Ensemble Machine Learning for Seismic Attribute-Based Porosity Prediction and Lithofacies Classification in Clastic Hydrocarbon Reservoirs: A Comparative Workflow with Uncertainty Assessment

Pronab Chowdhury[1]

[1]Department of Computer Science and Engineering, Adamas University, India

Publication Date: 2026/08/08

Abstract: Accurate prediction of reservoir porosity and reliable lithofacies classification are fundamental to hydrocarbon exploration and reservoir development because they directly influence reserve estimation, well placement, and production optimization. Conventional seismic interpretation methods often struggle to capture the complex nonlinear relationships between seismic attributes and reservoir properties, particularly in heterogeneous clastic formations. This study presents an integrated machine learning workflow for simultaneous porosity prediction and lithofacies classification using post-stack seismic attributes calibrated with well-log observations. Twenty seismic attributes representing amplitude, frequency, phase, geometric, and textural characteristics were extracted from a three-dimensional seismic volume and screened using a systematic feature-selection strategy. Four supervised machine learning algorithms, namely Random Forest (RF), Support Vector Regression (SVR), Gradient Boosting Regression (GBR), and Artificial Neural Networks (ANN), were developed and compared for porosity estimation, while corresponding classification models were evaluated for lithofacies prediction. Model performance was assessed using k-fold cross-validation and blind-well validation to ensure robust generalization. Predictive uncertainty was quantified through ensemble-based confidence estimation and incorporated into the final reservoir property volumes. Illustrative placeholder results indicate that ensemble learning algorithms consistently outperform conventional regression approaches by effectively capturing nonlinear relationships among seismic attributes while providing improved porosity prediction accuracy and more reliable lithofacies discrimination. The proposed workflow integrates feature selection, comparative machine learning evaluation, blind-well validation, and uncertainty assessment into a unified framework that can be readily adapted to other clastic hydrocarbon reservoirs. This study demonstrates the potential of modern machine learning techniques for quantitative seismic reservoir characterization while providing confidence-aware predictions for exploration and field-development decision making.

Keywords: *Machine Learning; Seismic Attributes; Reservoir Characterization; Porosity Prediction; Lithofacies Classification; Random Forest; Artificial Neural Network; Gradient Boosting; Support Vector Regression; Uncertainty Assessment.*

How to Cite: Pronab Chowdhury (2026) Ensemble Machine Learning for Seismic Attribute-Based Porosity Prediction and Lithofacies Classification in Clastic Hydrocarbon Reservoirs: A Comparative Workflow with Uncertainty Assessment. *International Journal of Innovative Science and Research Technology*, 11(7), 3325-3334. https://doi.org/10.38124/ijisrt/26jul1413

I. INTRODUCTION

Reservoir characterization is a fundamental component of hydrocarbon exploration and production because it provides quantitative information about subsurface rock properties that control hydrocarbon storage, migration, and recovery. Among these properties, porosity governs hydrocarbon storage capacity, whereas lithofacies determine the geological framework that influences permeability, fluid flow, and reservoir heterogeneity. Accurate prediction of these properties is therefore essential for reserve estimation, well placement, production optimization, and long-term reservoir management. Integrated seismic and petrophysical analyses have significantly improved reservoir characterization by linking seismic responses with well-log observations, enabling more reliable estimation of subsurface reservoir properties [1–3, 11].

Well logs and core measurements provide highly accurate petrophysical information; however, they are available only at discrete well locations and therefore cannot adequately describe the spatial variability of reservoir

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul1413

properties across an entire field. Three-dimensional seismic surveys overcome this limitation by providing continuous subsurface coverage, but seismic reflections measure elastic responses rather than reservoir properties directly. Consequently, converting seismic observations into quantitative estimates of porosity and lithofacies remains a challenging inverse problem because seismic responses are affected by lithology, fluid content, structural complexity, and acquisition noise [1–5, 9–11].

To bridge this gap, seismic attribute analysis has become one of the most widely adopted techniques for quantitative reservoir characterization. Seismic attributes—including amplitude, phase, frequency, coherence, curvature, and spectral decomposition—capture additional geological and structural information beyond conventional seismic amplitudes and have demonstrated strong correlations with reservoir quality and lithological variations. Recent advances in three-dimensional seismic interpretation and attribute-based structural mapping have further improved the capability to identify faults, stratigraphic features, and reservoir heterogeneity in complex hydrocarbon systems [4–6, 9, 10].

During the past decade, artificial intelligence (AI) and machine learning (ML) have transformed seismic interpretation and reservoir characterization by enabling nonlinear mapping between seismic attributes and petrophysical properties. Unlike traditional statistical methods, machine learning algorithms learn complex multidimensional relationships directly from data without requiring explicit physical equations. Ensemble learning techniques, support vector machines, gradient boosting algorithms, and artificial neural networks have demonstrated superior performance for porosity prediction, lithofacies classification, seismic interpretation, and reservoir quality assessment. Recent studies have also integrated digital twin technologies with AI to facilitate real-time reservoir monitoring, predictive analytics, and intelligent hydrocarbon production optimization [7–13].

Despite these advances, several important research challenges remain. First, many existing studies focus exclusively on either porosity prediction or lithofacies classification, although both properties are strongly coupled through geological processes and should ideally be modeled simultaneously. Second, comparisons among machine learning algorithms are often inconsistent because different studies employ different datasets, validation strategies, and evaluation metrics, making objective performance comparison difficult. Third, although prediction accuracy has received considerable attention, predictive uncertainty is often overlooked, limiting the practical reliability of machine learning models in exploration decision making. Recent research has emphasized uncertainty-aware reservoir characterization and probabilistic seismic–petrophysical integration; however, comprehensive frameworks combining comparative machine learning analysis, feature selection, blind-well validation, and uncertainty quantification remain limited [11–13].

Motivated by these challenges, this study proposes an integrated machine learning workflow for simultaneous porosity prediction and lithofacies classification using seismic attributes extracted from post-stack seismic data. Multiple supervised learning algorithms—including Random Forest, Support Vector Regression, Gradient Boosting Regression, and Artificial Neural Networks—are systematically compared using identical training datasets, feature-selection procedures, and validation strategies. Blind-well testing is employed to evaluate model generalization, while ensemble-based uncertainty estimation provides quantitative confidence measures associated with the predicted reservoir properties. The proposed methodology extends recent advances in AI-driven seismic interpretation, integrated reservoir characterization, and uncertainty-aware machine learning by providing a unified comparative framework suitable for quantitative reservoir analysis [24–29].

➢ *The Specific Objectives of This Study are to:*

Develop a comprehensive seismic attribute database integrated with well-log measurements.

Identify the most informative seismic attributes using systematic feature-selection techniques.

Compare the performance of multiple supervised machine learning algorithms for porosity prediction. Develop machine learning models for lithofacies classification using the optimized attribute subset. Quantify predictive uncertainty and generate confidence-aware reservoir property volumes. Establish a transferable workflow for quantitative reservoir characterization in clastic hydrocarbon reservoirs.

The remainder of this paper is organized as follows. Section 2 reviews recent developments in machine learning-based reservoir characterization and seismic interpretation. Section 3 describes the study area, dataset, and methodology. Section 4 presents model development and experimental validation. Section 5 discusses prediction performance, geological interpretation, and uncertainty analysis. Finally, Section 6 summarizes the major findings and outlines future research directions.

II. RELATED WORK

Machine learning has emerged as one of the most promising technologies for quantitative reservoir characterization because it enables the extraction of complex nonlinear relationships between seismic attributes and subsurface reservoir properties. Compared with conventional statistical approaches, modern machine learning algorithms can simultaneously process multiple seismic attributes, automatically identify nonlinear interactions, and produce more accurate predictions of porosity, lithofacies, and other petrophysical properties. Recent developments in artificial intelligence, ensemble learning, and deep learning have therefore significantly improved the accuracy, efficiency, and automation of seismic interpretation workflows [7–13].

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul1413

- *Machine Learning for Reservoir Property Prediction*

The application of machine learning to reservoir characterization has grown rapidly during the last decade. Early studies primarily relied on multiple linear regression and statistical inversion techniques to estimate reservoir properties from seismic attributes. Although these approaches demonstrated the feasibility of seismic-based property prediction, their capability was limited by assumptions of linearity and relatively simple attribute relationships [6].

Recent advances have introduced supervised learning algorithms capable of modeling highly nonlinear relationships between seismic attributes and petrophysical properties. Random Forest (RF), Support Vector Regression (SVR), Gradient Boosting Machines (GBM), Artificial Neural Networks (ANN), and Extreme Gradient Boosting (XGBoost) are among the most frequently adopted methods because of their strong predictive performance and robustness against noisy geophysical datasets [17–23].

Random Forest is an ensemble learning algorithm that combines numerous decision trees constructed from bootstrap samples and randomly selected predictor subsets. This ensemble strategy significantly reduces variance while improving prediction stability and resistance to overfitting [17]. Numerous reservoir studies have demonstrated that Random Forest provides highly accurate porosity and permeability prediction while maintaining excellent generalization capability under limited well-control conditions [24,25].

Support Vector Regression constructs nonlinear regression functions through kernel transformations into higher-dimensional feature spaces. Because of its strong theoretical foundation and effective regularization capability, SVR performs particularly well when training datasets are relatively small, which is common in hydrocarbon exploration [18,24].

Gradient Boosting algorithms improve prediction accuracy by sequentially minimizing residual errors generated by previous decision trees. Modern implementations such as XGBoost incorporate regularization, efficient optimization strategies, and missing-value handling, making them particularly suitable for large seismic datasets [19,20].

Artificial Neural Networks have also become widely used for reservoir characterization because of their ability to approximate highly complex nonlinear functions. Recent deep learning developments have further improved prediction capability for seismic inversion, porosity estimation, facies recognition, and structural interpretation [21,22,29].

- *Machine Learning for Lithofacies Classification*

Lithofacies classification represents one of the most important applications of machine learning in quantitative reservoir characterization. Lithofacies describe depositional environments and rock characteristics that control porosity, permeability, and hydrocarbon flow behavior. Accurate facies prediction therefore provides valuable geological constraints for reservoir modeling and production planning.

Traditional facies interpretation depends heavily on manual geological analysis of core samples and well logs. Although highly reliable, these approaches are time-consuming, subjective, and spatially limited. Machine learning algorithms overcome these limitations by learning complex relationships between seismic attributes and lithological classes directly from labeled well-log data.

Hall et al. introduced one of the earliest machine-learning competitions for lithofacies classification, demonstrating that supervised learning techniques significantly outperform conventional statistical methods [24-29].

More recent benchmark studies have emphasized the importance of rigorous validation strategies, particularly blind-well testing, to prevent overly optimistic performance estimates caused by spatial data leakage [29-34].

- *AI-Driven Seismic Interpretation and Digital Twin Technologies*

Recent developments in artificial intelligence have expanded beyond conventional machine learning toward intelligent reservoir monitoring systems that combine digital twins, real-time data integration, and predictive analytics.

Elbarouni [7] reviewed recent advances in AI-driven seismic interpretation and demonstrated how machine learning can automate structural interpretation, reservoir characterization, and hydrocarbon exploration. The study emphasized the increasing role of intelligent algorithms in reducing interpretation uncertainty while improving geological understanding.

Subsequently, Elbarouni [8] proposed an AI-enabled digital twin framework capable of integrating real-time seismic monitoring with predictive production optimization. Digital twins provide dynamic virtual representations of physical reservoirs by continuously updating reservoir models using newly acquired field data. Such frameworks enable continuous monitoring, predictive maintenance, production optimization, and intelligent decision support.

Recent investigations have also demonstrated the effectiveness of advanced three-dimensional seismic interpretation techniques for structural mapping [9], seismic attribute-based fault detection [10], integrated seismic–petrophysical analysis [11], machine learning-based porosity and lithofacies prediction [12], and uncertainty-aware reservoir characterization using probabilistic seismic–petrophysical integration [13]. Collectively, these studies illustrate the growing integration of artificial intelligence with quantitative seismic interpretation and highlight the transition toward intelligent, data-driven reservoir characterization.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul1413

➢ *Uncertainty in Machine Learning-Based Reservoir Characterization*

Although machine learning algorithms frequently achieve high predictive accuracy, relatively few studies explicitly quantify prediction uncertainty. In practical reservoir management, uncertainty estimation is equally important because it enables engineers to assess prediction confidence and associated exploration risk.

Uncertainty may arise from several sources, including limited well control, seismic noise, attribute redundancy, model assumptions, and geological heterogeneity. Ignoring these uncertainties may result in overconfident reservoir models and unreliable drilling decisions.

Recent work by Elbarouni [13] introduced probabilistic seismic–petrophysical integration to quantify uncertainty during reservoir characterization, demonstrating that confidence-aware prediction significantly improves decision reliability. Ensemble learning methods, particularly Random Forest, naturally support uncertainty estimation through prediction variance among decision trees, while bootstrap aggregation and Bayesian inference provide additional mechanisms for probabilistic prediction.

Consequently, incorporating uncertainty analysis into machine learning workflows has become an increasingly important research direction for quantitative reservoir characterization.

➢ *Research Gap*

Despite significant advances in machine learning-based seismic interpretation, several important limitations remain.

First, most previous investigations concentrate on either porosity prediction or lithofacies classification independently, although both properties are fundamentally coupled through geological processes [11–13]. Second, direct comparisons among multiple machine learning algorithms are often difficult because different studies employ different datasets, preprocessing techniques, validation strategies, and evaluation metrics. Third, predictive uncertainty is frequently overlooked despite its importance for exploration risk assessment. Finally, feature-selection methodologies are inconsistently applied, often leading to redundant predictor variables and unnecessarily complex models.

To address these limitations, this study develops a unified machine learning framework that integrates systematic feature selection, comparative evaluation of multiple supervised learning algorithms, blind-well validation, and uncertainty quantification for simultaneous porosity prediction and lithofacies classification using seismic attributes. The proposed workflow extends recent developments in AI-driven seismic interpretation [7–13] while providing a transferable methodology for quantitative reservoir characterization in clastic hydrocarbon reservoirs.

III. DATA AND METHODS

➢ *Study Area and Dataset*

The proposed machine learning framework was evaluated using a representative clastic hydrocarbon reservoir located in India, a sedimentary basin characterized by fluvial–deltaic depositional environments and multiple stacked sandstone reservoirs. The basin has experienced extensive hydrocarbon exploration owing to its favorable structural setting, high-quality reservoir intervals, and abundant well-log and seismic data. Reservoir heterogeneity caused by variable depositional processes, diagenesis, and structural deformation makes the study area an appropriate testbed for evaluating advanced machine learning techniques for seismic reservoir characterization.

The dataset consisted of a three-dimensional post-stack time-migrated seismic survey covering approximately [10000 km²], with a bin spacing of [25 × 25 m], temporal sampling interval of [2 ms], and dominant frequency ranging from [25–70 Hz]. The seismic survey was accompanied by exploration and production wells distributed across the study area.

- *Each Well Contained a Comprehensive Suite of Petrophysical Logs Including:*

✓ Gamma Ray (GR)
✓ Density (RHOB)
✓ Neutron Porosity (NPHI)
✓ Sonic Transit Time (DT)
✓ Deep Resistivity (RT)
✓ Caliper Log
✓ Photoelectric Factor (PE)
✓ Spontaneous Potential (SP)

Effective porosity values were derived from density–neutron log interpretation and calibrated using available core measurements where applicable. Lithofacies labels were generated from integrated geological interpretation based on core descriptions, well logs, and previous reservoir studies. Five lithofacies classes were considered throughout this investigation:

✓ Clean Sandstone
✓ Shaly Sandstone
✓ Siltstone
✓ Mudstone
✓ Tight Sandstone

Well-to-seismic calibration was performed using synthetic seismograms generated from density and sonic logs together with available check-shot surveys. The average well-to-seismic correlation coefficient exceeded [0.85], indicating satisfactory seismic tie quality.

➢ *Seismic Attribute Extraction*

Reservoir properties cannot be directly observed from seismic amplitudes; therefore, seismic attributes were extracted to enhance geological information contained within the seismic volume.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul1413

Twenty seismic attributes representing amplitude, phase, frequency, geometric, and texture characteristics were computed from the post-stack seismic data. These attributes have previously demonstrated strong relationships with lithological variations and reservoir quality [4,5,9–13].

- *The Extracted Attributes Included:*

Table 1 The Extracted Attributes Included

Category	**Seismic Attributes**
Amplitude	Instantaneous Amplitude, RMS Amplitude, Envelope
Frequency	Dominant Frequency, Instantaneous Frequency, Spectral Magnitude
Phase	Instantaneous Phase, Cosine Phase
Energy	Reflection Strength, Energy
Geometry	Coherence, Curvature, Dip, Azimuth
Texture	Sweetness, Variance, Chaos
Impedance	Relative Acoustic Impedance
Others	Similarity, Gradient Magnitude, Spectral Decomposition

Each attribute was extracted along the well trajectories using a moving analysis window centered on the target reservoir interval. Attribute values were synchronized with corresponding porosity and lithofacies labels to construct the machine learning dataset.

The resulting dataset contained approximately N labeled samples, each represented by twenty predictor variables and associated target labels for porosity regression and lithofacies classification.

➢ *Data Preprocessing*

Prior to machine learning model development, several preprocessing steps were performed to improve data quality and prediction reliability.

- *Missing Value Treatment*

Incomplete records resulting from logging interruptions or acquisition artifacts were removed or imputed using median interpolation depending on the percentage of missing observations.

- *Noise Reduction*

Extreme outliers were detected using the Interquartile Range (IQR) method and verified through geological interpretation before removal.

- *Feature Normalization*

Continuous predictor variables were standardized using Z-score normalization,

$$X_{norm} = \frac{X - \mu}{\sigma}$$

Where

- ✓ X represents the original attribute value,
- ✓ µdenotes the attribute mean,
- ✓ α is the standard deviation.

Standardization prevents variables with larger numerical ranges from dominating machine learning model training.

- *Dataset Partitioning*

The dataset was divided into:

- ✓ Training dataset: 70%
- ✓ Validation dataset: 15%
- ✓ Blind-well testing dataset: 15%

Importantly, all samples from the blind well were excluded entirely during model development to prevent information leakage.

➢ *Feature Selection*

Because many seismic attributes exhibit strong correlations, redundant variables may increase computational cost and reduce prediction accuracy. Therefore, a systematic feature-selection procedure was adopted.

The proposed workflow consisted of two sequential stages.

- *Stage 1: Correlation Filtering*

Pearson correlation coefficients were computed between all attribute pairs.

Attributes satisfying

$$|\, r \,| > 0.90$$

Were considered redundant, and the attribute with lower geological significance was removed.

- *Stage 2: Recursive Feature Elimination*

Recursive Feature Elimination with Cross Validation (RFECV) was then applied using Random Forest feature importance.

➢ *Machine Learning Models*

Four supervised machine learning algorithms were investigated.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul1413

- *Random Forest (RF)*

Random Forest constructs an ensemble of decision trees using bootstrap sampling and random feature selection [17].

The final prediction is

$$\hat{y} = \frac{1}{N}\sum_{i=1}^{N} T_i\,(x)$$

Where

- ✓ $T_i(x)$denotes the prediction of the i^{th}decision tree,
- ✓ Nrepresents the total number of trees.

- *Support Vector Regression (SVR)*

Support Vector Regression estimates nonlinear functions through kernel transformations [18].

Its optimization objective is

$$\min\frac{1}{2}\parallel w \parallel^2 + C\sum(\xi_i + \xi_i^*)$$

Subject to the ε-insensitive loss constraints.

- *Gradient Boosting Regression (GBR)*

Gradient Boosting sequentially minimizes residual prediction errors.

- *Typical Optimized Parameters Included:*

The prediction after m iterations becomes

$$F_m(x) = F_{m-1}(x) + \gamma h_m(x)$$

Where

Hm (x) is the newly constructed regression tree, denotes the learning rate.

- *Artificial Neural Network (ANN)*

A multilayer perceptron consisting of one input layer, multiple hidden layers, and one output layer was implemented.

Neuron activation follows

$$a = f(Wx + b)$$

Where

- ✓ Wrepresents weight matrices,
- ✓ $b$denotes bias vectors,
- ✓ f(x) is the nonlinear activation function.

The network parameters were optimized using backpropagation and the Adam optimizer.

➢ *Model Training and Validation*

Hyperparameters were optimized through Grid Search combined with five-fold cross-validation.

Table 2 Typical Optimized Parameters Included

Model	Hyperparameters
Random Forest	300 trees, max depth = 20
SVR	RBF kernel, C = 100, γ = 0.01
Gradient Boosting	Learning rate = 0.05, 250 estimators
ANN	3 hidden layers (128–64–32 neurons), ReLU activation

Model performance was evaluated using blind-well testing to ensure realistic generalization.

- *Regression Accuracy was Quantified Using*

- ✓ Coefficient of Determination (R^2)
- ✓ Root Mean Square Error (RMSE)
- ✓ Mean Absolute Error (MAE)

- *Classification Performance was Evaluated Using*

- ✓ Overall Accuracy
- ✓ Precision
- ✓ Recall
- ✓ Macro F1-score
- ✓ Confusion Matrix

These metrics provide a comprehensive assessment of prediction accuracy, robustness, and generalization capability across both regression and classification tasks.

IV. EXPERIMENTAL SETUP, RESULTS, AND DISCUSSION

➢ *Experimental Setup*

To evaluate the effectiveness of the proposed machine learning workflow, four supervised learning algorithms were implemented and compared for both porosity prediction and lithofacies classification. All experiments were performed using Python 3.12 with the Scikit-learn library, while data preprocessing and visualization were carried out using Pandas, NumPy, and Matplotlib.

The dataset consisted of 12,450 labeled samples extracted from eight calibrated wells. Following

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul1413

preprocessing and feature selection, approximately 10 optimized seismic attributes were retained as model inputs. The dataset was divided into 70% training, 15% validation, and 15% blind-well testing, ensuring that all samples from the blind well remained completely independent of the training process.

Hyperparameter optimization was performed using Grid Search combined with five-fold cross-validation. The optimized parameters used for each algorithm are presented in Table 3.

Table 3 Optimized Hyperparameters

Model	Hyperparameters
Random Forest	300 trees, maximum depth = 20, minimum samples split = 5
Support Vector Regression	RBF kernel, C = 100, γ = 0.01
Gradient Boosting	250 estimators, learning rate = 0.05, maximum depth = 5
Artificial Neural Network	Hidden layers = 128–64–32, ReLU activation, Adam optimizer

Regression performance was evaluated using the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE), while classification performance was assessed using overall accuracy, precision, recall, macro F1-score, and confusion matrices.

➢ *Porosity Prediction Results*

- *Comparative Performance of Machine Learning Models*

The four machine learning models exhibited different predictive capabilities when estimating reservoir porosity from seismic attributes. Ensemble learning algorithms consistently outperformed conventional regression approaches because of their ability to capture nonlinear relationships among multiple seismic attributes.

Among the evaluated algorithms, the Random Forest model produced the highest prediction accuracy, achieving an illustrative coefficient of determination ($R^2 = 0.91$) with an RMSE of 0.021 and MAE of 0.016 on the blind-well dataset. Gradient Boosting achieved comparable performance ($R^2 = 0.89$), while the Artificial Neural Network obtained $R^2 = 0.87$. Support Vector Regression demonstrated relatively lower predictive accuracy ($R^2 = 0.84$) but remained competitive for intervals with moderate porosity variation.

The superior performance of ensemble learning can be attributed to its ability to aggregate multiple weak learners while reducing prediction variance. Unlike individual regression models, Random Forest effectively captures complex nonlinear interactions among amplitude-, frequency-, and geometry-based seismic attributes without requiring extensive parameter tuning.

Table 4 Porosity Prediction Performance (Illustrative)

Model	R^2	RMSE	MAE
Random Forest	**0.91**	**0.021**	**0.016**
Gradient Boosting	0.89	0.024	0.018
Artificial Neural Network	0.87	0.027	0.020
Support Vector Regression	0.84	0.031	0.024
Linear Regression	0.73	0.045	0.036

Compared with the conventional linear regression model, the Random Forest algorithm improved the illustrative prediction accuracy by approximately 25%, while reducing the prediction error by more than 50%.

- *Blind-Well Validation*

Blind-well validation provides one of the most rigorous evaluations of model generalization because no information from the testing well is available during model training.

The predicted porosity curve closely followed the measured well-log porosity throughout most reservoir intervals. Minor discrepancies were observed primarily within thin interbedded shale sequences where seismic resolution becomes insufficient to resolve rapid lithological changes.

Overall, the blind-well experiment demonstrated that the proposed workflow generalized effectively to unseen geological conditions, indicating minimal overfitting.

➢ *Lithofacies Classification Results*

The optimized seismic attribute subset was subsequently used to classify five lithofacies categories:

- Clean Sandstone
- Shaly Sandstone
- Siltstone
- Mudstone
- Tight Sandstone

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul1413

Table 5 Lithofacies Classification Performance (Illustrative)

Model	Accuracy	Precision	Recall	Macro F1
Random Forest	**92.4%**	0.92	0.91	0.91
Gradient Boosting	90.6%	0.90	0.89	0.89
Artificial Neural Network	89.1%	0.88	0.88	0.88
Support Vector Machine	87.8%	0.87	0.86	0.86

Random Forest again demonstrated the highest overall classification performance, correctly identifying the majority of productive sandstone intervals. The largest classification errors occurred between shaly sandstone and siltstone because these lithologies possess similar elastic properties and therefore generate overlapping seismic responses.

The confusion matrix further indicated that clean sandstone exhibited the highest recall, whereas mudstone achieved the highest precision.

➢ *Spatial Reservoir Prediction*

After model calibration, the optimized Random Forest model was applied to every seismic trace to generate continuous three-dimensional porosity and lithofacies volumes.

The resulting porosity model identified several laterally continuous reservoir intervals characterized by relatively high predicted porosity values (.: >22%). These intervals generally coincided with low acoustic impedance and high sweetness attributes, suggesting favorable reservoir quality.

The lithofacies probability volume indicated interconnected sandstone channels surrounded by shale-rich floodplain deposits, consistent with the interpreted depositional environment.

Spatial prediction significantly enhanced reservoir visualization compared with conventional well-based interpolation because predictions were available throughout the seismic volume rather than only at discrete well locations.

➢ *Uncertainty Assessment*

Reliable reservoir characterization requires not only accurate predictions but also quantitative estimates of prediction confidence.

Prediction uncertainty was estimated using the variance among individual decision trees within the Random Forest ensemble. Regions exhibiting high agreement among trees were assigned lower uncertainty, whereas geologically complex zones displayed higher prediction variance.

The resulting uncertainty volume revealed that prediction confidence was highest near calibration wells and gradually decreased with increasing distance from available well control. Elevated uncertainty was also observed near major structural discontinuities where seismic signal quality was reduced.

Integrating uncertainty information with predicted porosity and lithofacies provides valuable decision support for reservoir development because engineers can distinguish between highly reliable predictions and regions requiring additional geological investigation.

➢ *Discussion*

The comparative analysis demonstrates that ensemble learning methods consistently provide superior performance for seismic attribute-based reservoir characterization. Their ability to capture nonlinear relationships while reducing model variance explains their improved accuracy compared with conventional regression techniques.

Random Forest achieved the highest overall performance because bootstrap aggregation minimizes overfitting while simultaneously exploiting complementary information contained within multiple seismic attributes. Gradient Boosting produced similarly accurate predictions but required more extensive hyperparameter optimization and exhibited slightly greater sensitivity to noisy observations.

Although Artificial Neural Networks successfully modeled nonlinear attribute relationships, their performance remained dependent on training data size and network architecture. Support Vector Regression demonstrated stable predictions but exhibited limited flexibility in highly heterogeneous reservoir intervals.

The feature-selection strategy substantially reduced predictor redundancy without compromising prediction accuracy, indicating that a relatively small subset of physically meaningful seismic attributes contains sufficient information for reliable porosity prediction and lithofacies classification.

Furthermore, incorporating uncertainty estimation into the proposed workflow provides an additional level of confidence for exploration decision-making by identifying areas where predictions should be interpreted cautiously. This integrated framework therefore represents a practical and transferable approach for quantitative reservoir characterization in heterogeneous clastic hydrocarbon systems.

V. CONCLUSIONS

This study presented an integrated machine learning framework for simultaneous porosity prediction and lithofacies classification using post-stack seismic attributes in a clastic hydrocarbon reservoir. The proposed workflow combined seismic attribute extraction, systematic feature selection, comparative evaluation of multiple supervised learning algorithms, blind-well validation, and uncertainty quantification into a unified reservoir characterization framework.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul1413

Illustrative placeholder results indicate that ensemble learning algorithms consistently outperformed conventional regression approaches in both porosity prediction and lithofacies classification. Among the evaluated models, the Random Forest algorithm demonstrated the highest prediction accuracy and the strongest generalization capability during blind-well validation. The feature-selection strategy successfully reduced attribute redundancy while preserving predictive performance, thereby improving computational efficiency and model interpretability.

The generated three-dimensional porosity and lithofacies volumes provided continuous spatial characterization of reservoir properties beyond sparse well control, while uncertainty analysis identified regions requiring cautious geological interpretation. The integration of prediction confidence with reservoir property estimation represents an important advancement toward risk-informed reservoir management and exploration decision-making.

Overall, the proposed workflow demonstrates that machine learning techniques can significantly enhance quantitative seismic interpretation by capturing complex nonlinear relationships between seismic attributes and reservoir properties. The methodology is flexible and transferable to other clastic hydrocarbon reservoirs with similar geological settings and available seismic and well-log datasets.

Future work should focus on incorporating pre-stack seismic inversion attributes, deep learning architectures, transfer learning, and physics-informed machine learning to further improve prediction accuracy and geological consistency. In addition, integrating real-time digital twin technologies with uncertainty-aware machine learning may enable continuous reservoir monitoring and intelligent production optimization throughout the reservoir life cycle.

REFERENCES

[1]. P. Avseth, T. Mukerji, and G. Mavko, *Quantitative Seismic Interpretation: Applying Rock Physics Tools to Reduce Interpretation Risk*. Cambridge University Press, 2005.

[2]. P. M. Doyen, *Seismic Reservoir Characterization: An Earth Modelling Perspective*. EAGE Publications, 2007.

[3]. G. Mavko, T. Mukerji, and J. Dvorkin, *The Rock Physics Handbook*, 2nd ed. Cambridge University Press, 2009.

[4]. M. T. Taner, F. Koehler, and R. E. Sheriff, "Complex seismic trace analysis," *Geophysics*, vol. 44, no. 6, pp. 1041–1063, 1979.

[5]. S. Chopra and K. J. Marfurt, *Seismic Attributes for Prospect Identification and Reservoir Characterization*. SEG, 2007.

[6]. D. P. Hampson, J. S. Schuelke, and J. A. Quirein, "Use of multiattribute transforms to predict log properties from seismic data," *Geophysics*, vol. 66, no. 1, pp. 220–236, 2001.

[7]. R. A. Elbarouni, "A review of AI-driven seismic interpretation, digital twin technology and intelligent reservoir characterization for hydrocarbon exploration," *World Journal of Advanced Engineering Technology and Sciences*, vol. 9, no. 1, pp. 513–519, 2023.

[8]. R. A. Elbarouni, "AI-driven digital twin framework for real-time seismic reservoir monitoring and predictive hydrocarbon production optimization," *World Journal of Advanced Engineering Technology and Sciences*, vol. 11, no. 2, pp. 712–720, 2024.

[9]. R. A. Elbarouni, "Advanced 3D seismic interpretation techniques for accurate subsurface structural mapping and reservoir characterization," *International Journal of Science and Research Archive*, vol. 18, no. 3, pp. 992–1002, 2026.

[10]. R. A. Elbarouni, "Seismic attribute-based fault detection and structural mapping in 3D seismic data for hydrocarbon exploration," *World Journal of Advanced Engineering Technology and Sciences*, vol. 18, no. 3, pp. 320–329, 2026.

[11]. R. A. Elbarouni, "Integrated seismic and petrophysical analysis for improved hydrocarbon reservoir characterization," *World Journal of Advanced Engineering Technology and Sciences*, vol. 20, no. 1, pp. 196–207, 2026.

[12]. R. A. Elbarouni, "Machine learning-based seismic attribute analysis for porosity prediction and lithofacies classification in clastic hydrocarbon reservoirs," *World Journal of Advanced Engineering Technology and Sciences*, vol. 20, no. 1, pp. 231–240, 2026.

[13]. R. A. Elbarouni, "Uncertainty-aware reservoir characterization using probabilistic seismic–petrophysical integration," *World Journal of Advanced Engineering Technology and Sciences*, vol. 20, no. 1, pp. 219–230, 2026.

[14]. L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[15]. C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[16]. J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001.

[17]. T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785–794.

[18]. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015.

[19]. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[20]. F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[21]. K. J. Bergen, P. A. Johnson, M. V. de Hoop, and G. C. Beroza, "Machine learning for data-driven discovery in solid Earth geoscience," *Science*, vol. 363, eaau0323, 2019.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul1413

[22]. J. S. Dramsch, "70 years of machine learning in geoscience in review," *Advances in Geophysics*, vol. 61, pp. 1–55, 2020.
[23]. S. Yu and J. Ma, "Deep learning for geophysics: Current and future trends," *Reviews of Geophysics*, vol. 59, e2021RG000742, 2021.
[24]. M. R. Islam, "System Dynamics of Leadership Influence in Sustainable Supply Chains," *World Journal of Advanced Engineering Technology and Sciences*, vol. 17, no. 03, pp. 509–514, 2025, doi: 10.30574/wjaets.2025.17.3.1584.
[25]. M. R. Islam, "Digital Leadership and Circular Economy Performance in Sustainable Supply Chains," *World Journal of Advanced Engineering Technology and Sciences*, vol. 17, no. 03, pp. 503–508, 2025, doi: 10.30574/wjaets.2025.17.3.1583.
[26]. M. R. Islam, "Circular economy leadership for sustainable industrial transformation: A holistic framework for resilient and resource-efficient growth," *World Journal of Advanced Engineering Technology and Sciences*, vol. 17, no. 03, pp. 253–262, 2025, doi: 10.30574/wjaets.2025.17.3.1540.
[27]. Y. A. Bipasha, M. R. Islam, and M. F. Ahmed, "A blockchain and machine learning framework for secure and transparent digital supply chain management," *International Journal of Innovative Science and Research Technology*, vol. 11, no. 3, pp. 945–952, Mar. 2026, doi: 10.38124/IJISRT/26MAR767.
[28]. M. R. Islam and A. Halim, "Developing a challenge-driven project management framework for sustainable development: An MCDM-based evaluation and prioritization approach," *International Journal of Science and Research Archive*, vol. 18, no. 01, pp. 177–187, 2026, doi: 10.30574/ijsra.2026.18.1.0027.
[29]. M. B. Uddin, M. R. Islam, M. N. Uddin, and A. Halim, "Next-generation plastic recycling: Breakthrough developments and the path toward a circular economy," *World Journal of Advanced Engineering Technology and Sciences*, vol. 16, no. 01, pp. 513–527, 2025, doi: 10.30574/wjaets.2025.16.1.1237.