition, remains zero before proppant injection (at the beginning of the fracturing operation when pure fluid is injected to fracture the formation with the fracture networks, as shown in Figs. 3 and 5). The data during this period, therefore, may contribute little to the predictions in this study. This segment of data is difficult to interpret because of the insufficiently recognized geological circumstances and the randomness of rock failure [54,55], thus requiring separate studies.

Notably, the performance of the new data-driven workflow highly depends on the consistency of the data sources between training and prediction. The application of a trained workflow is recommended to be restricted within the same region as the source of training data, in order to mitigate the uncertainties induced by geological and operational differences.

5. Conclusions

The mechanics of proppant injection is quantitatively defined through an evaluation of the proportion of proppant-filled fractures using a new workflow based on machine learning and numerical models. A total of 63 shale gas fracturing cases are collected for data processing, including 47 screen-out and near screen-out cases for algorithm training, 4 more cases for verification, and 12 regular cases for applications. A new data-driven workflow, integrating ensemble learning algorithms and numerical models, is established to process the field measurements and predict the PFI for regular fracturing cases. The predictions are boosted by optimizing the model inputs based on VIM analysis and backward elimination. A quantitative evaluation of proppant injection and qualitative perception of fracture evolution are defined for field practice, which are significant for post-fracturing analysis and characterizing artificially-fractured reservoirs. The major conclusions are:

(1) A new proppant filling index (PFI) is defined considering the effects of the pump rate, proppant accumulation, and fracture volume/capacity. It is a universal parameter that is adequate for both screen-out and regular fracturing cases. A higher PFI means more fractures filled by proppant, and also a higher risk of screen-out. By calculating the PFI, the experiences of screen-out and near screen-out cases are extracted for training the machine learning algorithms, then deployed to evaluate the proppant injection for regular cases. The evolution of predicted PFI reveals the dynamic matching relation between proppant injection and fracture propagation in underground reservoirs (Fig. 3).
(2) Ten regular fracturing cases are evaluated by PFI and the workflow, then compared with the proppant intensity, which results in a similar trend. Two more representative cases are analyzed for additional applications, in which the pressure descending case is considered to be successful regarding the smooth-growing PFI (approaching 80 by the end) and relatively safe pressure. The PFI of the pressure ascending case approaches 100, indicating a high risk of screen-out. Lower proppant concentration and longer injection slug are suggested by analyzing the PFI variation as a mitigation to reduce the hazard/risk. Interpreting the slope of the PFI may provide a basis for real-time adjustments.
(3) The advanced numerical models may determine the performance of the workflow, according to the higher contribution of the extracted features than that of improvement of the algorithm. Fundamental research is crucial, especially when a relatively strong learning algorithm has been employed. The pressure-

related calculations are mediocre based on the VIM analyses and feature optimization, especially for the flat pattern of pressure variations that may be difficult to interpret, thus requiring improvements in techniques. Other potential boosting approaches include the estimation of fracture complexity and real-time propagation, the effect of proppant size and concentration, new data for training, and new algorithms.

Credit author statement

Lei Hou: Conceptualization, Methodology, Investigation, Writing, Data curation; **Derek Elsworth**: Conceptualization, Writing – review & editing; **Fengshou Zhang**: Data curation, Methodology; **Zhiyuan Wang**: Data curation, Methodology; **Jianbo Zhang**: Writing – review & editing.

Declaration of competing interest

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

Data availability

The authors do not have permission to share data.

Acknowledgments

This research has received fundings from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 846775 and the National Natural Science Foundation of China under the grant 42077247. DE acknowledges support from the G. Albert Shoemaker endowment. LH would like to thank Jennifer L. Miskimins for her pre-review and comments.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.energy.2022.126122.

References

[1] Huang F, et al. Study of a low-damage efficient-imbibition fracturing fluid without flowback used for low-pressure tight reservoirs. Energy 2021:222.
[2] Pahari S, et al. Optimal pumping schedule with high-viscosity gel for uniform distribution of proppant in unconventional reservoirs. Energy 2021;216:119231.
[3] Hou L, Elsworth D, Geng X. Swelling and embedment induced by sub- and super-critical-CO2 on the permeability of propped fractures in shale. Int J Coal Geol 2020;225:103496.
[4] Rahimi-Aghdam S, et al. Branching of hydraulic cracks enabling permeability of gas or oil shale with closed natural fractures. Proc Natl Acad Sci U S A 2019;116(5):1532–7.
[5] Manchanda R, et al. Integrating reservoir geomechanics with multiple fracture propagation and proppant placement. SPE J 2020;25(2):662–91.
[6] Chaturvedi KR, Trivedi J, Sharma T. Single-step silica nanofluid for improved carbon dioxide flow and reduced formation damage in porous media for carbon utilization. Energy 2020:197.
[7] Hou L, et al. Effect of slickwater-alternate-slurry injection on proppant transport at field scales: a hybrid approach combining experiments and deep learning. Energy 2022:242.
[8] Zhao L, et al. Experimental Study on a new type of self-propping fracturing technology. Energy 2019;183:249–61.
[9] Siddhamshetty P, Wu K, Kwon JS-I. Modeling and control of proppant distribution of multistage hydraulic fracturing in horizontal shale wells. Ind Eng Chem Res 2019;58(8):3159–69.
[10] Bangi MSF, Kwon JS-I. Deep hybrid modeling of chemical process: application to hydraulic fracturing. Comput Chem Eng 2020;134:106696.
[11] Narasingam A, Kwon JS-I. Development of local dynamic mode decomposition with control: application to model predictive control of hydraulic fracturing. Comput Chem Eng 2017;106:501–11.
[12] Mao S, et al. Field-scale numerical investigation of proppant transport among multicluster hydraulic fractures. SPE J 2021;26(1):307–23.
[13] Dahi-Taleghani A, Olson JE. Numerical modeling of multistranded-hydraulic-fracture propagation: accounting for the interaction between induced and natural fractures. SPE J 2011;16(3):575–81.
[14] Yew CH, Weng X. Mechanics of hydraulic fracturing. Gulf Professional Publishing; 2014.
[15] Ugueto C GA, et al. Perforation cluster efficiency of cemented plug and perf limited entry completions; Insights from fiber optics diagnostics. In: Spe hydraulic fracturing technology conference. OnePetro; 2016.
[16] Sun F, Du S, Zhao Y-P. Fluctuation of fracturing curves indicates in-situ brittleness and reservoir fracturing characteristics in unconventional energy exploitation. Energy 2022:252.
[17] Nolte K. A general analysis of fracturing pressure decline with application to three models. SPE Form Eval 1986;1(6):571–83.
[18] Nolte KG. Fracturing-pressure analysis for nonideal behavior. J Petrol Technol 1991;43(2):210–8.
[19] Nolte K. Principles for fracture design based on pressure analysis. SPE Prod Eng 1988;3(1):22–30.
[20] Elliott SJ, Gale JFW. Analysis and distribution of proppant recovered from fracture faces in the HFTS slant core drilled through a stimulated reservoir. In: Proceedings of the 6th unconventional resources technology conference; 2018.

[21] Maity D, Ciezobka J, Eisenlord S. Assessment of in-situ proppant placement in SRV using through-fracture core sampling at HFTS. In: SPE/AAPG/SEG unconventional resources technology conference; 2018 [OnePetro].
[22] Hu J, et al. Data-driven early warning model for screenout scenarios in shale gas fracturing operation. Comput Chem Eng 2020:143.
[23] Maity D, Ciezobka J. Designing a robust proppant detection and classification workflow using machine learning for subsurface fractured rock samples post hydraulic fracturing operations. J Petrol Sci Eng 2019;172:588–606.
[24] Hou L, et al. Evaluating essential features of proppant transport at engineering scales combining field measurements with machine learning algorithms. J Nat Gas Sci Eng 2022;107.
[25] Hosseini N, Khoei AR. Numerical simulation of proppant transport and tip screen-out in hydraulic fracturing with the extended finite element method. Int J Rock Mech Min Sci 2020;128.
[26] Ben Y, et al. Real time hydraulic fracturing pressure prediction with machine learning. In: SPE hydraulic fracturing technology conference and exhibition. Society of Petroleum Engineers; 2020.
[27] Sun JJ, et al. Application of both physics-based and data-driven techniques for real-time screen-out prediction with high frequency data. In: Proceedings of the 8th unconventional resources technology conference; 2020.
[28] Yu X, Trainor-Guitton W, Miskimins J. A data driven approach in screenout detection for horizontal wells. In: SPE hydraulic fracturing technology conference and exhibition; 2020.
[29] Hou L, et al. Prediction of the continuous probability of sand screenout based on a deep learning workflow. SPE J 2022:1–11.
[30] Pandey A, et al. Experimental investigation on effect of reservoir conditions on stability and rheology of carbon dioxide foams of nonionic surfactant and polymer: implications of carbon geo-storage. Energy 2021:235.
[31] Kim TH, Cho J, Lee KS. Evaluation of CO2 injection in shale gas reservoirs with multi-component transport and geomechanical effects. Appl Energy 2017;190: 1195–206.
[32] Singh H. Hydrogen storage in inactive horizontal shale gas wells: techno-economic analysis for Haynesville shale. Appl Energy 2022:313.
[33] Guo T, et al. Experimental study of hydraulic fracturing for shale by stimulated reservoir volume. Fuel 2014;128:373–80.
[34] Wang J, et al. Bi-power law correlations for sediment transport in pressure driven channel flows. Int J Multiphas Flow 2003;29(3):475–94.
[35] Patankar NA, et al. Power law correlations for sediment transport in pressure driven channel flows. Int J Multiphas Flow 2002;28(8):1269–92.
[36] Massaras LV, Massaras DV. Real-time advanced warning of screenouts with the inverse slope method. In: SPE formation damage control international symposium. LA: Lafayette; 2012.
[37] Dontsov EV, Peirce AP. Slurry flow, gravitational settling and a proppant transport model for hydraulic fractures. J Fluid Mech 2014;760:567–90.
[38] Willingham J, Tan H, Norman L. Perforation friction pressure of fracturing fluid slurries. In: Low permeability reservoirs symposium. OnePetro; 1993.
[39] Cho K, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078. 2014.
[40] Biau G, Scornet E. A random forest guided tour. Test 2016;25(2):197–227.
[41] Bergstra J, Bengio Y. Random search for hyper-parameter optimization. J Mach Learn Res 2012;13(2).
[42] Hu MY, et al. A cross-validation analysis of neural network out-of-sample performance in exchange rate forecasting. Decis Sci J 1999;30(1):197–216.
[43] Gal Y, Ghahramani Z. A theoretically grounded application of dropout in recurrent neural networks. arXiv preprint arXiv:1512.05287. 2015.
[44] Kingma DP, Ba J. Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980. 2014.
[45] Cheng Y, et al. Sand screen-out prediction and VIM analyses based on Random Forest algorithm and field measurements of shale gas fracturing. Acta Pet Sin 2022. Submitted for publication.
[46] Webb GI, Zheng Z. Multistrategy ensemble learning: reducing error by combining ensemble learning techniques. IEEE Trans Knowl Data Eng 2004;16(8):980–91.
[47] Grömping U. Variable importance assessment in regression: linear regression versus random forest. Am Statistician 2009;63(4):308–19.
[48] Svetnik V, et al. Random forest: a classification and regression tool for compound classification and QSAR modeling. J Chem Inf Comput Sci 2003;43(6):1947–58.
[49] Altmann A, et al. Permutation importance: a corrected feature importance measure. Bioinformatics 2010;26(10):1340–7.
[50] Guevara JL, Patel R, Trivedi J. Optimization of steam injection in SAGD using reinforcement learning. J Petrol Sci Eng 2021:206.
[51] Goodfellow I, Bengio Y, Courville A. Deep learning. MIT press; 2016.
[52] Otchere DA, et al. Data analytics and Bayesian Optimised Extreme Gradient Boosting approach to estimate cut-offs from wireline logs for net reservoir and pay classification. Appl Soft Comput 2022:120.
[53] Ma L, et al. In-situ synchrotron characterisation of fracture initiation and propagation in shales during indentation. Energy 2021:215.
[54] Hou L, et al. Data-driven optimization of brittleness index for hydraulic fracturing. Int J Rock Mech Min Sci 2022;159.
[55] Hou B, et al. Integration analysis of 3D fractures network reconstruction and frac hits response in shale wells. Energy 2022:260.