Petroleum Science 23 (2026) 5543–5560

Contents lists available at ScienceDirect

Petroleum Science

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

Original Paper

Machine-learning predictions of high-pressure high-temperature filtrate loss in oil-based drilling fluids from well-site measurements

Shadfar Davoodi [a,b,*], Arslan Gulniyazov [a], David A. Wood [c], Mohammed Al-Shargabi [a], Nikita Makarov [a], Evgeny Burnaev [b,d]

[a] School of Earth Sciences & Engineering, Tomsk Polytechnic University, Lenin Avenue, Tomsk, Russia
[b] Artificial Intelligence Center, Skolkovo Institute of Science and Technology, Bolshoy Boulevard 30, Bld. 1, 121205, Moscow, Russia
[c] DWA Energy Limited, Lincoln, United Kingdom
[d] Autonomous Non-Profit Organization Artificial Intelligence Research Institute (AIRI), Moscow, Russia

ARTICLE INFO

Article history:
Received 15 May 2025
Received in revised form
20 November 2025
Accepted 24 March 2026
Available online 28 March 2026

Edited by Jia-Jia Fei

Keywords:
High-pressure high-temperature (HPHT)
Oil-based drilling fluids (OBDF)
Rapid filtrate-loss prediction
Drilling optimization

ABSTRACT

Precise and reliable predictions of filtrate loss (FL) from drilling fluids/muds under high-pressure-high-temperature (HPHT) conditions are required to prevent formation damage, maintain wellbore stability, and optimize drilling efficiency in complex reservoirs. Traditional laboratory-based measurements are too time consuming to provide HPHT FL values that are useable for decision-making while drilling. This study, therefore, develops machine learning (ML) models that enable frequent and prompt predictions of HPHT FL with high accuracy and reliable precision. To this end, a large dataset was compiled containing five input parameters, namely mud temperature (MT), mud weight (MW), funnel viscosity (FV), mud alkalinity (MA), and electrical stability (ES), and a single target variable HPHT FL. Following meticulous data preprocessing, four predictive models were developed using established ML algorithms: multilayer perceptron neural network (MLPNN), support vector regression (SVR), extreme gradient boosting (XGBoost), and extreme learning machine (ELM). For each ML algorithm, five separate model instances were developed and evaluated, with XGBoost providing the best performance in predicting the target parameter (root mean square error (RMSE) = 0.096 cc/30 min for the testing subset). The superior performance of the XGBoost model was further confirmed by residual, uncertainty, overfitting, and robustness evaluations, indicating its excellent generalization capability. Shapley additive explanations (SHAP) analysis identified ES as the most influential input parameter and FV as the least influential for XGBoost's HPHT FL predictions. The XGBoost model as developed offers a substantial improvement compared to laboratory FL analysis, enabling more efficient and quicker monitoring of FL in challenging HPHT downhole environments.

1. Introduction

Oil-based drilling fluids/muds (OBDFs), particularly invert emulsion systems, are favored for drilling in challenging environments of high pressures and temperatures (HPHT), shale formations, as well as difficult operational conditions associated with highly deviated wells. The reason for this is mainly due to the superior thermal stability, excellent lubricity, and inherent shale inhibition achieved by OBDF (Bridges and Robinson, 2020; Sun et al., 2024). Controlling filtration of drilling fluids is a critical determinant of wellbore stability, prevention of formation damage, and overall drilling efficiency (Al-Shargabi et al., 2022). The filtration behavior of OBDFs is not a fixed property but is highly dependent on the fluid's specific compositional formulation, including the base oil (e.g., diesel, mineral oil, or synthetic), the oil-to-water ratio, and the concentration of key additives such as emulsifiers, wetting agents, and fluid-loss controllers (Abdelaal et al., 2023; Caenn et al., 2016; Leporini et al., 2019; Soares et al., 2020).

* Corresponding author.
E-mail address: davoodis@hw.tpu.ru (S. Davoodi).
Peer review under the responsibility of China University of Petroleum (Beijing).

https://doi.org/10.1016/j.petsci.2026.03.053

The industry's standard method for characterizing filtrate loss is the HPHT FL test, performed according to API standards (Ali Khan et al., 2018; API 13B-2, 2005; Davoodi et al., 2025). Although this laboratory test provides reliable measurement of HPHT FL, it requires at least 1–2 h (depending on the intended test conditions) to complete after sample collection. The extensive test time required creates a significant time lag that renders the results reactive rather than proactive for fluid management decisions at the wellsite. This delay is particularly problematic given that the fluid composition is dynamically managed and can be changed in response to downhole conditions, underscoring a pressing need for rapid assessment techniques that can keep pace with operational demands (Davoodi et al., 2023).

Machine learning (ML) offers a promising pathway to overcome this temporal limitation by enabling real-time, data-driven prediction of HPHT FL (Davoodi et al., 2025; Gul and van Oort, 2020; Zhong et al., 2022). Previous studies have attempted to predict various drilling fluid properties using ML, but these efforts have faced consistent limitations, many related to compositional complexity. Early work by Elkatatny et al. (2016) pioneered the use of an artificial neural network (ANN) to predict OBDF rheology under low-pressure-low-temperature (LPLT) conditions. While demonstrating feasibility, their model was tailored to a single, specific fluid formulation, limiting its generalizability. Subsequent models predicting density under HPHT conditions, such as those by Osman and Aggour (2003) and the hybrid PSO-ANN model by Ahmadi et al. (2018), were trained on small, aggregated experimental datasets from the literature (234 and 664 data points, respectively). These datasets lack the consistency required to model the specific compositional nuances of a field-deployed OBDF. Specifically, they increase the risk of overfitting and reducing practical utility across different fluid systems. The only published study focused specifically on predicting HPHT FL for OBDFs was conducted by Gul and van Oort (2020). Their study, while a significant step forward, was constrained by a small dataset of 105 records compiled from 17 wells. Their best-performing random forest model showed relatively low prediction performance (coefficient of determination (R^2) = 0.840, mean absolute error (MAE) = 0.785 mL) and, depended on input variables such as oil and water content. These measurements are obtained from time-consuming retort tests, which are typically performed only once or twice daily at the wellsite, creating a fundamental temporal mismatch that prevents the near-to-real-time HPHT FL monitoring needed for proactive fluid management.

These collective shortcomings highlight fundamental research gaps, including: (1) fluid-specific design constraints that ignore compositional variability, (2) small training datasets that cannot capture the full range of fluid formulations and operational conditions, (3) a reliance on infrequent measurements which require time-consuming experiments rather than frequent easy-to-measure inputs available more frequently at the wellsite, and (4) no existing capability for robust, near-real-time FL prediction.

This study proposes a novel approach to drilling fluid properties management to overcome the difficulties mentioned. It does so by developing a near-real-time HP-HT FL prediction model for a major class of OBDFs (mineral oil-based invert emulsion fluids) using only five easy-to-measure input variables routinely measured at the wellsite: mud weight (MW), mud temperature (MT), mud alkalinity (MA), electrical stability (ES), and funnel viscosity (FV). These five variables are typically measured at drilling sites every 15–20 min, requiring just 1–5 min per test to be measured. To accomplish this, a large field dataset from multiple wells drilled in two different fields, including the five mentioned influential drilling fluid dependent variables has been compiled, with HP-HT FL as the independent variable. Four established ML algorithms, namely multilayer perceptron neural network (MLPNN), support vector regression (SVR), extreme gradient boosting (XGBoost), and extreme learning machine (ELM) are employed to develop robust models to predict HP-HT FL in mineral oil-based invert emulsion drilling fluids. This study then evaluates the HPHT FL models' reproducibility and prediction uncertainty, something that has not been provided in previously published OBDF-FL prediction models (Gul and van Oort, 2020). The development of reliable and rapid (near-real-time), semi-automated estimation of HP-HT FL offers the potential to empower well-site-based drilling fluid engineers with an effective data-driven tool for monitoring the target parameter. This eliminates/decreases drilling fluid crew dependency on slow laboratory-dependent measurements and enables prompt drilling fluid management and decision-making.

2. Data characterization

Table 1 presents the statistical summary of the 4462-record dataset evaluated in this study, compiled from 58 wells drilled in two oil and gas fields located in Russia. The table presents key characteristics of drilling fluid parameters and their HPHT filtration performance, revealing important operational trends and variability. The data reveal mud weights (MW) averaging 11.14 lb/gal (standard deviation (SD) = 1.16) within an 8.5–13.6 lb/gal operational range, while mud temperatures (MT) averaged 114.4 °F (SD = 18) with surprisingly wide variation (70–167 °F). Marsh funnel viscosity (FV) averaged 66.6 s/qt (SD = 17.9) but exhibited significant right-skewness (max 323 s/qt), suggesting occasional high-viscosity treatments. Electrical stability (ES) measurements average 608.6 V (SD = 180.8) with extreme values (165–1680 V) indicating both broken and exceptionally stable emulsions. The target variable, HPHT filtrate loss (HPHT FL), averages 2.31 cc/30 min (SD = 0.57) within a 0.4–5 cc/30 min range, where the 25th–75th percentile range (2.0–2.6 cc) suggests most operations achieve acceptable filtration control. Notable outliers include mud alkalinity (MA) values up to 285 (compared to a median value of 2.7), potentially reflecting drilling fluid system transitions or contamination events. The dataset captures substantial operational variability suitable for identifying robust correlations between fluid properties and filtration performance under HPHT conditions. Filtration tests were performed at a temperature of 250 °F and a pressure differential of 500 psi.

The OBDF used across all wells was a mineral oil-based invert emulsion fluid. The consistent use of this specific OBDF type ensures that the developed models are trained on a homogeneous dataset. While the filtration performance of OBDFs is fundamentally governed by their specific chemical composition (Caenn et al., 2016), the precise concentrations of emulsifiers and fluid-loss additives are not available as high-frequency, real-time measurements at the wellsite. Therefore, this study utilizes five parameters (MW, MT, FV, MA, ES) that are derived from rapid, routine wellsite measurements. These five inputs serve as effective proxies for the fluid's functional state.

To clarify the compositional baseline, Table 2 summarizes the typical formulation and primary function of the key additives used in this specific OBDF. That information is derived from the fluid inventory records from the studied wells. This structured formulation provides a stable compositional baseline, allowing the readily measurable macroscopic properties to reliably indicate changes in filtration performance. For instance, electrical stability is a direct indicator of the health of the emulsion, as it is a function of the primary emulsifier package (e.g., MEGAMUL, VERSAWET). The model successfully learns the complex, non-linear relationships between these proxy measurements and the resulting filtration performance. Consequently, the model's predictions are most directly applicable to fluids of this specific classification, which represents one of the most widely used and industrially

Table 1
Statistical summary of the 4462-record dataset evaluated in this study.

Variables	MW, lb/gal	MT, °F	FV, s/qt	MA	ES, V	HPHT FL, cc/30 min
Mean	11.14	114	66.60	2.72	608.55	2.31
SD	1.16	18	17.90	4.31	180.83	0.57
Min	8.50	70	37.00	0.50	165	0.40
25%	10.80	101	56.00	2.20	512.00	2.00
50%	11.35	115	62.00	2.70	591.50	2.20
75%	11.90	126	72.00	3.10	743.00	2.60
Max	13.60	167	323.00	285.00	1680.00	5.00

Note: lb/gal: pounds per gallon; °F: degrees Fahrenheit; s/qt: seconds per quart; cc/30 min: cubic centimeters per 30 min.

Table 2
Typical composition and functional role of the oil-based drilling fluid system.

Additive category	Typical additive name	Primary function	Influence on macroscopic properties
Base fluid	ESCAID 110	Continuous phase determining initial fluid density and rheology.	Directly influences MW, FV, and high-temperature rheological stability (Li et al., 2024).
Primary emulsifier	MEGAMUL, VERSAWET	Stabilizes the water-in-oil emulsion, creates a strong interfacial film.	Primary driver of ES. A stable emulsion is crucial for low filtrate loss.
Secondary emulsifier/ wetting agent	VERSACOAT HF	Ensures solids are oil-wet and aids emulsion stability.	Supports high ES and stable rheology.
Fluid-loss control agent	VERSATROL M	Forms a low-permeability filter cake by plugging pore throats.	Directly reduces HPHT FL. Works synergistically with emulsifiers. This mechanism is validated by novel additives including pharmaceutical waste biomass (Khodja et al., 2022), graphene nanoplatelets (Arain et al., 2022), and PLA/Henna composites (Hsu et al., 2025).
Viscosifier	VG-PLUS, DUO-VIS	Builds gel structure and suspension characteristics.	Primary driver of FV, gel strength, and thixotropy. Organophilic clays like Claytone-II significantly enhance yield point and gel strength in OBDFs (Li et al., 2024; Mahmoud et al., 2025).
Alkalinity control	Lime	Maintains high alkalinity to saponify oils, stabilizes emulsifiers, and scavenges CO_2/H_2S.	Directly measured as MA. Lime provides superior thermal stability compared to caustic soda, preventing gelation at high temperature (Vivas and Salehi, 2021).
Weighting agent	Barite, SAFE-CARB	Increases density to control formation pressure.	Primary driver of MW. High solids loading can increase FL if not properly encapsulated.
Water phase	$CaCl_2$ Brine	Internal phase; salinity controls osmotic balance with formations.	Affects ES and overall fluid chemistry.

relevant categories of OBDFs for HPHT drilling. Therefore, while the model is not presented as a universal predictor for all possible OBDF formulations, its applicability to a major class of field-deployed fluids ensures substantial practical significance. The methodology itself, using high-frequency wellsite measurements to predict HPHT FL, is universally transferable.

Fig. 1 shows the correlation matrix among the target variable (HPHT FL) and input parameters. The distributions displayed in that figure's diagonal reveal non-normal patterns, particularly for FV and MA, suggesting complex underlying behaviors. The strong negative correlation between ES and HPHT FL ($r = -0.72$) reflects fundamental emulsion physics—electrical stability directly measures the emulsion's resistance to breakdown under electric fields, where higher ES values indicate more stable emulsions that better prevent fluid migration through tighter droplet packing and stronger interfacial films. This electrochemical stabilization is crucial for maintaining filter cake integrity under HPHT conditions. The moderate positive MW-HPHT FL correlation ($r = 0.45$) arises from competing mechanisms: While increased MW provides greater hydrostatic pressure that could reduce FL, it typically requires higher solids loading, which may increase cake permeability and colloidal solids concentration—factors that dominate under HPHT conditions. The weak FV–HPHT FL correlation ($|r| < 0.15$) highlights the limitations of ambient viscosity measurements (Marsh funnel) in predicting downhole behavior, as drilling fluid rheology is strongly temperature-dependent and often exhibits thermal thinning.

The FV-MT negative correlation ($r = -0.34$) supports this, while the near-zero FV-MW correlation ($r = -0.06$) suggests that weight materials in this system are primarily non-viscosifying. MA's weak correlation may reflect its indirect role in stabilizing emulsions via soap formation, rather than directly influencing filtration. These relationships emphasize that HPHT FL is governed primarily by: (1) electrical stability (ES), (2) solids characteristics (MW), and (3) thermal-rheological effects that ambient measurements may not capture, highlighting ES as the dominant control in this dataset.

3. Methodology

Fig. 2 summarizes the HPHT-FL-prediction workflow adopted. It begins by compiling data from field and laboratory sources, incorporating five well-site derived input parameters (MW, MT, MA, ES, and FV). This selection ensures the model is based on data available for near-real-time decision-making, using functional proxies for the fluid's physicochemical state rather than relying on infrequent laboratory-based compositional analyses. The dataset is then preprocessed, which involves splitting it into training and testing subsets, normalization, and outlier detection using the leverage technique. Different k-fold values are evaluated using k-fold cross-validation to determine an appropriate portion of the training subset to allocate as the validation subset. The four ML models (XGBoost, SVR, MLPNN, and ELM) are then executed using five randomly selected subsets to ensure FL-prediction stability and reproducibility. ML model performance is assessed using five statistical metrics: root mean square error (RMSE), average relative error (ARE), average absolute relative error (AARE), SD, and R^2. Residual analysis uncertainty assessment, overfitting behavior, and robustness evaluation are performed to determine the most

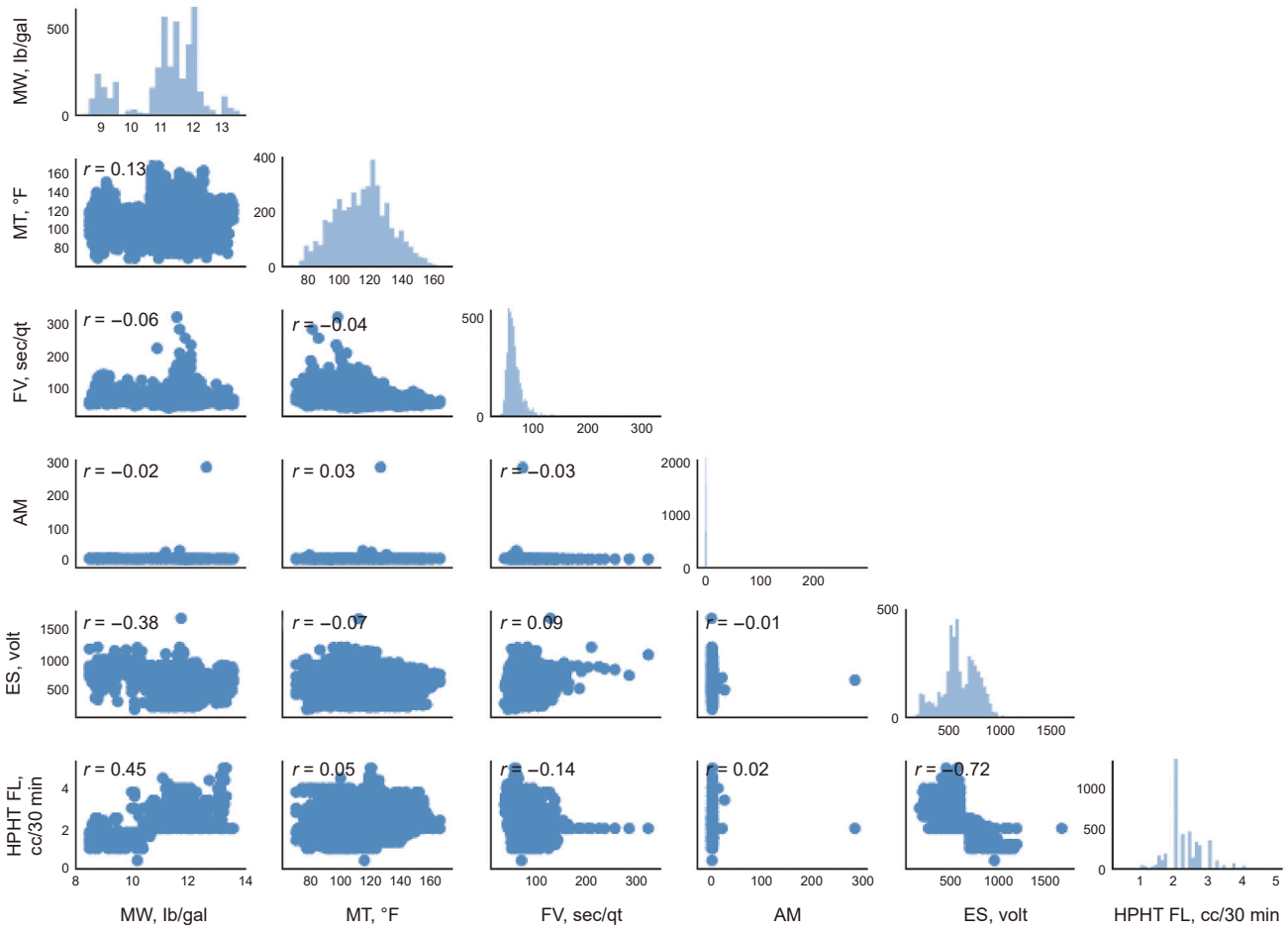

Fig. 1. Correlation matrix plot among the target variable, HPHT FL, and the model input parameters. The diagonal elements show univariate distributions using histograms. The lower triangle displays scatter plots with corresponding Pearson correlation coefficients.

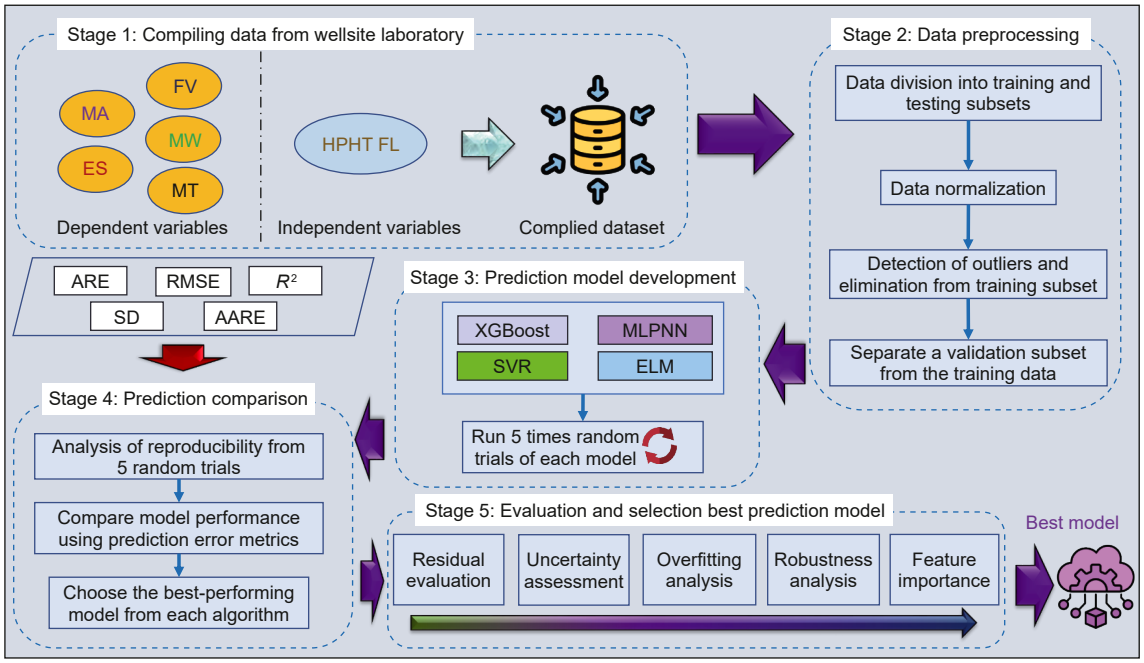

Fig. 2. Workflow of the proposed methodology for predicting the target parameter using machine learning models. Input variables include MW, MT, MA, ES and FV. Evaluation metrics used are RMSE, ARE, AARE, SD and R^2. Developed models are XGBoost, MLPNN, SVR and ELM.

reliable FL-prediction model. Feature importance analysis is then conducted using the best-performing model. Each of these analytical steps is discussed in detail in the following sections.

3.1. Data preprocessing

Preprocessing steps include data normalization to standardize feature scales, stratified splitting into training and testing subsets to maintain distributional consistency, and outlier detection/removal to mitigate undue influences. These steps enhance FL prediction performance while ensuring generalizability when trained models are applied to previously unseen data.

3.1.1. Data division and normalization

A trained model's prediction performance when applied to previously unseen data is critical, as it reflects its ability to generalize beyond the training dataset. To properly assess generalizability, a portion of the collected data must be reserved as an independent testing subset, completely excluded from the ML training process. Determining the optimal training-testing subset split ratio is essential, as it directly impacts model performance. There is no universally "correct" ratio—instead, a sensitivity analysis must be conducted considering different split ratios (e.g., 70:30, 80:20, and 90:10) to identify the most robust partitioning for a specific dataset. The training data must also exhibit sufficient volume and diversity to ensure the model captures the underlying relationships between input and output features effectively. The SVR was specifically selected for this sensitivity analysis due to its inherent robustness to outliers—a crucial advantage given that data partitioning occurs prior to outlier detection in our analytical pipeline. The split ratio yielding the best generalization performance based on RMSE or R^2 values achieved when the trained model is applied to the testing subset and used to select the optimal partitioning strategy.

Data normalization ensures that all features contribute proportionally during ML model training by mitigating scale-related biases. It is especially important for distance-based models like SVR and those relying on gradient descent optimization (Géron, 2022; Theodoridis, 2015). Normalization not only enhances training stability and convergence but also helps to provide clearer interpretations of feature significance (Bishop and Nasrabadi, 2006). Commonly applied normalization techniques include z-score scaling, decimal scaling, and logarithmic transformation, but this study adopts min-max normalization. This approach rescales each feature to a [0, 1] range based on its minimum ($X_{\min}$) and maximum ($X_{\max}$) values using Eq. (1), which is particularly suited to preserving the original distribution of features confined to bounded intervals (Raschka and Mirjalili, 2019). Min-max normalization is selected for its computational simplicity and alignment with the leverage-based outlier detection method used later, which benefits from uniformly scaled inputs (Aggarwal, 2017).

$$X_{\text{norm}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{1}$$

3.1.2. Outlier detection

The accuracy of ML models is contingent on the quality of the training data. Recognizing that routine wellsite measurements can be affected by human error, instrument drift, or transient operational conditions, a systematic outlier detection process was implemented to prevent these anomalies from distorting input-output relationships and compromising model generalization. Additionally, drilling-fluid experts reviewed the preprocessed dataset to ensure data consistency and credibility prior to modeling. This study adopts the leverage technique—a statistically robust approach for identifying influential outliers in predictive modeling (Kutner et al., 2005; Montgomery et al., 2021). Leverage quantifies the influence of individual data points based on their positional extremity within the input feature space, calculated via the diagonal elements of the hat matrix (Chatterjee and Hadi, 1986; Kutner et al., 2005; Montgomery et al., 2021). Crucially, it identifies high-leverage points that may disproportionately skew model predictions, a nuance often missed by conventional methods. The threshold applied for flagging influential points is defined by Eq. (2).

$$\text{Threshold} = \frac{2p}{N} \tag{2}$$

where p = predictor count (number of input variables plus intercepts where appropriate) and N = total observations (number of data records). Applied after data splitting and normalization, this technique ensures that outliers are identified and removed exclusively from the training subset (not the testing subset), enhancing model generalizability and prediction reliability.

3.1.3. Model validation applying k-fold cross-validation approach

To ensure robust model selection while maintaining computational efficiency, this study assesses k-fold cross-validation with three distinct k values (4, 5, or 10 folds), using SVR to determine the optimal fold count based on minimal mean validation error. In 10-fold validation, each iteration allocates 10% of the training data to the validation subset, cycling through all folds to ensure unbiased performance estimation. The SVR model was configured with the radial basis kernel function (RBF)-kernel evaluated prediction errors across all folds. A validation subset was only selected as representative if its final error was within 5% of the mean. Otherwise, the process to select a representative validation subset was repeated until this stability criterion was met.

During training, termination conditions—(1) maximum iterations or (2) no validation error improvement for 10 steps—were enforced for SVR and XGBoost to balance performance and efficiency. However, these conditions did not apply to ELM, as its training is non-iterative. Instead, 1000 ELM models were generated from the training data, and the one with the lowest validation RMSE was selected, leveraging ELM's inherent randomness and rapid execution for optimal performance without iterative tuning.

3.2. Predictive algorithms

Four ML models (ELM, MLPNN, SVR, and XGBoost) are developed to predict HPHT FL. Detailed descriptions of the algorithms of these models are provided in the Supplementary Information file. Table 3 compares the key advantages and disadvantages of these models, highlighting the trade-offs required to achieve computational efficiency, generalization performance, and model interpretability.

3.3. Assessment of developed models

The evaluation of the developed models is conducted in two phases. First, the reproducibility of the models is assessed. Second, the best-performing model for each algorithm is selected, and a detailed comparison is made between those selected models based on specific criteria.

Table 3
Summary of the key advantages and disadvantages of the four ML algorithms developed to predict HPHT FL.

Algorithm	Advantages	Disadvantages
XGBoost	Efficiently handles missing data (Chen and C, 2016). Delivers high performance using parallel processing and built-in regularization to reduce overfitting (Chen and C, 2016). Performs reliably on both small and moderately large datasets.	Computationally intensive for very large datasets (Bentéjac et al., 2021). Requires careful tuning of hyperparameters to achieve optimal performance (Probst et al., 2019).
ELM	Offers extremely fast training due to random assignment of hidden layer weights (Huang et al., 2006). Provides good generalization with minimal need for parameter tuning (Huang et al., 2015). Avoids issues with local minima common in traditional training algorithms (Huang et al., 2006).	May require a large number of hidden neurons to handle complex problems (Huang et al., 2015). Less interpretable than conventional neural networks (Ding et al., 2014).
SVR	Performs well in high-dimensional feature spaces (Smola and Schölkopf, 2004). Robust to outliers due to the use of an epsilon-insensitive loss function (Drucker et al., 1996) Effective with small datasets (Awad et al., 2015).	Becomes computationally expensive as dataset size increases (Awad et al., 2015). Requires careful selection of kernel type and tuning of hyperparameters (Chang and Lin, 2011; Cherkassky and Ma, 2004).
MLPNN	Capable of approximating any continuous function (Hornik et al., 1989). Flexible architecture suitable for a variety of tasks (Goodfellow et al., 2016). Effectively captures non-linear relationships (Bishop and Nasrabadi, 2006).	Susceptible to overfitting if not properly regularized (Bishop and Nasrabadi, 2006). Requires extensive tuning of architecture and parameters (Bengio et al., 2017). Training may be slow, especially for deep networks (LeCun et al., 2015).

3.3.1. ML model reproducibility analysis

The results generated by ML algorithms depend on their control-parameter values, which are adjusted to determine optimal trainable-parameters values. Since the data subset divisions are randomly selected for each model trial, the FL prediction outcomes can vary with each trial. To evaluate the consistency of the results, each algorithm is therefore evaluated with five separate random trials, generating five RMSE values for the training, validation, and testing subsets selected for each trial. The RMSE means and standard deviations for the predictions of these five trials are assessed to indicate the reproducibility of each ML model. A smaller standard deviation indicates greater consistency across multiple trials, suggesting more reproducible results. A lower mean RMSE reflects better overall model prediction performance compared to other models.

$$R^2 = 1 - \frac{\sum_{i=1}^{N}\left(\mathrm{HPHTFL}_{i,\mathrm{exp}} - \mathrm{HPHTFL}_{i,\mathrm{pred}}\right)^2}{\sum_{i=1}^{N}\left(\mathrm{HPHTFL}_{i.\mathrm{exp}} - \overline{\mathrm{HPHTFL}}\right)^2} \tag{3}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\mathrm{HPHTFL}_{i,\mathrm{exp}} - \mathrm{HPHTFL}_{i,\mathrm{pred}}\right)^2} \tag{4}$$

$$\mathrm{SD} = \sqrt{\frac{\sum_{i=1}^{N}\left(\mathrm{HPHTFL}_{i,\mathrm{exp}} - \mathrm{HPHTFL}_{i,\mathrm{pred}} - \frac{1}{N}\sum_{i=1}^{N}\mathrm{HPHTFL}_{i,\mathrm{exp}} - \mathrm{HPHTFL}_{i,\mathrm{pred}}\right)}{N-1}} \tag{5}$$

$$\mathrm{ARE} = \frac{1}{N}\sum_{i=1}^{N}\frac{\mathrm{HPHTFL}_{i,\mathrm{exp}} - \mathrm{HPHTFL}_{i,\mathrm{pred}}}{\mathrm{HPHTFL}_{i,\mathrm{exp}}} \tag{6}$$

$$\mathrm{AARE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{\mathrm{HPHTFL}_{i,\mathrm{exp}} - \mathrm{HPHTFL}_{i,\mathrm{pred}}}{\mathrm{HPHTFL}_{i,\mathrm{exp}}}\right| \times 100 \tag{7}$$

3.3.2. Statistical evaluation matrix

Five statistical prediction-performance metrics are used to evaluate the FL-prediction performance of the ML models developed. These metrics help quantify the agreement between the experimental measure ($\mathrm{HPHTFL_{exp}}$) and predicted ($\mathrm{HPHTFL_{pred}}$) values of HPHT FL. These metrics are R^2, RMSE, SD, ARE, and AARE. R^2 measures the proportion of variance in the dependent variable that is predictable from the independent variables (Eq. (3)). RMSE measures the squared differences between predicted and measured values, with lower values indicating better model performance (Eq. (4)). SD quantifies the amount of variation or dispersion in the predicted values (Eq. (5)). ARE determines the mean of the relative errors (Eq. (6)), while AARE provides the mean of the absolute values of the relative errors (Eq. (7)), offering a more robust measure of error magnitude.

4. Results

The findings of this study are organized into three subsections corresponding to the workflow sequence: data preprocessing, ML model tuning, and the execution of the developed HPHT FL prediction models.

4.1. Data preprocessing

To establish the optimal training-testing split ratio, first, 100 uniformly distributed data points were reserved as a fixed

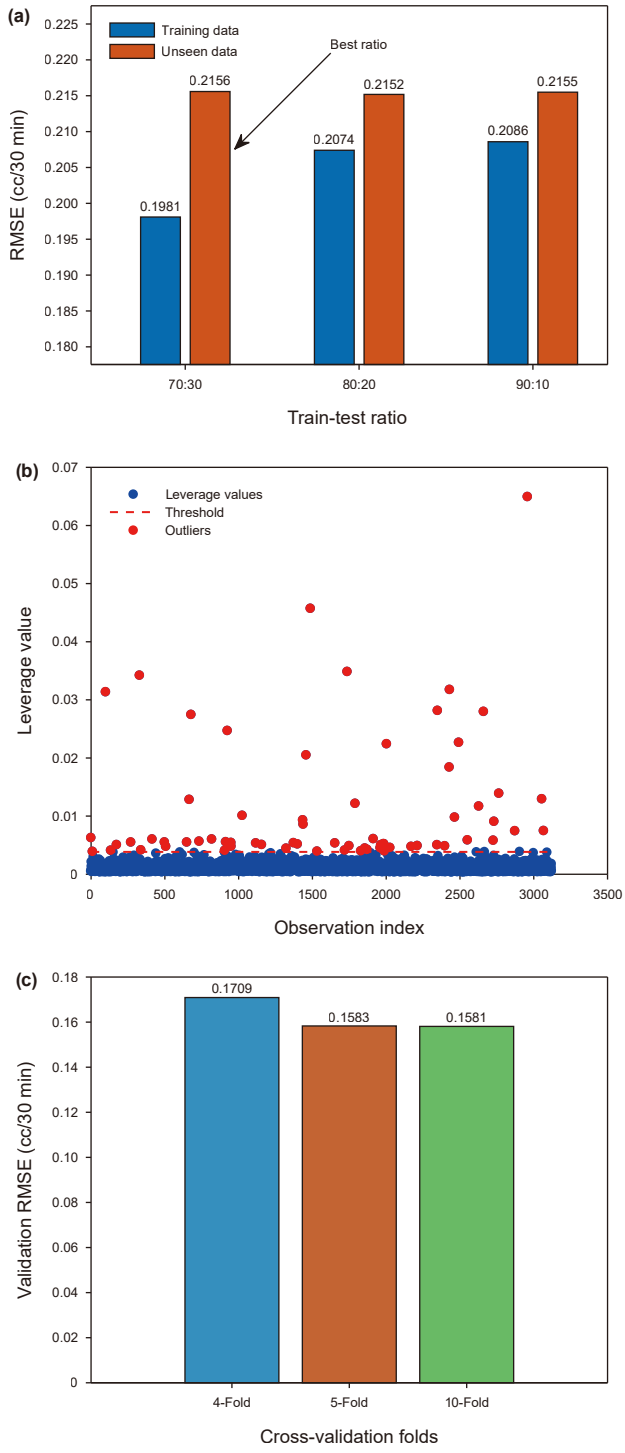

Fig. 3. (**a**) RMSE comparison of RBF-SVR models across training-testing-subset splits (70:30, 80:20, 90:10) for training and unseen data, highlighting 70:30 as the optimum split; (**b**) Leverage values for the training subset, used to identify outliers based on the specified threshold criterion, and (**c**) comparison of cross-validation performance across different fold configurations using RMSE as the evaluation metric. The results were obtained using an SVR model configured with an RBF.

verification subset. The remaining data were then partitioned into training and testing subsets according to three split ratios (70:30, 80:20, and 90:10). All models were subsequently evaluated against the same verification subset to ensure consistent assessment conditions. This approach provides a controlled comparison of training efficiency across different data allocation strategies while maintaining constant evaluation criteria. Fig. 3(a) shows the training and unseen data RMSE performance of an SVR model configured with an RBF-kernel for HPHT FL prediction applying three training-testing splits (70:30, 80:20, 90:10), with all models evaluated on a fixed unseen dataset (100 samples). As can be seen in this figure, the 70:30 ratio demonstrates optimal performance, achieving the lowest training RMSE (0.1981) while maintaining near-identical testing RMSE (0.2156) compared to other splits. This advantage arises from three factors: (1) Computational efficiency—larger training sets (90:10) significantly increase SVR processing time; (2) Noise susceptibility—expanding beyond 70% training data introduces more measurement noise and outliers, destabilizing hyperparameter optimization; and (3) The law of diminishing returns—the marginal improvement in testing RMSE (<0.5% difference between splits) fails to justify the 28.6% increase in training data volume from 70% to 90%. These results confirm the 70:30 split (3123 training, 1339 test samples) optimally balances prediction performance, computational cost, and robustness for HPHT FL prediction.

Following data partitioning into 70% training and 30% testing subsets, the leverage technique was applied to the training data to identify potential outliers. Fig. 3(b) displays the leverage versus data-record indices, revealing 69 observations with values exceeding the predetermined threshold. These high-leverage points were subsequently removed from the training set to improve model robustness.

Fig. 3(c) illustrates the results of the *k*-fold cross-validation process, also conducted using the SVR-RBF-configured model, applied to the HPHT FL dataset comprising 3054 data points. This analysis was performed to determine the most appropriate proportion of the training data subset to allocate to the validation set by evaluating the impact of different *k*-fold configurations (4-fold, 5-fold, and 10-fold) on model performance. The corresponding validation RMSEs were 0.1709, 0.1583, and 0.1581, respectively. While 10-fold cross-validation achieved the lowest error, the improvement over 5-fold was marginal (only 0.0002). Therefore, 5-fold cross-validation was selected as the most efficient and balanced option. Based on this configuration, 80% of the training-data subset (2443 samples) was allocated for training, and the remaining 20% (611 samples) was allocated to the validation subset.

4.2. Tuning the four ML models

The performance of machine learning models heavily depends on their hyperparameters. To achieve high prediction performance, selecting optimal hyperparameter values is essential. In this study, key hyperparameters, including the number of neurons and activation functions for MLPNN, the kernel function for SVR, and the number of neurons for ELM, were optimized through sensitivity analysis to develop effective HPHT FL prediction models.

Fig. 4 displays boxplots of training-subset RMSE values (in FL units of cc/30 min) obtained for different hyperparameter settings applied to the XGBoost model. Fig. 4(a) illustrates the impact of varying the learning rate (η), where lower values between 0.06 and 0.11 resulted in reduced median RMSE and narrower spread, indicating better and more stable performance, whereas higher η values led to performance deterioration. Fig. 4(b) demonstrates the effect of tree depth, showing that depths of 5 and 7 achieved lower RMSE values compared to depths of 3 and 9, suggesting that moderate tree complexity is more favorable. Fig. 4(c) plot examines the number of estimators (individual trees evaluated), with 100 and 150 estimators yielding slightly lower RMSE values than 50 or 200 estimators. Fig. 4(d) presents the influence of subsample

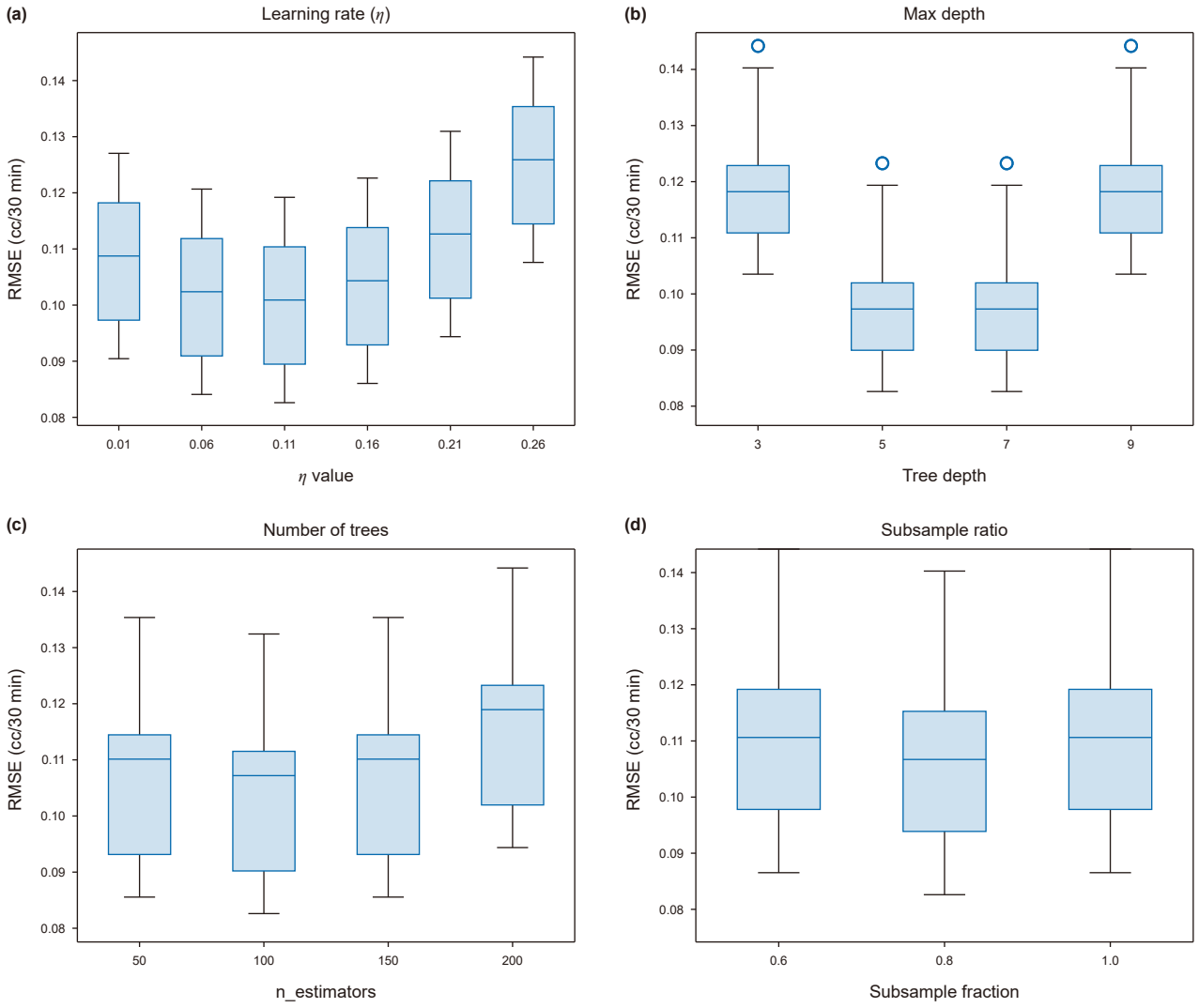

Fig. 4. Boxplots of RMSE (cc/30 min) values for different hyperparameter settings of the XGBoost model using the training subset for HPHT FL prediction. Each subplot illustrates the influence of a specific hyperparameter—(**a**) learning rate (η), (**b**) tree depth, (**c**) number of estimators, and (**d**) subsample fraction—while the other hyperparameters varied across their respective ranges. The boxplots capture the combined effects of hyperparameter interactions on model performance, providing insights into the stability and effectiveness of different parameter choices.

fraction, where using 0.8 or 1.0 provided marginally better RMSE performance compared to a value of 0.6. Fig. 4 results suggest that an "η" value of 0.11, a tree depth of 5, 100 estimators, and a subsample fraction of 0.8 are the most appropriate settings for developing a reliable XGBoost model for HPHT FL prediction.

Fig. 5 summarizes the sensitivity analysis conducted to optimize the hyperparameters of the other three ML models for predicting HPHT FL evaluated in this study. Fig. 5(a) reveals the influence of the number of neurons in the first and second hidden layers of the MLPNN model. A configuration with two hidden layers produced the minimum RMSE value of 0.1322 cc/30 min. Fig. 5(b) identifies that the Tanh activation function outperforms the alternative functions considered (sigmoid, Linear, and ReLU) for the MLPNN, achieving an RMSE of 0.0779 cc/30 min, followed closely by the Sigmoid function. Fig. 5(c) establishes that the SVR model configured with RBF kernel is the most effective (RMSE = 0.1612 cc/30 min), outperforming the linear, polynomial, and MLP kernels. Fig. 5(d) shows that the ELM model configured with 17 neurons generates the minimum RMSE value of 0.1937 cc/30 min. The Levenberg-Marquardt training algorithm was employed with the MLPNN model, gradient boosting with the XGBoost model, and the sequential minimal optimization (SMO) with the SVR model.

4.2.1. Reproducibility analysis of the FL predictions generated

Each algorithm was executed five times with different randomly selected subsets, and the RMSE values for training, validation, and testing subsets were recorded, providing an assessment of the reproducibility and stability of each ML model in terms of its FL predictions. Fig. 6 displays a raincloud plot of RMSE values across five trials for each ML algorithm for the training, validation, and testing subsets. The distributions reveal notable differences in model performance. XGBoost displays the narrowest and most sharply peaked RMSE distributions for all three subsets, indicating high reproducibility and low variance in the prediction performance. Those RMSE distributions are slightly left-skewed, suggesting that most RMSE values are concentrated in the lower-value side of the distribution, which reflects strong predictive performance.

The SVR model achieves relatively symmetric and moderately spread distributions, with a slight left skew, particularly in the validation and testing subsets, indicating stable performance with occasional higher RMSE values. The MLPNN model exhibits

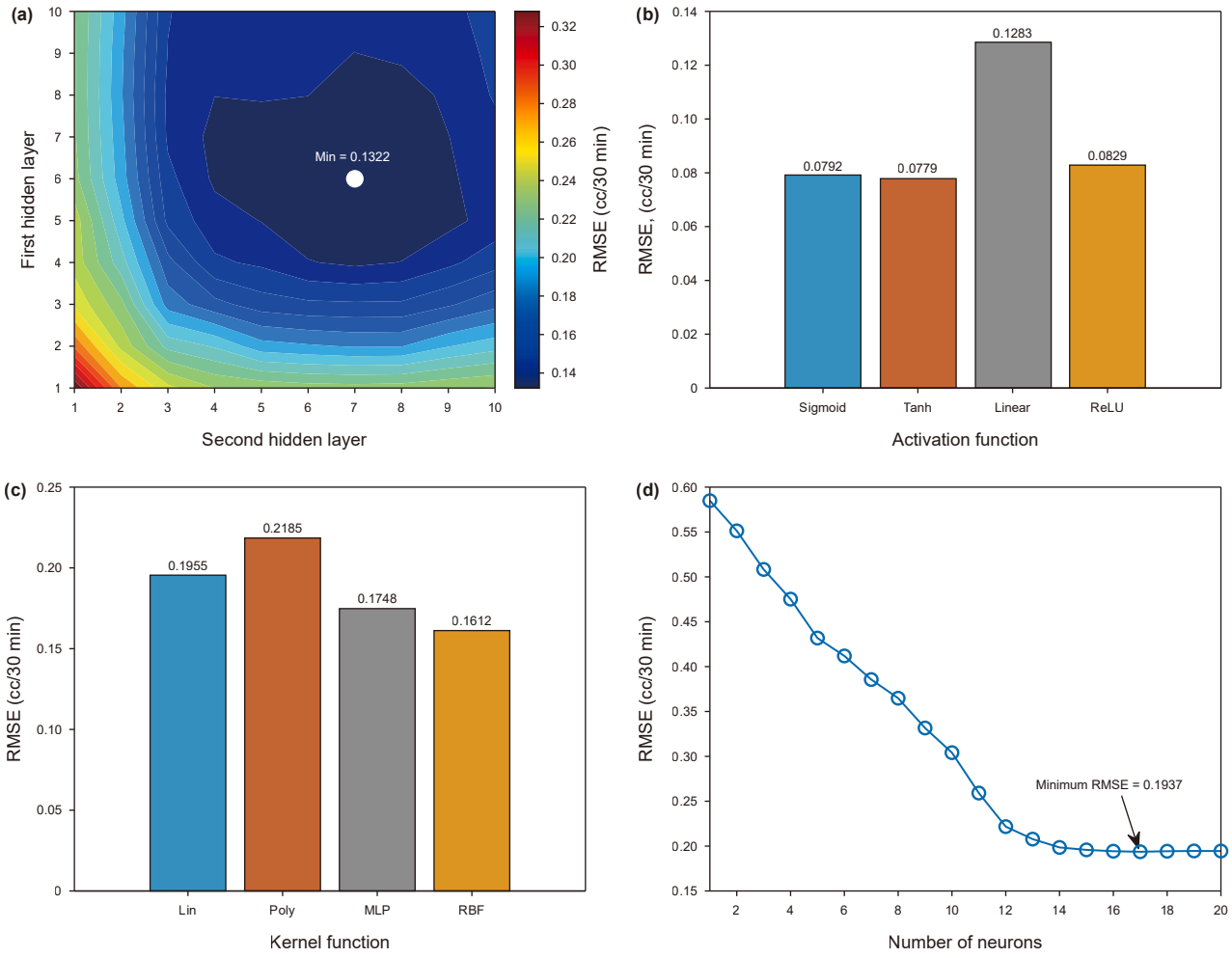

Fig. 5. Sensitivity analysis results for optimizing hyperparameters for MLPNN, SVR, and ELM models used to predict high pressure high temperature fluid loss (HPHT FL). (**a**) Determination of the optimal number of neurons in the first and second hidden layers of MLPNN, showing that two hidden layers generate the minimum FL-prediction RMSE; (**b**) Evaluation of different activation functions for the MLPNN model, with the Tanh function achieving the best performance; (**c**) Selection of the appropriate kernel function for SVR, where the RBF kernel produced the lowest FL-prediction RMSE; (**d**) Optimization of the number of neurons in the single hidden layer of the ELM model, identifying 17 neurons as the optimal number for minimizing RMSE. Tanh: hyperbolic tangent; ReLU: rectified linear unit; RMSE: root mean square error; Lin: linear kernel; Poly: polynomial kernel; MLP: multi-layer perceptron kernel; and RBF: radial basis function kernel.

broader and slightly right-skewed distributions, especially in the validation and testing subsets, implying greater variability and possible sensitivity to initial conditions or overfitting. The ELM model delivers a moderate spread and approximately symmetric distributions, reflecting consistent but slightly poorer FL prediction performance compared to XGBoost and SVR. It is apparent from Fig. 6 that XGBoost not only achieves the lowest RMSE but also demonstrates the most consistent and reliable behavior across multiple trials. Table 4 presents the mean and standard deviation of RMSE values calculated from five independent runs of each ML algorithm—MLPNN, XGBoost, SVR, and ELM—for the training, validation, and testing subset predictions of FL. XGBoost achieved the lowest average RMSE for all three subsets, with values of 0.0910, 0.1025, and 0.1117 cc/30 min for training, validation, and testing, respectively, accompanied by low standard deviations, indicating consistently high FL-prediction performance. The MLPNN model followed with moderate RMSE values and higher variability, especially during validation (0.0354 cc/30 min) and testing (0.0307 cc/30 min). SVR and ELM exhibited higher mean RMSEs, with ELM performing the poorest overall; however, it is marked as the best-performing model in this context due to additional criteria such as computational efficiency or simplicity, which may have been considered in the final evaluation. Notably, the standard deviations across models provide insight into performance stability, with XGBoost demonstrating the most reliable outcomes. This numerical analysis complements the visual distribution patterns shown in the previous figure and highlights quantitative distinctions among the models' predictive capabilities.

4.2.2. FL prediction performance of the top-performing models

The ELM, XGBoost, SVR, and MLPNN were evaluated through iterative training-validation cycles to identify the top-performing model for predicting FL in unseen (testing-subset) data. Fig. 7 displays scatter plots comparing the predicted and measured values of HPHT FL (cc/30 min) for the four ML models applied to training and validation subsets. In all the cross plots, the black diagonal line represents the ideal case where predictions perfectly match measurements ($Y = X$), while the orange line indicates the best linear fit through the data points. For MLPNN (Fig. 7(a)), the distribution of points is closely aligned with the $Y = X$ line for both training and validation subsets, indicating high FL prediction performance across its entire value range. For XGBoost (Fig. 7(b)), the data also follow the $Y = X$ line closely. Both MLPNN and

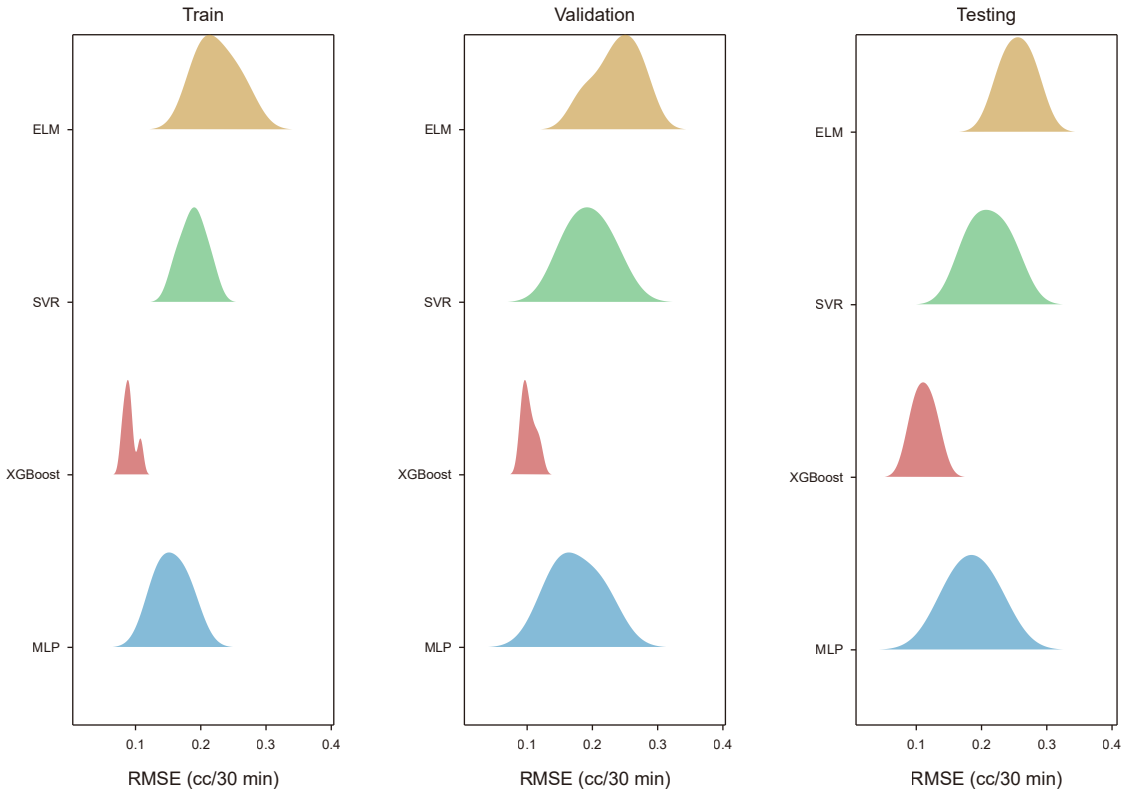

Fig. 6. Raincloud plot of the distributions of RMSE values for five trials executed for each algorithm and evaluated for training, validation, and testing subsets assessing HP-HT FL prediction performance.

Table 4
Mean and standard deviation of RMSE values (for FL in cc/30 min) from five runs of each algorithm for training, validation, and testing subsets to predict HPHT FL.

Models	Phase	RMSE	RMSE
		Mean (based on 5 trials)	SD (based on 5 trials)
MLP	Training	0.1554	0.0239
	Validation	0.1750	0.0354
	Testing	0.1853	0.0307
XGBoost[a]	Training	0.0910	0.0104
	Validation	0.1025	0.0105
	Testing	0.1117	0.0150
SVR	Training	0.1886	0.0204
	Validation	0.1940	0.0315
	Testing	0.2111	0.0310
ELM	Training	0.2251	0.0307
	Validation	0.2380	0.0358
	Testing	0.2552	0.0223

[a] Represents the best-performing model.

XGBoost models show slight underestimation at higher FL values, especially in the training subset. For SVR (Fig. 7(c)), there is a strong concentration of points near the $Y = X$ line at low to moderate values, but at higher values, the model tends to slightly underestimate FL. For ELM (Fig. 7(d)), the training and validation data points exhibit a good fit along the $Y = X$ line for low to moderate values; however, at higher values, the model consistently underestimates FL. The best-fit lines in all subplots closely match the ideal line for low to moderate FL values, indicating low prediction errors. However, the degree of underestimation at high FL values varies from ML model to model.

The trained models were then applied to the testing subset (Fig. 8). For MLPNN (Fig. 8(a)), the predicted values for the testing data closely align with the $Y = X$ line, with the best-fit line nearly overlapping it, although most of the high FL values are underestimated. This pattern is consistent with the training and validation results for this model. For XGBoost (Fig. 8(b)), the model displays a close fit to the $Y = X$ line, with a slight underestimation at high FL values. For SVR (Fig. 8(c)), there is distinct underestimation at higher FL values of HPHT FL, which again aligns with what was observed during SVR training (Fig. 7(c))—suggesting that this bias is inherent in the model's predictions. For ELM (Fig. 8(d)), there is a more substantial tendency to overestimate at lower FL values and underestimate at higher FL values. This imbalance was also observed during ELM training (Fig. 7(d)), and its persistence with the testing subset highlights the model's limited ability to generalize across the full range of HPHT FL values.

Table 5 compares the HPHT FL prediction performances for the best-performing ML models across the training, validation, and testing subsets using error metrics and R^2. The results reflect the models' abilities to generalize beyond the training data and maintain prediction performance when applied to new, unseen data. Among the evaluated models, XGBoost consistently outperformed the other ML models across all subsets, achieving the lowest error metrics and the highest R^2 values. This strong and stable performance across phases highlights XGBoost's superior generalization and robustness. The MLPNN model followed, showing robust FL prediction capabilities but with slightly higher errors, especially when applied to the testing subset. SVR and ELM models showed comparatively weaker performance, with the ELM model exhibiting the largest increase in error and drop in R^2 from training to testing data. This trend may indicate that ELM overfitted the training data and consequently struggled to generalize its predictions when applied to unseen data. In general, all models experienced a slight increase in FL prediction error when transitioning from training to testing.

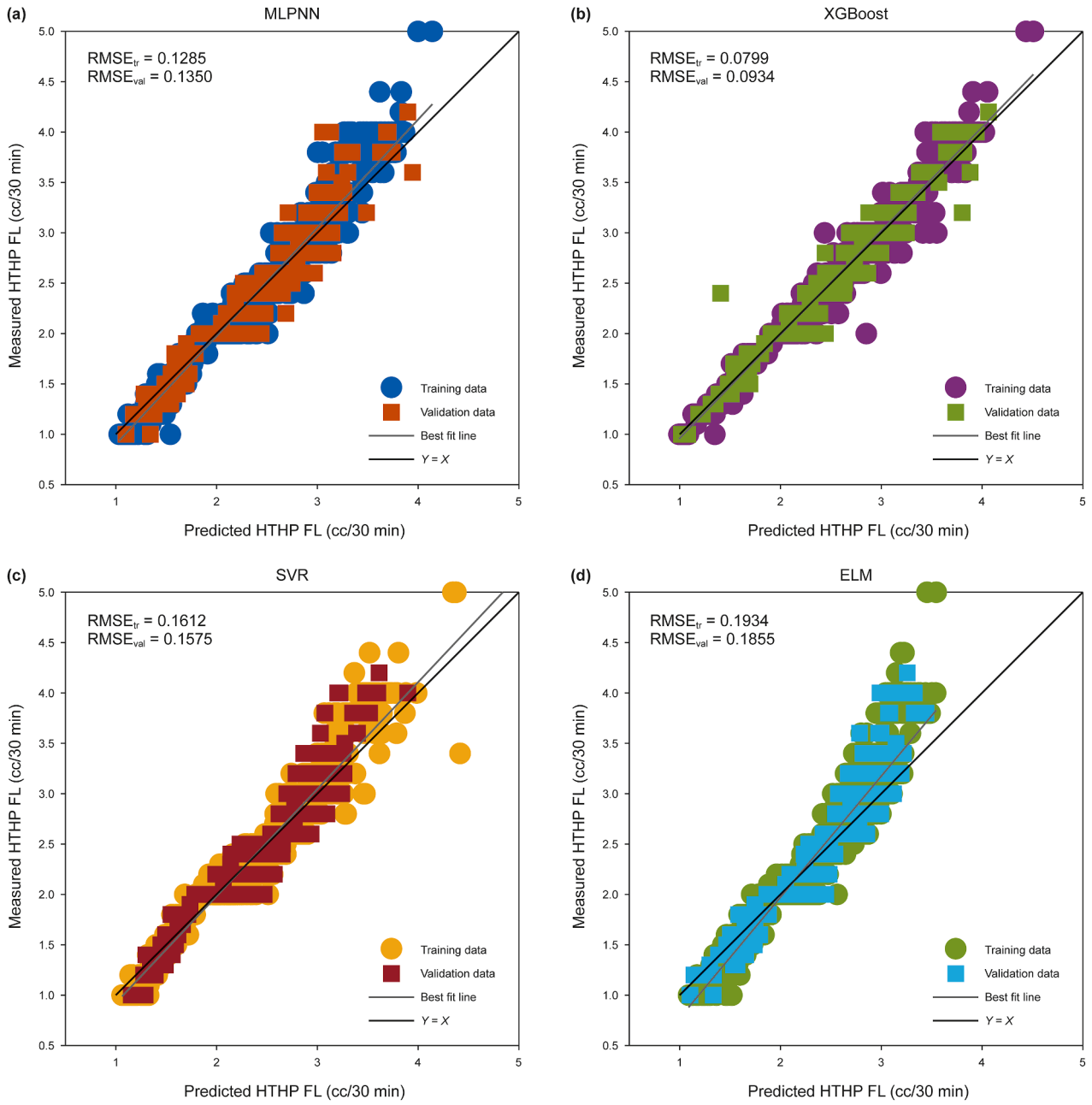

Fig. 7. Cross plots of measured versus predicted HPHT FL values for the top-performing ML models applied to the training and validation subsets. The $Y = X$ line represents perfect prediction. Deviation from the best-fit line to the right from the $Y = X$ line indicates overestimation, while deviation to the left indicates underestimation. The dispersion of data points relative to the $Y = X$ line reflects the prediction performance and consistency of the models. The terms $RMSE_{tr}$ and $RMSE_{val}$ represent training and validation RMSE, respectively.

5. Discussion

This section evaluates the FL prediction performance by applying residual analysis, uncertainty quantification, overfitting detection, and robustness sensitivity tests. Shapley additive explanations (SHAP) analysis determines feature importance, guiding optimal model selection while revealing key input-output relationships for drilling-fluid-prediction applications.

5.1. Residual analysis

Fig. 9 displays error-residual plots for the four developed ML models, revealing differences between the measured and predicted HPHT FL values across the entire dataset.

From Fig. 9, it is possible to identify potential biases, trends, and/or FL prediction outliers. The MLPNN model (top-left) displays a dense cluster of residuals around zero, although a slight upward spread is observed at higher HPHT FL values, indicative of underestimation at high FL values. The XGBoost model (top-right plot) exhibits a relatively uniform distribution of residuals with a tendency to underestimate in the FL value range between 0.3 and 0.5 cc/30 min. The SVR model (bottom-left plot) exhibits a general distribution of residuals around zero; however, a few large deviations in the FL value range 2 to 3.5 cc/30 min highlight the model's sensitivity to specific data points and a tendency to underestimate FL values > 3.5 cc/30 min. The ELM model (bottom-right plot) shows a more pronounced upward trend in residuals as values increase across the entire FL-value range, implying

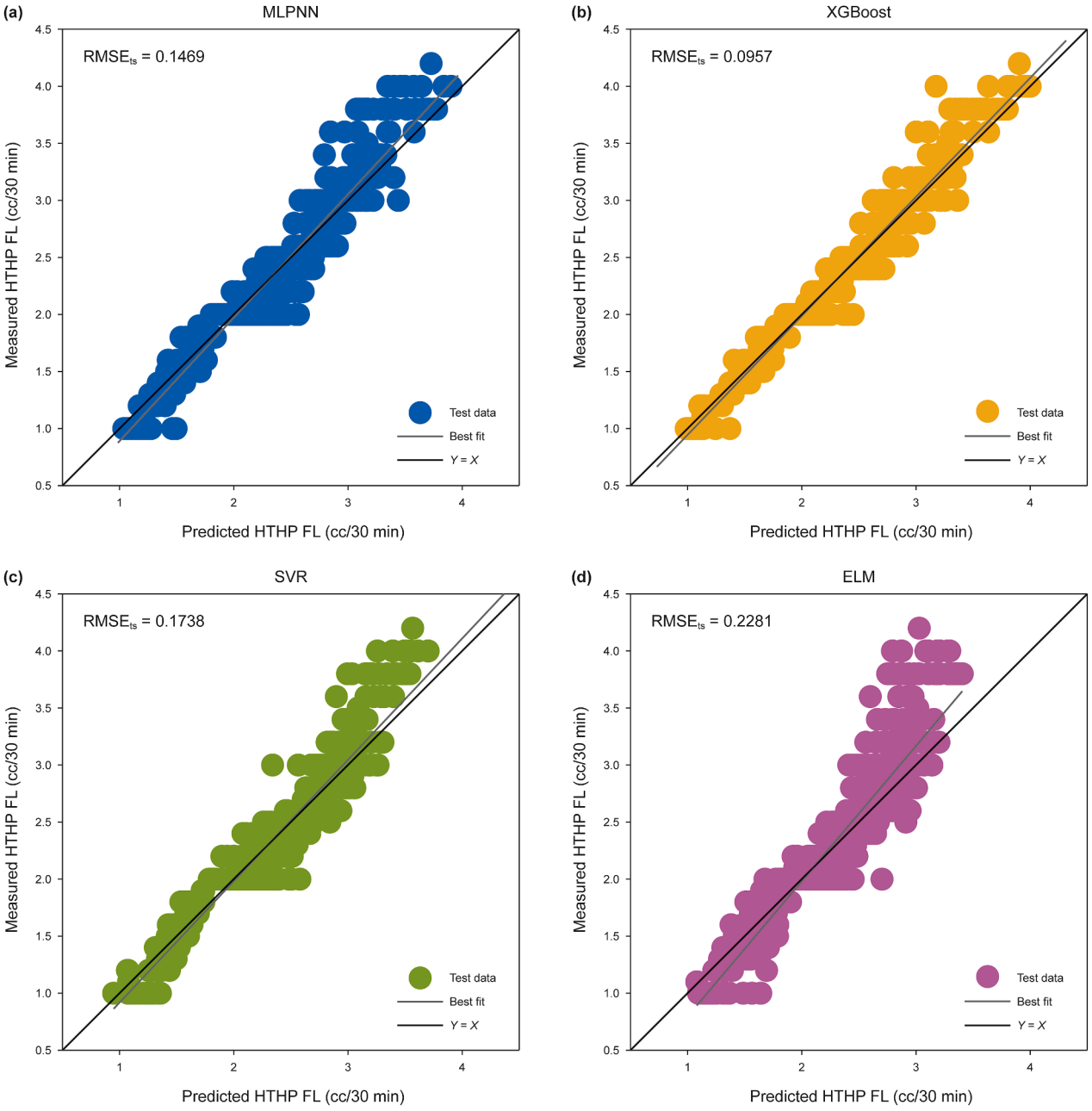

Fig. 8. Cross-plot of measured versus predicted high pressure high temperature fluid loss (HPHT FL) values for the top-performing machine learning models on the testing dataset. The $Y = X$ line represents perfect prediction. Deviation of the best-fit line to the right from the $Y = X$ line indicates overestimation at lower HPHT values, while deviation to the left indicates underestimation. The dispersion of data points relative to the $Y = X$ line reflects the prediction performance and consistency of the models. The term $RMSE_{ts}$ refers to the RMSE achieved by the testing subset.

systematic underestimation by the model for higher FL values and overestimation at the low end of the FL value range. These residual plots confirm that the XGBoost model offers superior FL predictive performance with minimal bias, despite a tendency to slightly underestimate at the high end of the FL value range, while the other ML models display larger and more systematic FL prediction errors.

5.2. Uncertainty assessment

The confidence bound width (WCB) quantifies prediction uncertainty by defining the FL value range where model outputs are statistically expected to lie. A narrow WCB reflects precise predictions with tightly clustered values, while a broad WCB indicates higher uncertainty and dispersed results. The absolute errors (ε) between the measured and predicted values are computed using Eq. (8) to determine the WCB.

$$\varepsilon_i = \left|\mathrm{HPHTFL}_{i,\mathrm{exp}} - \mathrm{HPHTFL}_{i,\mathrm{pred}}\right| \tag{8}$$

These errors are then analyzed to derive:

- SD of errors via Eq. (9), measuring error dispersion:

Table 5
Comparison of HPHT FL prediction performances of the top-performing ML models with the training, validation, and testing subsets.

Model	Phase	Error metric				
		ARE	AARE, %	R^2	SD	RMSE, cc/30 min
MLPNN	Training	0.0066	3.64	0.9547	0.1285	0.1285
	Validation	0.0072	3.88	0.9378	0.1351	0.1350
	Testing	0.0099	4.35	0.9348	0.1469	0.1469
XGBoost[a]	Training	0.0028	1.98	0.9814	0.0800	0.0799
	Validation	0.0021	2.28	0.9693	0.0935	0.0934
	Testing	0.0039	2.57	0.9717	0.0957	0.0957
SVR	Training	0.0056	4.44	0.9245	0.1612	0.1612
	Validation	0.0047	4.88	0.9145	0.1576	0.1575
	Testing	0.0080	5.24	0.9045	0.1739	0.1738
ELM	Training	−0.0004	5.34	0.9150	0.1911	0.1934
	Validation	0.0000	5.18	0.8980	0.1843	0.1855
	Testing	0.0020	6.75	0.8536	0.2262	0.2281

[a] Represents the best-performing model of the four ML models evaluated.

$$\mathrm{SD} = \sqrt{\frac{\sum_{i=1}^{N} (\varepsilon_i - \mathrm{MOE})^2}{N-1}} \tag{9}$$

- Standard error (SE) using Eq. (10), where smaller SE values indicate higher prediction reliability:

$$\mathrm{SE} = \frac{\mathrm{SD}}{\sqrt{N-1}} \tag{10}$$

For a 95% confidence level ($z = 1.96$), the margin of error (ME) and mean of error (MOE) are calculated with Eqs. (11) and (12).

$$\mathrm{ME} = z \times \mathrm{SE} \tag{11}$$

$$\mathrm{MOE} = \frac{\sum_{i=1}^{N} \varepsilon_i}{N} \tag{12}$$

The confidence bounds (lower (LB) and upper (UB)) and WCB are derived from Eqs. (13)–(15).

$$\mathrm{LB} = \mathrm{MOE} - \mathrm{ME} \tag{13}$$

$$\mathrm{UB} = \mathrm{MOE} + \mathrm{ME} \tag{14}$$

$$\mathrm{WCB} = \mathrm{UB} - \mathrm{LB} \tag{15}$$

Table 6 presents uncertainty analysis of FL predictions for the MLPNN, XGBoost, SVR, and ELM models, quantifying performance across six statistical metrics. This analysis confirms that XGBoost is the best-performing model, achieving the lowest MOE (0.0588) and the narrowest WCB (0.0081), which reflect its exceptional predictive consistency and robustness. In contrast, ELM performs

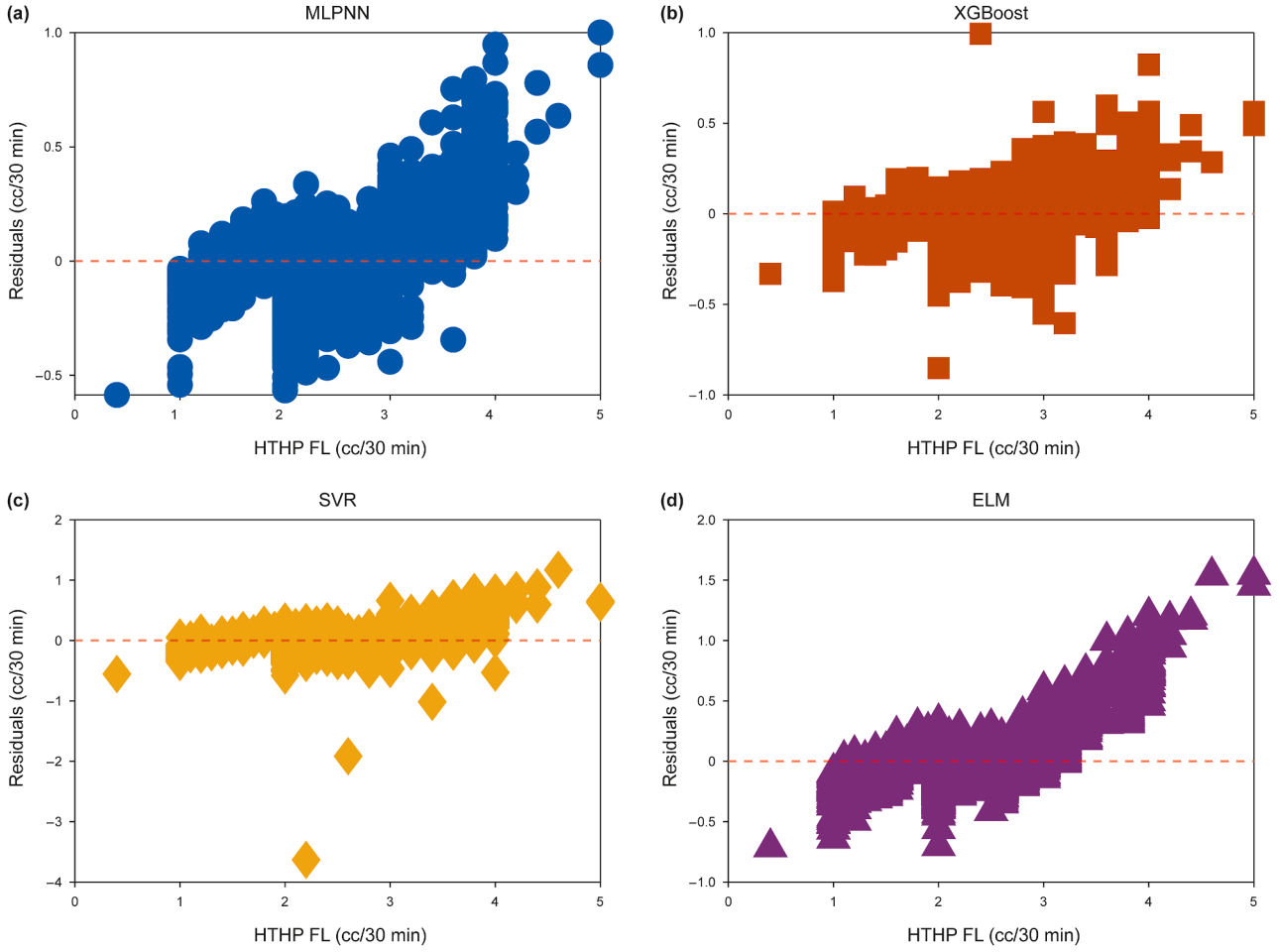

Fig. 9. Residual plots for the developed models that identify the differences between measured and predicted HPHT FL values across the entire dataset. The plots provide insights into the distribution and patterns of residuals, highlighting potential biases, trends, or outliers in model FL predictions. A high-performing prediction model is expected to exhibit residuals randomly scattered around zero (red dashed line), indicating minimal systematic error and robust predictive performance.

Table 6
The results of the uncertainty analysis conducted for the FL predictions of four ML models.

Model	MOE	SD	SE	ME	LB	UB	WCB
MLPNN	0.0978	0.1097	0.0030	0.0059	0.0919	0.1036	0.0118
XGBoost[a]	0.0588	0.0755	0.0021	0.0040	0.0547	0.0628	0.0081
SVR	0.1195	0.1262	0.0035	0.0068	0.1127	0.1263	0.0135
ELM	0.1564	0.1661	0.0045	0.0089	0.1475	0.1653	0.0178

[a] Represents the best-performing model.

the least favorably, with the highest MOE (0.1564) and WCB (0.0178). MLPNN and SVR exhibit intermediate performance, with WCB values of 0.0118 and 0.0135, respectively. Thus, based on the WCB values, the models can be ranked as XGBoost > MLPNN > SVR > ELM.

Fig. 10 illustrates the uncertainty analysis results presented in Table 6. It highlights XGBoost's dominance via its shortest bars (WCB ≈ 0.02–0.04) and reveals ELM's pronounced variability through the tallest bars (WCB ≈ 0.14–0.16).

5.3. ML model confidence limit and overfitting assessments

The bootstrapping technique was employed to further assess the uncertainty of the FL prediction models. Bootstrapping is a resampling method that involves repeatedly drawing samples from a dataset with replacement to generate multiple synthetic datasets. By training and testing ML models on multiple resampled datasets, the variability and stability of the prediction models can be quantified. This approach allows for the calculation of confidence intervals without relying on strong parametric assumptions, making it particularly suitable for datasets involving substantially non-linear relationships among their variables. Confidence limits derived in this way also provide valuable insights into an ML model's generalization ability and tendencies to overfit or underfit their trained subsets by comparing results from training, validation, and testing subsets.

Table 7 summarizes the 95% confidence intervals calculated by the bootstrapping technique to evaluate the overfitting behavior of the four ML models developed to predict HPHT FL. The LB and UB results derived from Eqs. (13) and (14) indicate that MLPNN, XGBoost, and SVR models generate narrow confidence intervals centered close to zero for training, validation, and testing subsets indicative of stable and unbiased FL predictions. In contrast, the ELM model shows noticeably higher and lower upper bounds, particularly in the training and testing phases, indicating greater prediction bias and overfitting compared to the other models.

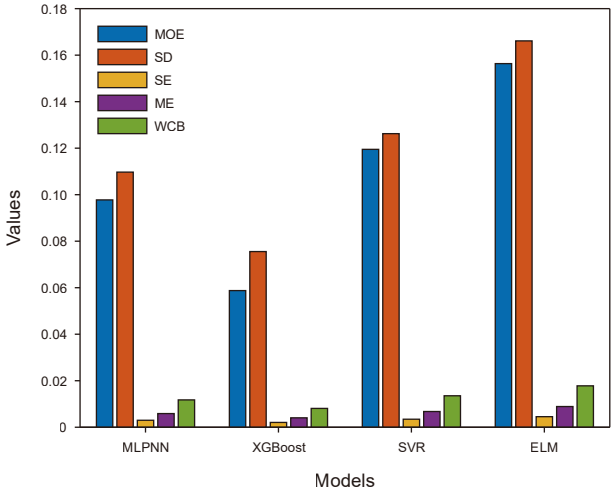

Fig. 10. Performance comparison for the developed models based on five uncertainty metrics: WCB, ME, SE, SD and MOE. For all metrics, lower values indicate superior model performance.

Fig. 11 displays the mean FL prediction error with the 95% confidence limits expressed as black error bars for the MLPNN, XGBoost, SVR, and ELM models applied to the training, validation, and testing subsets. The blue, red, and orange bars represent the mean errors for the training, validation, and testing subsets, respectively. The superior performance of the XGBoost model and the inferior performance of the ELM model are clearly apparent in Fig. 6.

5.4. Robustness analysis

To assess the robustness of the developed models, a noise-injection analysis was conducted. This method involves adding white noise to the input-variable value distributions to simulate real-world data variability and evaluate how well the models maintain their predictive performance under noisy conditions. The noise was generated by adjusting the variance parameter "s" in the probability density function $P(x)$, as defined by Eq. (16).

$$P_x(x) = \frac{1}{s\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2s}} \tag{16}$$

where μ represents the mean (set to zero) and "s" controls the standard deviation of the noise. By systematically increasing "s" from 0 to 0.3, different levels of noise intensity were introduced to the input-variable distributions.

At each noise level, the models were applied to the noisy input features, and their FL prediction performance assessed using the R^2 metric. More robust models exhibit smaller declines in R^2 values as noise levels increase.

Fig. 12 illustrates the results of the noise-injection analysis, distinguishing the XGBoost as the most robust of the four ML models evaluated, due to its low decline in R^2 (shallowest slope) as noise intensity grows. XGBoost's sophisticated error regularization and its stochastic gradient boosting help it to mitigate overfitting to a degree when confronted with noisy features. In contrast, the MLPNN and ELM display substantial and progressive negative impacts on their FL prediction performance as the level of injected noise increases.

5.5. Feature influence and importance analysis

SHAP analysis is a powerful interpretability tool that explains ML predictions by quantifying the contribution of each input variable to the prediction of each data record in a dataset. It facilitates the identification of key features, clarifies their influence on model outcomes, and visualizes how individual feature values impact specific predictions. The analysis provides global insights into overall ML model behavior with a specific dataset, as well as insights into individual data-record predictions. SHAP improves model transparency, enhances trust in model outputs, and aids in feature selection and model refinement.

SHAP analysis was conducted on the best-performing XGBoost model to determine the influence of each input feature on HPHT FL predictions and to provide a deeper understanding of the model's prediction mechanism. Fig. 13(a) displays the SHAP values for each predictor variable applied to each data record, with the color scaled from blue (low feature value) to pink (high feature value). Positive SHAP values in this plot (horizontal axis) indicate that a feature's value improves the FL prediction of that data record,

Table 7
Computed 95% confidence intervals using a bootstrapping technique for overfitting assessment of developed HPHT FL predictive models by comparing training, validation, and testing subset results. LB represents the lower bound, and UB represents the upper bound of the confidence intervals.

Subset		Models			
		MLPNN	XGBoost	SVR	ELM
Training	LB	−0.0057	−0.0040	−0.0034	0.0223
	UB	0.0045	0.0024	0.0097	0.0377
Validation	LB	−0.0158	−0.0073	−0.0096	0.0081
	UB	0.0061	0.0083	0.0148	0.0360
Testing	LB	−0.0117	−0.0052	−0.0094	0.0168
	UB	0.0034	0.0051	0.0102	0.0426

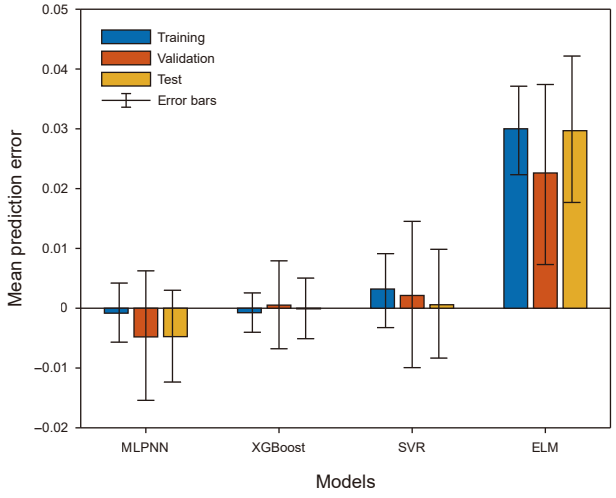

Fig. 11. 95% confidence intervals for mean prediction errors associated with training, validation, and testing subsets computed with a bootstrapping technique to assess overfitting in ML models of HPHT FL. Where both ends of the error bars (UB and LB) have positive values, it indicates that a model is substantially underfitting the data (e.g. ELM). Where both ends of the error bars (UB and LB) have negative values, it indicates that a model is substantially overfitting the data, but none of the models evaluated display that circumstance. Error bar length reflects prediction variability, with shorter bars indicating greater consistency. Models with narrower error-bar intervals and smaller mean errors are more reliable, with XGBoost models clearly outperforming the other three ML models for FL prediction.

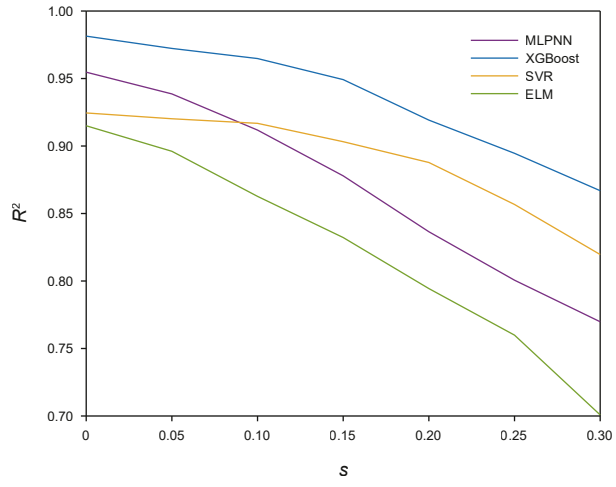

Fig. 12. Robustness analysis of the four ML models developed to predict FL expressed in terms of R^2 as increasing levels of synthetic white noise are introduced to the input-variable value distributions.

whereas negative SHAP values indicate that a feature's value degrades the FL prediction of that data record. Notably, the ES variable exhibits the widest spread and a distinctive bimodal distribution of SHAP values, with the lowest ES values generating the highest SHAP values and the highest ES values generating the lowest SHAP values. The other input features (FV, MT, MW, and MA) show narrower SHAP distributions with comparatively smaller impacts on FL predictions. Although the MW variable exhibits a much narrower spread of SHAP values than ES, its impacts on FL predictions are in the opposite directions, with the highest MW values generating the highest SHAP values and the lowest MW values generating the lowest SHAP values. Overall, it is apparent from Fig. 13(a) that ES has the dominant influence on FL predictions.

Fig. 13(b) provides a bar chart summarizing the mean absolute SHAP values for each feature, ranking them (most influential at the top) based on their overall influence on FL predictions. Consistent with the summary plot, ES emerges as the most influential predictor, followed distantly by MW, MT, MA, and FV. This data-driven insight from SHAP aligns directly with the mechanistic framework established in Table 2. The dominance of ES as the most influential parameter directly reflects the critical role of the emulsifier package (e.g., MEGAMUL, VERSAWET) in controlling filtration loss. A high ES value indicates a robust, tightly-packed emulsion with strong interfacial films, which is the primary barrier to filtrate invasion. Consequently, the model correctly identifies the functional outcome of effective emulsification as the most critical factor for low FL. The secondary influence of MW relates to the overall solids content, including weighting material (e.g., barite) and colloidal particles, which contribute to building a low-permeability filter cake. Therefore, the model successfully leverages macroscopic, readily-measurable properties that serve as proxies for the underlying drilling fluids chemical composition and its effectiveness.

6. Limitations and recommendations for future study

Although the outcomes of the comprehensive evaluation firmly confirm the significant predictive ability of the ML models for determining HPHT FL for the major fluid type (mineral oil-based invert emulsion fluids) covered in the applied dataset, the following limitations should be considered:

- The model's development focused specifically on the high-frequency measurements available at the wellsite to maintain real-time applicability. Consequently, it does not include detailed compositional parameters such as precise additive concentrations or particle size distributions, which are known to influence filtrate loss but are measured infrequently. To

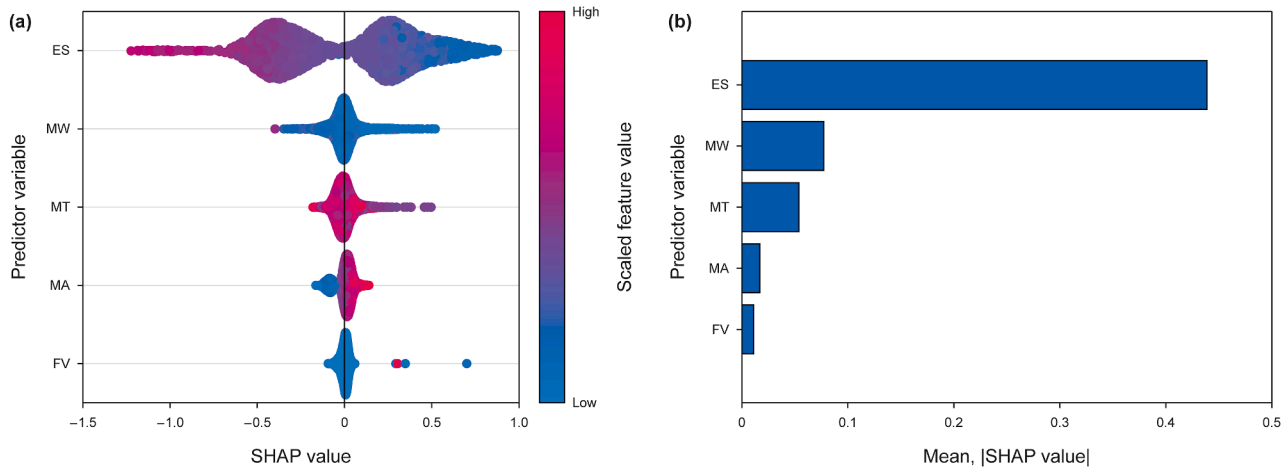

Fig. 13. Visualizing the influence of each input feature on HPHT FL predictions with SHAP values for the XGBoost model: (**a**) SHAP detailed feature impact plot and (**b**) SHAP summary plot distinguishing overall relative feature importance.

address this challenge, it is recommended to integrate the continuous stream of high-frequency wellsite data with periodic, in-depth laboratory measurements (e.g., precise additive concentrations and particle size distributions). Such a hybrid approach could significantly enhance HPHT FL prediction performance and, more importantly, provide deeper mechanistic insights and calibration into the chemical drivers of filtration loss.

- The ML models were developed and validated specifically for mineral oil-based invert emulsion fluids. Their performance on other OBDF systems (e.g., synthetic-based fluids, diesel-based fluids, or all-oil fluids) has not been established and cannot be guaranteed without further testing and potential ML retraining. Moreover, the dataset used for modeling was collected from a limited number of wells drilled in two fields, which may not capture the full geological and operational variability encountered in global HPHT drilling scenarios. To expand HPHT FL model applicability to diverse OBDF systems, future work should involve compiling datasets from various OBDF types (e.g., synthetic-based, diesel-based, and all-oil fluids) to validate the transferability of the presented methodology and to develop new, robust models tailored to those specific fluid classifications. Additionally, incorporation of multi-field datasets from wells drilled through diverse geological formations and drilling conditions is highly recommended to improve model generalizability.
- The current input parameters (MW, MT, FV, MA, and ES) were selected based on domain knowledge rather than systematic feature selection methods, potentially missing important predictors available at the well-site or including redundant ones. Consequently, future studies could apply reliable feature selection methods to identify the most influential subset of features to enhance prediction performance of the HPHT FL model.
- The training data relies on API-standard static HPHT filtration tests, which do not fully replicate dynamic downhole conditions or prolonged OBDF exposure effects. This limitation could be addressed by combining laboratory HPHT FL data with real-time downhole sensor measurements. Such an approach would better capture the dynamic interactions among the variables enabling the development of more reliable models for predicting HPHT FL in real-time. Moreover, incorporating the models into drilling fluid digital twins that can simulate and optimize fluid performance throughout the well construction process could help minimize drilling fluid-related problems during drilling operations.

7. Conclusions

This study develops robust machine learning (ML) models to predict high pressure high temperature fluid loss (HPHT FL) in oil-based drilling fluids (OBDFs). It does so by configuring and comparing the performance of four ML algorithms: multi-layer perceptron neural network (MLPNN), extreme learning machine (ELM), support vector regression (SVR), and extreme gradient boosting (XGBoost). Field data from drilling wellsite laboratories were collated, containing property measurements of mineral oil-based invert emulsion fluids to compile the modeled dataset. Following the multiphase preprocessing and division of the collected dataset into subsets, the training data were analyzed using *k*-fold cross-validation to determine the appropriate validation subset size. To ensure statistical reliability, each algorithm was executed with five randomly selected subsets for training and validation. The developed top-performing models were comprehensively assessed using four evaluation techniques: residual analysis, uncertainty quantification, overfitting detection, and robustness testing. The best-performing ML model was further examined by Shapley additive explanations (SHAP) to determine the relative influences of the five input parameters on the prediction target, HP-HT FL. The key findings of this study are as follows:

- The identification of an optimal 70:30 data split and the use of 5-fold cross-validation, coupled with the removal of 69 data outliers, created a robust foundation for model development. This satisfies the objective of ensuring model generalizability and reliability by rigorously addressing data quality and partitioning.
- Comparative analysis revealed that the XGBoost algorithm delivered superior accuracy and the lowest prediction error, achieving root mean square error (RMSE) values of 0.0957, 0.0799, and 0.0934 cc/30 min for the testing, training, and validation phases, respectively. This model also exhibited the highest reproducibility across multiple trials, which directly fulfills one of the primary objectives of the current study, namely the development of a highly accurate and precise predictive model for HPHT FL.
- Comprehensive evaluations, including residual analysis, uncertainty quantification (showing the narrowest confidence bound width), and overfitting checks, proved the XGBoost model makes precise, unbiased predictions of HPHT FL on new data. This satisfies the critical objective of ensuring the model's

practical robustness and trustworthiness for real-world decision-making.

- The noise-injection analysis demonstrated that XGBoost's performance degraded the least in the presence of simulated data noise. This meets the objective of creating a robust model capable of handling the inherent variability and uncertainty of routine wellsite measurements.
- SHAP analysis confirmed electrical stability (ES) as the most critical input parameter, validating the model's alignment with physical principles. This fulfills the study's objective to not only predict but also to interpret the results, providing fluid engineers with a scientifically sound basis for proactive fluid management.

It is apparent from the results and interpretation presented that an ML model has been successfully developed to provide near-real-time prediction of HPHT FL. The XGBoost model leverages only five drilling fluid input parameters (mud weight (MW), mud temperature (MT), mud alkalinity (MA), electrical stability (ES), Marsh funnel viscosity (FV)) that are routinely measured every 15–20 min at the wellsite. This directly satisfies the objective of developing a near-real-time prediction model for HPHT FL, effectively eliminating the 1–2-h delay of laboratory tests. By integrating high accuracy, proven robustness, and clear interpretability, this model provides a substantial improvement over conventional methods, empowering drilling engineers with a powerful tool for frequent monitoring of HPHT FL of the considered type of OBDF during drilling operations.

CRediT authorship contribution statement

Shadfar Davoodi: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Arslan Gulniyazov:** Writing – review & editing, Writing – original draft, Visualization, Software, Investigation, Formal analysis, Conceptualization. **David A. Wood:** Writing – review & editing, Visualization, Validation, Methodology, Formal analysis. **Mohammed Al-Shargabi:** Writing – review & editing, Visualization, Project administration, Investigation, Formal analysis. **Nikita Makarov:** Writing – original draft, Visualization, Formal analysis, Conceptualization. **Evgeny Burnaev:** Writing – review & editing, Validation, Supervision, Resources, Formal analysis.

Acknowledgements

The work was supported by the grant for research centers in the field of AI provided by the Ministry of Economic Development of the Russian Federation in accordance with the agreement 000000C313925P4F0002 and the agreement with Skoltech N№139-10-2025-033.

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.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.petsci.2026.03.053.

References

Abdelaal, A., Ibrahim, A.F., Elkatatny, S., 2023. Data-driven framework for real-time rheological properties prediction of flat rheology synthetic oil-based drilling fluids. ACS Omega 8, 14371–14386. https://doi.org/10.1021/ACSOMEGA.2C06656.

Aggarwal, C.C., 2017. An Introduction to Outlier Analysis. Springer. https://doi.org/10.1007/978-3-319-47578-3_1.

Ahmadi, M.A., Shadizadeh, S.R., Shah, K., Bahadori, A., 2018. An accurate model to predict drilling fluid density at wellbore conditions. Egypt. J. Pet. 27, 1–10. https://doi.org/10.1016/j.ejpe.2016.12.002.

Ali Khan, M., Khan, M.A., Al-Salim, H.S., Arsanjani, L.N., Akademia Baru, P., 2018. Development of high temperature high pressure (HTHP) water based drilling mud using synthetic polymers, and nanoparticles. J. Adv. Res. Fluid Mech. Therm. Sci. 45, 99–108.

Al-Shargabi, M., Davoodi, S., Wood, D.A., Al-Musai, A., Rukavishnikov, V.S., Minaev, K.M., 2022. Nanoparticle applications as beneficial oil and gas drilling fluid additives: A review. J. Mol. Liq. 352, 118725. https://doi.org/10.1016/J.MOLLIQ.2022.118725.

API 13B-2, 2005. Recommended Practice for Field Testing of Oil-based Drilling Fluids.

Arain, A.H., Ridha, S., Ilyas, S.U., Mohyaldinn, M.E., Suppiah, R.R., 2022. Evaluating the influence of graphene nanoplatelets on the performance of invert emulsion drilling fluid in high-temperature wells. J. Pet. Explor. Prod. Technol. 12, 2467–2491. https://doi.org/10.1007/s13202-022-01501-5.

Awad, M., Khanna, R., Awad, M., Khanna, R., 2015. Support Vector Regression. Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers 67–80. https://doi.org/10.1007/978-1-4302-5990-9_4.

Bengio, Y., Goodfellow, I., Courville, A., 2017. Deep Learning. MIT press, Cambridge, MA, USA.

Bentéjac, C., Csörgő, A., Martínez-Muñoz, G., 2021. A comparative analysis of gradient boosting algorithms. Artif. Intell. Rev. 54, 1937–1967. https://doi.org/10.1007/s10462-020-09896-5.

Bishop, C.M., Nasrabadi, N.M., 2006. Pattern Recognition and Machine Learning. Springer.

Bridges, S., Robinson, L.H., 2020. A practical handbook for drilling fluids processing. Gulf Professional. https://doi.org/10.1016/C2019-0-00458-X.

Caenn, R., Darley, H.C.H., Gray, G.R., 2016. Composition and Properties of Drilling and Completion Fluids, seventh ed. Gulf Professional Publishing. https://doi.org/10.1016/c2015-0-04159-4.

Chang, C.C., Lin, C.-J., 2011. LIBSVM: A library for support vector machines. ACM Trans. Intell. Syst. Technol. 2, 1–27. https://doi.org/10.1145/1961189.1961199.

Chatterjee, S., Hadi, A.S., 1986. Influential observations, high leverage points, and outliers in linear regression. Stat. Sci. 379–393.

Chen, T., Guestrin, C., 2016. A scalable tree boosting system. In: Proc 22nd ACM SIGKDD Int Conf Knowl Discov Data Min, pp. 785–794. https://doi.org/10.1145/2939672.2939785.

Cherkassky, V., Ma, Y., 2004. Practical selection of SVM parameters and noise estimation for SVM regression. Neural Netw. 17, 113–126. https://doi.org/10.1016/S0893-6080(03)00169-2.

Davoodi, S., Geri, M.B., Wood, D.A., Al-Shargabi, M., Mehrad, M., Soleimanian, A., 2025. Predicting water-based drilling fluid filtrate volume in close to real time from routine fluid property measurements. Petroleum 11 (2), 174–187. https://doi.org/10.1016/j.petlm.2025.03.002.

Davoodi, S., Mehrad, M., Wood, D.A., Ghorbani, H., Rukavishnikov, V.S., 2023. Hybridized machine-learning for prompt prediction of rheology and filtration properties of water-based drilling fluids. Eng. Appl. Artif. Intell. 123, 106459. https://doi.org/10.1016/j.engappai.2023.106459.

Ding, S., Xu, X., Nie, R., 2014. Extreme learning machine and its applications. Neural Comput. Appl. 25, 549–556. https://doi.org/10.1007/s00521-013-1522-8.

Drucker, H., Burges, C.J., Kaufman, L., Smola, A., Vapnik, V., 1996. Support vector regression machines. Adv. Neural Inf. Process. Syst. 9.

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

Géron, A., 2022. Hands-on Machine Learning with Scikit-Learn, Keras, and Tensorflow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media, Inc.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT press.

Gul, S., van Oort, E., 2020. A machine learning approach to filtrate loss determination and test automation for drilling and completion fluids. J. Pet. Sci. Eng. 186. https://doi.org/10.1016/j.petrol.2019.106727.

Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. Neural Netw. 2, 359–366. https://doi.org/10.1016/0893-6080(89)90020-8.

Hsu, C.Y., Kumar, A., Hussain, M., Yajid, M.S.A., Beemkumar, N., Ojha, M.K., Singh, V., Jain, V., Singh, P., Sherzod, S., 2025. Synthesis of polylactic acid/Henna polymer composite and its application in optimizing drilling fluid rheology and filtration performance. Sci. Rep. 15, 38942. https://doi.org/10.1038/s41598-025-22807-4.

Huang, G. Bin, Zhu, Q.Y., Siew, C.K., 2006. Extreme learning machine: Theory and applications. Neurocomputing 70, 489–501. https://doi.org/10.1016/j.neucom.2005.12.126.

Huang, G., Huang, G.-B., Song, S., You, K., 2015. Trends in extreme learning machines: A review. Neural Netw. 61, 32–48. https://doi.org/10.1016/j.neunet.2014.10.001.
Khodja, M., Debih, H., Lebtahi, H., Amish, M.B., 2022. New HTHP fluid loss control agent for oil-based drilling fluid from pharmaceutical waste. Clean. Eng. Technol. 8, 100476. https://doi.org/10.1016/j.clet.2022.100476.
Kutner, M.H., Nachtsheim, C.J., Neter, J., Li, W., 2005. Applied Linear Statistical Models. McGraw-hill.
LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539.
Leporini, M., Marchetti, B., Corvaro, F., di Giovine, G., Polonara, F., Terenzi, A., 2019. Sand transport in multiphase flow mixtures in a horizontal pipeline: An experimental investigation. Petroleum 5, 161–170. https://doi.org/10.1016/J.PETLM.2018.04.004.
Li, Q., de Viguerie, L., Laporte, L., Berraud-Pache, R., Zhuang, G., Souprayen, C., Jaber, M., 2024. Oily bioorganoclays in drilling fluids: Micro and macroscopic properties. Appl. Clay Sci. 247, 107186. https://doi.org/10.1016/j.clay.2023.107186.
Mahmoud, A., Gajbhiye, R., Elkatatny, S., 2025. Enhanced performance of oil-based drilling fluids under HPHT conditions using an organophilic phyllosilicate. Sci. Rep. 15, 22447. https://doi.org/10.1038/s41598-025-05864-7.
Montgomery, D.C., Peck, E.A., Vining, G.G., 2021. Introduction to Linear Regression Analysis. John Wiley & Sons.
Osman, E.A., Aggour, M.A., 2003. Determination of drilling mud density change with pressure and temperature made simple and accurate by ANN. Proc. Middle East Oil Show 13, 115–126. https://doi.org/10.2118/81422-ms.
Probst, P., Wright, M.N., Boulesteix, A., 2019. Hyperparameters and tuning strategies for random forest. Wiley Interdiscip. Rev., Data Min. Knowl. Discov. 9, e1301. https://doi.org/10.1002/widm.1301.
Raschka, S., Mirjalili, V., 2019. Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and Tensorflow 2. Packt Publishing Ltd.
Smola, A.J., Schölkopf, B., 2004. A tutorial on support vector regression. Stat. Comput. 14, 199–222. https://doi.org/10.1023/B:STCO.0000035301.49549.88.
Soares, A.S.F., Scheid, C.M., Marques, M.R.C., Calçada, L.A., 2020. Effect of solid particle size on the filtration properties of suspension viscosified with carboxymethylcellulose and xantham gum. J. Pet. Sci. Eng. 185, 106615. https://doi.org/10.1016/J.PETROL.2019.106615.
Sun, J., Yang, J., Bai, Y., Lyu, K., Liu, F., 2024. Research progress and development of deep and ultra-deep drilling fluid technology. Petrol. Explor. Dev. 51, 1022–1034. https://doi.org/10.1016/S1876-3804(24)60522-7.
Theodoridis, S., 2015. Machine Learning: A Bayesian and Optimization Perspective. Academic press.
Vivas, C., Salehi, S., 2021. Rheological investigation of effect of high temperature on geothermal drilling fluids additives and lost circulation materials. Geothermics 96, 102219. https://doi.org/10.1016/j.geothermics.2021.102219.
Zhong, R., Salehi, C., Johnson, R., 2022. Machine learning for drilling applications: A review. J. Nat. Gas Sci. Eng. 108, 104807. https://doi.org/10.1016/j.jngse.2022.104807.