Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Explainable Physics-Guided Machine Learning for Modeling and Prediction of Facies-Supported Unconfined Compressive Strength: Implication for CO 2 Storage

Ali A. Abdulazeez
Basrah University for Oil and Gas
Abdulsayed H. Abdulsayed
Basrah University for Oil and Gas
Watheq J. Al-Mudhafar
watheq.almudhafar@utexas.edu

Basrah Oil Company
Mojtaba M. Hasak
Basrah University for Oil and Gas
Mohammed H. Abdulredha
Basrah University for Oil and Gas
Raed H. Allawi
Thi-Qar Oil Company

Article

Keywords: Unconfined Compressive Strength, geomechanical bearing capacity, CO2 storage, boosting regression, physics-guided machine learning, explainable artificial intelligence

Posted Date: November 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10080825/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Explainable Physics-Guided Machine Learning for Modeling and Prediction of Facies-Supported Unconfined Compressive Strength: Implication for CO_2 Storage

Ali A. Abdulazeez[1], Abdulsayed H. Abdulsayed[1], Watheq J. Al-Mudhafar[2*], Mojtaba M. Hasak[1], Mohammed H. Abdulredha[1], and Raed H. Allawi[3]

[1]Basrah University for Oil and Gas, Basrah, Iraq
[2]Basrah Oil Company, Basrah, Iraq
[2]Thi-Qar Oil Company, Thi-Qar, Iraq
*Corresponding Author: Watheq J. Al-Mudhafar, Email: watheq.almudhafar@utexas.edu ORCID: 0000-0002-5327-8300

Abstract

The geomechanical stability of CO_2 storage reservoirs, particularly the Unconfined Compressive Strength (UCS) of caprocks and reservoir formations, is critical for ensuring long-term containment security. Traditional UCS estimation methods, laboratory testing and empirical correlations, are often cost-prohibitive, time-intensive, or insufficiently accurate for heterogeneous subsurface systems. This study introduces a novel Explainable Physics-Guided Machine Learning (PGML) framework that, for the first time, combines facies-specific empirical equations with five advanced machine learning algorithms, AdaBoost, Gradient Boosting (GBT), Random Forest (RF), CatBoost, and Radial Basis Function Neural Network (RBFNN), in a discrepancy learning mode. Rather than predicting UCS directly, the models learn the residual between empirical predictions and laboratory-measured values using conventional well logs (porosity, density, gamma ray, resistivity, slowness, and ultrasonic velocity) from 45 core samples across seven wells in the Zubair Oil Field, Iraq. This physics-anchored approach preserves physical interpretability while capturing complex data-driven patterns.

Results demonstrate substantial improvements across all algorithms: PGML increased the coefficient of determination (R^2) from 0.10–0.84 (conventional ML) to 0.82–0.98, with the AdaBoost-PGML model achieving the highest accuracy (R^2 = 0.975, RMSE = 5.9 MPa), a 72% reduction in prediction error compared to conventional AdaBoost (R^2 = 0.68, RMSE = 21 MPa). SHapley Additive exPlanations (SHAP) analysis reveals that depth, porosity, bulk density, and gamma ray are the primary controls on the empirical-model discrepancy, indicating that compaction trends, pore structure, and clay content, absent from simple velocity-based equations, critically influence UCS. Applied to CO_2 storage assessment in the Zubair Field, the framework identifies a thick sandstone reservoir (φ = 16.9%, k = 428 mD) with high-integrity caprock (UCS = 156 MPa) as highly suitable for secure sequestration, while flagging a deeper interval where moderate caprock strength (UCS < 100 MPa) poses leakage risks. This hybrid PGML-XAI approach provides a robust, interpretable, and practically implementable tool for geomechanical characterization, directly supporting risk-informed decision-making in carbon storage projects.

Keywords: Unconfined Compressive Strength; geomechanical bearing capacity; CO2 storage; boosting regression; physics-guided machine learning; explainable artificial intelligence.

1. Introduction

The petroleum industry remains a cornerstone of global energy production but is also a major contributor to greenhouse gas emissions. Carbon dioxide (CO_2), one of the principal greenhouse gases, is released continuously throughout the stages of exploration, extraction, refining, and petrochemical processing. These emissions intensify the greenhouse effect by trapping heat in Earth's atmosphere, thereby accelerating global warming and climate change. The sector's contribution includes both direct emissions from fuel combustion and fugitive leaks, as well as indirect emissions associated with electricity and heat generation. Growing awareness of climate change has emphasized the urgent need to adopt cleaner and more sustainable practices within the petroleum industry. In this context, advancements in carbon capture, utilization, and storage (CCUS) technologies, combined with improved energy efficiency and renewable energy integration, represent essential pathways for mitigating CO_2 emissions (Yan et al., 2021).

Carbon Capture and Storage (CCS) technologies represent a vital area of research in efforts to mitigate climate change. CCS is increasingly recognized as an essential strategy for reducing greenhouse gas emissions by capturing and permanently storing carbon dioxide (CO_2) within deep geological formations. The successful implementation of CCS, however, relies heavily on a comprehensive understanding of the geomechanical stability of storage reservoirs, particularly their bearing capacity. Bearing capacity, a fundamental geomechanical parameter, defines the ability of reservoir rocks to sustain loads without undergoing significant deformation or failure. These subsurface formations must withstand the pressure exerted by injected CO_2 while maintaining long-term structural integrity. Key geomechanical indicators such as shear wave velocity and elastic moduli are instrumental in assessing the mechanical properties and deformation behavior of the storage media. Ensuring reservoir stability is therefore crucial to minimizing risks such as leakage or subsurface collapse, which could compromise the efficiency and safety of CCS operations (Al-Hawary et al., 2023).

In recent years, CO_2 sequestration has attracted increasing attention as a dual-purpose strategy for enhancing oil recovery and mitigating climate change. Among the key geomechanical factors governing the safe and long-term storage of CO_2 is the Unconfined Compressive Strength (UCS) of reservoir rocks. Traditionally, UCS has been estimated using two main approaches: direct laboratory testing and indirect empirical correlations derived from previous studies. However, both methods are often costly, time-consuming, and limited by data availability. In response, machine learning (ML) has emerged as a powerful alternative for efficiently estimating geomechanical parameters. By integrating large datasets from prior measurements with established physical relationships, ML models can rapidly analyze and predict these parameters with accuracy comparable to, and sometimes exceeding, conventional methods. Accordingly, this study includes a critical literature review focusing on UCS as a primary parameter controlling the bearing capacity of formations intended for CO_2 storage.

Several approaches have been developed to estimate the Unconfined Compressive Strength (UCS) of rocks. Direct methods, such as laboratory testing procedures standardized by the International Society for Rock Mechanics (ISRM) and the American Society for Testing and Materials (ASTM), are the most reliable but also the most resource-intensive. Although conceptually straightforward, these tests are expensive, time-consuming, and require carefully prepared core specimens. Obtaining cores of sufficient quality can be particularly challenging in highly fractured, weak, or weathered formations. Indirect methods, by contrast, are simpler and more economical because they generally do not require physical sample preparation. For this reason, they are often preferred in practical applications. Conventionally, empirical correlations have been developed to relate UCS to parameters obtained from indirect tests, such as the point load test, indentation test, and Schmidt hammer rebound test, using regression techniques including linear, power-law, logarithmic, and exponential models (Li et al., 2024). Furthermore, several empirical relationships have been proposed to estimate rock strength based on geophysical well-log data. Such correlations are especially valuable when core samples are unavailable, as many factors influencing rock strength also affect physical properties like wave velocity, elastic moduli, and porosity (Chang et al., 2006). More recently, advances in computing have enabled the adoption of intelligent modeling techniques, including machine learning (ML) algorithms, artificial neural networks (ANNs), fuzzy inference systems (FIS), and hybrid approaches, for more accurate and efficient UCS prediction.

Numerous researchers have developed predictive models to estimate the Unconfined Compressive Strength (UCS) of rocks. Ceryan et al. (2012) proposed a Levenberg–Marquardt algorithm-based artificial neural network (LM-ANN) model to predict the UCS of carbonate rocks formed in different facies. Their results demonstrated that the LM-ANN algorithm effectively modeled UCS and could also be applied to other geotechnical parameters, outperforming the traditional multiple linear regression (MLR) approach. In the same study, Ceryan et al. (2012) developed generalized regression neural networks (GRNNs) using various input combinations of one, two, and three parameters. Across all configurations, the GRNN models exhibited superior performance compared with MLR and feed-forward back-propagation neural networks (FFNNs). Çelik (2019) predicted the UCS values of five carbonate rock types, marble, dolomite, two limestone varieties, and travertine, and found that the least squares support vector machine (LS-SVM) method achieved more accurate UCS predictions than multivariate regression analysis (MR). Similarly, He et al. (2019) integrated a deep convolutional neural network (DCNN) technique into the drilling process, demonstrating that the DCNN provided higher accuracy in UCS estimation across various rock types compared with the traditional Mohr–Coulomb criterion. This approach effectively addressed key limitations associated with the limit equilibrium-based drilling method. Barzegar et al. (2019) developed and compared three standalone tree-based machine learning models, Random Forest (RF), M5 Model Tree, and Multivariate Adaptive Regression Splines (MARS), for predicting

the UCS of travertine rocks. Additionally, they introduced an ensemble committee-based ANN model that combined the strengths of the individual models, yielding further improvements in UCS prediction accuracy.

Gowida et al. (2021) developed three predictive models using different algorithms, Artificial Neural Networks (ANN), Adaptive Neuro-Fuzzy Inference System (ANFIS), and Support Vector Machine (SVM), to estimate UCS from drilling parameters. Among these, the ANN model produced the most accurate results, showing excellent agreement between predicted and measured UCS values, with a correlation coefficient (R) exceeding 0.98 and an average absolute percentage error (AAPE) of 4.2%. Hassan and Arman (2022) compared the performance of four machine learning techniques, Principal Component Regression (PCR), Sequential Artificial Neural Networks (SANN), Multiple Linear Regression (MLR), and ANFIS, for predicting the UCS of carbonate rocks. Their results indicated no statistically significant differences among the models; however, based on overall accuracy metrics (RMSE, MAE, and R^2), the SANN model demonstrated a slight advantage over the others. Arachchilage et al. (2023) employed two ensemble machine learning methods, Gradient Boosting Regression (GBR) and Random Forest (RF), alongside two single models, Support Vector Regression (SVR) and ANN, to predict the UCS of alkali-activated slag (AAS)-based cemented paste backfill (CPB). Their findings revealed that the tree-based ensemble methods (GBR and RF) outperformed the single algorithms (SVR and ANN). Similarly, Afolagboye et al. (2023) applied four ML models, RF, Relevance Vector Machine (RVM), SVM, and ANN, to predict the UCS of Precambrian basement rocks, concluding that the SVM model achieved the highest prediction accuracy. Feature importance analysis further showed that the point load strength index was the most influential parameter across all models.

Wen et al. (2024) constructed three nonlinear models, Back Propagation (BP), Particle Swarm Optimization Back Propagation (PSO-BP), and Particle Swarm Optimization Least Squares Support Vector Machine (PSO-LSSVM), to predict UCS using six index-based datasets for various rock types. Their error analysis demonstrated that the PSO-LSSVM model outperformed the other two approaches. Niu et al. (2024) proposed a K-Nearest Neighbor (KNN)-based UCS prediction model that leverages the algorithm's instance-based learning capability, which requires minimal training and enables real-time adaptability to dynamic geological conditions. The study further enhanced KNN performance by integrating two metaheuristic optimization algorithms, Alibaba and the Forty Thieves (AFT) and Improved Manta-Ray Foraging Optimizer (IMRFO). The optimized KNN model (KNIM), developed using IMRFO, achieved higher accuracy than both the standard KNN and the AFT-optimized variant (KNAF), confirming the robustness and interpretability of the proposed approach.

Despite the proliferation of machine learning applications for UCS prediction, a critical review of the most recent literature reveals several persistent gaps that limit the practical utility of existing models for CO_2 storage assessment. Wang et al. (2025) demonstrated that Random Trees (RT) and Support Vector Regression (SVR) achieve high accuracy ($R^2 = 0.959$) using measurement-while-drilling data, yet their purely data-driven approach lacks physical constraints and the MWD inputs may not be available in legacy wells or CO_2 storage characterization campaigns. Miah et al. (2025) conducted rigorous feature importance analysis using LSSVM and RF on well-log data from sandstone reservoirs, identifying gamma ray and compressional wave velocity as most influential; however, their models remain purely data-driven and do not incorporate established geomechanical relationships. Rahaman and Miah (2025) advanced the field by applying Transformer architectures and SHAP analysis to UCS prediction, achieving excellent accuracy ($R^2 = 0.99$), but their reliance on Schmidt hammer rebound and point load tests, which require field campaigns, limits applicability to subsurface characterization where only well logs are available. Khan et al. (2025) developed sophisticated composite XGBoost models for organic-rich shales incorporating wettability and fabric metrics ($R^2 = 0.981$), yet their focus on a single lithology and specialized inputs restricts generalizability to heterogeneous reservoir sequences. Zhao and Shi (2025) achieved remarkable accuracy ($R^2 = 0.995$) using optimized KNN for stabilized sands, but their focus on engineered materials rather than natural reservoir rocks limits relevance to geological storage. Collectively, these studies demonstrate the power of modern ML techniques but reveal critical gaps: (1) absence of physics integration to ensure physical consistency; (2) limited attention to facies-dependent behavior; (3) narrow application domains that do not address CO_2 storage requirements; and (4) insufficient interpretability connecting predictions to underlying geological processes. This study addresses these gaps through a novel PGML framework that combines facies-specific empirical equations with five ML algorithms in a discrepancy learning mode, applies SHAP to interpret the physical meaning of the modeled discrepancy, and demonstrates practical application to CO_2 storage site screening in a heterogeneous carbonate-clastic system.

Table 1 provides a summary of representative studies that have developed predictive models for estimating the Unconfined Compressive Strength (UCS) of various rock types. The table outlines the algorithms or modeling

techniques employed in each study, ranging from empirical correlations and statistical regression to advanced machine learning approaches such as ANN, SVM, RF, and ensemble methods, along with the key input parameters used and their reported accuracy metrics. This compilation highlights the evolution of UCS prediction methodologies from traditional empirical formulations toward intelligent, data-driven approaches that achieve greater precision and efficiency.

Table 1: Comprehensive summary of representative studies for UCS prediction using machine learning models, including recent advances (2025) and identification of research gaps addressed by this study.

Author	Method	Inputs	Data Sets	Lithology	R^2	Key Gap/limitation addressed by this study
Ceryan et al. (2012)	LM-ANN	V_m , n	56	Carbonate Rocks	0.8126	Single algorithm; no physics integration; limited inputs (only V~m~ and porosity); no facies consideration; black-box model with no interpretability
Ceryan et al. (2012)	GRNNs	Vm , n, Id	56	Carbonate Rocks	0.7664	Limited feature set; no physics guidance; moderate accuracy (R^2 < 0.80); no uncertainty quantification; no application to CO_2 storage
Çelik (2019)	LS-SVM	V_p , SHR, L	90	Carbonate Rocks	0.867	Requires non-destructive tests (Schmidt hammer) not always available in wells; single lithology; no physics constraints; no interpretability
He et al. (2019)	DCNN	Fn, ROP, T, RPM	76	various rock types	0.998	Requires drilling parameters (not available in legacy wells); purely data-driven; no physical principles; black-box deep learning with no explanation of predictions
Barzegar et al. (2019)	MARS, RF, M5 tree, ensemble ANN-committee model	V_p, R_n, n, I_s	93	travertine	0.689 0.23 0.327 0.792	Moderate to low accuracy; single lithology (travertine only); no physics integration; limited feature importance

						analysis
Gowida et al. (2021)	ANN, ANFIS, SVM	ROP, GPM, SPP, RPM, T, WOB	1771	various rock types	0.846 0.941 0.980	Requires real-time drilling parameters; purely data-driven; no facies consideration; black-box models with no interpretability; no validation for CO_2 storage context
Hassan &Arman (2022)	SANN, PCR, MLR, ANFIS	SHVC, γn	93	various rock types	0.402 0.402 0.425 0.450	Low accuracy (R^2 < 0.45); limited feature set; no physics guidance; statistical methods unable to capture complex nonlinear relationships
Arachchilage et al. (2023)	GBR, SVR, RF, ANN	PSD characteristics of tailings, mixture design characteristics, chemical composition of the activator, tailings, and slag.	N/A	(AAS)-based (CPB)	0.887 0.889 0.907 0.969	Focus on engineered cemented paste backfill (not natural rocks); laboratory-based inputs not available in wells; no CO_2 storage application
Afolagboye et al. (2023)	RF, RVM,, SVM, ANN, MLR	ACV, AIV, LAAV, SHR, Is (50)	N/A	Precambrian basement rocks	0.87 0.87 0.90 0.90 0.90	Requires aggregate tests (not available in wells); purely data-driven; no physics integration; limited to basement rocks (not reservoir lithologies)
Wen et al. (2024)	BP, PSO-BP, PSOLSSVM	Is (50), V_p, Rn, SSH, BTS	N/A	shale, sandstone, volcanic rock, and granite	0.92 0.95 0.96	Uses point load and Brazilian tests (not routine well logs); purely data-driven; no facies-specific modeling; no application to CO_2 storage
Niu et al. (2024)	KNN, KNIM, KNAF	SRn, BD, BTS, DD, Vp, Is(50)	106	hydrothermally altered soft rocks	0.970 0.986 0.980	Requires multiple lab tests; focuses on altered soft rocks (not typical reservoirs); no physics guidance;

						black-box KNN with limited interpretability
Wang et al. (2025)	MLP, SVR, CNN, RT, LSTM	MWD parameters	197	Multiple	0.959	Purely data-driven; MWD inputs not always available; no physics integration
Miah et al. (2025)	LSSVM-CSA, RF	GR, Vp, Vs, ρb, E	183	Sandstone	0.96	Purely data-driven; single lithology; no explainable AI
Rahaman & Miah (2025)	CNN, Transformer, SVM, FNN	SHR, point load, porosity, Vp	N/A	Sedimentary	0.99 (training), 0.95 (testing)	Inputs require field/lab tests; no physics; no CO_2 application
Khan et al. (2025)	XGBoost (simple/composite)	UPV, fabric metrics, wettability	1217	Organic shale	0.981 (training), 0.91 (validation)	Single lithology; specialized inputs; no physics integration
Zhao & Shi (2025)	KNN-GTO, KNN-TSA	LL, PL, PI, cement/lime	N/A	Stabilized sand	0.995	Engineered materials; not natural rocks; no CO_2 relevance
This study	AdaBoost, GBT, RF, CatBoost, RBFNN with PGML	Depth, ρb, GR, RESD, porosity, ΔT, Vp, facies	**45 cores, 7 wells**	Sandstone, limestone, siltstone, shale	0.975 (AdaBoost-PGML)	First to integrate facies-specific physics with discrepancy learning; SHAP for physical interpretation; CO_2 storage application

The effectiveness of carbon dioxide (CO_2) geological storage is largely governed by the geomechanical properties of the storage formations, emphasizing the need to understand their bearing capacity, particularly across different geological facies. Traditional approaches for modeling geomechanical behavior often fail to capture the complexity of subsurface conditions, underscoring the importance of advanced techniques such as physics-guided machine learning (PGML) to enhance predictions of geomechanical bearing capacity in facies-supported environments. Variations in geological facies can substantially influence the mechanical response of storage formations, and an incomplete understanding of these effects introduces significant risks to the long-term stability of CO_2 storage sites. Moreover, a gap persists in integrating physics-based models with data-driven machine learning approaches, an integration essential for developing robust predictive frameworks that can inform decision-making in carbon storage operations. Addressing uncertainties in geomechanical behavior is therefore vital, not only to ensure the safety and feasibility of carbon storage projects but also to satisfy regulatory standards and strengthen public confidence in such technologies (Bachu et al., 2007).

A critical examination of the literature reveals significant variation in how different studies have leveraged petrophysical features for UCS prediction. Among the key well-log parameters most relevant to rock strength, porosity (ϕ), gamma ray (GR), acoustic slowness (ΔT), compressional wave velocity (Vp), and shear wave velocity (Vs), no single prior study has comprehensively integrated all these features within a physics-guided framework. Miah et al. (2020, 2025) demonstrated that GR and Vp are the most influential predictors for sandstone reservoirs using LSSVM and RF algorithms, yet their approach remained purely data-driven without embedding physical constraints. Similarly, Rahaman and Miah (2025) incorporated Vp and neutron porosity with transformer architectures, achieving high accuracy, but their reliance on Schmidt hammer rebound data, which requires field campaigns, limits applicability to subsurface characterization where only well logs are available. Wang et al. (2025) utilized measurement-while-drilling parameters exclusively, omitting petrophysical logs entirely, while Khan et al. (2025) focused on organic-rich shales using specialized fabric and wettability metrics not routinely available. Studies by Ceryan et al. (2012), Çelik (2019),

and Barzegar et al. (2019) incorporated Vp and porosity but omitted GR and resistivity, missing critical lithological and fluid saturation effects. This fragmented utilization of petrophysical features across the literature creates a fundamental gap: no existing framework simultaneously (1) leverages the full suite of readily available well logs (ϕ, GR, ΔT, Vp, resistivity, density), (2) embeds facies-specific physical principles, and (3) provides interpretable connections between predictions and underlying geological processes. The present study addresses this gap by integrating all key petrophysical features within a PGML framework that preserves physical interpretability while capturing complex data-driven patterns.

The primary objective of this study is to develop a sequential physics-guided machine learning framework that addresses the limitations identified in recent literature (Wang et al., 2025; Miah et al., 2025; Rahaman & Miah, 2025; Khan et al., 2025). Unlike purely data-driven approaches, our PGML framework embeds facies-specific empirical equations (Table 3) as physical anchors, learning only the discrepancy between these physics-based predictions and measured UCS. This approach preserves physical interpretability while leveraging data-driven flexibility, a fundamental advancement over the black-box models prevalent in recent studies. We systematically compare five diverse algorithms (boosting, bagging, and neural network architectures) within the same PGML framework, providing insights into which model types benefit most from physics guidance, a comprehensive evaluation absent from previous work. Furthermore, we apply SHAP analysis not merely to rank feature importance (as in Rahaman & Miah, 2025) but to interpret the physical meaning of the modeled discrepancy, revealing which geological parameters capture the limitations of empirical equations (Nauta et al., 2023). Finally, we demonstrate the practical utility of this framework for CO_2 storage decision-making by identifying suitable and risky injection intervals based on predicted UCS and caprock integrity, an application domain not addressed in recent UCS prediction studies. This integrated approach directly responds to Bachu et al.'s (2007) call for methods that combine physical principles with data-driven techniques for reliable CO_2 storage assessment.

The accurate determination of Unconfined Compressive Strength (UCS) is fundamental to geomechanical characterization for CO_2 storage, yet existing approaches face three fundamental challenges that this study directly addresses. First, the data availability paradox**:** laboratory testing provides the most reliable UCS measurements but requires intact core samples that are costly and often unavailable, particularly in heterogeneous formations or from legacy wells. Empirical correlations using well logs are widely applicable but lack accuracy for site-specific conditions. This creates a paradox where the most accurate methods are least available, while the most available methods are least accurate. Second, the physics-data disconnect: as demonstrated in Table 1, existing machine learning approaches achieve high accuracy but operate as black boxes, learning correlations without respecting physical principles. When applied outside their training distribution, inevitable in heterogeneous geological systems—these models can produce physically implausible predictions with no mechanism for correction. Conversely, purely physics-based models capture first-order behavior but miss complex, site-specific patterns that data could reveal. Third, the interpretability gap: even when ML models perform well, they provide little insight into *why* a particular UCS value is predicted. This lack of interpretability undermines confidence in predictions and prevents geoscientists from connecting model outputs to underlying geological processes, essential for risk assessment in CO_2 storage where decisions have long-term safety implications.

This study is crucial because it introduces a Physics-Guided Machine Learning (PGML) framework that simultaneously resolves all three challenges. First, it uses only conventional well logs (ρb, GR, RESD, porosity, ΔT, Vp), the most widely available subsurface data—ensuring broad applicability to existing well datasets without requiring additional field campaigns or laboratory tests. Second, it embeds facies-specific empirical equations as physical anchors, ensuring predictions remain grounded in geomechanical principles while allowing data to capture site-specific heterogeneity, bridging the physics-data disconnect. Third, it applies SHAP analysis to interpret the physical meaning of the modeled discrepancy, revealing which geological parameters (compaction from depth, clay content from GR, pore structure from porosity) explain the limitations of simple empirical models. By bridging the physics-data disconnect and providing geological interpretability, this framework enables reliable UCS prediction for CO_2 storage assessment where neither purely physics-based nor purely data-driven approaches alone would suffice. The framework's ability to identify suitable versus risky storage intervals (Section 5) demonstrates its practical value for real-world decision-making in carbon sequestration projects.

2. Field Description

The Zubair Oil Field, also known as Az-Zubayr Field, is located in southern Iraq, approximately 20 km southwest of Basrah Governorate (Lazim and Dawood, 2019). Discovered in 1949, the field covers an area of about 1,170 km², extending 65 km in length and 18 km in width. It lies within the Mesopotamian Basin, a region long recognized as one of the most prolific hydrocarbon provinces in the Middle East. The basin is characterized by extensive sandstone reservoirs, with the Zubair, Nahr Umr, and Mishrif formations representing the principal oil-producing intervals (Z.F.O.D. Annual Report). Structurally, the field forms a gentle anticline resulting from tectonic inversion associated with the convergence between the Arabian and Eurasian plates (Sissakian et al., 2014). The Zubair structure comprises four main domes, Hammar, Shuaiba, Rafidiyya, and Safwan, arranged from north to south along a northwest–southeast-trending axis. The anticline displays asymmetrical geometry, with the western flank notably steeper than the eastern one at both the Shuaiba and Rafidiyya domes (Lazim and Dawood, 2019).

The Zubair reservoirs consist primarily of sandstones interbedded with shale and siltstone (Al-Jaberi and Al-Jafar, 2020). These lithologies exhibit high porosity and permeability, providing substantial pore space and efficient fluid pathways that favor hydrocarbon production. The reservoir is overlain by a thick shale cap rock, which serves as an effective seal preventing fluid migration (Al-Douri et al., 2024). This sealing capacity is critical for hydrocarbon retention and also positions the Zubair Field as a promising candidate for long-term geological CO_2 storage, as the cap rock can effectively trap injected CO_2 and mitigate leakage risks.

Figure 1. (A) Location map of the Zubair Oil Field in southern Iraq, illustrating its proximity to the Basrah Governorate and surrounding structural features. (B) Structural map of the Zubair Field showing the major domes. Hammar, Shuaiba, Rafidiyya, and Safwan, arranged along the northwest–southeast structural trend. The anticline geometry and associated structural elements, as depicted by Almallah et al. (2019), highlight the field's regional tectonic framework and reservoir compartmentalization.

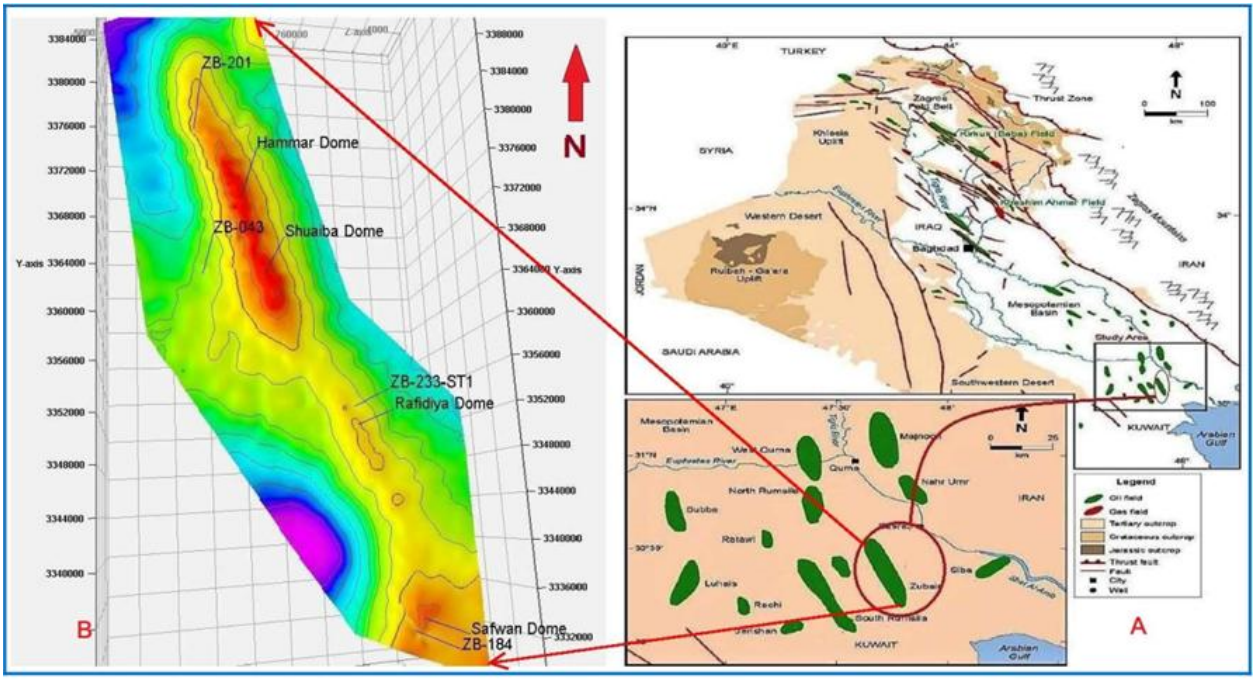


Figure 1: A) Location map of the Zubair oil field. (B) Structural map of Zubair oil field (Almallah et al., 2019).

In southern Iraq, the Early Cretaceous Zubair Formation represents a major hydrocarbon-bearing unit within a large anticline. It consists mainly of sandstones interbedded with siltstone, shale, and minor carbonates, lying stratigraphically between the underlying Ratawi Shales and the overlying Shuaiba Carbonates (Al-Muhailan et al., 2013; Al-Jafar and Al-Jaberi, 2021). The depositional environment of the formation reflects complex sedimentary dynamics: the upper and middle sandstone units exhibit tidal and estuarine influences, whereas the lower portion reflects fluvial and mouth-bar systems (Al-Mudhafar, 2015). Core descriptions reveal upward-coarsening sequences with cross-bedded and bioturbated fine sandstones, consistent with high-energy deltaic settings (Harris et al., 2012).

The formation is typically divided into five main elements: Upper Shale, Upper Sandstone (USS), Middle Shale, Lower Sandstone, and Lower Shale. Within the USS, five key subunits (A, B, H, L, and N) are identified as the principal oil-bearing zones, with the H unit being of particular significance due to its excellent reservoir quality (Al-Jafar and Al-Jaberi, 2019).

Figure 2. (A) Location of the study area within southern Iraq, emphasizing the regional geological context of the Zubair Formation. (B) Generalized stratigraphic column of the southern Iraq region, showing the main lithological units relevant to the study area. The column delineates the position of the Zubair Formation between the underlying Ratawi Shale and the overlying Shuaiba Carbonate, as described by Al-Jafar and Al-Jaberi (2021). This stratigraphic framework provides essential context for interpreting reservoir characteristics and depositional environments within the study field.

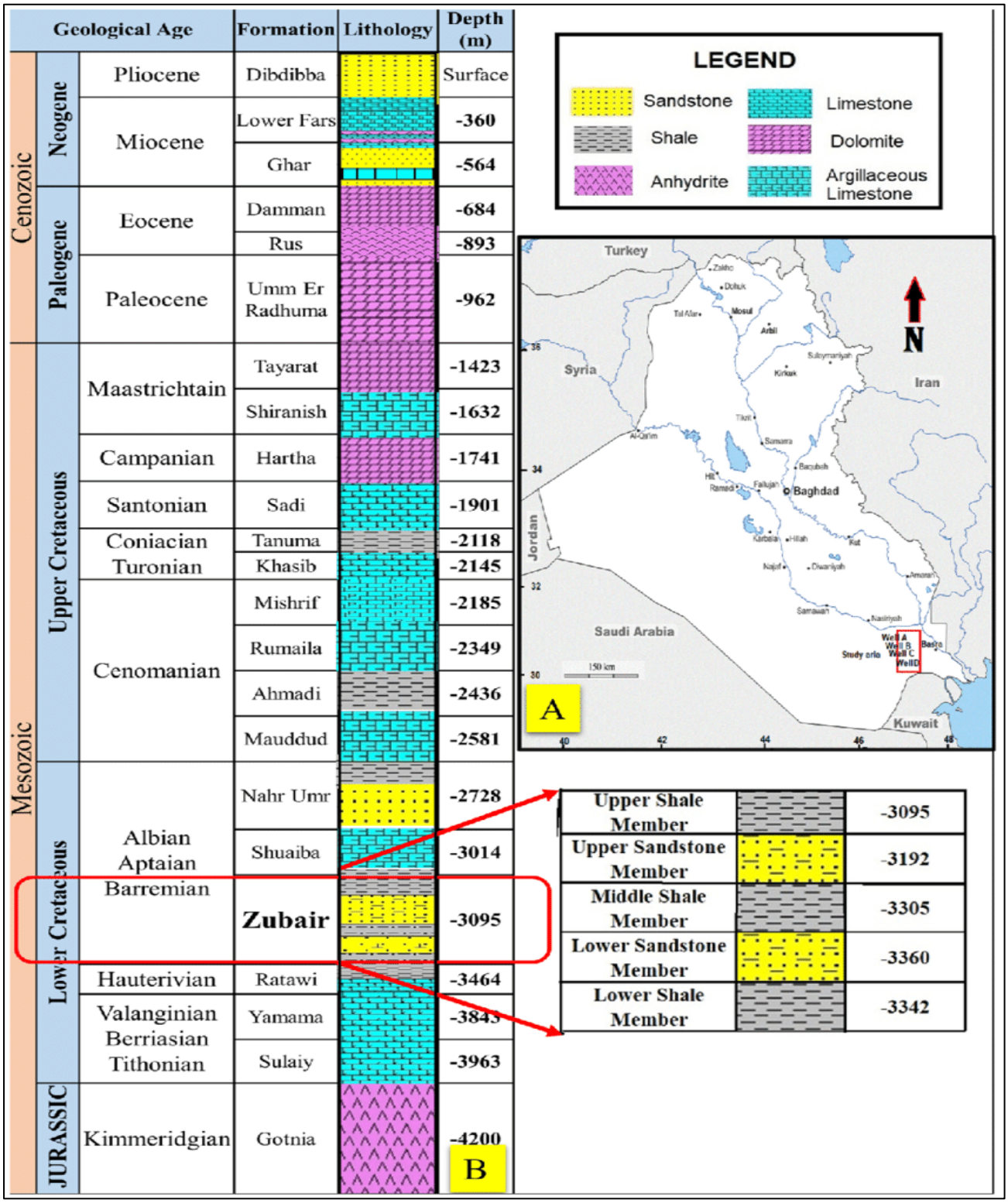


Figure 1: A. Location of the study area, southern Iraq, B. A stratigraphic column in southern Iraq for the study area (Al-Jafar and Al-Jaberi, 2021).

3. Methodology

The workflow begins with data collection and preprocessing, during which well-log and rock-property datasets are gathered, cleaned, analyzed, and imputed to address missing values and ensure overall data quality. Multiple machine learning algorithms, Adaptive Boosting (AdaBoost), Gradient Boosting (GBT), Random Forest (RF), Categorical Boosting (CatBoost), and Radial Basis Function Neural Network (RBFNN), are implemented under both Physics-Guided Machine Learning (PGML) and conventional machine learning frameworks to predict the Unconfined Compressive Strength (UCS) using well-log measurements as input parameters. Model performance is enhanced through hyperparameter optimization using Random Grid Search, and each model is trained, validated, and evaluated using performance metrics such as the coefficient of determination (R^2) and root mean square error (RMSE).

To ensure interpretability and reliability, the Explainable Artificial Intelligence (XAI) technique SHapley Additive exPlanations (SHAP) is applied to analyze model outputs and identify the most influential input features. This interpretive step enhances transparency and builds confidence in the predictive framework. Finally, the best-performing models are employed to predict UCS across a representative sector of the formation, allowing evaluation of the predicted geomechanical properties and their suitability for CO_2 storage feasibility assessment**.**

Figure 3 illustrates the overall workflow adopted in this study for developing and implementing the Physics-Guided Machine Learning (PGML) framework to predict the Unconfined Compressive Strength (UCS) and evaluate its implications for CO_2 storage. The process begins with data collection and preprocessing, where well-log and rock-property datasets are compiled, cleaned, and subjected to missing-value imputation to ensure data completeness and quality. The prepared dataset is then used to train and validate multiple machine learning models under both PGML and conventional ML paradigms.

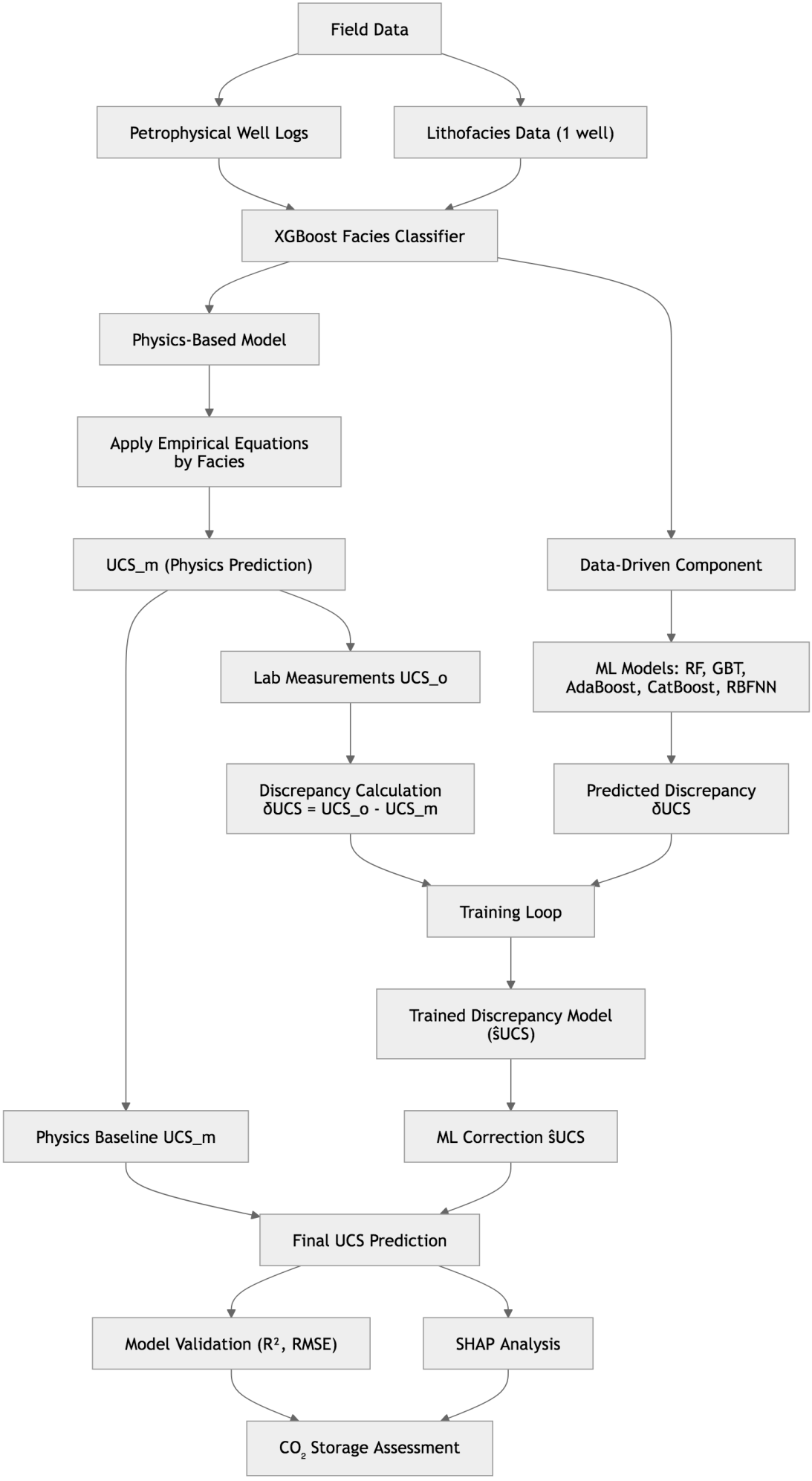

Field Data
Petrophysical Well Logs
Lithofacies Data (1 well)
XGBoost Facies Classifier
Physics-Based Model
Apply Empirical Equations by Facies
UCS_m (Physics Prediction)
Data-Driven Component
Lab Measurements UCS_o
ML Models: RF, GBT, AdaBoost, CatBoost, RBFNN
Discrepancy Calculation δUCS = UCS_o - UCS_m
Predicted Discrepancy δUCS
Training Loop
Trained Discrepancy Model (δ̂UCS)
Physics Baseline UCS_m
ML Correction δ̂UCS
Final UCS Prediction
Model Validation (R², RMSE)
SHAP Analysis
CO2 Storage Assessment

Figure 3: Generalized workflow of the proposed Physics-Guided Machine Learning (PGML) framework for Unconfined Compressive Strength (UCS) prediction. The framework operates in two stages: (1) Physics-based baseline: Facies-specific empirical equations (Table 3) generate initial UCS predictions (UCS_{pred}) from ultrasonic velocity (Vp). (2) Discrepancy learning: Five standard ML algorithms (AdaBoost, GBT, RF, CatBoost, RBFNN) are trained to predict the discrepancy (δUCS) between measured UCS (UCS_o) and physics-based predictions using the full suite of well-log features. Final predictions combine the physics baseline with ML-predicted discrepancies. The "physics-guided" designation applies to the overall framework architecture, not to modifications of the ML algorithms themselves.

3.1 Data Collection and Analytic

3.1.1 Data Collection

To evaluate the accuracy of the proposed Physics-Guided Machine Learning (PGML) framework, this study aimed to predict the Unconfined Compressive Strength (UCS) across different lithologies. A total of 45 core samples were collected from seven wells penetrating sandstone, limestone, siltstone, and shale formations within the Zubair Field. Each sample was subjected to laboratory UCS testing following ISRM suggested methods, providing measured UCS values (UCS_o) that serve as the target variable for model training and validation.
For each sample location, conventional well logs were acquired, including:

- Porosity (ϕ, %): Total porosity from density/neutron combination
- Bulk density (ρb, g/cc): Formation density from density log
- Gamma ray (GR, gAPI): Natural radioactivity, indicator of clay content and lithology
- Deep resistivity (RESD, Ω·m): Formation resistivity, indicator of fluid saturation and pore structure
- Compressional slowness (ΔT, μs/ft): Acoustic transit time from sonic log—the direct measurement of compressional wave travel time
- Ultrasonic P-wave velocity (Vp or USV**,** m**/s)**: Compressional wave velocity derived from ΔT (Vp = 1/ΔT, with appropriate unit conversion)

These log measurements serve as input features for the predictive models. **Note**: Both ΔT and Vp are included in the feature set. While they are mathematically related (inverse), providing both allows the ML models to capture nonlinear relationships and deviations from the simple inverse relationship that might be present in the data due to measurement physics or formation effects.

3.1.2 Data Analytic

The dataset was divided into two subsets: 80% for model training and 20% for testing. During data preparation, two major challenges were encountered.

Missing Value Imputation: Missing values in some parameters were addressed using two imputation techniques: Random Forest Regressor (RF) and Multiple Imputation by Chained Equations (MICE). These methods were chosen for their effectiveness in handling incomplete datasets while maintaining statistical integrity.

Facies Classification: Facies data were available for only one well, limiting direct use in model training. To overcome this limitation, an XGBoost classifier was trained using available well-log data (GR, resistivity, density, porosity, ΔT, Vp) to predict facies (sandstone, limestone, siltstone, shale) for the remaining six wells. The model achieved 97% accuracy on the test set (20% of the dataset), as illustrated in Figure 4. These predicted facies serve two purposes:

1. Selecting the appropriate facies-specific empirical equation (Table 3) for physics-based baseline prediction.
2. Serving as an input feature in the ML models to capture lithology-dependent geomechanical behavior.

A critical aspect of data preparation was the determination of porosity (PHI) from well logs. The porosity values reported in Table 2 represent effective porosity, the interconnected pore space relevant to fluid flow and geomechanical behavior, obtained through a systematic log interpretation workflow:

1. Density porosity (ϕD) was calculated using lithology-specific matrix densities (2.65 g/cc for sandstone, 2.71 g/cc for limestone, 2.68-2.77 g/cc for siltstone/shale) based on facies classification.
2. Clay volume (Vshale) was estimated from gamma ray using the Larionov (1969) method for tertiary rocks.
3. Effective porosity (ϕe) was obtained by applying clay correction to either the density-neutron crossplot porosity (for moderately shaly formations) or the density porosity (for shaly formations) $\phi e = \phi_{DN} - V_{shale} \times \phi_{N,sh}$ where ϕDN is the density-neutron crossplot porosity and $\phi N,sh$ is the neutron porosity in adjacent shale zones.
4. All calculated porosities were validated against available core measurements (where available), achieving correlation coefficients (R^2) exceeding 0.85, confirming interpretation reliability.

This approach ensures that the porosity values used as input features represent the effective pore space that controls mechanical behavior, rather than total porosity, which includes isolated and clay-bound pores with different mechanical properties.

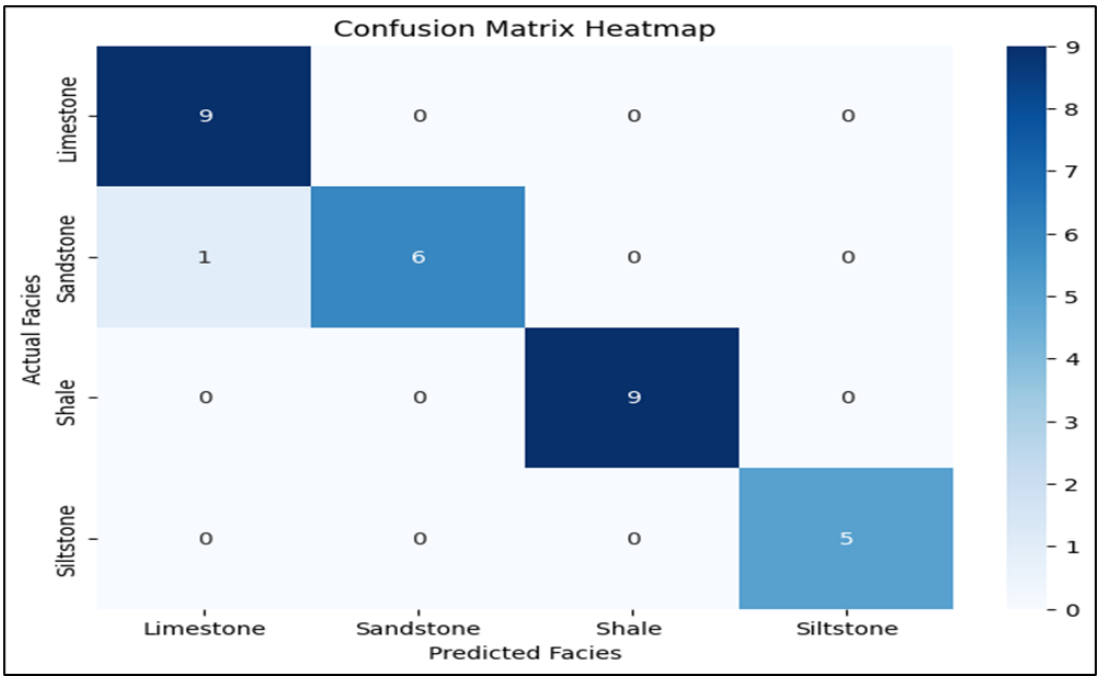


Figure 4: Heatmap confusion matrix for the test result of the classification model.

3.1.3 UCS Measurement vs. Prediction

It is essential to distinguish between two uses of "UCS" in this study:

- Measured UCS (UCS_o): Obtained directly from laboratory tests on 45 core samples. These are the true values used for training, discrepancy calculation, and validation.
- Physics-based UCS (UCS_m): Calculated from Vp using facies-specific empirical equations (Table 3). These are baseline predictions that intentionally omit other geological factors to isolate their effects.
- ML-predicted discrepancy (δUCS): The output of machine learning algorithms trained to predict the difference between measured and physics-based UCS using the full suite of well-log features.
- Final predicted UCS (UCS_final): The sum of physics-based baseline and ML-predicted discrepancy, representing the PGML framework's estimate for intervals without core data.

During model development, the measured UCS values (UCS_o) are used to calculate the discrepancy ($\delta UCS = UCS_o - UCS_m$) and to train the ML models. Once validated, the PGML framework can estimate UCS for any interval with available well logs, even without core samples, by applying the two-stage process illustrated in Figure 3.

Table 2 presents a summary of the core-based Unconfined Compressive Strength (UCS) measurements collected from the studied wells along with the input features of well logging attributes. The table outlines the range, mean, and variability of UCS values across different lithologies, including sandstone, limestone, siltstone, and shale. These data provide a fundamental reference for evaluating the mechanical strength of the formations and serve as the target

variable for model training and validation. The observed variation in UCS magnitudes reflects the heterogeneity of the subsurface formations, emphasizing the need for advanced modeling approaches such as Physics-Guided Machine Learning (PGML) to capture these complex geomechanical behaviors.

The target variable (UCS) also exhibits wide variability, ranging from 37.11 to 258.28 MPa. This heterogeneity underscores the complexity of the subsurface and reinforces the need for advanced modeling techniques to accurately predict geomechanical properties such as UCS in heterogeneous formations. Table 3 summarizes the statistical characteristics of the well-log data used as input features in the machine learning models. The parameters include depth, bulk density, porosity, gamma ray (GR), slowness (ΔT), resistivity, and ultrasonic P-wave velocity (USV). For each parameter, the table reports the minimum, maximum, mean, and standard deviation values, providing an overview of the data distribution and variability within the studied interval. These statistics highlight the wide range of lithological and petrophysical conditions across the formation, which directly influence the geomechanical response and, consequently, the prediction of UCS.

Table 2: Statistical summary of core-derived Unconfined Compressive Strength (UCS) and well-log parameters employed as input features for machine learning models, including depth, bulk density, porosity, gamma ray (GR), slowness (ΔT), resistivity, and ultrasonic P-wave velocity (USV).

Parameters	Mean	Std Dev	Minimum	25th Percentile	50th Percentile (Median)	75th Percentile	Maximum
Target Variable							
UCS (MPa)	142.59	43.58	37.11	111.03	149.35	170.83	258.28
Input Features							
Depth, MD (m)	3314.74	392.06	2388.96	3285.8	3395.28	3462.1	4003.4
Corrected Bulk Density, ρ (g/cc)	2.39	0.13	1.98	2.34	2.39	2.44	2.63
Slowness, Delta-T (μs/ft)	69.76	6.93	50.74	66.11	71.08	74.43	82.25
Gamma Ray, GR (gAPI)	40.69	22.41	10.4	19.12	37.24	51.83	106.13
Deep Resistivity, RESD (Ω·m)	13.15	21.04	0.47	3.64	7.04	14.27	134.15
Porosity, PHI (v/v)	15.38	5.53	1	13.5	16.92	19.5	24.98
Compressional wave, Vp(m/s)	4417.1	486.3	3706.86	4095.98	4289.4	4611.48	6008.21

3.2 Overview of models' selection and principles

3.2.1 Adaptive Boosting (AdaBoosting)

The Adaptive Boosting (AdaBoost) regressor is a meta-estimator that iteratively fits multiple weak learners to improve predictive accuracy. Initially, all training samples are assigned equal weights. After each iteration, mispredicted samples are given higher weights, forcing subsequent learners to focus on difficult cases. Each weak learner's contribution to the final ensemble is weighted based on its performance. The overall prediction is a weighted sum of the outputs from all weak regressors. Unlike Random Forest, which trains models independently, AdaBoost trains them sequentially, where each new model corrects the errors of the previous one.

3.2.2 Gradient boosting

Gradient Boosting builds an additive model in a stage-wise manner, allowing the optimization of arbitrary differentiable loss functions. In each stage, a weak learner (typically a shallow decision tree) is trained on the negative gradient of the loss function, effectively modeling the residuals of the previous ensemble. The final prediction is

obtained by summing the contributions of all weak learners. Unlike AdaBoost, which adjusts sample weights, Gradient Boosting employs gradient descent to minimize the chosen loss function (e.g., Mean Squared Error), leading to improved predictive performance over sequential iterations.

3.2.3 Random Forest

The Random Forest algorithm is an ensemble method that builds multiple independent decision trees on different subsets of the dataset and averages their predictions to reduce overfitting. Each tree is trained on a bootstrap sample (random sampling with replacement), while a random subset of features is considered at each split to decrease correlation between trees. Typically, about two-thirds of the data are used for training and the remaining one-third for out-of-bag (OOB) validation. The final output is the mean prediction from all trees, providing a balance between bias and variance and enhancing model robustness.

3.2.4 Categorical Boosting (CatBoost)

Categorical Boosting (CatBoost) is an open-source, high-performance gradient-boosting framework developed by Yandex. It is designed for classification, regression, and ranking tasks and is particularly effective in handling categorical features. CatBoost constructs decision trees sequentially, where each new tree is trained to correct the residual errors of the preceding ensemble. This iterative process continues until a predefined number of trees is reached. CatBoost's strength lies in its efficient encoding of categorical variables and its ability to reduce overfitting while maintaining high accuracy and computational efficiency.

3.2.5 Radial Basis Function Neural Network

The Radial Basis Function Neural Network (RBFNN) is a type of feed-forward neural network that employs Gaussian radial basis functions as activation functions. It is particularly effective for modeling nonlinear relationships. The network architecture consists of three layers: an input layer that receives the feature data, a hidden layer (radial basis layer) where each neuron computes a similarity measure between input data and a specific center using Gaussian functions, and an output layer that aggregates the activations to produce the final prediction. RBFNNs are well suited for capturing complex, localized patterns in multidimensional data.

3.3 Physics-Guided Machine Learning (PGML) Framework

Figure 3 illustrates the complete PGML framework adopted in this study, which differs fundamentally from conventional machine learning approaches. The framework comprises four key components:

3.3.1 Conceptual Foundation

The Physics-Guided Machine Learning (PGML) framework adopted in this study differs fundamentally from purely data-driven approaches. Rather than using machine learning algorithms to predict Unconfined Compressive Strength (UCS) directly from well-log data, we employ a discrepancy modeling strategy wherein ML algorithms learn the residual between physics-based predictions and laboratory measurements. This approach, illustrated in Figure 3, ensures that all final predictions remain anchored to established geomechanical principles while leveraging the flexibility of ML to capture site-specific complexities that simple empirical equations cannot represent.

3.3.2 Physics-Based Baseline: Rationale and Selection

For each core sample, a baseline UCS prediction (UCS_{pred}) is generated using facies-specific empirical equations (Table 3) that relate compressional wave velocity (Vp) to compressive strength. Vp is derived from the measured compressional slowness (ΔT) as $Vp = 1/\Delta T$ (with appropriate unit conversion). These equations represent the current state of physical understanding of the Vp-UCS relationship for different lithologies.

Table 3: Facies-specific empirical equations used to generate physics-based baseline UCS predictions (UCS_{pred}) from compressional wave velocity (Vp). These equations represent established physical relationships between elastic wave velocity and unconfined compressive strength for different lithologies. The appropriate equation is selected based on facies classification (XGBoost classifier, Section 3.1.2).

Parameter	Facies	Formula	References
UCS	Sandstone & Shale	$UCS = 0.0642 * V_p - 117.99$	Sharma and Singh (2008)

Siltstone	$UCS = 0.033 * V_p - 34.83$	Khandelwal (2013)
Limestone	$UCS = 0.0199 * V_p + 36.303$	Aladejare et al. (2024)

The relationship between compressional wave velocity (Vp) and unconfined compressive strength (UCS) is grounded in fundamental rock physics. The same microstructural features control both properties: grain contact stiffness, porosity, microcrack density, and mineral composition. Rocks with well-cemented grain contacts, low porosity, and few microcracks exhibit both higher elastic wave velocities (due to greater stiffness) and higher compressive strength (due to fewer failure initiation points). This shared microstructural control creates a physical basis for empirical Vp-UCS correlations—they are not arbitrary statistical artifacts but reflections of underlying physical reality.

The three equations in Table 3 were selected based on four criteria:

1. **Lithology-specificity**: Each equation is calibrated for a specific facies (sandstone/shale, siltstone, limestone) rather than providing a general relationship that might obscure lithological differences.
2. **Independence from study data**: All equations were developed using datasets completely independent from our Zubair Field samples, ensuring that the physics baseline is not "contaminated" and that the learned discrepancy represents genuine site-specific effects.
3. **Geological relevance**: The source studies used rock samples with characteristics similar to our formations—sedimentary sequences from comparable geological settings.
4. **Coverage of all lithologies**: The combination provides coverage for all facies present in our study area, including siltstone for which specific equations are rare.

These physics-informed equations serve as the anchor in our PGML framework. They provide a baseline prediction (UCS_{pred}) that is physically reasonable and ensures that all final predictions remain within physically plausible ranges. The machine learning component then learns only the discrepancy—the component of strength that these simple equations cannot capture due to site-specific factors such as clay content (reflected in GR), pore structure (porosity and resistivity), and compaction trends (depth). This combination of physics-based anchoring and data-driven refinement is the essence of the PGML approach.

The physics-based baseline equations in Table 3 were selected over alternatives such as Chang et al. (2006) and Miah et al. (2020) based on four criteria: lithology-specificity, independence from our dataset, geological relevance to the Middle East, and complete facies coverage. Chang et al. (2006), while seminal, provides generalized trends that combine multiple lithologies rather than the facies-specific relationships required for our framework. Miah et al. (2020), though excellent, developed site-specific correlations for the North Sea Volve Field using machine learning, using these as the physics baseline would compromise the independence required for genuine discrepancy learning in our PGML framework. The selected equations satisfy all criteria and ensure that the physics baseline is both lithology-appropriate and independent of our ML component.

3.3.3 Discrepancy Calculation

The discrepancy (δUCS) between the laboratory-measured UCS (UCS_o) and the physics-based prediction (UCS_m) is calculated as:

$$\delta UCS = UCS_o - UCS_m \quad (1)$$

This δUCS represents the component of rock strength that the empirical equations fail to capture. Positive δUCS indicates that the empirical equation underestimates the true strength (the rock is stronger than predicted by velocity alone), while negative δUCS indicates overestimation (the rock is weaker than predicted). These discrepancies arise from geological factors not represented in simple velocity-based equations, including:

- Clay content and mineralogy (captured by gamma ray, GR)
- Pore structure and fabric (captured by porosity and resistivity)
- Compaction and diagenetic trends (captured by depth)
- Textural variations (captured by slowness, ΔT)

3.3.4 Machine Learning for Discrepancy Prediction

Five standard machine learning algorithms, Adaptive Boosting (AdaBoost), Gradient Boosting Tree (GBT), Random Forest (RF), Categorical Boosting (CatBoost), and Radial Basis Function Neural Network (RBFNN), are trained to predict δUCS using the full suite of well-log features (depth, bulk density, GR, deep resistivity, porosity, ΔT, and Vp). Importantly, Vp is included despite being used in the empirical equations, allowing the ML models to learn nonlinear corrections to the Vp-UCS relationship that the linear empirical equations cannot represent.

Importantly, these algorithms are not themselves "physics-guided" in their internal operations**.** They remain standard implementations of boosting, bagging, and neural network architectures. The "physics-guided" designation applies to the framework within which they operate, specifically, their role in predicting the physics-model discrepancy rather than the target variable directly. This distinction is critical: the ML algorithms serve as flexible function approximates that learn to correct the limitations of simplified physical models, combining the interpretability of physics with the adaptability of data-driven methods.

3.3.5 Final Prediction and Model Integration

The final UCS prediction is obtained by summing the physics-based baseline and the ML-predicted discrepancy:

$$UCS_{final} = UCS_m - \delta UCS_{pred} \qquad (2)$$

where δUCS_{pred} is the ML model's prediction of the discrepancy. This additive combination ensures that:

1. All predictions are anchored to the physical baseline (UCS)
2. ML provides corrections only where data supports deviation from physics
3. The framework remains interpretable, large discrepancies highlight where physical understanding is incomplete

3.3.6 Comparison with Conventional ML

To evaluate the value of physics guidance, we compare the PGML framework against conventional ML approaches where the same five algorithms are trained to predict UCS directly from well-log features, without the physics-based baseline. This comparison (Section 4.1) isolates the contribution of physics integration from the inherent capabilities of the ML algorithms themselves.

3.4 Validation Metrics

The coefficient of determination (R^2) measures the proportion of variance in the target variable that can be explained by the model's predictors. It ranges from 0 to 1, with values closer to 1 indicating better fit. The metric is expressed as:

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \qquad (3)$$

where y is the observed value, $\hat{y}$ is the predicted value, and $\bar{y}$ is the mean of the observed values.

While R^2 is a useful indicator of overall fit, it does not guarantee causal accuracy and can be misleading when comparing models trained on different datasets or when regression assumptions are violated (Rose & McGuire, 2019). Therefore, it is best interpreted alongside complementary metrics such as RMSE and feature-importance analyses

(e.g., SHAP), which together provide a more comprehensive assessment of model reliability (Huang et al., 2020; Yang et al., 2024).

The Root Mean Square Error (RMSE) quantifies the average magnitude of prediction errors in the same units as the target variable:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2} \qquad (4)$$

RMSE provides an absolute measure of predictive accuracy, indicating how closely the predicted values match observed data (Hsieh, 2023). While R^2 evaluates the proportion of explained variance, RMSE reveals the practical magnitude of residual errors. Reporting both metrics ensures a balanced evaluation of model performance, as a high R^2 may still coincide with a large RMSE if systematic biases exist (Aptula et al., 2005; Hodson, 2022).

3.5 Hyperparameters optimization using Random Grid Search

Hyperparameter tuning is the process of identifying and selecting the optimal hyperparameters for training a machine learning model. When performed effectively, it minimizes the model's loss function, thereby enhancing predictive accuracy. Traditionally, determining the ideal hyperparameter values has been a manual and experience-driven task. Researchers often rely on their prior knowledge and familiarity with machine learning algorithms to guide the selection process. However, even with expertise, the optimal hyperparameter settings can vary significantly across different datasets, making it difficult to generalize values based solely on previous experience.

To address this challenge, automated and systematic methods have been developed to explore alternative model configurations. One common approach is grid search, where each hyperparameter is discretized into a range of possible values, and models are trained and evaluated across all possible combinations. Although straightforward to implement, grid search can be computationally expensive, as its cost increases exponentially with the number of hyperparameters and their possible values.

A simpler yet often more efficient alternative is random search, which samples random combinations of hyperparameter values for model evaluation. As demonstrated by Bergstra and Bengio (2012), model performance is often more sensitive to certain hyperparameters than others; thus, randomly exploring a wider range of values can increase the likelihood of finding effective configurations. Moreover, since some hyperparameters may have minimal impact on model performance for specific datasets, random search avoids unnecessary computations associated with evaluating every possible combination, particularly in high-dimensional hyperparameter spaces. Nonetheless, the success of random search still depends on the number and diversity of sampled configurations.

3.6 Explainable Artificial Intelligence (XAI)

In recent years, machine learning (ML) has become a key enabling technology across scientific and industrial domains. Advances in algorithms, the availability of large datasets, and increased computational power have enabled modern ML models to achieve remarkable performance on a growing number of complex tasks. However, due to their highly nonlinear and nested architectures, most ML models are often regarded as "black boxes," offering limited insight into how predictions are generated.

In scientific research, understanding why a model makes a particular prediction is often as important as the prediction itself. Interpreting and explaining model behavior helps uncover the underlying physical, chemical, or biological mechanisms that drive the observed outcomes, potentially leading to new scientific discoveries (Samek & Müller, 2019). To achieve interpretability, several techniques have been developed, including SHapley Additive exPlanations (SHAP), LIME (Local Interpretable Model-agnostic Explanations), occlusion sensitivity, and Layer-wise Relevance Propagation (LRP), among others.

In this research, we employ SHapley Additive exPlanations (SHAP) to interpret the predictions of the Physics-Guided Machine Learning (PGML) models. The method is grounded in the concept of Shapley values from cooperative game theory, originally introduced by Lloyd Shapley (1953) as cited in Lundberg and Lee (2017). The Shapley value framework is based on four fairness properties, efficiency, symmetry, additively, and null player, and provides the only set of values that satisfy all these conditions. In this context, each player (feature) is assigned a Shapley value (ϕ_j), representing its fair contribution to the total prediction output (V).

$$\phi_j = \sum_{S \subseteq N \setminus (j)} \frac{|S|!\,(|N| - |S| - 1)!}{|N|!} \big(V(\,\mathrm{SU}\{j\}) - V(S)\big) \qquad (5)$$

where S is a subset of all features N, and $V(S)$ represents the model output for subset S.

Lundberg, S.M., and Lee, S.I. (2017) unified several existing interpretability methods by introducing SHAP (SHapley Additive exPlanations), which computes Shapley values efficiently for complex models. SHAP provides both local (instance-level) and global (model-level) interpretability, enabling transparent identification of the most influential input parameters affecting UCS predictions. The method has been widely adopted in geoscientific applications due to its consistency with game-theoretic principles and its model-agnostic nature (Molnar, 2022).

4. Results and Discussion

4.1 PGML Modeling

The results indicate a substantial improvement in predictive performance when using the Physics-Guided Machine Learning (PGML) approach compared to conventional machine learning (ML). Across all five tested algorithms, PGML models consistently produced higher R^2 values and lower RMSE values, confirming superior predictive accuracy and robustness.

Adaptive Boosting (AdaBoost): The PGML-AdaBoost model achieved an R^2 of 0.975 and an RMSE of 5.9 MPa, demonstrating strong predictive capability with minimal error. In contrast, the conventional ML model attained a much lower R^2 of 0.68 and a higher RMSE of 21 MPa, indicating reduced reliability (Figure X).

Gradient Boosting (GBT): The PGML-GBT model outperformed the conventional ML approach, achieving an R^2 of 0.8321 versus 0.53, and a significantly lower RMSE (15.34 MPa vs. 25.4 MPa), reflecting greater model accuracy and stability.

Random Forest (RF): The PGML-RF model achieved an R^2 of 0.9333 and an RMSE of 9.67 MPa, whereas the standard ML version recorded $R^2 = 0.75$ and RMSE = 18.72 MPa, confirming PGML's improved generalization.

Categorical Boosting (CatBoost): The PGML-CatBoost model achieved an R^2 of 0.9493 and an RMSE of 8.42 MPa, outperforming the conventional ML model ($R^2 = 0.8433$, RMSE = 14.82 MPa).

Radial Basis Function Neural Network (RBFNN): The PGML-RBFNN model achieved an R^2 of 0.824 and an RMSE of 15.70 MPa, significantly outperforming the conventional model ($R^2 = 0.101$, RMSE = 35.50 MPa) as summarized in Table 4.

Table 4: Comparison between the test results of PGML approach and Conventional ML approach for five different Algorithms.

Algorithm	ML		PGML		Differences in RMSE
	R^2	RMSE	R^2	RMSE	
AdaBoost	0.68	21	**0.975**	**5.9**	**-72%**
Gradient Boosting	0.53	25.4	0.8321	15.34	-39.6%

Random Forest	0.75	18.72	0.9333	9.668	-48.35%
CatBoost	0.8433	14.82	0.9493	8.42	-43.2%
RBFNN	0.101	35.498	0.824	15.696	-55.8%

Figures 5 to 9 illustrate the comparison between measured and predicted UCS values for the five models, Adaptive Boosting (AdaBoost), Gradient Boosting (GBT), Random Forest (RF), Categorical Boosting (CatBoost), and Radial Basis Function Neural Network (RBFNN), with the 1:1 reference line indicating perfect prediction. The clustering of data points around this line clearly demonstrates the enhanced accuracy of the PGML approach relative to conventional ML models.

It should be noted that all hyperparameters not explicitly listed were retained at their default values, as they did not have a significant impact on improving model performance. Table 5 presents the optimized hyperparameters for all PGML models obtained through the random grid search procedure, along with the search ranges considered for each parameter.

Table 5: Optimized hyperparameters for the five PGML algorithms (AdaBoost, Gradient Boosting, Random Forest, CatBoost, and RBFNN) obtained through random grid search with 5-fold cross-validation. The search range column indicates the values explored during optimization. Hyperparameters not listed were retained at their default values, as initial screening showed minimal impact on model performance for this dataset.

Algorithm	Hyperparameter	Optimized Value	Search Range Considered
AdaBoost	Base Estimator	Decision Tree	—
	Max Depth	3	[1, 3, 5]
	Number of Estimators	50	[10, 50, 100, 200]
	Learning Rate	0.1	[0.01, 0.05, 0.1, 0.5]
Gradient Boosting	Loss Function	Squared Error	—
	Max Depth	3	[2, 3, 5, 7]
	Number of Estimators	100	[50, 100, 200]
	Learning Rate	0.01	[0.001, 0.01, 0.1]
Random Forest	Criterion	Squared Error	—
	Max Depth	10	[5, 10, 20, None]
	Number of Estimators	50	[10, 50, 100, 200]
	Min Samples Split	2	[2, 5, 10]
CatBoost	Depth	8	[4, 6, 8, 10]
	Iterations	100	[50, 100, 200]
	Learning Rate	0.7	[0.01, 0.05, 0.1, 0.7]
	L2 Regularization	3	[1, 3, 5]
RBFNN	Number of Centers	75	[25, 50, 75, 100]
	Spread Parameter	0.6	[0.1, 0.3, 0.6, 1.0]
	Regularization Parameter	0.3	[0.01, 0.1, 0.3, 0.5]

Figure 5 illustrates the comparative performance of the AdaBoost-based models and the empirical UCS equation. The PGML-enhanced AdaBoost model (Figure 5a) demonstrates a strong alignment between the predicted and measured UCS values; with data points closely distributed along the 1:1 reference line, indicating excellent predictive accuracy. In contrast, the conventional AdaBoost ML model (Figure 5b) shows greater scatter and deviation from the reference line, reflecting reduced reliability. The empirical model (Figure 5c) exhibits the weakest correlation, underestimating UCS values for higher strength ranges. The improvement observed in the PGML model is attributed to the incorporation of physical constraints that guide the learning process, minimizing overfitting and enhancing generalization across lithological variations.

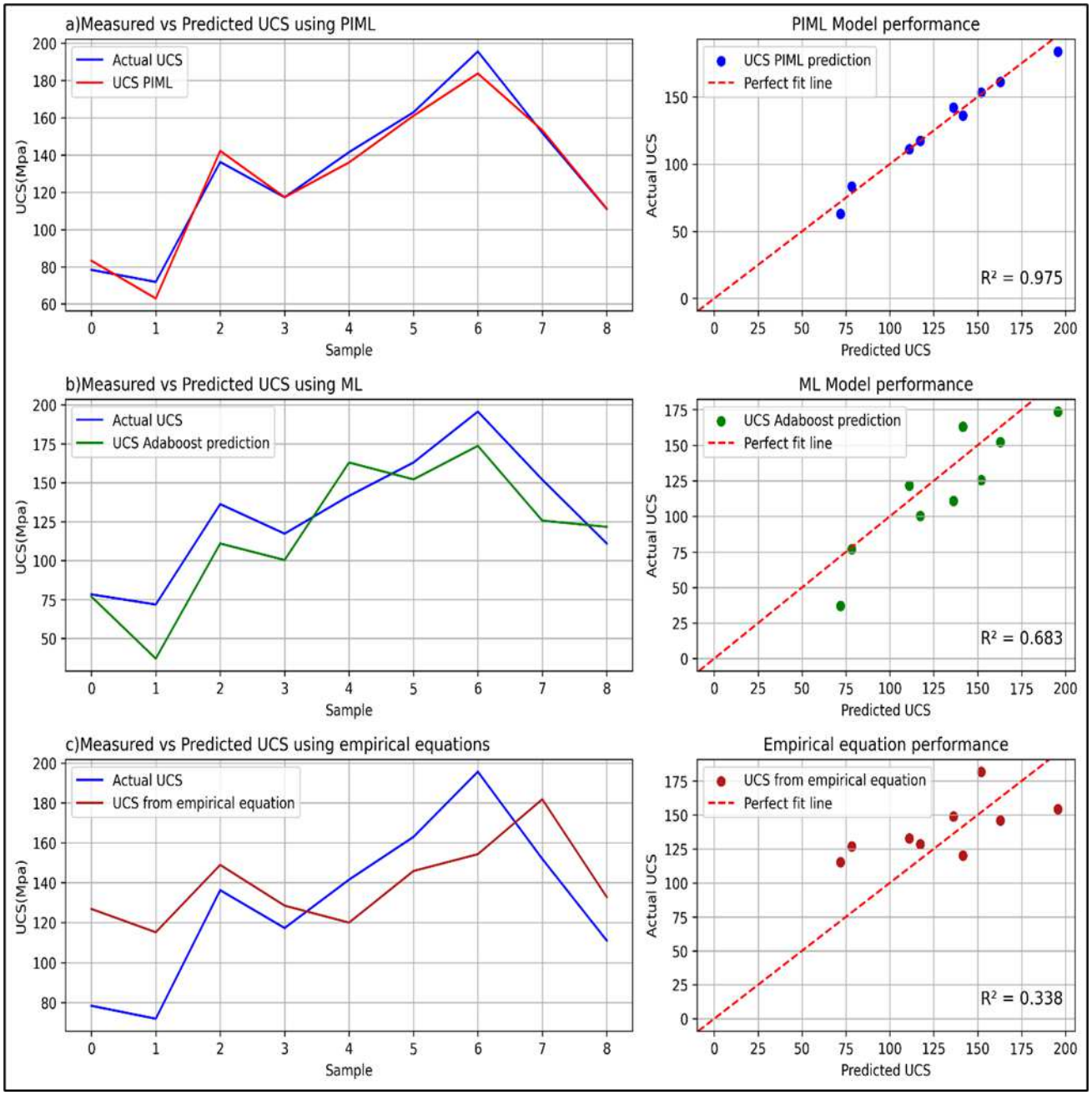


Figure 5: Comparison for the prediction accuracy of confined compressive strength (UCS) among physics guided machine learning, conventional machine learning, and empirical model. a-AdaBoost PGML model for UCS prediction, b- AdaBoost ML model for UCS prediction, c-Empirical equation for UCS estimation.

Figure 6 compares the results of the Gradient Boosting models and the empirical UCS estimation. The PGML-enhanced Gradient Boosting model (Figure 6a) achieves a high degree of correspondence between predicted and measured UCS values, with data points tightly clustered near the reference line. This demonstrates the model's ability to capture nonlinear relationships between input features and UCS, even in heterogeneous formations. The conventional ML model (Figure 6b), however, reveals wider dispersion and a noticeable tendency to underestimate UCS in the upper range. The empirical model (Figure 6c) again shows poor agreement, reinforcing the limitations of static regression-based approaches for complex geomechanical behaviors. The improved accuracy of the PGML model underscores the value of embedding physical principles within the learning process to constrain and correct purely data-driven predictions.

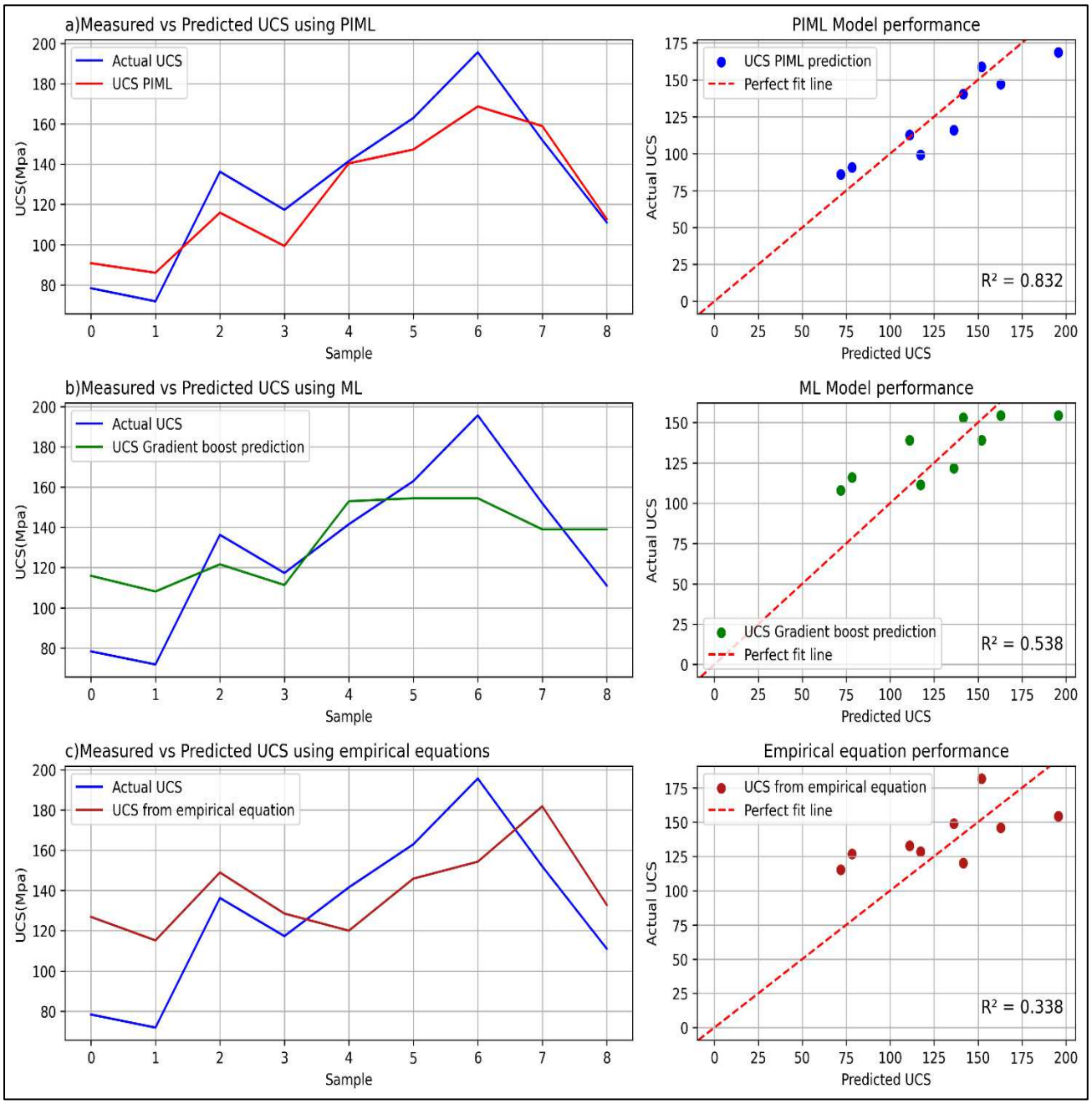


Figure 6: Comparison for the prediction accuracy of confined compressive strength (UCS) among physics-guided machine learning, conventional machine learning, and empirical model. a- Gradient boost PGML model for UCS prediction ,b- Gradient boost ML model for UCS prediction, c- Empirical equation for UCS estimation.

Figure 7 presents the performance comparison for the Random Forest-based models and the empirical UCS equation. The PGML-integrated Random Forest model (Figure 7a) yields highly accurate predictions, with most points lying along the 1:1 correlation line, indicating an R^2 value exceeding 0.93. The incorporation of physics-guided corrections effectively mitigates biases common in purely data-driven Random Forest models. Conversely, the standard ML model (Figure 7b) shows noticeable prediction variability, particularly at intermediate UCS values. The empirical model (Figure 7c) produces the largest deviations, particularly underestimating UCS in high-strength formations. The enhanced performance of the PGML model confirms the robustness of combining ensemble learning with physical constraints to improve generalization and reduce error propagation.

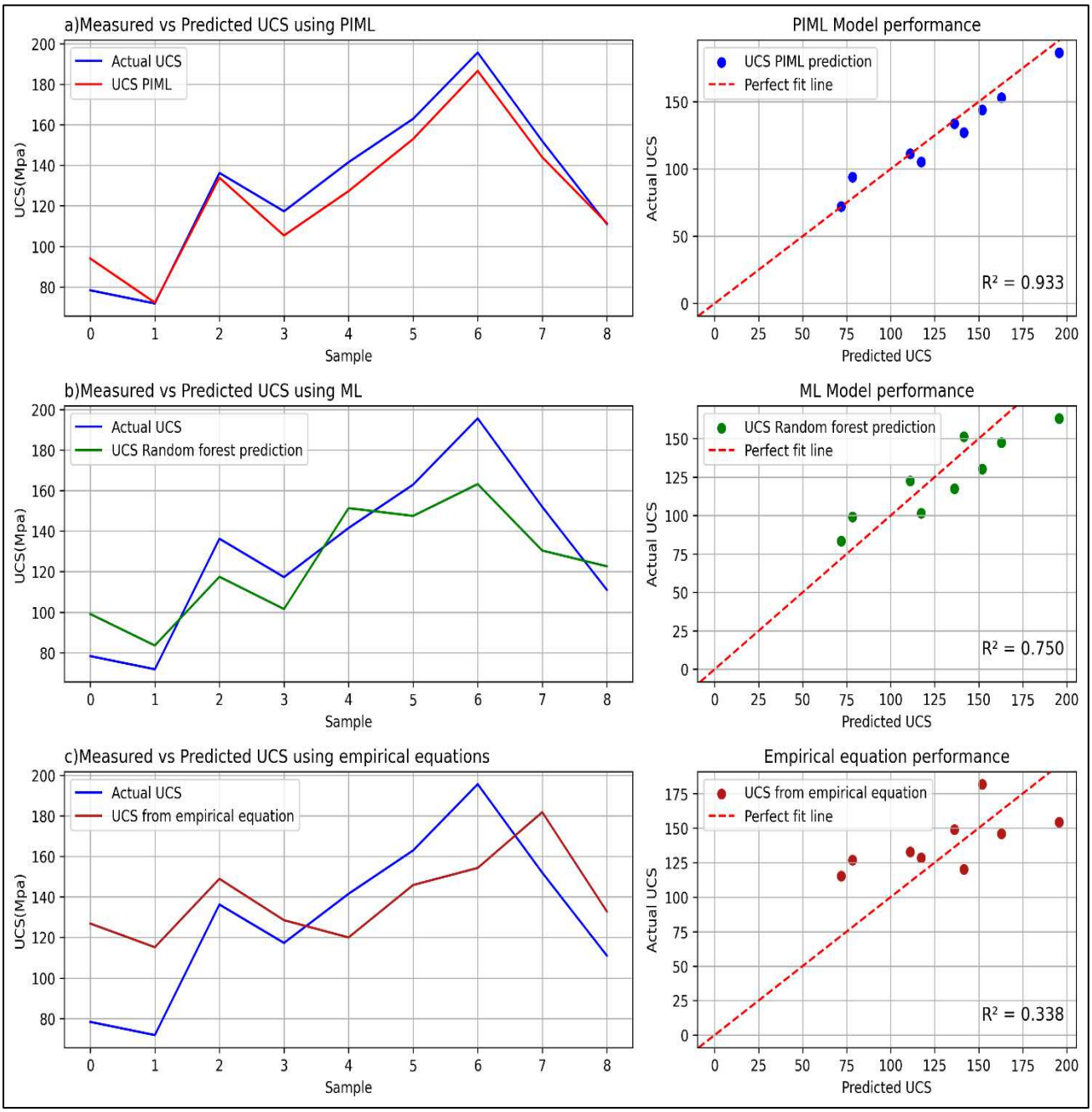


Figure 7: Comparison for the prediction accuracy of confined compressive strength (UCS) among physics guided machine learning, conventional machine learning, and empirical model. a-Random forest PGML model for UCS prediction, b-Random forest ML model for UCS prediction, c-Empirical equation for UCS estimation.

Figure 8 illustrates the comparison between CatBoost-based models and the empirical equation for UCS prediction. The CatBoost-PGML model (Figure 8a) displays excellent predictive capability, with data points closely aligned along the reference line and minimal residual spread, corresponding to an R^2 value of approximately 0.95. The conventional CatBoost ML model (Figure 8b), while still performing better than the empirical model, shows broader scatter and occasional under prediction. The empirical model (Figure 8c) exhibits a clear nonlinear bias, unable to account for the complex relationships among the geophysical and geomechanical parameters. The results highlight that CatBoost, when enhanced by physics-based guidance, can efficiently captures both categorical and continuous feature interactions, producing more physically consistent predictions.

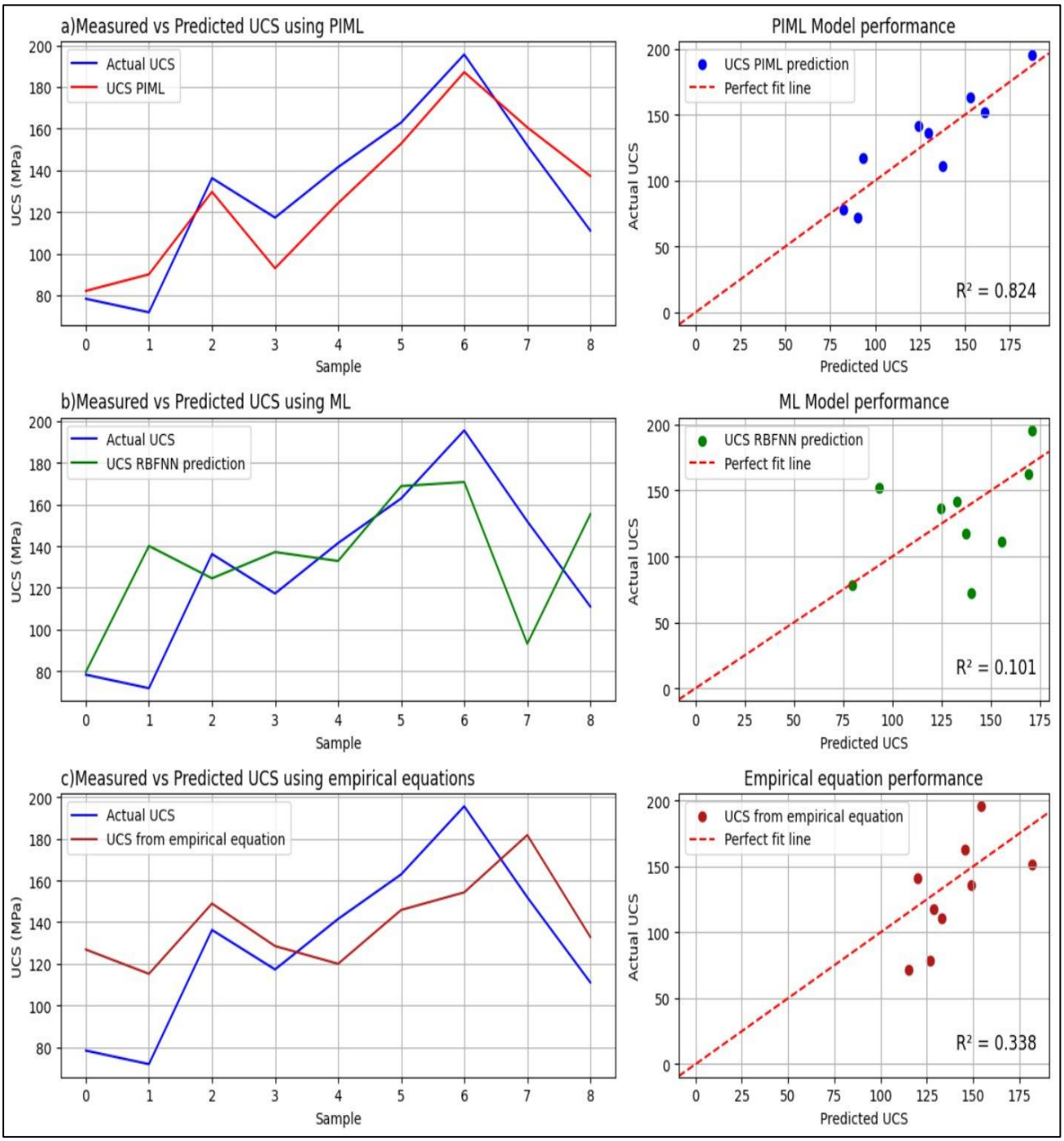


Figure 8: Comparison for the prediction accuracy of confined compressive strength (UCS) among physics guided machine learning, conventional machine learning, and empirical model. a-CatBoost PGML model for UCS prediction, b- CatBoost ML model for UCS prediction, c-Empirical equation for UCS estimation.

Figure 9 compares the performance of the RBFNN-based models and the empirical UCS equation. The PGML-augmented RBFNN model (Figure 9a) significantly improves prediction accuracy, with data points densely concentrated along the 1:1 line, reflecting an R^2 of 0.824 and a notably reduced RMSE. The conventional RBFNN model (Figure 9b) performs considerably worse, with wide dispersion and poor correlation ($R^2 = 0.10$), demonstrating the instability of purely data-driven neural networks when applied to complex, nonlinear geological systems. The empirical model (Figure 9c) again shows limited reliability and systematic bias. These results demonstrate that integrating physical knowledge into neural network architectures can substantially enhance model interpretability, stability, and extrapolation capability in predicting UCS across heterogeneous lithologies.

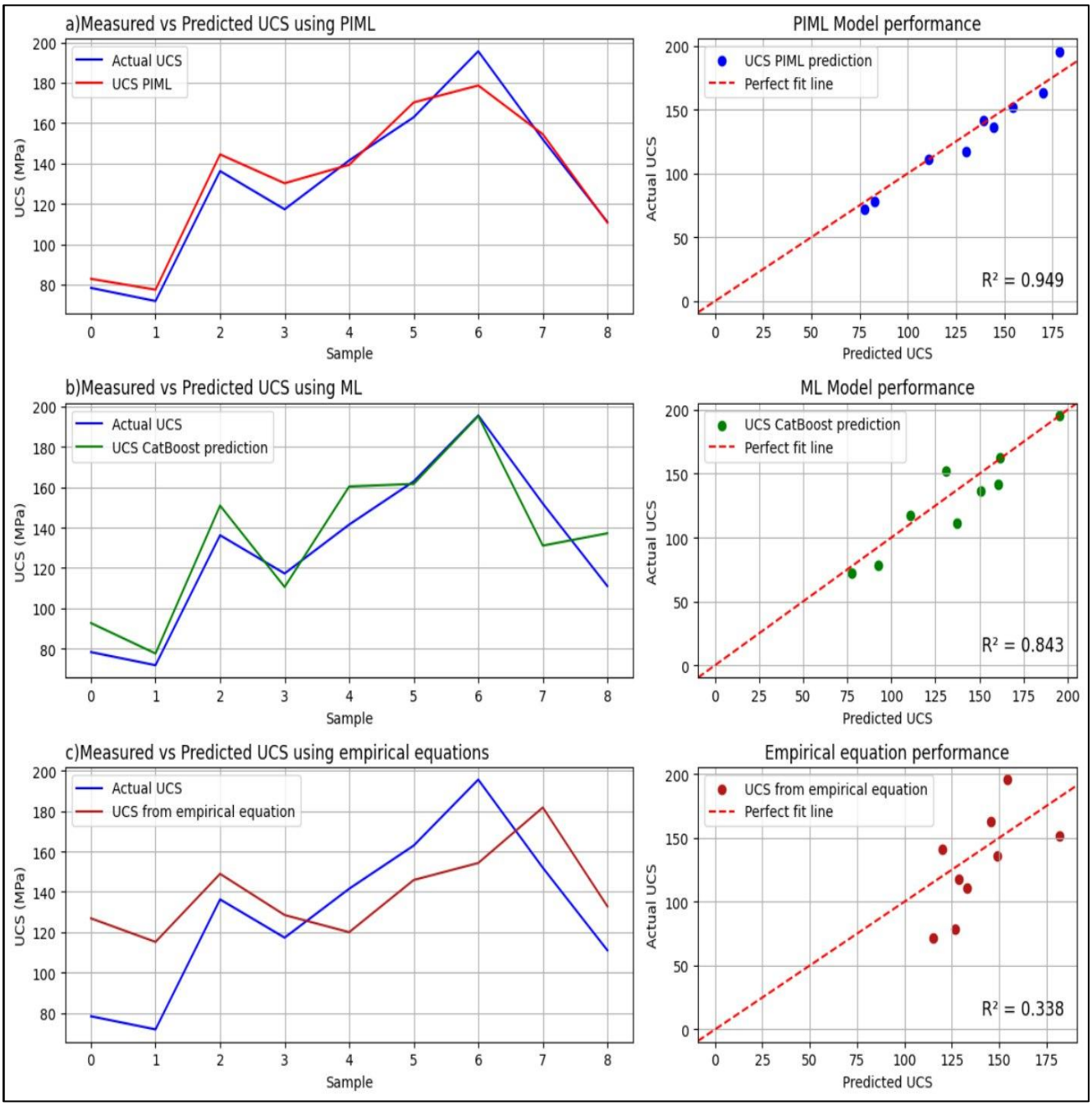


Figure 9: Comparison for the prediction accuracy of confined compressive strength (UCS) among physics guided machine learning, conventional machine learning, and empirical model. a-CatBoost PGML model for UCS prediction, b- CatBoost ML model for UCS prediction, c-Empirical equation for UCS estimation.

4.2 Shapley Additive exPlanations (SHAP)

4.2.1 SHAP Analysis for AdaBoost-PGML

Figure 10 presents the SHAP beeswarm plot for the AdaBoost-PGML model, illustrating how each of the seven input features contributes to the predicted discrepancy (δUCS). Each point represents a SHAP value for a specific feature and prediction, with color indicating feature value (red = high, blue = low). Features are ordered by their mean absolute SHAP value (global importance).

The plot reveals that depth, gamma ray, and bulk density are the most influential features controlling the discrepancy. Higher depth values (red points) show predominantly positive SHAP values, indicating that the empirical equations underestimate UCS for deeper samples—likely due to compaction and diagenetic strengthening not captured by Vp-based equations. Higher gamma ray values (clay-rich samples) also show positive contributions, suggesting clay-bearing lithologies are stronger than predicted, possibly due to clay-bound water effects or the mechanical

properties of clay minerals. Higher bulk density shows negative SHAP contributions, indicating overestimation of UCS for dense samples, potentially because high density can reflect heavy minerals or cementation that affect velocity differently than strength.

Figure 11 quantifies this through mean absolute SHAP values, confirming depth (0.82), gamma ray (0.58), and bulk density (0.45) as the top three features, with porosity (0.32), Vp (0.21), ΔT (0.12), and resistivity (0.08) showing progressively lower importance.

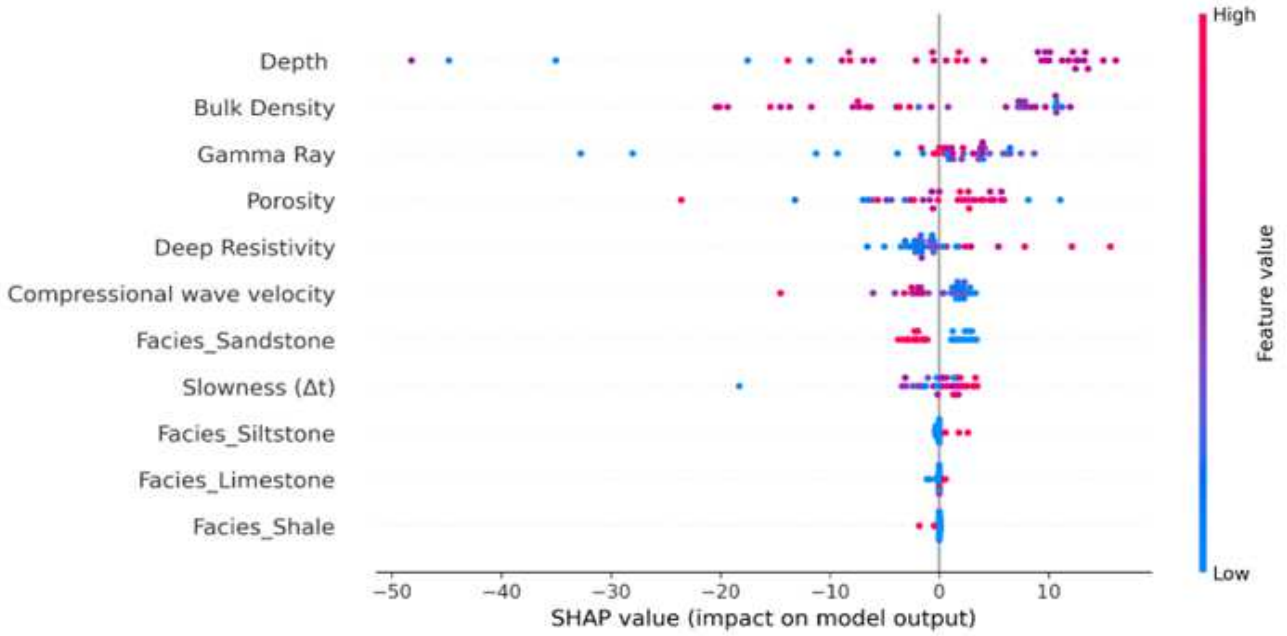


Figure 10: Beeswarm plot for UCS PGML model using AdaBoost algorithm.

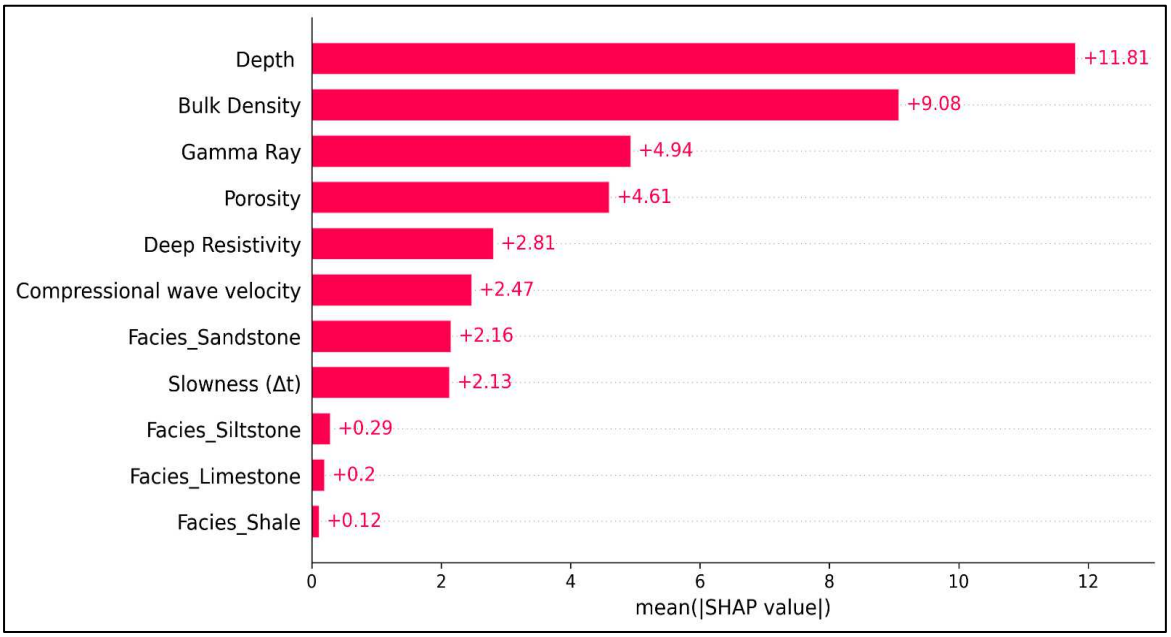


Figure 11: Bar chart for UCS PGML model using AdaBoost algorithm.

4.2.2 Cross-Algorithm SHAP Comparison

To assess the robustness of these findings, we extended SHAP analysis to all five PGML algorithms (AdaBoost, GBT, RF, CatBoost, and RBFNN). The key findings from cross-algorithm comparison are:

1. Strong consensus on top features: All five algorithms rank depth, bulk density, and gamma ray as the top three features controlling the physics-model discrepancy (Table 6). Depth is ranked #1 by four algorithms and #3 by RBFNN; bulk density and gamma ray consistently occupy ranks 2-3. This consensus across fundamentally different algorithmic families (boosting, bagging, neural networks) provides robust evidence that these features capture genuine physical processes rather than algorithmic artifacts.
2. Consistent ranking of moderate and low importance features: Porosity and Vp consistently rank 4-5 across all algorithms, while ΔT and resistivity consistently rank lowest (6-7). This stability further supports the interpretation that these features contribute minimally to explaining why simple Vp-based equations fail.
3. Algorithm-specific variations: RBFNN shows slightly different absolute rankings (depth #3 instead of #1) but still places depth, bulk density, and GR in the top three. This minor variation likely reflects the different mathematical structure of neural networks compared to tree-based methods and does not undermine the overall consensus.

4.2.3 Physical Interpretation of SHAP Consensus

The strong cross-algorithm agreement that depth, bulk density, and gamma ray are the primary controls on the physics-model discrepancy has clear physical meaning:

- Depth captures compaction and diagenetic trends—processes that strengthen rocks through mechanical and chemical changes during burial. Simple Vp-based equations, typically developed from surface or near-surface samples, cannot account for these depth-dependent strengthening effects. The positive SHAP contribution of depth indicates that deeper samples are stronger than Vp alone predicts.
- Bulk density reflects the combined effects of mineral composition and porosity. Its importance suggests that the Vp-UCS relationship is modulated by density in ways the empirical equations don't capture—dense rocks may have different failure mechanisms or stress distributions than less dense rocks with similar velocities.
- Gamma ray (clay content) affects both elastic wave propagation and strength through complex, nonlinear relationships. Clays can either strengthen (through binding and cohesion) or weaken (through plasticity and swelling) depending on their type, distribution, and diagenetic state. The consistent importance of GR indicates that the empirical equations miss these lithology-specific effects.

The features with low importance (ΔT and resistivity) are either redundant with Vp (ΔT is inversely related) or reflect fluid effects (resistivity) that may be secondary to mechanical behavior in these formations.

This cross-algorithm SHAP analysis significantly strengthens our confidence in the physical interpretation of the PGML framework and demonstrates that the identified feature importance is robust across different machine learning approaches.

Our SHAP analysis identified depth, porosity, bulk density, and gamma ray as the most influential features controlling the empirical-model discrepancy. This aligns partially with Miah et al. (2025), who found GR and Vp most important for sandstone UCS using correlation-based ranking, and Rahaman & Miah (2025), who identified Schmidt hammer rebound as dominant using SHAP. The divergence reflects our discrepancy modeling approach: we identify features that explain why empirical equations fail, whereas previous studies identified features that best predict UCS directly. The prominence of depth in our analysis (absent from previous studies) highlights the importance of compaction and diagenesis, processes not captured by Vp-based empirical equations, demonstrating the unique insights enabled by physics-guided discrepancy learning.

5. Carbon Storage Criteria and implications

5.1 Carbon Storage Criteria

The geomechanical stability of a CO_2 storage site is essential to ensuring the long-term containment of the injected gas. The Unconfined Compressive Strength (UCS) serves as a key parameter for assessing the mechanical integrity of

both the reservoir and the overlying caprock. UCS directly measures a rock's strength and its resistance to failure under axial stress. Higher UCS values indicate stronger rock formations, thereby reducing the likelihood of caprock fracturing or shear failure that may result from pressure buildup during CO_2 injection (Haug et al., 2007). According to Mohanty (2013), the classification of intact and jointed rocks based on UCS provides a useful reference for evaluating rock strength, as summarized in Table 6**.**

Table 6: Strength classification of intact and jointed rocks (Mohanty 2013).

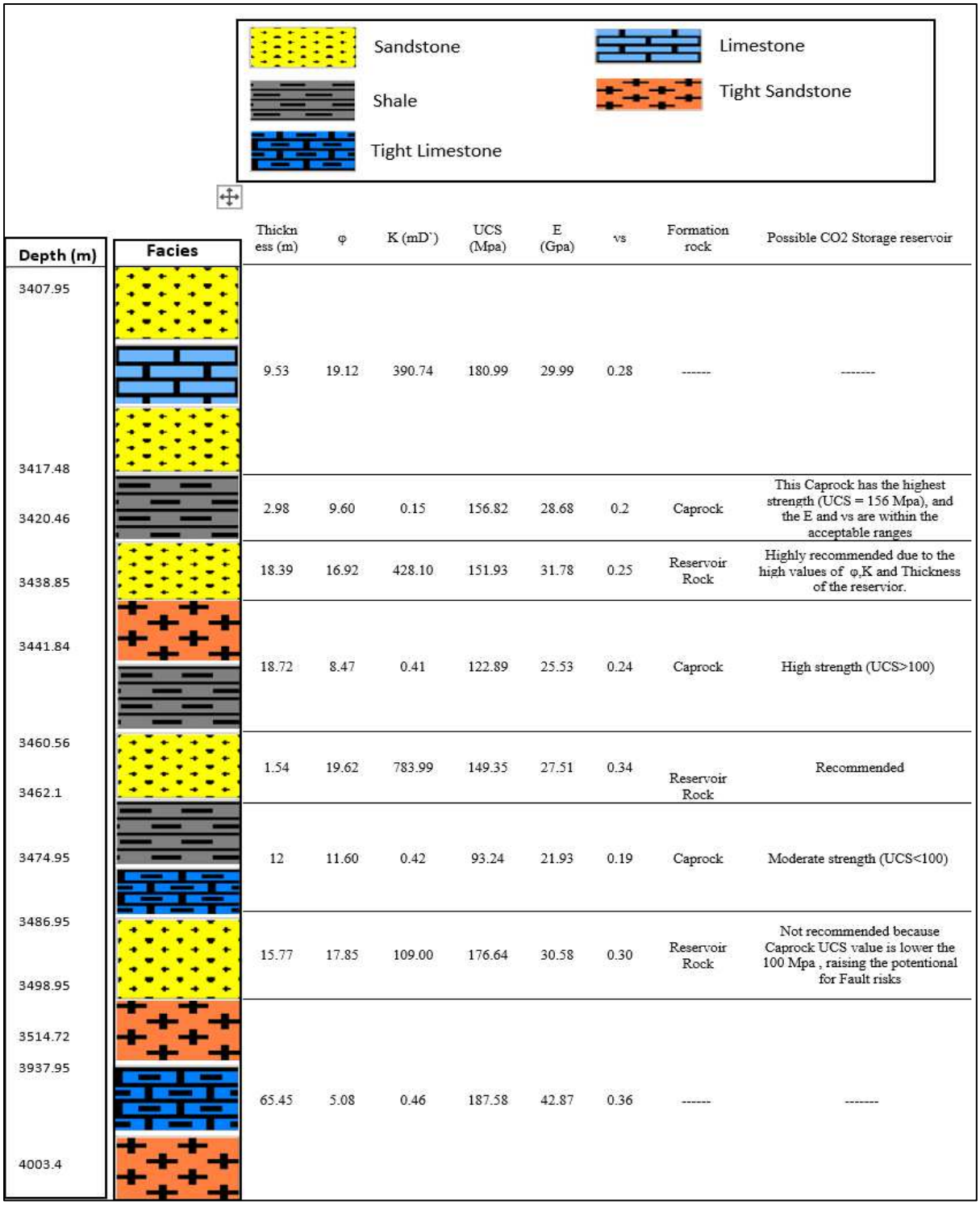
Class	Description	UCS (Mpa)
A	Very high strength	>250
B	High strength	100-250
C	Moderate strength	50-100
D	Medium strength	25-50
E	Low strength	5-25
F	Very low strength	<5

5.2 Carbon Storage implications

Figure 12 provides a comprehensive overview of the three reservoir formations and their associated geomechanical and petrophysical characteristics. The first reservoir formation represents the most promising interval for CO_2 storage. This sandstone unit exhibits high porosity (16.92%) and permeability (428.10 mD), indicating excellent injectivity and ample pore space for CO_2 containment. With a thickness exceeding 18 meters, it offers significant storage potential. Its overlying caprock also demonstrates strong geomechanical integrity, with a UCS of approximately 156 MPa, the highest among the three caprocks, indicating robust sealing capacity and low risk of fracturing under injection pressures. Accordingly, this reservoir is considered highly suitable for secure and long-term CO_2 storage.

The second reservoir formation, located at depths between 3,460.56 and 3,462.10 meters, is relatively thin (1.54 meters) but features exceptionally high permeability (783.99 mD) and favorable porosity (19.62%). These properties make it an attractive target from an injectivity standpoint. The caprock above this interval exhibits a UCS of 122.89 MPa, which falls within the acceptable range for safe CO_2 containment. However, due to its limited thickness, this formation may offer lower overall storage capacity and is thus more suitable for smaller-scale storage operations or as part of a stacked storage system.

The third and deepest sandstone reservoir also shows encouraging petrophysical characteristics, with porosity of 17.85%, permeability of 109 mD, and a thickness exceeding 15 meters. It possesses strong intrinsic geomechanical strength, reflected by a UCS of 176.64 MPa. However, concerns arise regarding the integrity of its overlying caprock. As shown in Figure 12, this caprock has a UCS value below 100 MPa, suggesting moderate strength and potential susceptibility to fracturing under elevated injection pressures. Such conditions could compromise its sealing capacity and increase the risk of CO_2 leakage. Therefore, despite the reservoir's favorable properties, it is not recommended for CO_2 storage without further caprock integrity evaluation and the implementation of risk mitigation measures.



Depth (m)	Facies	Thickness (m)	φ	K (mD`)	UCS (Mpa)	E (Gpa)	vs	Formation rock	Possible CO2 Storage reservoir
3407.95 – 3417.48		9.53	19.12	390.74	180.99	29.99	0.28	------	-------
3417.48 – 3420.46		2.98	9.60	0.15	156.82	28.68	0.2	Caprock	This Caprock has the highest strength (UCS = 156 Mpa), and the E and vs are within the acceptable ranges
3420.46 – 3438.85		18.39	16.92	428.10	151.93	31.78	0.25	Reservoir Rock	Highly recommended due to the high values of φ,K and Thickness of the reservior.
3438.85 – 3441.84		18.72	8.47	0.41	122.89	25.53	0.24	Caprock	High strength (UCS>100)
3441.84 – 3460.56		1.54	19.62	783.99	149.35	27.51	0.34	Reservoir Rock	Recommended
3460.56 – 3462.1		12	11.60	0.42	93.24	21.93	0.19	Caprock	Moderate strength (UCS<100)
3462.1 – 3474.95		15.77	17.85	109.00	176.64	30.58	0.30	Reservoir Rock	Not recommended because Caprock UCS value is lower the 100 Mpa , raising the potentional for Fault risks
3474.95 – 3486.95		65.45	5.08	0.46	187.58	42.87	0.36	------	-------
3486.95 – 3498.95									
3498.95 – 3514.72									
3514.72 – 3937.95									
3937.95 – 4003.4									

Figure 12: Stratigraphic column of the studied field in southern Iraq, highlighting the main reservoir and caprock formations assessed for geological CO_2 sequestration. The figure summarizes the lithological characteristics, petrophysical properties, and geomechanical strengths that determine each interval's suitability for safe and long-term CO_2 storage.

6. Conclusion

This study introduced a Physics-Guided Machine Learning (PGML) framework for predicting the Unconfined Compressive Strength (UCS), a key geomechanical property used to evaluate the bearing capacity and stability of subsurface reservoirs for CO_2 storage applications. The proposed approach integrates five advanced machine learning algorithms, Adaptive Boosting (AdaBoost), Gradient Boosting (GBT), Random Forest (RF), Categorical Boosting (CatBoost), and Radial Basis Function Neural Network (RBFNN), with physics-based empirical correlations to enhance predictive reliability and physical consistency. Conventional well log data, including porosity, bulk density,

gamma ray (GR), deep resistivity (RESD), slowness (ΔT), ultrasonic P-wave velocity (Vp), and lithofacies classification, were utilized as input features to train and validate the models.

A comprehensive comparison between the different ML and PGML models revealed the superior performance of the physics-guided approach in reproducing realistic geomechanical behavior. The integration of domain-specific physics not only improved accuracy but also enhanced model generalization across heterogeneous lithologies.

6.1 Key Findings

1. The proposed PGML framework consistently outperformed conventional machine learning models in predicting UCS, achieving substantial improvements in both R^2 and RMSE values across all algorithms. The AdaBoost-PGML model delivered the best performance, with an R^2 of 0.975, indicating excellent agreement with measured UCS values.
2. The inclusion of physics-based correlations significantly improved model reliability, with the average predictive performance across all models increasing from 59.5% to 92%, demonstrating the added value of integrating physical constraints into data-driven workflows.
3. Model interpretability was achieved using Shapley Additive Explanations (SHAP), which identified depth, porosity (Φ), bulk density, and compressional wave velocity (Vp) as the most influential features affecting UCS predictions. This insight provides a physically meaningful understanding of how these variables control geomechanical strength.
4. The developed PGML-UCS framework demonstrates that coupling physical knowledge with machine learning techniques can mitigate the limitations of purely data-driven models, offering a more robust, interpretable, and generalizable solution for subsurface geomechanical characterization.

6.2 Recommendations for Future Work

1. To further improve predictive accuracy and reduce uncertainty, it is recommended to expand the dataset by incorporating additional high-resolution well logs, core analyses, and drilling parameters from diverse lithologies and stress regimes.
2. The implementation of stacked or ensemble PGML models should be explored to enhance performance robustness and generalization across different reservoir conditions.
3. The developed PGML approach can be extended to predict other reservoir geomechanical and petrophysical properties, such as shear strength, elastic moduli, or fracture gradient.
4. Future research should also investigate alternative physics-informed frameworks, including Physics-Informed Neural Networks (PINNs)**,** constrained optimization approaches**,** or hybrid surrogate-physics models, to further strengthen the coupling between data-driven learning and fundamental physical laws.

Nomenclature

ΔT: Acoustic slowness (travel time per unit length), μs/ft
ϕ: Porosity, %
ρb:Bulk density, g/cm³
Vp: Compressional (P-wave) velocity, m/s
UCS: Unconfined Compressive Strength, MPa
UCS: Observed (measured) unconfined compressive strength, MPa
UCS: Modeled or predicted unconfined compressive strength, MPa
δUCS: Discrepancy between modeled and observed UCS, MPa
RESD: Deep resistivity, ohm·m
GR: Gamma ray log reading, gAPI
R^2: Coefficient of determination
RMSE: Root Mean Square Error
N: Number of observations

y: Actual (measured) value of the target variable
$\hat{y}$: Predicted value of the target variable
$\bar{y}$: Mean of actual values
V(S): Model output (payout) for coalition S in SHAP framework
ϕ_j Shapley value for feature j

Abbreviations and Acronyms

AI: Artificial Intelligence
ANN: Artificial Neural Network
AAPE: Average Absolute Percentage Error
AAS: Alkali-Activated Slag
AdaBoost: Adaptive Boosting
ANFIS: Adaptive Neuro-Fuzzy Inference System
ASTM: American Society for Testing and Materials
CCS: Carbon Capture and Storage
CCUS: Carbon Capture, Utilization, and Storage
CatBoost: Categorical Boosting Algorithm
CPB: Cemented Paste Backfill
DCNN: Deep Convolutional Neural Network
GBT: Gradient Boosting Tree
GRNN: Generalized Regression Neural Network
ISRM: International Society for Rock Mechanics
KNN: K-Nearest Neighbor
LRP: Layer-wise Relevance Propagation
LS-SVM: Least Squares Support Vector Machine
MAE: Mean Absolute Error
MARS: Multivariate Adaptive Regression Splines
MICE: Multiple Imputation by Chained Equations
ML: Machine Learning
MLR: Multiple Linear Regression
PGML: Physics-Guided Machine Learning
PINN: Physics-Informed Neural Network
PSO: Particle Swarm Optimization
RBFNN: Radial Basis Function Neural Network
RF: Random Forest
RVM: Relevance Vector Machine
SANN: Sequential Artificial Neural Network
SHAP: SHapley Additive exPlanations
SVM: Support Vector Machine
SVR: Support Vector Regression
USV: Ultrasonic P-wave Velocity
XAI: Explainable Artificial Intelligence
XGBoost: Extreme Gradient Boosting Algorithm

Statements and Declarations

Data Availability

The data that support the findings of this study are available from Basrah Oil Company but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the co-corresponding author **Watheq J. Al-Mudhafar** (watheq.almudhafar@utexas.edu) upon reasonable request and with permission of Basrah Oil Company.

Author Contributions

Ali A. Abdulazeez, Abdulsayed H. Abdulsayed, Mojtaba M. Hasak, and Mohammed H. Abdulredha Methodology, Formal analysis, investigation, Writing-original draft, Visualization. **Watheq J. Al-Mudhafar:** Conceptualization, Resources, Supervision, Validation, Visualization, Writing-review & editing. **Raed H. Allawi:** Investigation, Formal Analysis.

Declaration of Competing Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding Declaration

The authors declare that the research did not receive any financial support from funding agencies, commercial entities, or institutional grants. The study was entirely self-funded by the authors.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author used ChatGPT to generate some paragraphs. After using this tool/service, the authors reviewed and edited the content as needed and took full responsibility for the content of the publication.

References

Afolagboye, L.O., Ajayi, D.E., and Afolabi, I.O. (2023) 'Machine learning models for predicting unconfined compressive strength: A case study for Precambrian basement complex rocks from Ado-Ekiti, Southwestern Nigeria', Scientific African, 20, e01715. http://doi.org/10.1016/j.sciaf.2023.e01715.

Aladejare, A.E., Idowu, K.A. and Ozaji, T. (2024) 'Reliability of Monte Carlo simulation approach for estimating uniaxial compressive strength of intact rock', Earth Science Informatics, 17, pp. 2043–2053. Available at: https://doi.org/10.1007/s12145-024-01262-1

Al-Douri, M.A., Kareem, Y.S. & Al-Hazaa, S.H., 2024. Petrographic properties of Sandstone Rocks in the Zubair Formation (Tuba Oil Field) in Basrah Governorate, Southern Iraq. *Samarra Journal of Pure and Applied Science*, 6(2), pp.289–300. Available at: https://doi.org/10.54153/sjpas.2024.v6i2.824

Al-Hawary, S.I.S., Ali, E., Kamona, S.M.H., Saleh, L.H., Abdulwahid, A.S., Al-Saidi, D.N., Alhassan, M.S., Rasen, F.A., Abbas, H.A., Alawadi, A., Abbas, A.H. and Sina, M., 2023. Prediction of geomechanical bearing capacity using autoregressive deep neural network in carbon capture and storage systems. *Heliyon*, 9(9), p.e21913. https://doi.org/10.1016/j.heliyon.2023.e21913

Al-Jaberi, M.H. & A-Jafar, M.K., 2020. Elements distribution for the upper sandstone member of the Zubair Formation in Zubair oil field, southern Iraq. *Iraqi Geological Journal*, 53(1E), pp.55–74. Available at: https://doi.org/10.46717/igj.53.1E.5Ry-2020-07-05

Al-Jafar, M.K. and Al-Jaberi, M.H., 2021. Determination of clay minerals using gamma ray spectroscopy for the Zubair Formation in Southern Iraq. Journal of Petroleum Exploration and Production Technology. https://doi.org/10.1007/s13202-021-01371-3

Al-Mudhafar, W.J. and Mohamed, L. 2015. Incorporating lithofacies classification and well logs into statistical learning algorithms for comparative multisource permeability modeling. SPE North Africa Technical Conference and Exhibition, Cairo, Egypt. https://doi.org/10.2118/175776-MS

Al-Muhailan, M., Hussain, I., Maliekkal, H., Ghoneim, O., Nair, P. and Fayed, M., 2013. New HTHP Cutter Technology Coupled with FEA-Based Bit Selection System Improves ROP by 60% in Abrasive Zubair Formation. International Petroleum Technology Conference, Beijing, China. IPTC-17122.

Arachchilage, C.B., Fan, C., Zhao, J., Huang, G., and Liu, W.V. (2023) 'A machine learning model to predict unconfined compressive strength of alkali-activated slag-based cemented paste backfill', Journal of Rock Mechanics and Geotechnical Engineering, 15, 2803-2815. http://doi.org/10.1016/j.jrmge.2022.12.009.

Bachu, S., Bonijoly, D., Bradshaw, J., Burruss, R., Holloway, S., Christensen, N.P. & Mathiassen, O.M., 2007. CO2 storage capacity estimation: Methodology and gaps. *International Journal of Greenhouse Gas Control*, 1(4), pp.430–443. https://doi.org/10.1016/S1750-5836(07)00086-2

Barzegar, R., Sattarpour, M., Deo, R., Fijani, E. and Adamowski, J. (2019) 'An ensemble tree-based machine learning model for predicting the uniaxial compressive strength of travertine rocks', Neural Computing and Applications. http://doi.org/10.1007/s00521-019-04418-z.

Bergstra, J. and Bengio, Y. (2012). Random search for hyper-parameter optimization. The Journal of Machine Learning Research, 13(1), pp. 281–305. DOI: https://dl.acm.org/doi/10.5555/2503308.2188395

Çelik, S.B. (2019) 'Prediction of uniaxial compressive strength of carbonate rocks from nondestructive tests using multivariate regression and LS-SVM methods', Arabian Journal of Geosciences, 12(193).https://doi.org/10.1186/s44147-024-00472-0

Ceryan, N., Okkan, U. and Kesimal, A. (2012) 'Application of Generalized Regression Neural Networks in Predicting the Unconfined Compressive Strength of Carbonate Rocks', Rock Mechanics and Rock Engineering, 45(6), pp. 1055–1072. doi:10.1007/s00603-012-0239-9.

Ceryan, N., Okkan, U. and Kesimal, A. (2013) 'Prediction of unconfined compressive strength of carbonate rocks using artificial neural networks', Environmental Earth Sciences, 68(3), pp. 807–819.http://doi.org/10.1007/s12665-012-1783-z

Chang, C., Zoback, M.D. and Khaksar, A. (2006) 'Empirical relations between rock strength and physical properties in sedimentary rocks', Journal of Petroleum Science and Engineering, 51(3-4), pp. 223–237. doi:10.1016/j.petrol.2006.01.003.

Gowida, A., Elkatatny, S. and Gamal, H. (2021) 'Unconfined compressive strength (UCS) prediction in real-time while drilling using artificial intelligence tools', Neural Computing and Applications. http://doi.org/10.1007/s00521-020-05546-7.

Harris, G.D., Wellner, R.W., Catterall, V., Kairo, S., Liu, C. and Chen, Y., 2012. Stratigraphy and Depositional Environment of the Upper Zubair Sandstone ("Main Pay"), West Qurna 1 Field, Iraq. EAGE Workshop on Iraq: Hydrocarbon Exploration & Field Development, Istanbul, Turkey.

Hassan, M.Y. and Arman, H. (2022) 'Several machine learning techniques comparison for the prediction of the uniaxial compressive strength of carbonate rocks', Scientific Reports, 12(1), 20969. http://doi.org/10.1038/s41598-022-25633-0.

He, M., Zhang, Z., Ren, J., Huan, J., Li, G., Chen, Y., and Li, N. (2019) 'Deep convolutional neural network for fast determination of the rock strength parameters using drilling data', International Journal of Rock Mechanics and Mining Sciences, 123, 104084. http://doi.org/10.1016/j.ijrmms.2019.104084.

Howard, G. C. and Fast, C. Robert. (1970). Hydraulic fracturing. 203.

Hsieh, W.W., 2023. *Introduction to Environmental Data Science*. Cambridge: Cambridge University Press. Available at: https://doi.org/10.1017/9781107588493

Huang, X., Kroening, D., Ruan, W., Sharp, J., Sun, Y., Thamo, E., Wu, M., & Yi, X. (2020). A survey of safety and trustworthiness of deep neural networks: Verification, testing, adversarial attack and defence, and interpretability. Computer Science Review, 37. https://doi.org/10.1016/j.cosrev.2020.100270

Kaheman, K., Kaiser, E., Strom, B., Kutz, J. N., & Brunton, S. L. (2019) Learning discrepancy models from experimental data. Conference on Decision and Control 2019. DOI: https://doi.org/10.48550/arXiv.1909.08574

Khan, M.N. et al. (2025). Machine learning-based prediction of unconfined compressive strength of organic-rich clay shales using hybrid destructive and non-destructive inputs, Scientific Reports, 15, 31855. https://doi.org/10.1038/s41598-025-15572-x

Khandelwal M (2013) Correlating P-wave velocity with the physico-mechanical properties of different rocks. Pure Appl Geophys 170:507–514.

Khandelwal, M., Singh, T.N., 2009. Correlating static properties of coal measures rocks with P-wave velocity. Int. J. Coal Geol. 79, 55–60.

Lazim, A.A. & Dawood, H.S., 2020. Comparison between Rafidhiya and Shuaiba Domes within the properties of Mishrif Formation in Zubair Field; the implication of structural Geology and petrophysical analyses. *Journal of Petroleum Research and Studies*, 10(3). Available at: https://doi.org/10.52716/jprs.v10i3.331

Li, S., Wang, Y. and Xie, X. (2021) 'Prediction of Uniaxial Compression Strength of Limestone Based on the Point Load Strength and SVM Model', Minerals, 11(12), p. 1387. Available at: https://doi.org/10.3390/min11121387

Lundberg, S.M. and Lee, S.I. (2017). A unified approach to interpreting model predictions, *Advances in Neural Information Processing Systems*, 30, pp. 4765–4774.

Nauta, M., Trienes, J., Pathak, S., Nguyen, E., Peters, M., Schmitt, Y., Schlötterer, J., van Keulen, M. & Seifert, C., 2023. From Anecdotal Evidence to Quantitative Evaluation Methods: A Systematic Review on Evaluating Explainable AI. *ACM Computing Surveys*, 55(13), pp.1-42. Available at: https://doi.org/10.1145/3583558

Miah, M.I. et al. (2025). Effects of Petrophysical Parameters on Sedimentary Rock Strength Prediction: Implications of Machine Learning Approaches, Applied Sciences, 15(16), 9158. https://doi.org/10.3390/app15169158

Molnar, C. (2022) Interpretable Machine Learning: A Guide for Making Black Box Models Explainable. 2nd edn. Available at: https://christophm.github.io/interpretable-ml-book/

Niu et al. (2024). Unconfined compressive strength prediction of rock materials based on machine learning, Journal of Engineering and Applied Science,71:137.https://doi.org/10.1186/s44147-024-00472-0

Rahaman, M.S. and Miah, M.I. (2025) 'Data-driven resilient model development and feature selection for rock compressive strength prediction using machine learning and transformer techniques', Earth Science Informatics, 18, 293. https://doi.org/10.1007/s12145-025-01803-2

Rose, S., & McGuire, T. G. (2019). Limitations of P-Values and R-squared for Stepwise Regression Building: A Fairness Demonstration in Health Policy Risk Adjustment. The American Statistician, 73(sup1), 152–156. https://doi.org/10.1080/00031305.2018.1518269

Samek, W., & Müller, K. R. (2019). Towards Explainable Artificial Intelligence. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 11700 LNCS, 5–22. https://doi.org/10.1007/978-3-030-28954-6_1

Shapley, L.S. (1953). A value for n-person games', in Kuhn, H.W. and Tucker, A.W. (eds.) *Contributions to the Theory of Games II*. Princeton: Princeton University Press, pp. 307–317.

Sharma PK, Singh TN (2008) A correlation between P-wave velocity, impact strength index, slake durability index and uniaxial compressive strength. Bull Eng Geol Environ 67:17–22.

Sissakian, V. , Al-Ansari, N. and Knutsson, S. (2014) Origin of Some Transversal Linear Features of NE-SW Trend in Iraq, and Their Geological Characters. *Natural Science*, **6**, 996-1011. https://doi.org/10.4236/ns.2014.612091

Wang, Y. et al. (2025). Comparison of machine learning models for rock UCS prediction using measurement while drilling data', Scientific Reports, 15, 8434. https://doi.org/10.1038/s41598-025-93111-4

Wen, T., LI, D., WANG, Y., HU, M., and TANG, R. (2024) 'Machine learning methods for predicting the uniaxial compressive strength of the rocks: a comparative study', Frontiers of Earth Science. http://doi.org/10.1007/s11707-024-1101-6.

Yan, Y., Borhani, T. N., Subraveti, S. G., Pai, K. N., Prasad, V., Rajendran, A., Nkulikiyinka, P., Asibor, J. O., Zhang, Z., Shao, D., Wang, L., Zhang, W., Yan, Y., Ampomah, W., You, J., Wang, M., Anthony, E. J., Manovic, V., & Clough, P. T. (2021). Harnessing the power of machine learning for carbon capture, utilization, and storage (CCUS)-a state-of-the-art review. In Energy and Environmental Science (Vol. 14, Issue 12, pp. 6122–6157). Royal Society of Chemistry. https://doi.org/10.1039/d1ee02395k

Yang, C., Guan, X., Xu, Q., Xing, W., Chen, X., Chen, J., & Jia, P. (2024). How can SHAP (SHapley Additive exPlanations) interpretations improve deep learning based urban cellular automata model? Computers, Environment and Urban Systems, Volume 111, 102133. http://doi.org/10.1016/j.compenvurbsys.2024.102133

Z.F.O.D. Zubair Field Operating Division, Annual Report of Zubair Formation. (internal Report unpublished), (2016).

Zhao, Q. and Shi, Y. (2025) 'Prediction of Unconfined Compressive Strength of Stabilized Sand Using Machine Learning Methods', Indian Geotechnical Journal, 55, 315–332. https://doi.org/10.1007/s40098-024-00924-7