Bulletin of Geophysics and Oceanography Vol. XX, n. X, pp. XXX-XXX; Xxxxx 20xx
DOI 10.4430/bgo00516

A weighted machine learning ensemble for predicting shear slowness in the Tensleep Formation, SE Wyoming (USA)

S. Amoura[1], R. Ketteb[1], K. Lounnas[2], S. Gaci[3], S. Barbosa[4], A.Boulassel[1], H.r Zairi[5] and Y. Kiche[6]

[1] *Geophysics Lab., University of Science and Technology Houari Boumediene, Algiers, Algeria*
[2] *Scientific and Technical Research Center for the Development of Arabic Language, Algiers, Algeria*
[3] *Sonatrach- Central Directorate of Research and Development, Boumerdes, Algeria*
[4] *Departamento de Ciencias da Terra e GeoBioTec, FCT Universidade NOVA, Lisbon, Portugal*
[5] *Biomedical Lab., University of Science and Technology Houari Boumediene, Algiers, Algeria*
[6] *Datageo Solutions, Pau, France*

(Received: 2 June 2025; accepted: 20 November 2025; published online: 19 December 2025)

ABSTRACT Accurate determination of shear slowness (*DTS*) is essential for well placement optimisation, mechanical rock property estimation, and reservoir characterisation. However, direct *DTS* measurement is expensive, limited in availability, and lacking repeatability. Traditional empirical methods require extensive calibration and are only valid for specific rock types. This study presents the voting regressor (VR), an ensemble machine learning (ML) technique combining Extra Trees, Random Forest, Gradient Boosting, LightGBM, support vector regressor, and multi-layer perceptron to enhance *DTS* prediction. The model uses well logs including gamma ray (*GR*), bulk density (*RHOB*), porosity index (*PHIX*), and compressional slowness (*DT*), applied to the Tensleep Formation in the Teapot Dome field, Wyoming, USA. Each model was evaluated using the coefficient of determination (R^2), mean absolute error (*MAE*), mean-squared error (*MSE*), and root mean-squared error (*RMSE*). A weighted averaging approach based on R^2 performance was used to build the VR model, achieving an R^2 of 0.96, *RMSE* of 0.19, *MAE* of 0.12, and *MSE* of 0.037. *DT*, *PHIX*, and *RHOB* were the most important features, while *GR* showed minimal impact. Validation on two unseen wells confirmed a strong generalisation of the VR (R^2 = 0.91-0.92). This work highlights the potential of ML for accurate *DTS* prediction and improved subsurface characterisation in data-limited settings.

Key words: shear slowness, voting regressor, machine learning, prediction, feature importance.

1. Introduction

Machine learning (ML) has become a transformative tool in geophysics, particularly in well logging and reservoir characterisation (Yazid and Gaci, 2022, 2024; Lawal *et al*., 2024). As subsurface formations grow increasingly complex, drilling costs rise, and challenges (such as missing well-logging data) persist, the need for accurate predictions of key logging parameters [such as shear slowness (*DTS*) and compressional slowness (*DT*)] has become critical. *DTS* plays a fundamental role in evaluating mechanical rock properties and identifying fracture zones (Boulassel *et al*., 2025a, 2025b; Ma *et al*., 2025) as well as determining fluid

types and characterising overall formation properties (Gaci and Hachay, 2017; Amoura *et al*., 2019, 2022; Boulassel *et al*., 2021; Meng *et al*., 2024). These insights are essential for optimising drilling operations, enhancing reservoir characterisation, and reducing exploration risks.

Given the high cost and technical challenges associated with direct *DTS* measurements, predictive techniques have gained considerable attention. Traditional mathematical correlations and regression models, such as those cited in Olayiwola *et al*. (2021), have been widely employed to estimate V_s (or *DTS*) from related parameters like compressional velocity (V_p), bulk density (*RHOB*), and resistivity. These techniques range from simple linear relationships, such as the Picket (1963) model ($V_S = V_P/1.9$), to more complex polynomial and multivariate equations, like the Brocher (2005) and Anemangly *et al*. (2019) formulations. While these approaches provide a foundation for *DTS* estimation, their applicability often hinges on specific lithological and depositional conditions, limiting their generalisability across diverse formations (Greenberg and Castagna, 1992; Brocher, 2005). Furthermore, formation-specific bias, wireline environmental effects, and variations in borehole conditions, which should be considered when interpreting results, can affect the applicability of any predictive model.

In contrast, ML approaches offer a powerful and flexible alternative by learning complex, nonlinear relationships directly from data. These methods leverage a wide range of well-logging inputs to predict *DTS* with improved accuracy and robustness. By effectively filling gaps in datasets and accounting for local geological variations, ML techniques can outperform traditional correlations in heterogeneous and complex reservoirs. The application of ML in geophysics can be traced back to the early use of neural networks for seismic data processing. Since then, its role has expanded significantly with the availability of big data and increased computational power. Currently, ML techniques are trained on historical well-logging data to predict critical logging parameters in new or under characterised wells, effectively reducing the need for expensive and time-consuming logging operations. This approach also offers the advantage of inferring missing data in older wells from neighbouring wells with similar geological settings (Dongapure, 2024; Nwankwo *et al*., 2024).

Various research efforts have demonstrated the efficiency of ML techniques in predicting *DTS* and *DT*. For instance, techniques such as Gradient Boosting (GBoost), Random Forest (RF), Extra Trees (ETR), Support Vector Regressor (SVR), and others have been employed individually to predict *DTS* based on available well-logging data, showing high accuracy (Yu *et al*., 2021; Cui *et al*., 2025). Furthermore, deep learning approaches, including artificial neural networks, and convolutional neural networks, have significantly enhanced prediction performance by capturing nonlinear connections and spatial dependencies in complicated reservoir formations (Kanfar *et al*., 2020; Zhang *et al*., 2022). In addition, unsupervised ML approaches, that include clustering algorithms and principal component analysis, have been utilised to categorise distinct lithofacies based on well data (Joshi and Raghuvanshi, 2021; El-Dabaa *et al*., 2024). Hybrid models, which combine ML with geostatistical methods, have also been developed to improve the spatial resolution of predictions across reservoir intervals (Al-Handhali *et al*., 2023; Zhang *et al*., 2025).

However, ML application continues to face challenges in geoscience, particularly in model generalisation and interpretation. Reservoir heterogeneities and the quality of training data can significantly affect model performance. While ML techniques can achieve high accuracy, they may be difficult to generalise across geological contexts. The evolution of ML algorithms has addressed many of these limitations. Although individual techniques such as RF, ETR, and GBoost are effective, recent advances in ML have demonstrated that combining multiple improved

algorithms into a single predictive framework can further enhance accuracy and robustness. This approach, known as voting regression (VR), integrates the strengths of different models, resulting in better generalisation and performance across diverse datasets (Drucker *et al*., 1996; Cui *et al*., 2025).

This work presents an ensemble-based approach that combines six ML models: GBoost, RF, ETR, LightGBM (LGBM), SVR, and multi-layer perceptron (MLP) to estimate *DTS* using well-logging data. Each method contributes its strengths, enabling the model to perform reliably in varying geological formations. This strategy is beneficial when direct *DTS* recordings are not available, offering a practical and efficient alternative for petrophysical analysis.

The structure of the paper is as follows: after the introduction, Section 2 describes the methodology and the approach proposed and Section 3 presents the outcomes of the suggested model, along with a discussion of the results from each executed step. The paper concludes with a final section that summarises the results.

2. Methodology and proposed approach

This study presents a hybrid ensemble framework to estimate *DTS* from well-logging data, utilising a VR that integrates six well-established ML models: ETR, RF, GBoost, LGBM, SVR, and MLP. The selection of these specific individual models was informed by their demonstrated performance in previous applications (Huang *et al*., 2020; Aziz *et al*., 2023). These models, widely recognised in geoscientific and petrophysical research (Anemangely *et al*., 2019; Xu *et al*., 2022; Cherana and Aliouane, 2024; Hassaan *et al*., 2024; Lawal *et al*., 2024), are combined in a unified ensemble to exploit their diverse learning capabilities and improve generalisation across heterogeneous subsurface settings.

The ensemble leverages a VR, which synthesises predictions from multiple models into a single output. A weighted voting strategy is employed, where the influence of each model on the final prediction is proportional to its performance during validation, thus allowing more reliable models to contribute more significantly. This avoids the unrealistic assumption of equal model efficacy inherent in simple averaging methods. By integrating tree-based, kernel-based, and neural-based learners, the models capture complex linear and nonlinear relationships within the logging data. This design minimises individual model biases and enhances prediction robustness across variables.

The contribution of each model to the final ensemble prediction is governed by a weighting scheme based on the average R^2_2 score across test wells. The individual average performance of model M_i, denoted $\overline{R}^2_i$, is computed as:

$$\overline{R_i^2} = \frac{1}{T}\sum_{j=1}^{T} R_{i,j}^2 \tag{1}$$

where T is the number of test wells and $R^2_{i,j}$ is the performance of model M_i on the j-th test set.

These averages are, then, normalised to form weights w_i satisfying $\Sigma_{i=1}^{m} w_i = 1$, where m is the number of models (Chen and Luc, 2022):

$$w_i = \frac{\overline{R_i^2}}{\sum_{k=1}^{m} \overline{R_k^2}}. \tag{2}$$

Weights are non-negative, normalised to sum to one, and computed from out-of-fold R^2 values. Near-uniform weights reflect the comparable strength of the base learners and help stabilise variance. The final ensemble prediction is obtained as a weighted sum of the individual model outputs, thereby enhancing robustness and favouring models with stronger generalisation capacity. As an avenue for future work, stacking approaches (e.g. non-negative least squares on out-of-fold predictions) could be used to estimate weights adaptively.

The model development process consists of two major phases: data preparation and ensemble construction and evaluation. In the data preparation phase, well logs were cleaned, normalised using the Yeo-Johnson power transform, and split into training and test sets with fixed random seeds to ensure reproducibility. During ensemble construction, each base model was trained individually and their outputs were aggregated under the voting scheme to yield the final regression output.

Fig. 1 illustrates the structured architecture of the ensemble workflow, which supports coordinated training and evaluation, yielding reliable and generalisable results for *DTS* prediction.

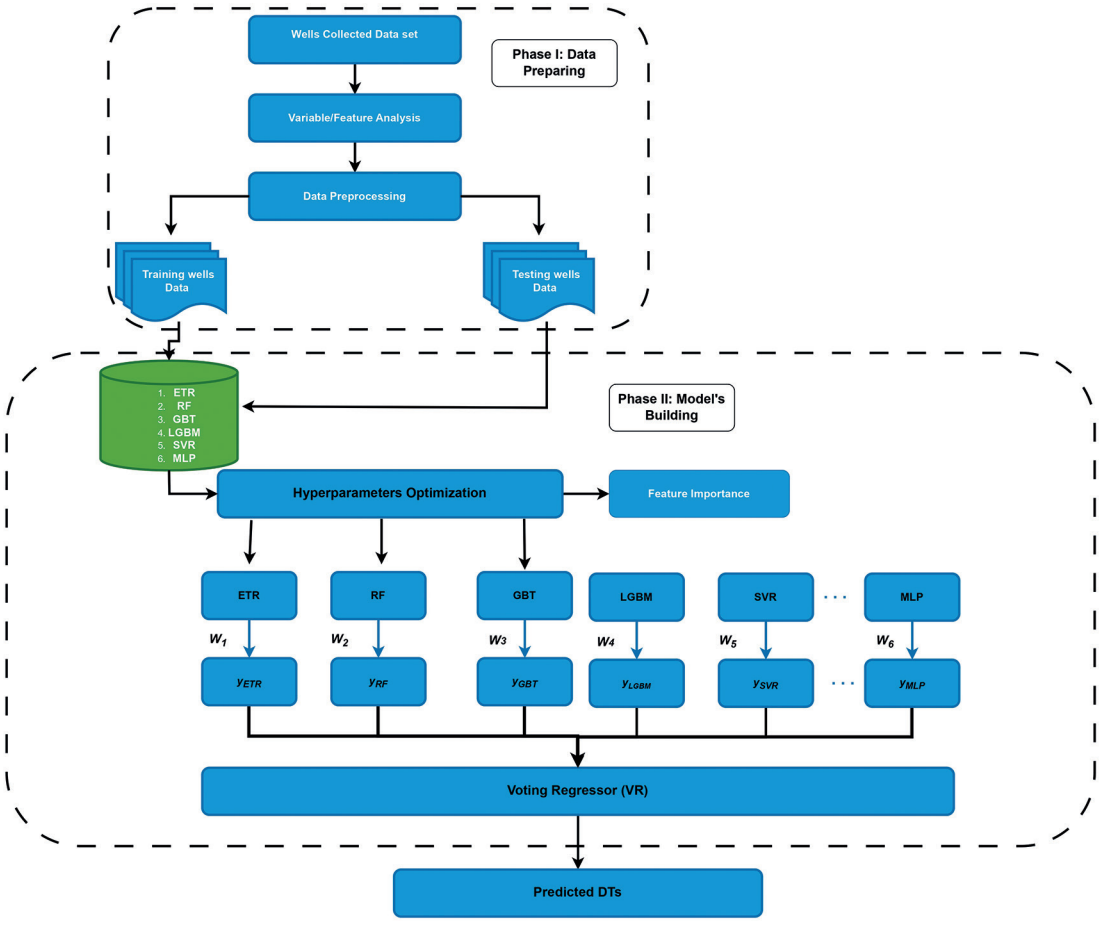

Fig. 1 - Model training and evaluation workflow.

2.1. Data collection and preprocessing

Following the acquisition of well-logging data, a systematic preprocessing pipeline is implemented to ensure data integrity and analytical reliability. The procedure includes the following steps:

a) variable selection: the input features used for model development include *GR*, *RHOB*, total porosity (*PHIX*), and *DT*, which are well-established indicators of lithological and elastic properties relevant to *DTS* prediction;
b) normalisation: to mitigate disparities in scale and units, the Yeo-Johnson power transform is applied. This method is particularly effective for stabilising variance and addressing skewed distributions, making it suitable for the typical characteristics of well-logging data;
c) statistical profiling and outlier detection: descriptive statistics (mean, standard deviation, quartiles) are computed for each feature. Box plots are used to visually detect and remove anomalous values that may distort the learning process;
d) correlation analysis: pairwise relationships among variables are explored using pair plots and correlation matrices to assess redundancy and identify the most influential predictors.

2.2. Model training and validation framework

To rigorously evaluate model performance, two validation schemes are employed. First, the models are trained on the combined data from all wells and tested individually on each well, to assess generalisation to unseen spatial domains. Second, the dataset is randomly partitioned into training (70%), test (15%), and validation (15%) subsets to evaluate temporal or intra-well variability. For reproducibility, the random seed was fixed (at 42) during partitioning.

The split was performed without stratification by well, allowing samples from all wells to be represented across training, validation, and test subsets. Because this random split was not stratified by well, the 70/15/15 results primarily assess within-field interpolation. Cross-well generalisation is, therefore, evaluated via the held-out wells (W9 and W10). Each of the six regressors is independently trained and fine-tuned using the training set. The predictions are subsequently aggregated using a weighted voting mechanism, informed by the relative predictive power of each model.

To quantify the prediction quality, four commonly used regression metrics are computed:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \tag{3}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{4}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{5}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|(y_i - \hat{y}_i)| \tag{6}$$

where y_i and $\hat{y}_i$ denote the observed and predicted *DTS* values, $\bar{y}$ is the mean of observed values, and n is the total number of samples.

3. Results and discussion

3.1. Data acquisition

We gathered well-logging data from 10 wells located in the Tensleep formation (refer to Table 1). The U.S. Department of Energy and the Rocky Mountain Oilfield Testing Center (RMOTC) manage the Teapot Dome field, which is under the ownership of the U.S. government. The Teapot Dome field, also known as Naval Petroleum Reserve-3, is an asymmetrical Laramide-age anticline on the south-western edge of the Powder River basin (Fig. 2). The direction of its axis is NW-SE (Beinkafne, 1986; Cooper *et al*., 2001). It was formed as a result of geological constraints created by the Laramide orogeny, a period of uplift and reverse subsurface movement on a W-SW thrust beginning in the Palaeocene and ending in the Lower Eocene (Fanshawe, 1971; Mankiewicz and Steidtmann, 1979).

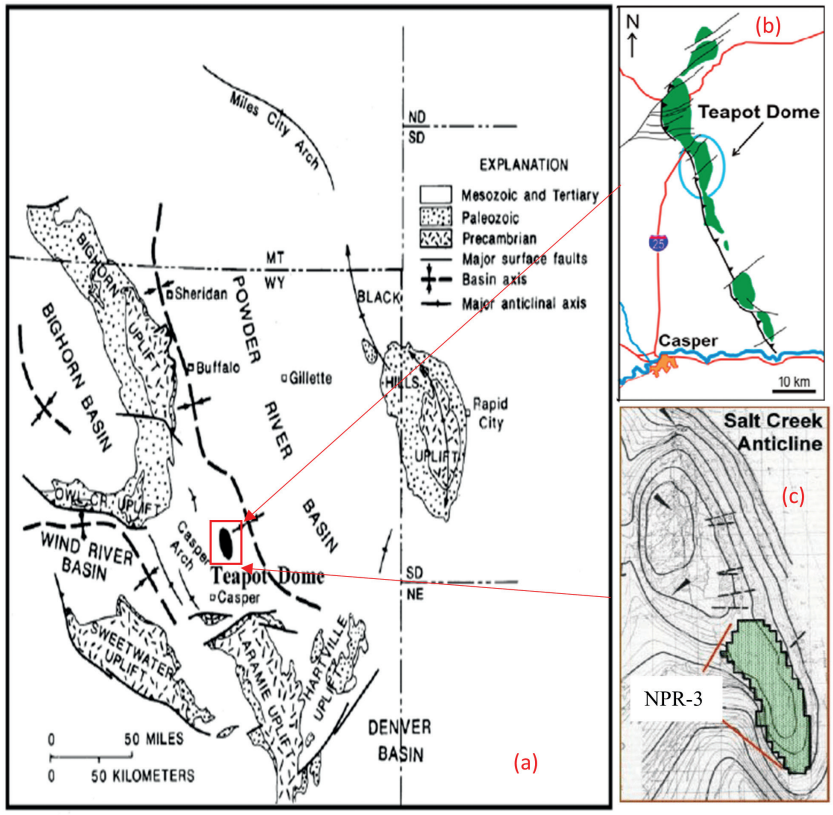

Fig. 2 - Location of the Teapot Dome field: a) map showing the geographical position of Teapot (Fox *et al.*, 1990); b) structural direction of Salt Creek (topographic relief in green); c) structural map at the top of the Teapot oilfields (RMOTC document).

Table 1 - Data acquisition and organisation for model development.

Well	Usage	Split	Features (inputs) and target (outputs)
W1 to W8	Training/testing/validation	70% training 15% testing	*GR, RHOB, PHIX, DT → DTS GR, RHOB, PHIX, DT → DTS GR, RHOB, PHIX, DT → DTS GR, RHOB, PHIX, DT → DTS*
W9 and W10	Unseen data	15% validation 100% testing	

3.2. Variable analysis

This section details the results of preprocessing of well-logging data, as well as statistical analysis and outlier removal. Fig. 3 shows an example of well-logging data for two wells (W1 and W4) in the studied depth interval. Table 2 presents the statistical description for the key features of the 10 wells studied, including *GR*, *PHIX*, *RHOB*, *DT*, and *DTS*.

Table 2 - Descriptive statistics of the key well features from 10 wells.

Statistic	***GR* (API)**	***PHIX* (v/v)**	***RHOB* (g/cm³)**	***DT* (μs/ft)**	***DTS* (μs/ft)**
Mean	91.995	0.135	2.547	76.947	138.895
Standard deviation	35.717	0.062	0.122	13.562	29.327
Minimum	8.372	0.000	2.158	45.587	86.000
25th percentile	70.393	0.084	2.467	64.851	114.083
Median (50%)	91.222	0.140	2.541	77.717	138.832
75th percentile	112.709	0.186	2.633	88.116	160.996
Maximum	357.762	0.347	2.996	137.331	286.941

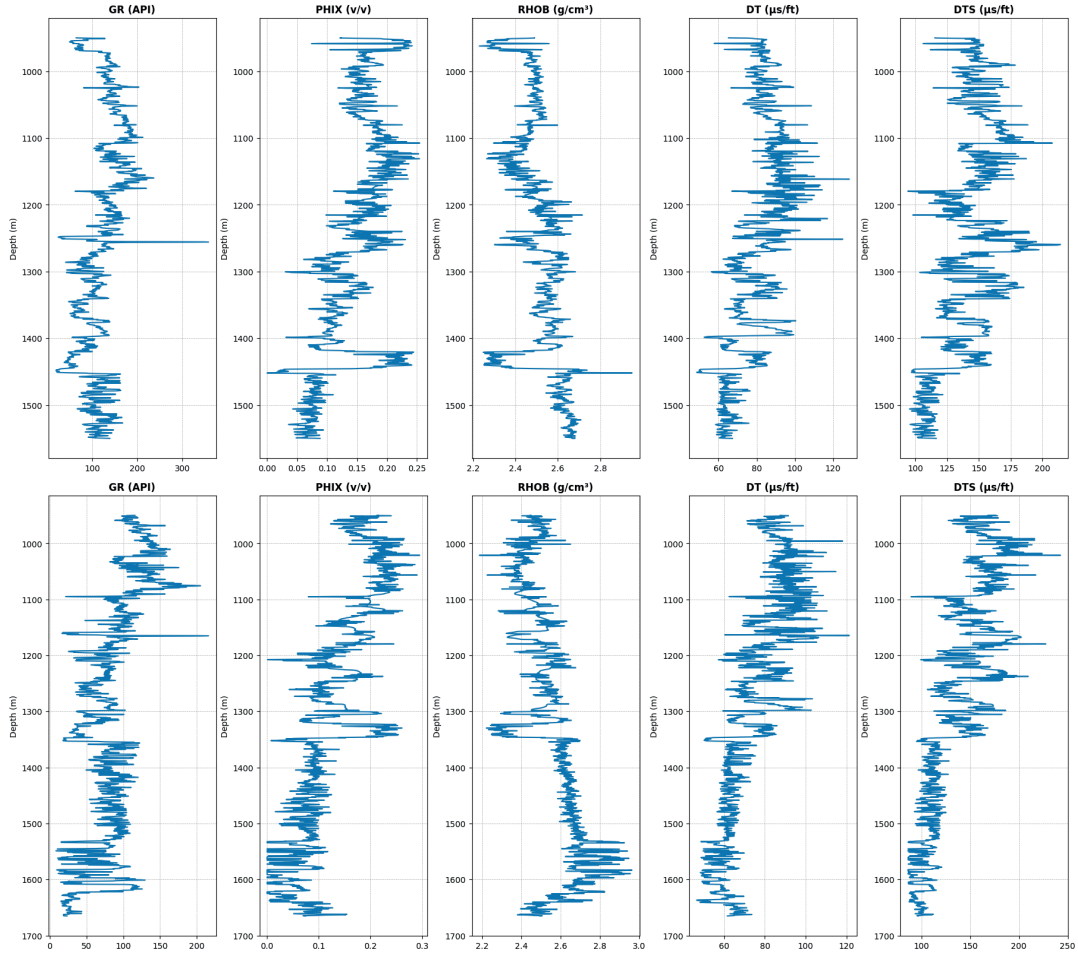

Fig. 3 -Example of well-logging data for W1 (top) and W4 (bottom) in the studied depth interval.

These data provide a complete understanding of the formations travelled, showing the variety of geological and physical parameters, which is critical for further study and predictive reservoir modelling.

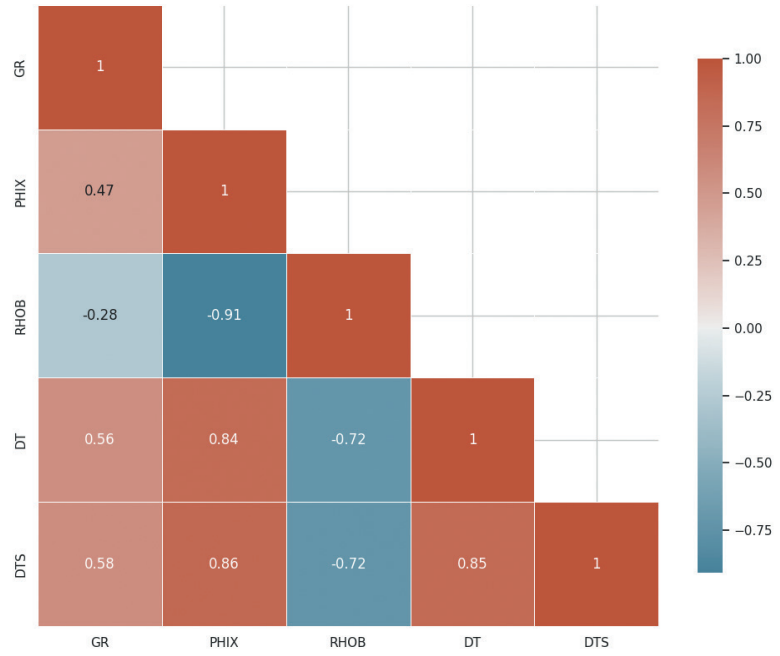

Fig. 4 - Correlation matrix of well-logging data.

The correlation matrix from well logging gives a summary of the linear relation among variables, ranging from -1 to 1 (Fig. 4). There is a high positive correlation (around 1) seen between *PHIX* and *DTS* (0.86), as well as between *DT* and *DTS* (0.85), and between *PHIX* and *DT* (0.84), showing that these features tend to rise together. Moderate positive correlations were also seen between *GR* and *DTS* (0.58), *GR* and *DT* (0.56), and *GR* and *PHIX* (0.47). On the other hand, a strong negative correlation exists between *PHIX* and *RHOB* (-0.91), *DT* and *RHOB* (-0.72), and *DTS* and *RHOB* (-0.72). Based on this analysis, it can be concluded that *DT*, *PHIX*, and *RHOB* exhibit stronger correlations with *DTS*, while the contribution of *GR* is moderate but comparatively less important. The presence of outliers in the database negatively impacts the execution of ML techniques and, therefore, to address this problem, we use several methods such as the One-Class SVR, Isolation Forest, Minimum Covariance Determinant, Local Outlier Factor, and Standard Deviation Filter. The provided boxplot illustrates the impact of these outlier removal methods on well-logging data; we compare the distribution of the log variables before and after outlier removal (Fig. 5). The original well-logging data used for this study, contain numerous outliers, as seen in the boxplot, with many points outside the whiskers, especially in the *GR* and *RHOB* logs. After applying the One-Class SVR method, a significant reduction in outliers appears across all logs and the data distribution becomes more compact and symmetric, which indicates effective outlier removal, with 31,168 samples out of the original 34,633. The Minimum Covariance Determinant method shows effectiveness similar to the One-Class SVR method and achieves a compact and symmetric distribution across all logs while retaining 31,169 samples. The Isolation Forest and Local Outlier Factor methods show a substantial reduction in outliers, particularly in the *GR* and *DTS* logs, but the *RHOB* logs still have some extreme values. These

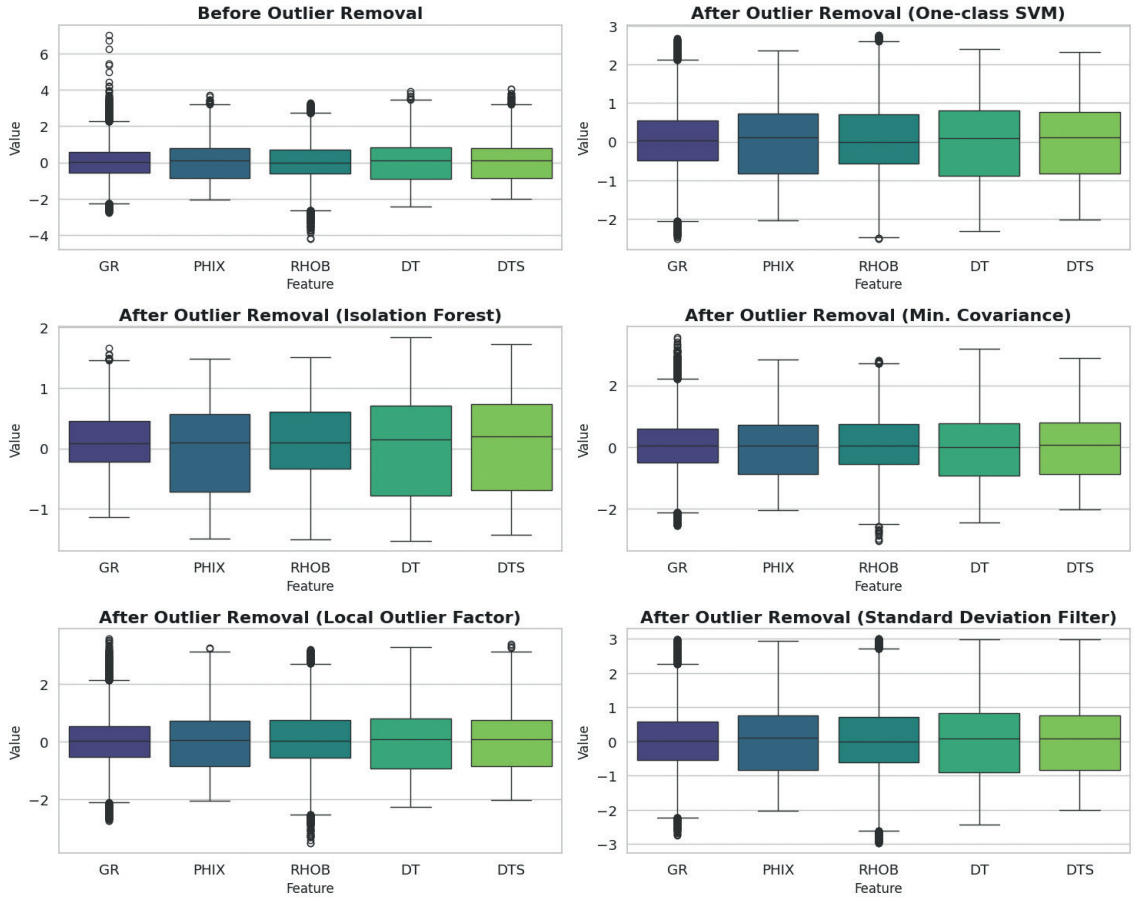

Fig. 5 - Effect of different outlier removal methods on well-logging data.

methods remove a large percentage of the data and retain only 17,317 and 24,243 samples, respectively. In contrast, the Standard Deviation Filter has the least impact, leaves most outliers in the data and retains 34,356 samples.

To choose the best method for outlier removal, it is crucial to balance outlier reduction with sample maintenance. Both the One-Class SVR and Minimum Covariance Determinant methods are effective, but One-Class SVR is slightly better due to its consistent performance, which makes it more suitable for accurate well-logging interpretation.

The pair plot shows the correlations between different scoring variables before and after using the One-Class SVR method to remove outliers. Before removal (left panel), the scatter plots indicate a significant presence of outliers that differ significantly from the overall data patterns, particularly in variables such as *GR* and *PHIX* (Fig. 6). After removal (right panel), the scatter plots indicate a tighter distribution, with fewer outliers, meaning that the method effectively identifies and removes outliers. The distribution of density on the diagonal reveals that the data distribution becomes more normalised when outliers are removed.

These findings highlight the critical role of choosing suitable outlier removal methods to preserve the accuracy and consistency of the dataset. Ensuring the quality of data through effective outlier handling remains a key focus of current research in the field (Jha *et al*., 2022; Ali *et al*., 2023; Dastjerdy *et al*., 2023).

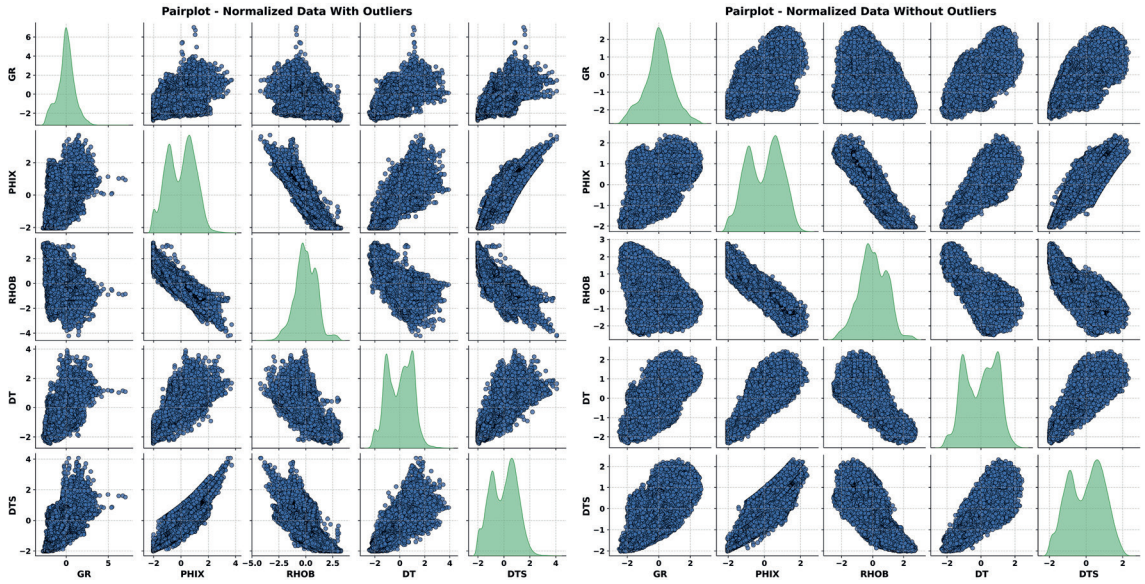

Fig. 6 - Well-logging data pair plots before and after outlier removal using the One-Class SVR method (31,168 of 34,633 samples retained).

3.3. Evaluation of machine learning models

Hyperparameter tuning was carried out for each ML model. These hyperparameters, which are critical model parameters, can significantly improve a model's accuracy and ability to generalise when appropriately fine-tuned. The dataset was split into training and test sets. To explore a range of hyperparameter values, GRid Search Cross-Validation (GridSearchCV) was employed. The optimal values identified through this process are summarised in Table 3, which also explains the meaning of each hyperparameter, the range of values tested, and the selected optimal settings.

Each model was trained using the selected features (depth, *GR*, *DT*, *PHIX*, and *RHOB*) and the target variable (*DTS*). The testing step was conducted across eight wells (test 1 to test 8). The performance metrics for each model are presented in Fig. 7 along with the results for the proposed voting model for comparison. These results highlight the strong performance of both individual techniques and the proposed VR ensemble model. The variation in performance across different wells may be attributed to the fact that the same techniques and hyperparameters were used across all wells. This suggests that the size of the data available for training in each well influences the model performance, with more data generally leading to better performance.

Consequently, the data splitting strategy adopted in this study, with 70% for training, 15% for validation, and 15% for testing, ensured the development of a robust model capable of generalising well to unseen data. Six individual ML techniques were evaluated for their ability to predict *DTS* from well-logging data.

Among these, GBoost achieved the best performance with an R^2 of 0.9623, *RMSE* of 0.1846, and *MAE* of 0.1055, closely followed by LGBM with an R^2 of 0.9601 and *RMSE* of 0.1899. RF and ETR techniques also exhibited strong performance, with R^2 values around 0.959, while the SVR and MLP showed comparatively accuracy lower than the previous models, particularly the MLP, which achieved an R^2 of 0.8572 and *RMSE* of 0.3593. The results are summarised in Tables 4 and 5.

Table 3 - Hyperparameters and optimal values for different models.

Algorithm	Hyperparameter	Search values	Optimal value
RF	Max_depth	(5, 10, 50, 100, None)	100
	Max_features	(3, 5, 10, 15)	3
	n_estimators	(200, 500, 1000, 1200)	1080
	Min_samples_split	(2, 5, 8, 10)	5
ETR	Max_depth	(5, 10, 50, 100, None)	100
	Max_features	(3, 5, 10, 15)	3
	n_estimators	(200, 500, 1000, 1200)	1087
	Min_samples_split	(2, 5, 8, 10)	5
GBoost	Learning_rate	(0.01, 0.05, 0.1)	0.01
	Max_depth	(5, 10, 50, 100)	10
	n_estimators	(500, 800, 1000, 1200)	1000
	Subsample	(0.5, 0.7, 0.9, 1.0)	0.7
LGBM	Learning_rate	(0.01, 0.05, 0.1)	0.01
	Max_depth	(5, 10, 50, 100)	10
	n_estimators	(500, 800, 1000, 1200)	1000
	Subsample	(0.5, 0.7, 0.9, 1.0)	0.7
	Random_state	(0, 21, 42)	42
SVR	Kernel	(linear, poly, rbf)	rbf
	C	(0.1, 1, 2, 10, 100)	2.0
	Epsilon	(0.01, 0.1, 0.5)	0.1
MLP	Hidden_layer_sizes	((15,), (50,), (100,), (150,))	15
	Max_iter	(500, 800, 1000)	1000
	Random_state	(0, 21, 42)	42

Table 4 - Prediction results on the test set for different models.

Model	R^2 (test)	*RMSE* (test)	*MAE* (test)	*MSE* (test)
RF	0.9591	0.1923	0.1108	0.0370
ETR	0.9595	0.1913	0.1108	0.0366
GBoost	0.9623	0.1846	0.1055	0.0341
LGBM	0.9601	0.1899	0.1117	0.0361
SVR	0.9436	0.2257	0.1411	0.0510
MLP	0.8572	0.3593	0.2667	0.1291
Linear regression (baseline)	0.8308	0.3909	0.2951	0.1528

To leverage the strengths of the previous models, a weighted voting model was developed, combining the individual techniques with weights proportional to their contributions.

The proposed weighted voting model demonstrates a good performance compared to individual models, effectively balancing accuracy and generalisation. By combining the predictions of six techniques through weighted averaging, the voting model achieves an R^2 of 0.9582 and an *RMSE* of 0.1943 on the test set (Tables 6 and 7). This performance is on par with

Table 5 - Prediction results on the validation set for different models.

Model	R^2 (validation)	*RMSE* (validation)	*MAE* (validation)	*MSE* (validation)	Execution time (s)
RF	0.9563	0.1972	0.1108	0.0389	207.8902
ETR	0.9581	0.1930	0.1092	0.0372	79.1228
GBoost	0.9596	0.1897	0.1052	0.0360	183.6903
LGBM	0.9587	0.1918	0.1114	0.0368	2.3130
SVR	0.9446	0.2219	0.1368	0.0493	48.9106
MLP	0.8632	0.3488	0.2595	0.1216	2.8410
Linear regression (baseline)	0.8320	0.3865	0.2929	0.03888	0.0000

Fig. 7 - Prediction evaluation results for each ML algorithm in each well.

Table 6 - Average weights for each model. Weights are computed from Eqs. (1) and (2) using the average per-well R^2. Because all models achieved similarly high R^2 values (≈ 0.94 – 0.96, Tables 4 and 5), the resulting weights are nearly uniform, indicating balanced contributions across models.

Model	RF	ETR	GBoost	LGBM	SVR	MLP
Average weight	0.17	0.170074	0.170563	0.170176	0.167257	0.15193

Table 7 - Performance assessment of the proposed voting model. Execution time for the ensemble includes aggregation overhead (units: s).

Voting model	R^2	*RMSE*	*MAE*	*MSE*	Execution time
Test	0.9582	0.1943	0.1223	0.0378	532.7050
Validation	0.9572	0.1951	0.1201	0.0381	-

the best individual models, such as GBoost, LGBM, ETR, and RF, while surpassing SVR and MLP. Similarly, on the validation set, the voting model maintains its high accuracy with an R^2 of 0.9572 and an *RMSE* of 0.1951, highlighting its capacity to effectively generalise to novel, unseen data. Relative to a linear regression baseline (*DTS* ~ *DT* + *RHOB* + *PHIX*), the ensemble improves by $\Delta R^2 \approx 0.13$ and reduces *RMSE* by ≈ 0.20 on the test set, underscoring the value of nonlinear learners and ensembling.

Execution time analysis reveals a clear trade-off between accuracy and computational cost. While the ensemble voting model required ~533 s due to its integration overhead, individual models such as LGBM (2.31 s), GBoost (183.69 s), RF, and ETR (tens to hundreds of seconds) delivered comparably high accuracy within substantially shorter timescales. This indicates that in time- or resource-constrained applications, single models may be the more practical choice. Nevertheless, the ensemble remains valuable for its robustness, as it combines complementary strengths of multiple learners, mitigates bias and variance, and ensures more consistent performance across heterogeneous wells.

Scatterplots of predicted versus actual *DTS* values further validate the model's performance (Fig. 8). The voting model shows minimal deviation from the ideal fit line, highlighting its

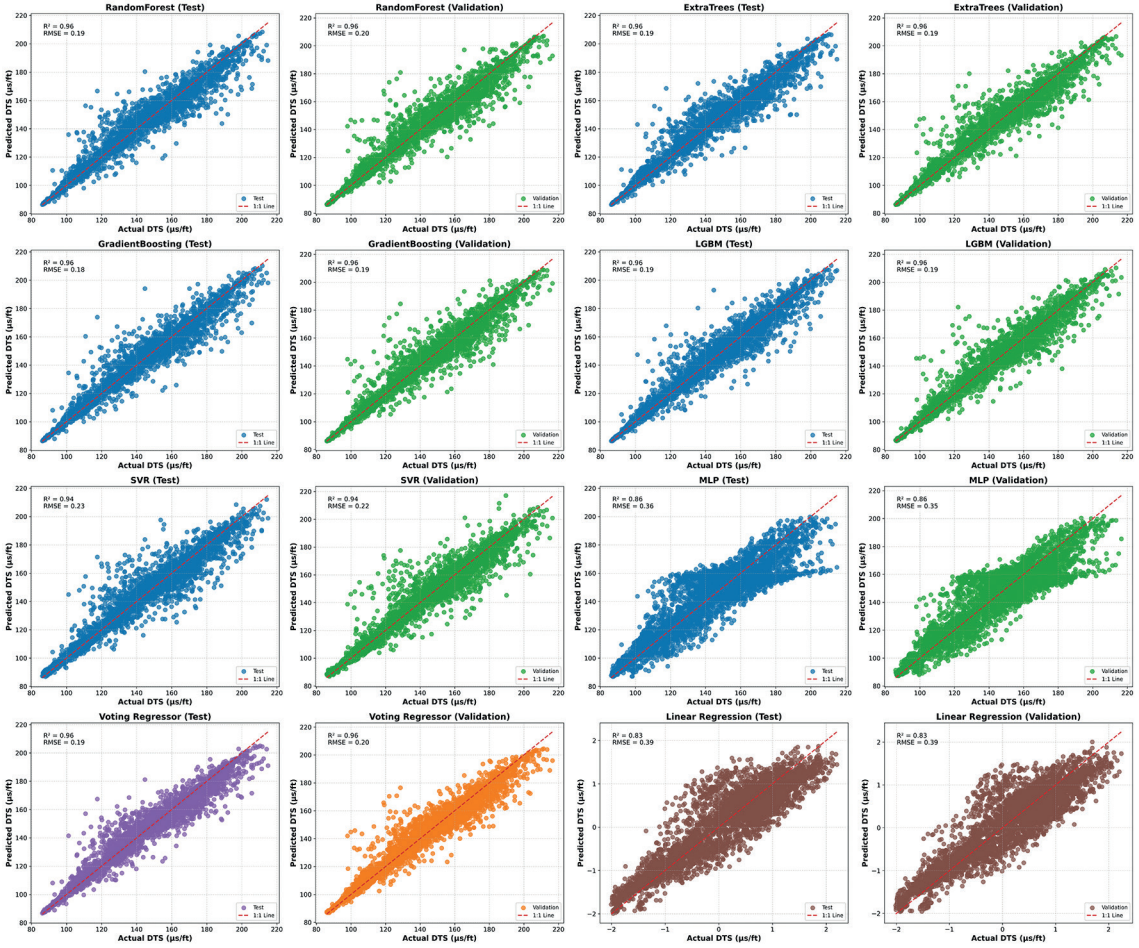

Fig. 8 - Scatter plots of actual *DTS* versus predicted *DTS* for six individual techniques and the proposed model compared with baseline classical linear regression, for both the test and validation sets.

accuracy and robustness. By comparison, SVR and MLP show greater spread, indicating weaker predictions. The weighted voting model is particularly advantageous because it combines the complementary strengths of GBoost, LGBM, and other techniques while minimising the influence of weaker techniques like MLP. This ensures competitive accuracy, improved stability and reliability, making the ensemble model more effective for predicting DTS in well-logging tasks.

To explain the performance of each model, the analysis of feature importance across various regression techniques provides valuable insights into how well-logging parameters influence the prediction of *DTS*, as presented in Fig. 9.

These relationships align with physical principles (Table 8). The key predictor in all techniques is *DT*, which demonstrates a strong correlation with *DTS* due to the fundamental relationship between compressional and shear wave in rock physics (Serra, 1984; Asquith *et al*., 2004).

Table 8 – Key equations used in the analysis of future importance.

Equation	Description	Reference
$DT = (1-\phi)\Delta t_{ma} + \phi \Delta t_f$	Compressional slowness as a function of porosity. *DT*: compressional slowness (μs/ft); ϕ: porosity (fraction); Δt_{ma}: matrix slowness (μs/ft); Δt_f: fluid slowness (μs/ft).	Serra (1984)
$V_p = \frac{10^6}{DT} = \sqrt{\frac{K + \frac{4}{3}\mu}{\rho_b}}$	Compressional (P-wave) velocity. V_p: velocity (m/s or ft/s); *K*: bulk modulus; μ: shear modulus; ρ_b: bulk density (g/cm^3).	Pirson (1963)
$V_s = \frac{10^6}{DTS} = \sqrt{\frac{\mu}{\rho_b}}$	Shear (S-wave) velocity. V_s: velocity (m/s or ft/s); *DTS*: shear slowness (μs/ft). Other symbols as above.	Asquith *et al.* (2004)
$\rho_b = \phi\rho_f + (1-\phi)\rho_{ma}$	Bulk density mixing law. ρ_b: bulk density; ρ_f: fluid density; ρ_{ma}: matrix density. Densities in (g/cm^3).	Serra (1984)
$GR = \frac{\rho \cdot V \cdot A}{\rho_b}$ $GR = a[K] + b[U] + c[Th]$	Gamma-ray response. First expression: theoretical activity concentration. ρ: mineral density (g/cm^3); *V*: volume fraction ; *A*: specific activity (Bq/g); ρ_b: bulk density (g/cm^3). Second expression: practical logging calibration. [*K*]: potassium (%);[*U*]: uranium (ppm); [*Th*]: thorium (ppm). *a, b, c*: tool-specific calibration constants (API per unit concentration).	Pirson (1963) and Serra (1984)

The relationship between *DT* and *DTS* is further explained through the V_p equation, highlighting how changes in *DT* significantly influence shear wave velocity. This makes *DT* the most influential predictor in *DTS* prediction. Shear velocity (V_s) is inversely related to *DTS* and reflects the critical role of rock properties, such as elasticity and density, in determining shear velocity and influencing *DTS* prediction. While *DT* and *DTS* both vary in response to similar elastic properties and *RHOB*, they differ in magnitude, with *DT* generally being lower than *DTS*. Total *PHIX* is another important feature, especially in nonlinear techniques like SVR and MLP. Higher *PHIX* typically reduces shear velocity, consistent with theoretical understanding. *RHOB* also impacts *DTS* prediction but is secondary to *DT* and *PHIX*. *GR* remains the least important feature, consistent with its role as a lithological indicator rather than a direct predictor of elastic properties (Serra, 1984; Asquith *et al*., 2004).

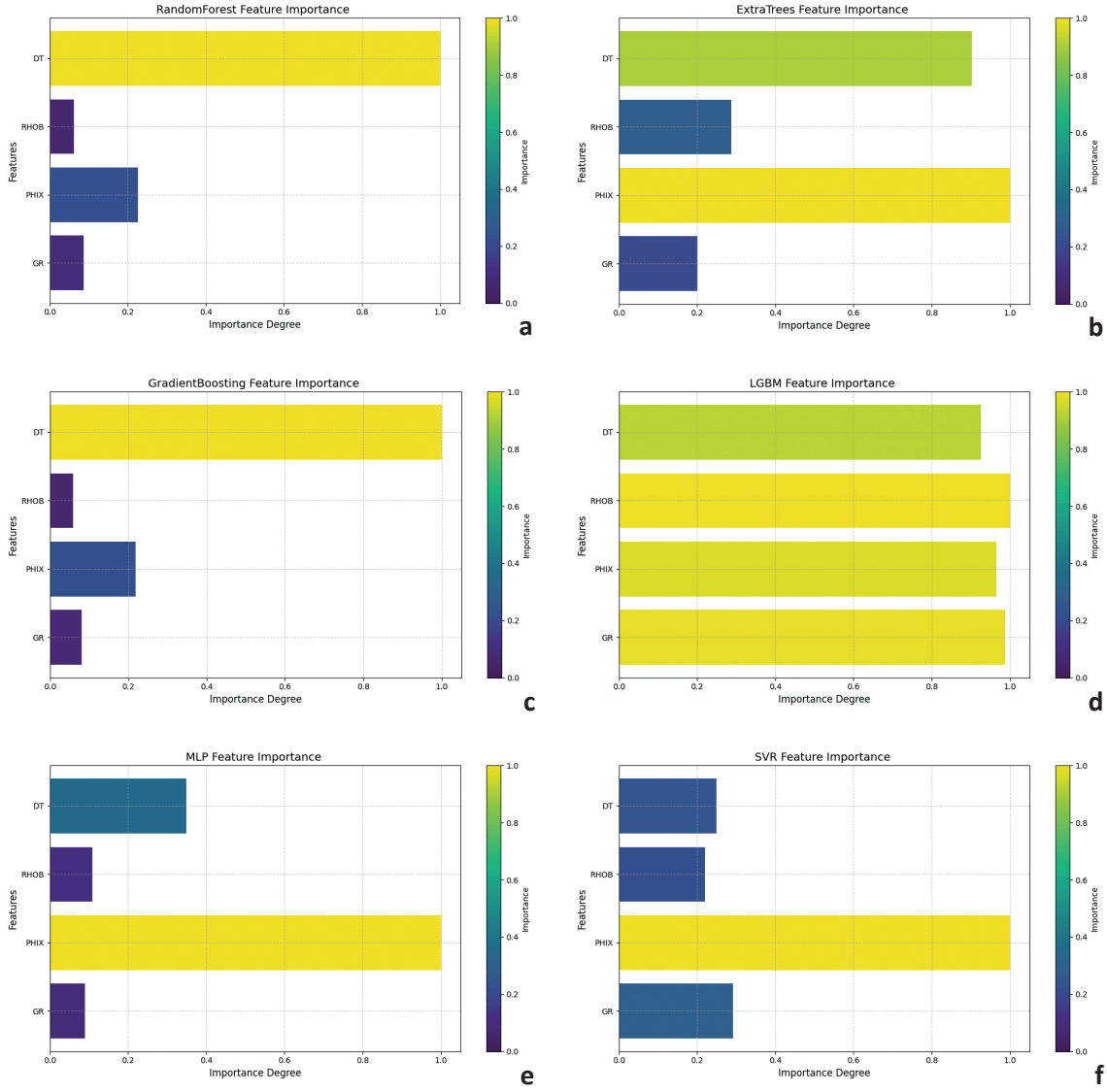

Fig. 9 - Feature importance of each individual model. Note: for tree-based models (RF, ETR, Gboost, and LGBM), importance corresponds to impurity-based measures, whereas for SVR and MLP it is obtained via permutation importance on a common held-out test set. Importances are not directly comparable across model families but indicate the relative influence of features within each model.

Tree-based techniques such as RF, ETR, and GBoost emphasise *DT* as the leading feature, with moderate consideration of *RHOB* and *PHIX*. LGBM assigns slightly more weight to *GR* due to lithological effects, whereas MLP and SVR techniques give greater importance to *PHIX*.

From Fig. 10, the feature importance of voting model analysis confirms that *DT* is the most influential predictor, followed by *PHIX*, *RHOB* and *GR* that have a moderate to lower influence. These findings align with earlier research (Castagna *et al*., 1985). Thus, tree-based techniques and the boosting model best capture the impact of acoustic travel time, density, and porosity on *DTS*, which explains their more accurate predictions compared to techniques emphasising secondary features.

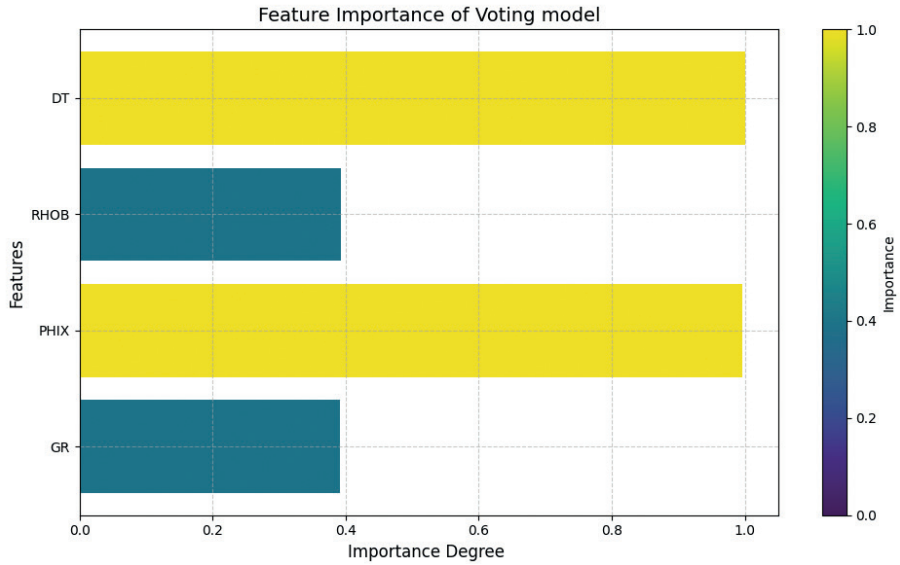

Fig. 10 - Feature importance of the voting model. The ensemble-level importance was derived by aggregating the contributions of base learners according to their weights.

3.4. Performance assessment of voting techniques on unseen data

In the final phase, the VR model was validated on two independent wells, W9 and W10, which were excluded from training and testing to provide an unbiased evaluation. On the normalised scale, the model achieved an R^2 of 0.92 for both wells, with low errors (*RMSE* ≈ 0.28; *MAE* ≈ 0.17-0.18). On the original (denormalised) scale, performance remained strong with R^2 = 0.91. For W9, *RMSE* was 9.02 ms/ft and *MAE* was 5.42 ms/ft, while for W10, *RMSE* was 8.54 ms/ft and *MAE* was 5.59 ms/ft. Visual inspection of the crossplots confirmed the close agreement between predicted and observed *DTS* values.

Residual analysis showed that for W9, the mean residual was 171 ms/ft with a standard deviation of 8.86 ms/ft, while for W10, the mean residual was -1.70 ms/ft with a standard deviation of 8.37 ms/ft. These results indicate that most predictions fall within ±9 ms/ft of the true values, reflecting minimal uncertainty relative to the *DTS* range and reinforcing the reliability of the VR model (Figs. 11 and 12 and Table 9).

Table 9 - Validation results of the VR model on wells W9 and W10 (left block = normalised scale; right block = original/ denormalised scale).

Well	R^2 (normalised)	*RMSE* (normalised)	*MAE* (normalised)	R^2 (original)	*RMSE* (original, µs/ft)	*MAE* (original, µs/ft)	Mean residual (µs/ft)	Std. dev. residual (µs/ft)
W9	0.92	0.28	0.17	0.91	9.02	5.42	1.71	8.86
W10	0.92	0.27	0.18	0.91	8.54	5.59	-1.70	8.37

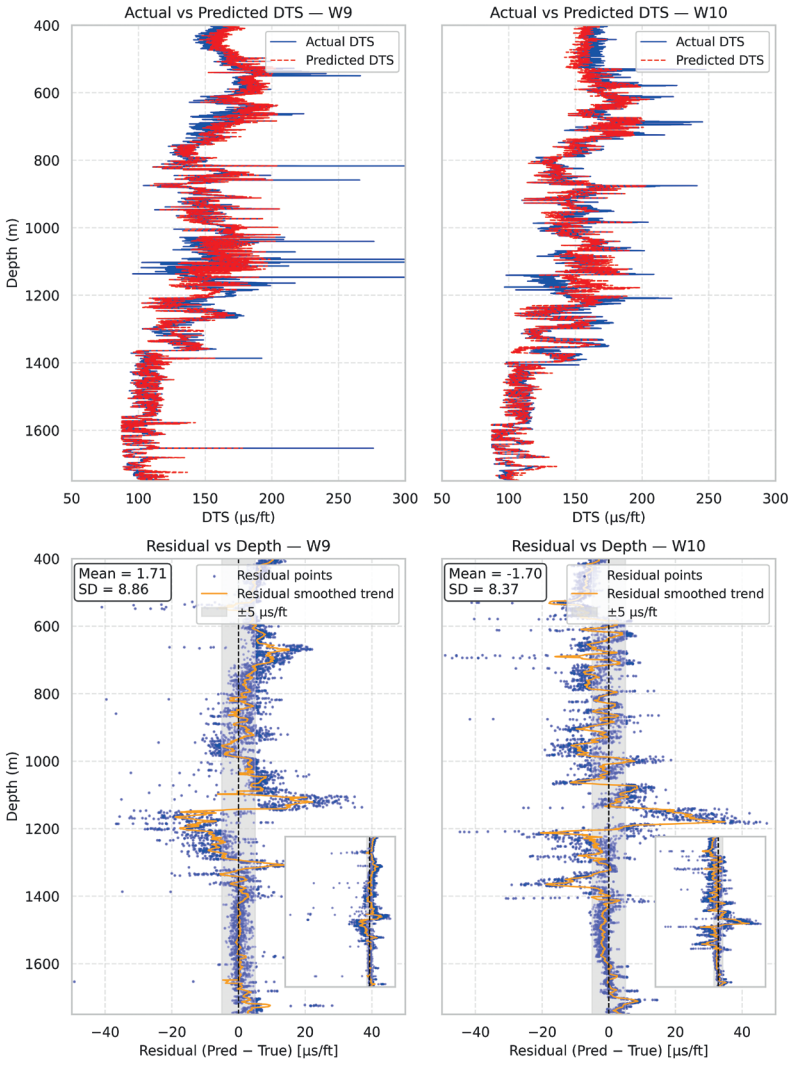

Fig. 11 - Comparison of actual versus predicted *DTS* for wells W9 and W10 with residual analysis.

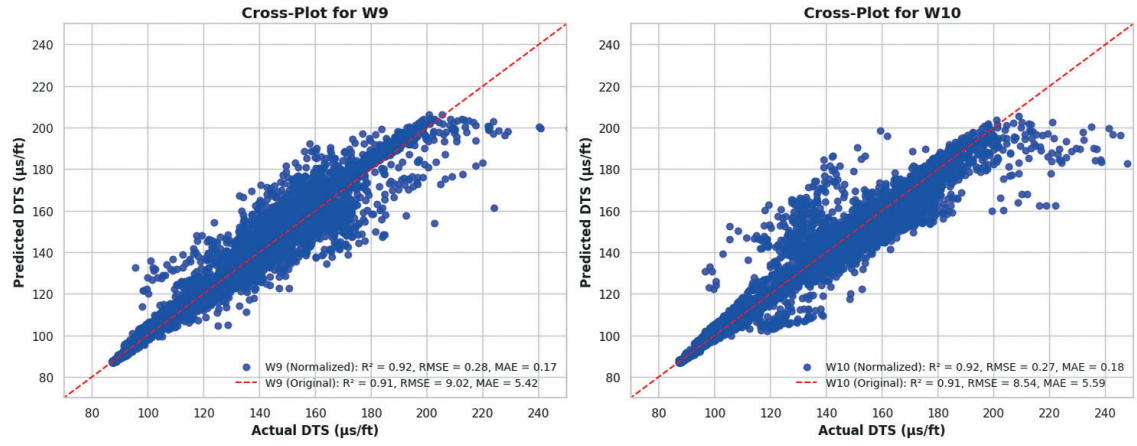

Fig. 12 - Scatter plot of actual versus predicted *DTS* with model performance metrics for wells W9 and W10.

4. Conclusions

This study introduced an ensemble ML approach to predict *DTS* using well-logging data from the Tensleep Formation, aiming to enhance reservoir characterisation. Data preprocessing techniques, including outlier detection methods such as One-Class SVR and Minimum Covariance Determinant, were applied to enhance model predictions by refining the input data, which is critical for accurate subsurface analysis and characterisation. Several ML algorithms, i.e. ETR, GBoost, RF, LGBM, SVR, and MPL, were employed to predict *DTS*, with performance evaluated through R^2, *MAE*, and *RMSE*. The VR approach, which integrates the strengths of multiple models, showed superior performance compared to individual algorithms, with tree-based and boosting methods delivering the best results in terms of accuracy and consistency. This highlights the advantage of ensemble techniques in improving prediction reliability. Feature importance analysis indicated that *DT*, total *PHIX*, and *RHOB* were the most influential variables for predicting *DTS*, while *GR* had minimal impact, in agreement with physical expectations. A limitation of this study is that the models were developed and validated on data from a single geological formation and region (Tensleep, SE Wyoming). Future work should, therefore, explore lithofacies or zonation-aware ensemble models to improve generalisation across different reservoirs and depositional settings. Overall, the ensemble methodology provides a robust framework for *DTS* prediction and offers valuable potential for broader geophysical applications in subsurface modelling.

Acknowledgments. The well-logging data used in this study are based on the Teapot Dome (RMOTC) dataset, with additional confidential extensions that cannot be shared. The RMOTC dataset is available under the original provider's licensing terms. The code developed for model training, evaluation, and figure generation is part of ongoing research; however, the corresponding author can provide access for research collaboration upon reasonable request. Full methodological details are described in the manuscript. All experiments were executed on a laptop equipped with an Intel® Core™ i7-8550U CPU @ 1.80 GHz, 8 GB RAM (2400 MHz), and a 2 GB GPU, running Windows 11 (version 24H2, build 26100.6584). Models were implemented in Python 3.11 using scikit-learn (version 1.7.2) and LGBM (version 4.6.0). Execution times reported in Tables 5 and 7 represent wall-clock measurements obtained on this system. Ensemble times explicitly include both base-learner training and aggregation overhead.

REFERENCES

Al-Handhali S., Al-Aamri M. and Sundararajan N.; 2023: *Beyond conventional approach: hybrid supervised learning and feature selection algorithms for predicting sonic logs - a study in a tight gas sand, north of Oman*. Int. J. Oil, Gas, Coal Tech., 34, 359-385.

Ali M., Ahmed K., Hassan R., Khan S. and Farooq A.; 2023: *A novel machine learning approach for detecting outliers, rebuilding well logs, and enhancing reservoir characterization*. Nat. Resour. Res., 32, 1047-1066.

Amoura S., Gaci S., Bounif M.-A. and Boussa L.; 2019: *On characterizing heterogeneities from velocity logs using Hölderian regularity analysis: a case study from Algerian tight Devonian reservoirs*. J. Appl. Geophys., 170, 103833, doi: 10.1016/j.jappgeo.2019.103833.

Amoura S., Gaci S., Barbosa S., Farfour M. and Bounif M.A.; 2022: *Investigation of lithological heterogeneities from velocity logs using EMD-Hölder technique combined with multifractal analysis and unsupervised statistical methods*. J. Pet. Sci. Eng., 208, 109588, doi: 10.1016/j.petrol.2021.109588.

Anemangely M., Ramezanzadeh A., Amiri H. and Hoseinpour S.A.; 2019: *Machine learning technique for the prediction of shear wave velocity using petrophysical logs*. J. Pet. Sci. Eng., 174, 306-327.

Asquith G., Krygowski D., Henderson S. and Hurley N.; 2004: *Basic relationships of well log interpretation*. In: Asquith G., Krygowski D., Henderson S. and Hurley N. (eds), Basic well log analysis - 2nd ed., The American Association of Petroleum Geologists, Tulsa, OK, USA, pp. 1-27.

Aziz T., Khan R., Liu H., Patel S. and Zhang L.; 2023: *REM-based indoor localization with an Extra-Trees Regressor*. Electron., 12, 4350.

Beinkafner K.J.; 1986: *Use of dipmeter logs to refine structural mapping of Teapot Dome, Wyoming*. Am. Assoc. Pet. Geol. Bull., 70, 7054194, <www.osti.gov/scitech/biblio/7054194>.

Boulassel A., Zaourar N., Gaci S. and Boudella A.; 2021: *A new multifractal analysis-based approach for identifying the reservoir fluid nature*. J. Appl. Geophys., 185, 104185, doi: 10.1016/j.jappgeo.2020.104185.

Boulassel A., Makhlouf S., Boumelit Z., Zegagh B., Boufenchouche S. and Cheddad F.A.; 2025a: *Unlocking deeper insights: using machine learning to predict dynamic shear wave slowness from well logs*. In: Geophysical exploration for hydrocarbon reservoirs, geothermal energy, and carbon storage: new technologies and aI-based approaches, pp. 323-342, doi: 10.1002/9781394261567.ch16.

Boulassel A., Makhlouf S., Cheddad F.A., Boumelit Z., Zegagh B. and Boufenchouche S.; 2025b: *Logging-data-driven fluid prediction in clastic reservoir based on fractal attributes and machine learning methods*. In: Geophysical exploration for hydrocarbon reservoirs, geothermal energy, and carbon storage: new technologies and aI-based approaches, pp. 303-322, doi: 10.1002/9781394261567.ch15.

Brocher T.M.; 2005: *Empirical relations between elastic wave speeds and density in the Earth's crust*. Bull. Seismol. Soc. Am., 95, 2081-2092.

Castagna J.P., Batzle M.L. and Eastwood R.L.; 1985: *Relationships between compressional-wave and shear-wave velocities in clastic silicate rocks*. Geophys., 50, 571-581.

Chen S. and Luc N.M.; 2022: *RRMSE voting regressor: a weighting function based improvement to ensemble regression*, arXiv:2207.04837, doi: 10.48550/arXiv.2207.04837.

Cherana A. and Aliouane L.; 2024: *Prediction model of reservoir porosity via incorporating particle swarm optimisation into an Adaptive Neuro-Fuzzy Inference System: application to Triassic reservoirs of the Hassi R'mel field (Algeria)*. Bull. Geoph. Ocean., 65, 97-114, doi: 104430/bgo00432.

Cooper S.P., Lorenz J.C. and Goodwin L.B.; 2001: *Lithologic and structural controls on natural fracture characteristics, Teapot Dome, Wyoming*. Sandia National Laboratories, Livermore, CA, USA, Report Sand2001-1786, 74 pp., doi: 10.2172/783091.

Cui F., Zhang H., Li Y., Wang S. and Chen X.; 2025: *Prediction of coal burst location and risk level in roadway using XGBoost with multi-element microseismic information and its application in steeply inclined ultra-thick coal seam*. Rock Mech. Rock Eng., 58, 4023-4042, doi: 10.1007/s00603-024-04371-x.

Dastjerdy E., Saeidi A. and Heidarzadeh S.; 2023: *Review of applicable outlier detection methods to treat geomechanical data*. Geotech., 3, 375-396.

Dongapure P.T.; 2024: *Examining the boundaries of machine learning and deep learning: a thorough review of the main obstacles in missing well log estimation*. Appl. Math. Sci. Eng., 1, 45-65.

Drucker H., Burges C.J.C., Kaufman L., Smola A.J. and Vapnik V.; 1996: *Support vector regression machines*. Adv. Neural Inf. Process. Syst., 28, 779-784.

El-Dabaa S.A., Metwalli F.I., Maher A. and Ismail A.; 2024: *Unsupervised machine learning-based multi-attributes analysis for enhancing gas channel detection and facies classification in the serpent field, offshore Nile Delta, Egypt*. Geomech. Geophys. Geo-Energ. and Geo-Resour., 10, 185, doi: 10.1007/s40948-024-00907-1.

Fanshawe J.R.; 1971: *Structural evolution of Big Horn Basin*. In: Proc. Symposium on Wyoming Tectonics and their economic significance, Guidebook 1971, 23rd Field Conference, Wyoming Geological Association, Casper, WY, USA, pp. 35-42.

Fox J.E., Dolton G.L. and Clayton J.L.; 1990: *Petroleum Geology of the Powder River Basin, Wyoming and Montana*. Open-file report 88-450-P, U.S. Geological Survey (USGS), Denver, CO, USA.

Gaci S. and Hachay O. (eds); 2017: *Oil and gas exploration: methods and application*. American Geophysical Union, Wiley Editions, Hoboken, NJ, USA, 311 pp.

Greenberg M.L. and Castagna J.P.; 1992: *Shear-wave velocity estimation in porous rocks: theoretical formulation, preliminary verification and applications*. Geophys. Prospect., 40, 195-209.

Hassaan S., Mohamed A., Ibrahim A.F. and Elkatatny S.; 2024: *Real-time prediction of petrophysical properties using machine learning based on drilling parameters*. ACS Omega, 9, 17066-17075.

Huang J.C., Tsai Y.C., Wu P.Y., Lien Y.H., Chien C.Y., Kuo C.F., Chen L.S. and Kuo C.H.; 2020: *Predictive modeling of blood pressure during hemodialysis: a comparison of linear model, random forest, support vector regression, XGBoost, LASSO regression and ensemble method*. Comput. Methods and Programs Biomed., 195, 105536.

Jha H.S., Khanal A., Seikh H.M.D. and Lee W.J.; 2022: *A comparative study on outlier detection techniques for noisy production data from unconventional shale reservoirs*. J. Nat. Gas Sci. Eng., 105, 104720.

Joshi P. and Raghuvanshi A.S.; 2021: *Hybrid approaches to address various challenges in wireless sensor networks for IoT applications: opportunities and open problems.* International Journal of Computer Networks and Applications, 8, 151-187.

Kanfar R., Shaikh O., Yousefzadeh M. and Mukerji T.; 2020: *Real-time well log prediction from drilling data using deep learning*. In: Proc. International Petroleum Technology Conference, Dhahran, Kingdom of Saudi Arabia, IPTC-19693-MS, doi: 10.2523/IPTC-19693-MS.

Lawal A., Yang Y., He H. and Baisa N.L.; 2024: *Machine learning in oil and gas exploration: a review*. IEEE Access, 12, 19035-19058, doi: 10.1109/ACCESS.2023.3349216.

Ma K., Shen Q., Zhang Z. and Wang T.; 2025: *Development and validation of intelligent models for predicting rockburst location based on microseismic monitoring and machine learning*. Rock Mech. and Rock Eng., 58, 5093-5113, doi: 10.1007/s00603-025-04388-w.

Mankiewicz D. and Steidtmann J.R.; 1979: *Depositional environments and diagenesis of the Tensleep Sandstone, eastern Big Horn Basin, Wyoming*. SEPM Special Publication, 26, pp. 319-336, doi: 10.2110/pec.79.26.0319.

Meng W., Xu N., Zhao Z. and Wu W.; 2024: *Excavation-induced fault instability: a machine learning perspective*. Rock Mech. Rock Eng., 57, 5251-5265.

Nwankwo C.G., Alade S.M., Agbakwuru O.A., Amanze B.C. and Akawuku G.I.; 2024: *Investigating the frontiers of deep learning and machine learning: a comprehensive overview of key challenges in missing well log estimation*. Int. Res. J. Innovations Eng. Tech., 8, 1-15, doi: 10.47001/IRJIET/2024.807001.

Olayiwola T., Tariq Z., Abdulraheem A. and Mahmoud M.; 2021: *Evolving strategies for shear wave velocity estimation: smart and ensemble modeling approach*. Neural Comput. Appl., 33, 17147-17159.

Pickett G.R.; 1963: *Acoustic character logs and their applications in formation evaluation*. J. Pet. Tech., 15, 659-667.

Pirson S.J.; 1963: *Handbook of well log analysis for oil and gas formation evaluation*. Prentice Hall, Hoboken, NJ, USA, 326 pp.

Serra O.E.; 1984: *Fundamentals of well-log interpretation*: *the acquisition of logging data - 1.* Elsevier Science Ltd, Amsterdam, The Netherlands, 436 pp.

Xu C., Fu L., Lin T., Li W. and Ma S.; 2022: *Machine learning in petrophysics: advantages and limitations*. Artif. Intell. Geosci., 3, 157-161, doi: 10.1016/j.aiig.2022.11.004.

Yazid H. and Gaci S.; 2022: *Machine learning-based techniques for reservoir characterization using rock-typing from well logs*. In: Proc. 7th International Conference on Image and Signal Processing and Their Applications (ISPA), Mostaganem, Algeria, pp. 1-6, doi: 10.1109/ISPA54004.2022.9786300.

Yazid H. and Gaci S.; 2024: *Integrating hybrid weak learners for lithofacies classification using well log data*. In: Proc. 8th International Conference on Image and Signal Processing and Their Applications (ISPA), Biskra, Algeria, pp. 1-5, doi: 10.1109/ISPA59904.2024.10536797.

Yu Y., Xu C., Misra S., Li W., Ashby M., Pan W. and Izadi H.; 2021: *Synthetic sonic log generation with machine learning: a contest summary from five methods*. Petrophys., 62, 393-406.

Zhang F., Deng S., Wang S. and Sun H.; 2022: *Convolutional neural network long short-term memory deep learning model for sonic well log generation for brittleness evaluation*. Interpretation, 10, T367-T378.

Zhang R., Li Y., Gui Y., Armaghani D.J. and Yari M.; 2025: *Adaptive weighted multi-kernel learning for blast-induced flyrock distance prediction*. Rock Mech. Rock Eng., 58, 679-695, doi: 10.1007/s00603-024-04166-0.

Corresponding author: Saliha Amoura
Geophysics Laboratory, Faculty of Earth Sciences, Geography and Territorial Planning, University of Science and Technology Houari Boumediene
BP 32, Bab Ezzouar, 16111,Algiers, Algeria
Phone: +213 552 205 568; e-mail: samoura@usthb.dz