Journal Homepage: http://ijcpe.uobaghdad.edu.iq
Iraqi Journal of Chemical and Petroleum Engineering
Vol. 27 No. 2 (June 2026) 45 – 58
EISSN: 2618-0707, PISSN: 1997-4884

Prediction of rock unconfined compressive strength for subsurface formation CO_2 storage using physics augmented machine learning: A comparative performance evaluation

Fatimah F. Mutashar [a, *], Hassan A. Abdul Hussein [a]

a Petroleum Engineering Department, College of Engineering, University of Baghdad, Baghdad, Iraq

Abstract

The paper introduces a state-of-the-art Physics Augmented Machine Learning framework to predict the critical geomechanical property – Unconfined Compressive Strength (UCS) – in the Zubair reservoir of the Zubair field for mitigating the risk of CO_2 storage. The geomechanical results and modeling workflow were obtained via an integrated software system that was developed in Python, while all computational processes and model training were conducted in the Deepnote interactive cloud computing environment; thus, all of the necessary libraries and the high-performance processing capacities were seamlessly integrated. Through the comparison of six ensemble algorithms which are Categorical Boosting (CatBoost), XGBoost (Extreme Gradient Boosting), RF(Random Forest Regressor), ExtraTrees (Extremely Randomized Trees), GBM (Gradient Boosting Machine), and Deep ANN (Deep Artificial Neural Networks), the research determined the most dependable models, while the Gradient Boosting model stood out by achieving UCS prediction with the best metrics of (R^2=0.9995, MAE=3.64). The models are not only providing a precise and reliable but also a scalable assessment tool for subsurface stability, containment integrity, and real-time reservoir monitoring.

Keywords: Physics-Augmented Machine Learning; Machine learning; Unconfined Compressive Strength; Zubair field; geomechanical property.

Received on 10/02/2026, Received in Revised Form on 01/04/2026, Accepted on 03/04/2026, Published on 30/06/2026

https://doi.org/10.31699/IJCPE.2026.2.4

1- Introduction

The environment-friendly and sustainable climate change mitigation measure of CO_2 storage in depleted reservoirs is gained significant momentum with the introduction of more carbon-sensitive policies worldwide, thus giving rise to its prominence [1]. Along these lines, UCS becomes a mainstay in ensuring that the reservoir will be stable and remain intact during and after CO_2 injection [2, 3] For accurate capacity predictions to be made, quite a detailed understanding of the UCS of the reservoir is needed, which mainly refers to knowing how the formation will behave in terms of stress changes as a result of the injection activities done without open confinement or risking caprock failure [4, 5].

In the past, the Unconfined Compressive Strength (UCS) that is a critical parameter has been determined through laboratory core testing. These direct methods are not only very precise but also costly, lengthy, and subject to the core sample availability in different reservoir intervals [6, 7]. Thus, using empirical correlations and numerical simulations is a common practice for many field-scale studies. Nevertheless, the traditional approaches do not always represent the real sub-surface complexities, since they commonly oversimplify the nonlinear relationship between the petrophysical logs and geomechanical properties [8, 9].

Physics Augmented Machine Learning has proven to be a major solution to the current difficulties in the field of rock mechanics by smoothly integrating the power of machine learning with the basic laws of physics [10, 11]. Physics Augmented Machine Learning models are not only capable of outputting data-based results but also ensuring that these output values are physically feasible through the use of geological information, well log datasets and the like [12, 13]. Recent developments disclosed the remarkable impact of integrating physical constraints—like elastic modulus coupling—on subsurface characterization to be more reliable compared to usual AI methods [14, 15].

An advanced Physics Augmented Machine Learning modeling technique is proposed by doing the UCS analysis over the Zubair reservoir of the Zubair field to forecast the geomechanical bearing capacity in the exhausted reservoirs. The parameter is very critical since it indicates with certainty that either a reservoir will undergo mechanical action of CO_2 injection without failure or it will fail leading to the formation of leakage pathways [16, 17]. To thoroughly assess, the paper investigates six most recent and sophisticated methods of ensemble and neural network algorithms: Random Forest (RF), Extremely Randomized Trees (ExtraTrees), Gradient Boosting Machines (GBM), XGBoost,

*Corresponding Author: Email: fatima.Falaah2208@coeng.uobaghdad.edu.iq

CatBoost, and Artificial Neural Networks (ANN) in terms of their efficacy.

The main aim of the study is to validate the Physics Augmented Machine Learning potential in delivering accurate geomechanical forecasts that do away with the necessity of using expensive and time-consuming traditional methods. This work not only intends to present a reliable but also a scalable instrument that measures the stability of a reservoir during the entire period of CO_2 storage operations by processing the Zubair reservoir of the Zubair field.

The entire procedure followed in this research is illustrated in Fig. 1, which represents the gradual transition from data collection to model evaluation and interpretation step-by-step.

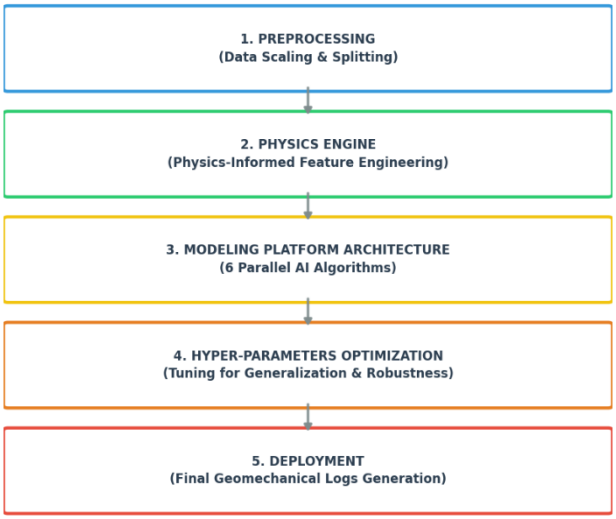

Fig. 1. The algorithms framework and methodological stages

2- Geological background

The Zubair oilfield located about 20 km southwest of Basra in southern Iraq is one of the most eminent hydrocarbon giants in the Middle East in terms of the total amount of reserves and historical cumulative production [18, 19]. The field, which is depicted in Fig. 2 below, is characterized, as a major northwest-southeast trending anticline, of approximately 60–67 km in length and is composed of four structural domes: Al-Hammar, Shuaiba, Rafidya, and Safwan [18, 20]. Its long-term production history since the 1950s has caused significant reservoir pressure depletion and in-situ stress change [1, 21]. This mature state simulates a real-world laboratory for testing the geomechanical response of exhausted formations during CO_2 sequestration and storage operations [19, 22].

The Zubair Formation (Lower Cretaceous), with its high lithological complexity, is the most promising reservoir for CO_2 storage. This complexity is due to its original deltaic to marginal marine depositional environments [23] The formation is made of pore-rich sandstone layers with shales and siltstones at regular intervals [6].

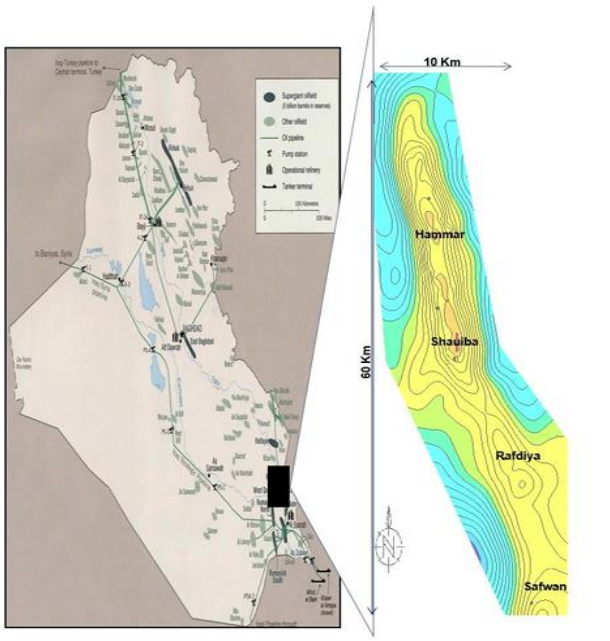

Fig. 2. Zubair oilfield map [personal communication]

The vertical migration of CO_2 is prevented by a "double-seal" system: the carbonate-rich Shuaiba Formation, which is the primary caprock, and the overlying shale-dominated Nahr Umr Formation that acts as a secondary barrier and is quite extensive [4, 21].

Based on the variations of the sand-to-shale ratios, which influence the hydraulic and mechanical behavior of the reservoir, the Zubair Formation is officially divided into five members, as shown in Table 1.

- Upper Shale Member (USM): It is the immediate flow baffle for the region and also the top-seal component.
- Upper Sandstone Member (USSM): Known as the "Third Pay," this unit is commercially celebrated for its excellent reservoir quality and being the main push target [13].
- Middle Shale Member (MSM): A major argillaceous section that acts as a hydraulic barrier between the upper and lower reservoirs [23].
- Lower Sandstone Member (LSSM): is known as the Fourth Pay because it functions as a secondary reservoir with distinct mechanical and petrophysical characteristics. [8].
- Lower Shale Member (LSM): serves as the lower boundary of the formation while it functions to separate this unit from the underground units that exist below it.

The mechanical difference between these five members -especially the different stiffness and strength- essentially governs the direction of stress and the extent of the fracture during the CO_2 injection [21,24]. The Zubair field, with its challenging geological situations, becomes a perfect place for trials of the Physics Augmented Machine Learning models to gain precise and physically consistent geomechanical predictions [22].

Table 1. Internal subdivision of the zubair field by different names

Formation	Member	Pool	SOC Zonation
ZUBAIR Fm	Upper Shale	UShale 5	Z/1
		UShale 4	
		UShale 3	
		UShale 2	
		UShale 1	
	Upper Sandstone (3rd Pay)	AB Sandstone	AB
		H Sandstone	CG
		H Sandstone	
		H Sandstone	H
		H Sandstone	
		H Sandstone	
		L Sandstone	IK
		L Sandstone	L
		L Sandstone	
		L Sandstone	
		N Sandstone	M
	Middle Shale	—	N
		—	Z/4
		—	
	Lower Sandstone (4th Pay)	Lower Sandstone	Z/5
		Lower Sandstone	
		Lower Sandstone	
		Lower Sandstone	
		Lower Sandstone	
	Lower Shale	—	Z/6

3- Methodology

Unified, high-quality databases are mandatory for the development and anchoring of a formidable Physics Augmented Machine Learning. The database of this research study consisted of data from nine vertical wells in the Zubair oilfield, including petrophysical logs (GR, DT, RHOB), pore space (PHIE) and permeability (K) data, and laboratory results for Unconfined Compressive Strength (UCS). The primary spatial reference for data alignment was "Measured Depth" (MD). The dataset was subdivided into three workable subsets: first, the Training Subset (6 Wells) which made up of Wells Zb-40, Zb-122, Zb-201, Zb-210, Zb-232, and Zb-233. These wells were used to build the "Master Data" matrix in order to assure that the proper incorporation of the broad spectrum of lithological facies and stress regimes takes place in the models.The decision to select the six identified wells was carried out strategically in order to ensure the best data quality and continuity, since the often-limited depth of the data available subsurface, limited more so by the veil of industry confidentiality and very draconian data-sharing policies. The resolution of well logs is at the highest level conceivable, with measurements made at intervals of less than one meter. With this much higher density training matrix, there is a certainty maintained with regard to sufficient statistical significance rendering the Physics Augmented Machine Learning framework particularly competent to document the complex stratigraphic heterogeneity distributed within the different members of the Zubair Formation. Second was the Testing and Validation Subset (1 Well) in which Well Zb-284 was set aside for blind testing to measure the generalization ability and for refining the hyperparameters; and third, the Blind Prediction Subset (2 Wells) were meant for the final prediction stage and consisted of Wells Zb-236 and Zb-229.

3.1. Data cleaning and imputation

Having discussed the nature of well logs associated with possible data losses due to harsh drilling conditions, the problem of missing values had to be curbed with the help of the custom "MissingValueCompensator" protocol. This approach was based on the work of imputation engineers who opted for the best method according to the type of missing feature and the extent of loss. [8]. The methods used in the present study are as below:

- Feature-wise Inclusion of Multivariate Imputation by Chained Equations (MICE - Iterative Imputer): This method is used to handle complex and multiple missing points by computing the missing value as a function of other features in a progressive manner [13].
- K-Nearest Neighbors Imputer (KNN Imputer): This replacement is done based on local "nearest neighbours" so as to conserve possible local interactions between rock properties [20].
- Linear Interpolation and Sequential Fills: For the continuity of the geophysical logs and to ditch the nonphysical jumps, these techniques were implemented on sequential data (Measured Depth - MD) [21].
- Statistical Analysis: Mean and Median values act as a last line of defense for gaps that are not entirely correlated to some sequential pattern [24].

3.2. Algorithm categories

Predictive model uses a system architecture in order to have realization on six different algorithms for parallel computing to get prediction facts with higher stability regarding the different geophysical complexities that the platform may ever encounter; plus, it discusses three general classes of numerical modellings:

- Boosting Models: In this category. The tree-based models in this case are CatBoost, XGBoost, and GBM models. They have a peculiar feature of the ability to reduce successive errors for processing high-resolution geomechanical data [25, 26].
- Bagging and Tree Ensembles: Samples being selected from Random Forest (RF) and Extremely Randomized Trees (ExtraTrees) are representatives of this subspace. These approaches run along the path of creating hundreds of independent tree structures in order to curb an-nihilation in this case [27].
- Neural Computation: In this where, deep ANN [99; 97], specifically MLPRegressor, has been integrated to capture deep non-linear relationships that tree-based models are not able to capture [28].

To avoid any bias, each model is first trained on an array of data and then subjected to validation through the use of the same 'Master Matrix' to compare and finalize results. The "Champion Model" for each geomechanical

property is defined in accordance with the most performance criteria (R^2, MAE).

3.3. Physics-augmented feature engineering

The following part is not only about outperforming normal ML but also about building a "Geomechanical Physics Engine" that combines the numerical data with rock engineering laws:

3.3.1. Level I: Guided features

The incorporation of geomechanical rules as extra raw inputs was meant to facilitate the learning. In place of just using the raw logs (DTC, DTS, RHOB), columns that represented the elastic properties were made through the use of the established physical equations. Also, more properties were introduced as supporting and reinforcing features due to their close linkage to the target parameters:

- Dynamic shear modulus (G_{dyn}): Represents rock's resistance to lateral deformation due to applied stresses. Mathematically, the dynamic shear modulus can be estimated based on Eq. 1 below [29, 30]:

$$G_{dyn} = 13474.45 \ \frac{ROHB}{(DTS^2)} \quad (1)$$

Where $G_{dyn}$is dynamic shear modulus (Mpsi); ρb= bulk density, g/cm³; DTS= shear slowness, µs/ft; 13474.45= conversion factor.

- Dynamic Bulk Modulus (k_{dyn}): The dynamic bulk modulus measures a rock's resistance to uniform volumetric compression under hydrostatic loading, where stresses are applied equally in all directions [19]. Mathematically, the dynamic bulk modulus can be estimated based on Eq. 2 below [29, 30]:

$$k_{dyn} = 13474.45 \ ROHB \ \frac{1}{(DTc^2)} - \left(\frac{4}{3}\right) G_{dyn} \quad (2)$$

Where $k_{dyn}$is dynamic Bulk modulus (Mpsi); ρb= bulk density, g/cm³; DTC= shear slowness, µs/ft; 13474.45= conversion factor.

- Young's Modulus **(**E**):** It is a rock stiffness characteristic under axial compressive stress E_{dyn} is estimated based on Eq. 3 [3]:

$$E_{dyn} = \frac{9G_{dyn}K_{dyn}}{G_{dyn}+3K_{dyn}} \quad (3)$$

Where $E_{dyn}$is dynamic Young's modulus (Mpsi);$k_{dyn}$is dynamic Bulk modulus (Mpsi);$G_{dyn}$is dynamic shear modulus (Mpsi).

Dynamic values are typically higher than static ones because of pore pressure, cementation, rate of stress-strain, and amplitude. Mathematically, the static young's modulus is defined as Eq. 4 by [31]:

$$E_{sta} = 0.032 * {E_{dyn}}^{1.632} \quad (4)$$

Where $Esta$ = Static Young's modulus, Mpsi.

- Poisson's ratio (ν) referred to as the ratio between lateral expansion due to longitudinal compression is computed the dynamic Poisson's ratio is obtained as show in Eq. 5 by [3] and Eq. 6 by [32]:

$$\nu_{dyn} = \frac{3K_{dyn}-2G_{dyn}}{6K_{dyn}+2G_{dyn}}. \quad (5)$$

$$\nu_{sta} = \nu_d * \nu_{multiplier}. \quad (6)$$

- Unconfined Compressive Strength (UCSHorsrud): This was integrated as a global physical reference in evaluating compressive strength as a function of wave travel time. Mathematically, the UCS is defined as Eq. 7 [2].

$$\text{UCS} = 0.77 \left(\frac{304.8}{\text{DTC}}\right)^{2.93} \quad (7)$$

Where UCS is in (psi).

- Wave Velocity Ratio (Vp/Vs): Calculated as a direct ratio between DTS and DTC, it serves as a primary indicator for identifying lithology and fluid saturation [29].

3.3.2. Level II: Hard constraints

The hard constraints were applied through the use of the *np.clip* function. At this stage, Physics takes on the role of an output "Guardrail" that imposes restrictions on the results based on logic (e.g., will never see Poisson's ratio greater than 0.5). The noisy data is filtered such that the actual physical logs of rock are maintained according to natural rock-laws remaining out of the noisy data [12, 33].

The sample shown in Table 2 is taken from the processed Master Matrix, which represents the integration of raw logs with engineered features derived from physical equations, and finally the dataset for model training.

Table 2. Sample of the integrated geomechanical master matrix featuring raw logs and physics engineered attributes

MD	3319	3327	3335	3342	3372	3414	3423	3433	3438	3439	3445
DTC	77.79	82.65	86.62	76.11	88.68	79.53	79.2	77.08	77	69.78	75.93
DTS	171	166	191	147	135	173	143	128	115	107	137
ROHB	2.5064	2.5071	2.5077	2.5083	2.5107	2.5141	2.5148	2.5156	2.516	2.5161	2.5166
GR	101.6	111	87.33	103.4	94.44	94.72	60.32	35.22	62.5	39.3	87.05
PHIE	0.268	0.191	0.28	0.172	0.262	0.17	0.172	0.231	0.173	0.177	0.116
K	2783	51.95	5043	19.91	2085	17.39	19.15	406.9	20.35	25.63	1.065
G_{dyn}	1.1493	1.2265	0.9234	1.5712	1.8512	1.1312	1.6545	2.058	2.5851	2.9618	1.8197
K_{dyn}	4.0481	3.31	3.272	3.7391	1.8334	3.8477	3.1966	2.9608	2.2717	3.0144	3.456
E_{dyn}	5.5804	4.9454	4.5032	5.8339	4.3016	5.356	5.4026	5.7048	5.7186	6.9635	5.8823
Esta	2.01	1.8	1.12	1.03	1.89	1.98	1.99	2.05	2.36	2.06	2.46
V_{dyn}	0.37	0.34	0.37	0.32	0.12	0.37	0.28	0.22	0.09	0.13	0.28
Vsta	0.22	0.19	0.15	0.25	0.21	0.225	0.154	0.18	0.15	0.252	0.204
UCS	9774	9191	8053	9087	7076	3002	5049	9124	4894	9706	4312
Vp/Vs	2.2036	2.008	2.2084	1.9269	1.5244	2.1759	1.807	1.6649	1.4873	1.5333	1.7979

3.4. Optimization and hyperparameter tuning

An efficient systematic, heuristic-based hyperparameter tuning method ensured the reliability of the proposed Physics Augmented Machine Learning framework's predictive performance and generalization stability [20]. In contrast to automated optimization methods,The hyperparameters in this study, through a series of iterative trial-and-error experiments, were carefully selected and from the very beginning, they were thoroughly validated by k-fold cross-validation to establish the most favorable balance between bias and variance. This method was crucial considering the extremely high heterogeneity and noise inherent in the Zubair Formation's geomechanical logs, and often standard default parameters suffer from overfitting due to this. The models were limited to learning deep non-linear physical patterns without memorizing data outliers by programmatically varying the 'max_depth' and 'n_estimators' over the six ensemble algorithms. Although it was manual, structured optimization nonetheless ensured that every single model, from tree-based CatBoost and XGBoost to intricate ANN architectures, was accurately matched to the physical range and mechanical complexity of the target property (UCS).

To standardize the architecture of the deep artificial neural networks for calculation reproducibility, we created a Multi-Layer Perceptron Regressor (or MLPRegressor). The multilayer perceptron comprised three hidden layers, having 128, 64, and 32 neurons each. The activation function being employed was the Rectified Linear Unit (ReLU) to handle nonlinear data, while 'L-BFGS' solver was chosen for weight optimization, owing to its fantastic performance on small yet complex tabular geomechanics-related datasets. The following Table 3 depicts the final configurations that were adopted in all stages of the project.

4- Results and discussion

The six Physics-Augmented Machine Learning models were compared for their predictive performance in the training, validation, and blind prediction stages for the target geomechanical property, which is unconfined compressive strength (UCS).

Table 3. Final optimized hyperparameters and structural configurations

Algorithm	Fixed Parameters	Methodological Stability Notes
CatBoost	iterations=1000, depth=8, verbose=0	Activation of loss_function='MAE 'to minimize absolute deviation in predictions.
XGBoost	n_estimators=1000, max_depth=8, learning_rate=0.03/0.05	Low learning rate ensures stable and precise convergence towards the optimal solution.
Gradient Boosting	n_estimators=500, max_depth=7, loss='huber'	Use of Huber loss to resist noise and outliers in geophysical logs.
Extra Trees	n_estimators=1000, max_depth=15	High tree depth to maximize diversity and minimize statistical variance.
Random Forest	n_estimators=1000, max_depth=12	Classical ensemble approach for stabilizing predictions through forest averaging.
ANN (MLP)	hidden_layer_sizes=(128, 64, 32), activation='relu', solver='lbfgs'	Deployment of L-BFGS solver, optimized for limited and complex tabular datasets.

4.1. Uniaxial compressive strength

The training phase for UCS showed a remarkable difference in the performance of tree-based ensemble models, especially Gradient Boosting and XGBoost, which reached an R^2 of 0.9995. In the scenario of blind testing (Well 284), the ExtraTrees model displayed the most stability in generalization with an R^2 of 0.9982 and the Least Mean Absolute Error (MAE) of 4.31. In the final prediction phase (Wells 236 and 229), the ensemble models managed to reconstruct continuous UCS logs and exhibited a great deal of difference over the optimized ANN, which was not able to deal with the formation's stratigraphic heterogeneity as shown in the comprehensive summary Table 4 below.

Table 4. Comprehensive performance comparison of machine learning models across training, validation, and field prediction stages for UCS

Algorithm	R^2 Train	MAE Train	R^2 Valid	MAE Valid	R^2 236 prediction	MAE 236	R^2 229 prediction	MAE 229
GradientBoosting	0.9995	3.64	0.9854	14.82	0.9799	19.89	0.9299	52.74
RandomForest	0.9993	16.13	0.9963	22.20	0.9811	37.55	0.9488	70.54
XGBoost	0.9995	9.10	0.9947	25.28	0.9567	60.23	0.8997	89.29
CatBoost	0.9882	115.12	0.9854	127.71	0.9759	108.66	0.9067	238.96
ExtraTrees	0.9731	147.73	0.9649	229.22	0.9600	145.88	0.7368	448.94
ANN (Deep)	0.6511	838.97	0.6000	838.97	0.7472	610.29	0.3546	945.36

The high R^2 values may seem surprising in the context of the stratigraphic heterogeneity of the Zubair Formation, which is often subject to noisy well-log data. However, the models were trained in a Physics-Augmented manner, with geomechanical constraints and geophysically derived features guiding the learning process. Such an approach confines the solution space to physically reasonable relationships as well as minimizes the influence of random fluctuations in the data. On a more speculative assumption, utilizing techniques like RobustScaler and ensemble learning with Gradient Boosting stabilized the model and minimized the impact of outliers so that these models could learn from structural deterministic patterns of rock strength predominantly worth targeting.

Table 5 and Table 6 below illustrate the prediction results of one of the models as a typical sample. The outputs given by the final prediction of the Gradient

Boosting model indicate the deep physical comprehension the model has gained during the training phase.

In Well 229, the model managed to perfectly link the decrease of the acoustic transit time and the increase of density with the rise of UCS values, reaching 6784 psi. Well 236 was the place where the model very accurately predicted a stiffness value of 9774 psi, which was a high figure considering the Gamma Ray readings were already quite elevated. This is the strong proof of the model's excellent capability in geological heterogeneity handling with high flexibility.

Table 5. Detailed numerical prediction results for UCS in well 229 using the gradient boosting model

MD (m)	DTC (us/ft)	DTS (us/ft)	ROHB (g/c)	GR (gAPI)	PHIE	K (mD)	Predicted UCS (psi)
3131.058	60.42401	118.1613	2.589227	24.18188	0.240525	671.222	6784.002
3186.074	84.91709	118.5127	2.599343	41.14477	0.114672	1.007943	6783.996
3212.592	82.02201	133.3799	2.604187	91.48536	0.14903	5.946143	4604.61
3238.652	80.92654	125.3563	2.608938	62.50134	0.145966	5.075706	6053.999
3250.692	74.13524	128.0558	2.611129	66.8155	0.159955	10.45516	4201.986
3260.903	57.0216	122.8674	2.612984	39.66681	0.19993	82.44109	4573
3268.37	83.26128	118.1986	2.61434	32.70226	0.209569	135.6402	4573.129
3283.001	70.87578	121.7319	2.616988	44.3044	0.131149	2.36095	9096
3329.33	70.36388	119.8696	2.625322	31.54666	0.253561	1316.195	8953.999
3367.278	93.1288	128.4631	2.632113	81.6859	0.080255	0.170339	4893.751
3371.24	68.10043	127.8214	2.63282	69.84205	0.076535	0.140559	3001.888
3379.775	89.13388	132.42	2.634343	91.21533	0.141144	3.956563	3001.878
3390.138	90.48466	133.2109	2.63619	96.25735	0.107772	0.705731	7638.97
3395.472	72.41743	131.5339	2.637141	85.84751	0.055931	0.048486	2310.031
3400.654	78.13216	132.6423	2.638063	90.55305	0.0853	0.221052	5371

Table 6. Detailed numerical prediction results for UCS in well 236 using the gradient boosting model

MD (m)	DTC (us/ft)	DTS (us/ft)	ROHB (g/c)	GR (gAPI)	PHIE	K (mD)	Predicted UCS (psi)
3319	77.79472	171.4257	2.50643	101.6429	0.268058	2783.243	9774.1
3327.4	82.65001	165.9608	2.507115	111.0217	0.190992	51.95478	9774.1
3334.6	86.62318	191.2952	2.507702	87.32872	0.279564	5042.905	9191.28
3342.3	76.11443	146.6687	2.508328	103.3731	0.17242	19.90558	8053
3371.6	88.68201	135.186	2.510701	94.44135	0.262468	2085.178	7076
3403.7	73.56364	128.4169	2.513286	69.48779	0.199408	80.24778	8816.11
3413.8	79.52914	173.0489	2.514096	94.7244	0.169802	17.38768	8816.11
3422.7	79.1967	143.1105	2.514808	60.3221	0.171673	19.1521	5049
3433.2	77.08356	128.3401	2.515646	35.2182	0.230836	406.9113	9123.72
3437.7	76.99587	114.5173	2.516005	62.50152	0.172852	20.35478	9123.72
3439.2	69.77605	106.9895	2.516124	39.30351	0.17731	25.62588	9706
3444.9	75.92549	136.5087	2.516578	87.05248	0.115737	1.064948	9706

The vertical distribution profiles (Fig. 3, Fig. 4, Fig. 5, Fig. 6, Fig. 7, and Fig. 8) of the predicted UCS values indicate a perfect match to the laboratory reference values at different depths. The digital logs illustrate the ability of the models to adjust the stiffness of the rock formations very closely and to flexibly the best. This close visual alignment affirms the models' success in coping with geological variation and overcoming generalization problems.

There is a clear performance stratification noticed where the ensemble trees (Gradient Boosting, XGBoost, and RF) achieved R2 scores close to one (0.99) and the animated Deep ANN-got an up to 0.65. This lack of compatibility in the prediction accuracy can be accounted for scientifically due to stratigraphic heterogeneity within the Zubair Formation. Tree-based algorithms easily take care of formatted data and readily capture distinct lithological switches of a decision tree basis for making them fit well in well-log analysis. GR and DTC are the features that may exhibit wild fluctuations. To be honest, although theoretically capable of non-linear mapping, Deep ANN often needs a sufficiently vast training sample, for the ANN neural networks converge/normally cannot handle highly variable geological settings. Also, the relatively weakening patterns revealed by artiness in the geophysical logging layers could have been of help as the ANN failed to simplify the intricate heterogeneity existing among the Zubair members compared to the ensemble boosting and bagging methodologies.

5- Geomechanical properties affect CO_2 storage

The most important consideration that makes such properties relevant for CO_2 storage is the potential of the mechanical properties of the permeability in the Zubair formation, which controls the essential factors of how complex and how quickly a boundaries system can respond to injection dynamics [6].

Extremely large Unconfined Compressive Strength (UCS) values found in some of the Zubair formation intervals signify a very strong rock skeleton that Nevertheless, very high UCS could be a problem during injection when the hot and stiff rock together with flowing liquid would create high localized stresses that might even make minute cracks in the rock [6].

In order to offer a complete picture of the reservoir's mechanical response, an integrated geomechanical log for Wells 229 and 236 is shown in Fig. 9 which, Tracks show the predicted UCS, based on the best machine learning

models (GradientBoosting model). The zones that are highlighted in red represent "Potential Leakage Zones" with a striking decrease in rock strength and changes in elastic properties. This composite map serves as a diagnostic tool for detection of scarily weak mechanical behavior, which might actually cause sealing to take less injection pressure than it normally would, for proper CO_2 containment.

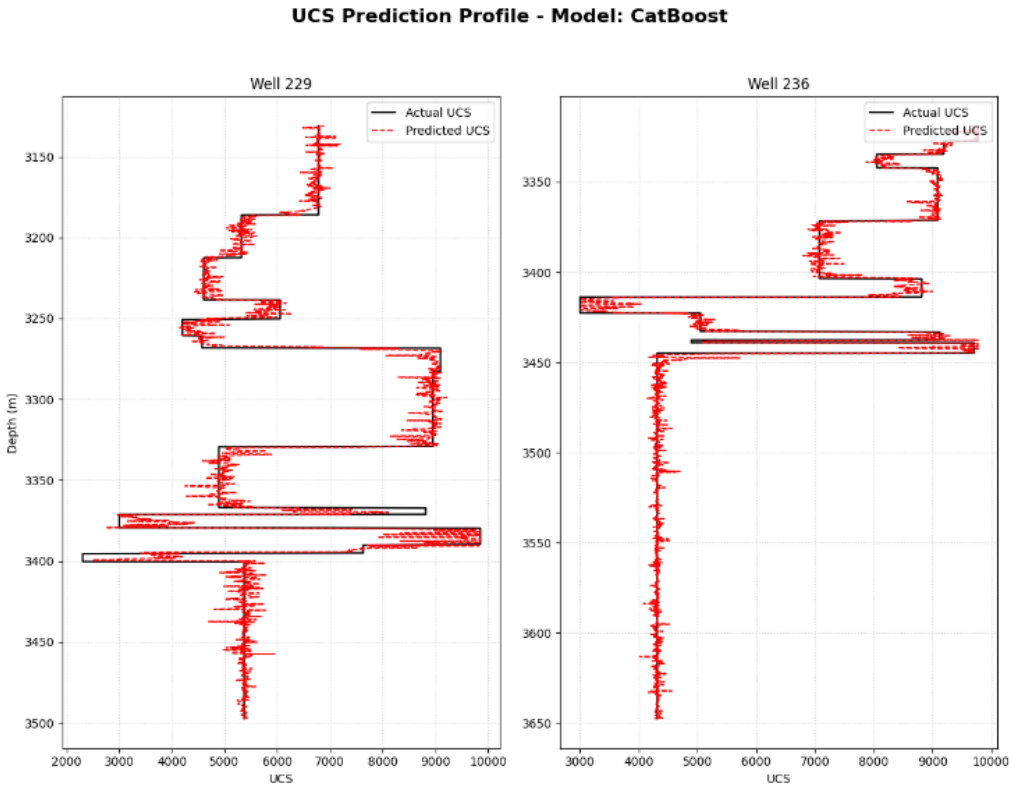

Fig. 3. UCS prediction Vs actual profile - model: Catboost

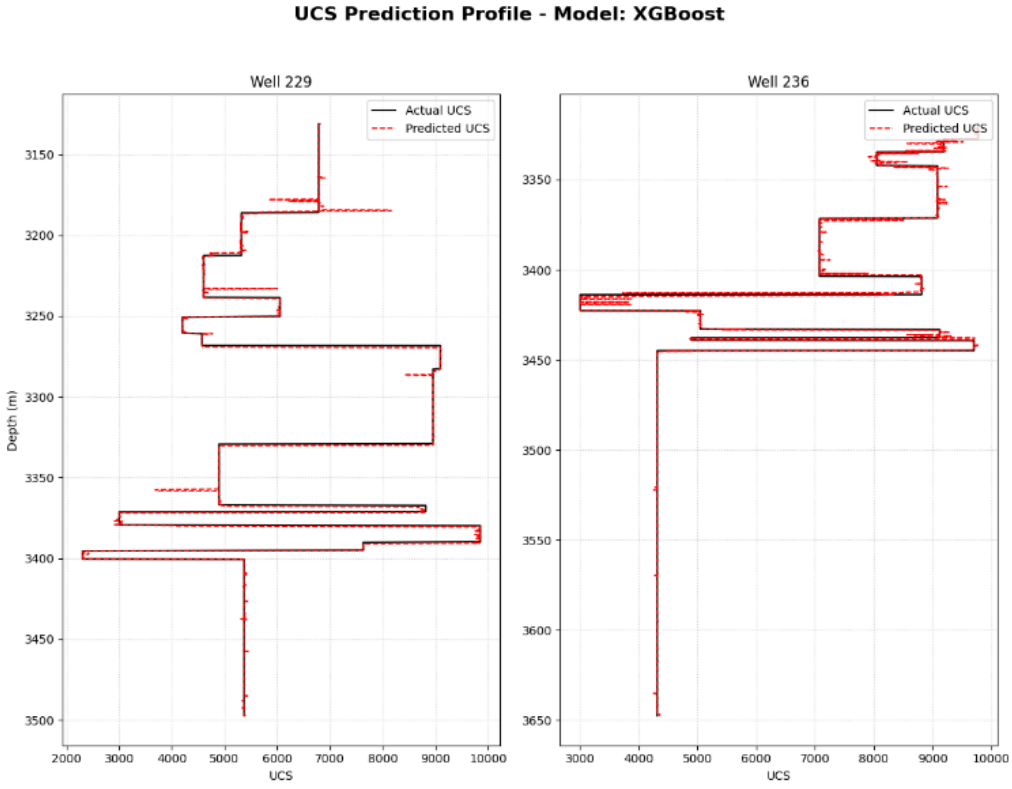

Fig. 4. UCS prediction profile Vs actual - model: XGboost

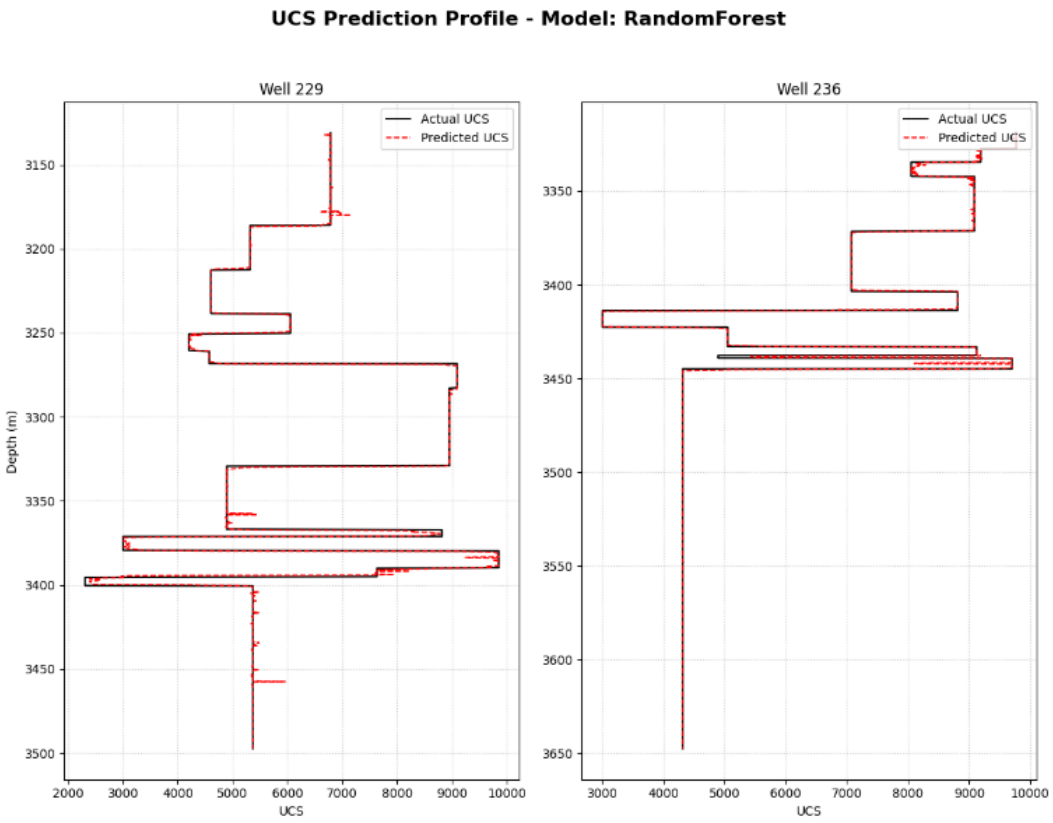

Fig. 5. UCS prediction profile Vs actual - model: random forest

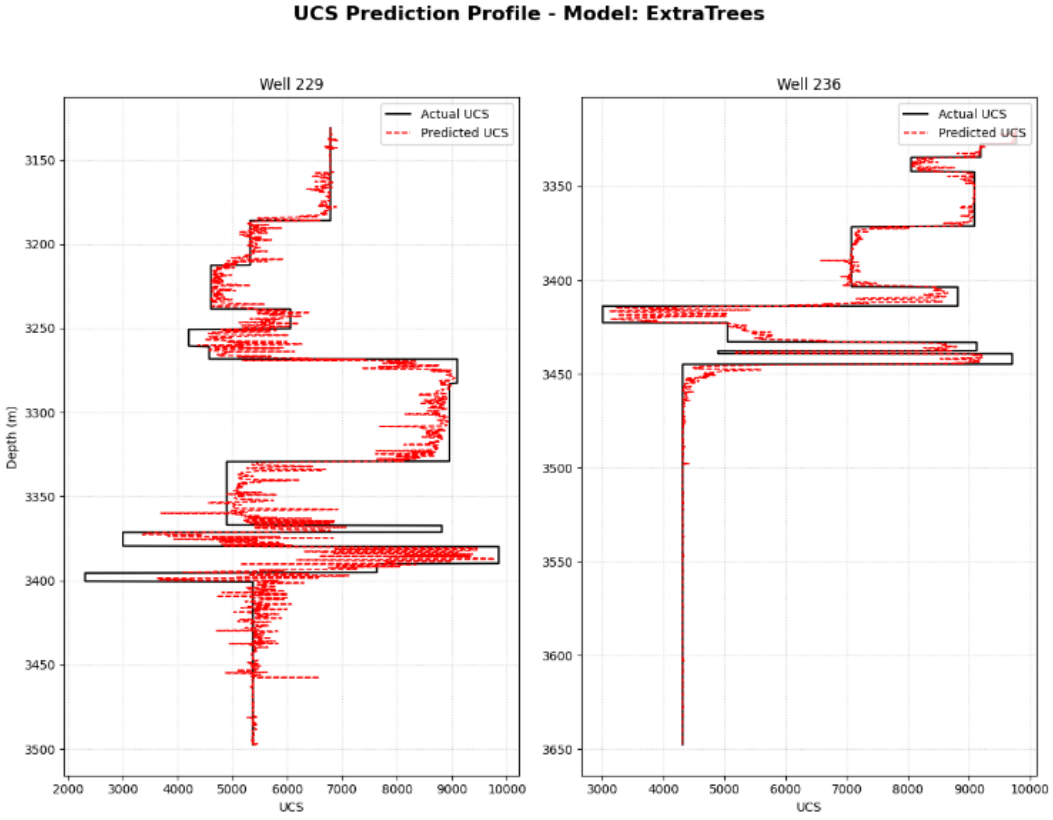

Fig. 6. UCS prediction profile Vs actual - model: extra trees

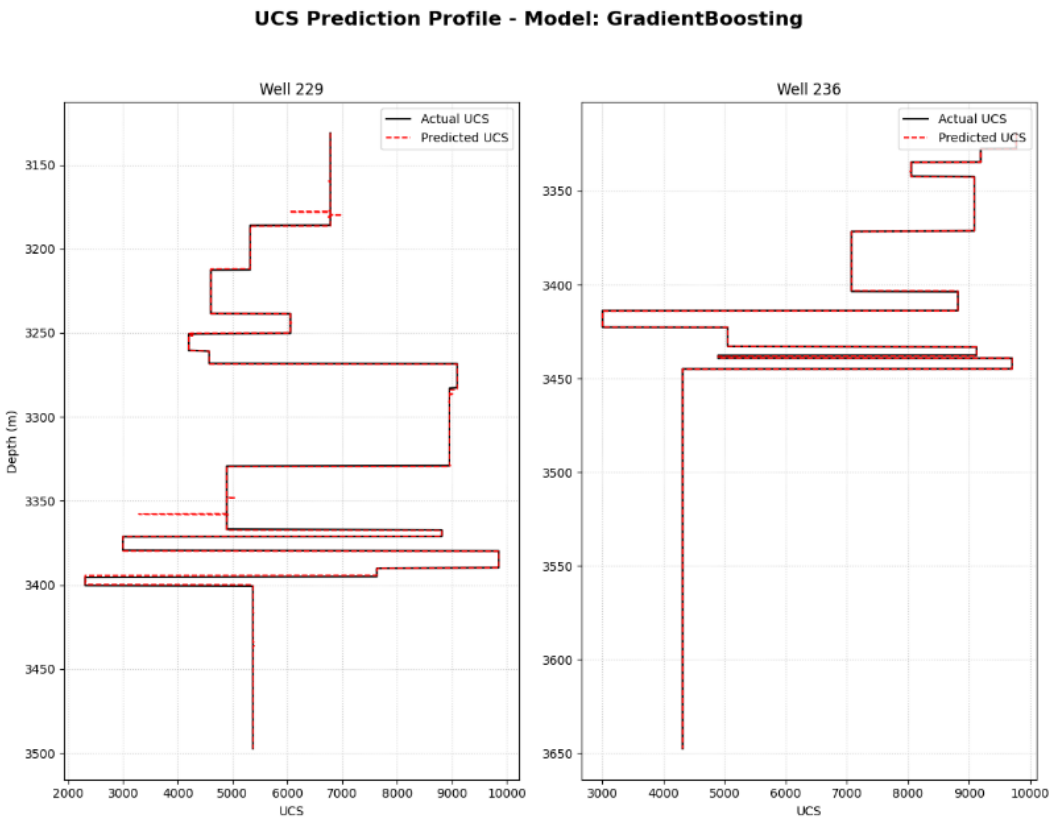

Fig. 7. UCS prediction profile Vs actual - model: gradient boosting

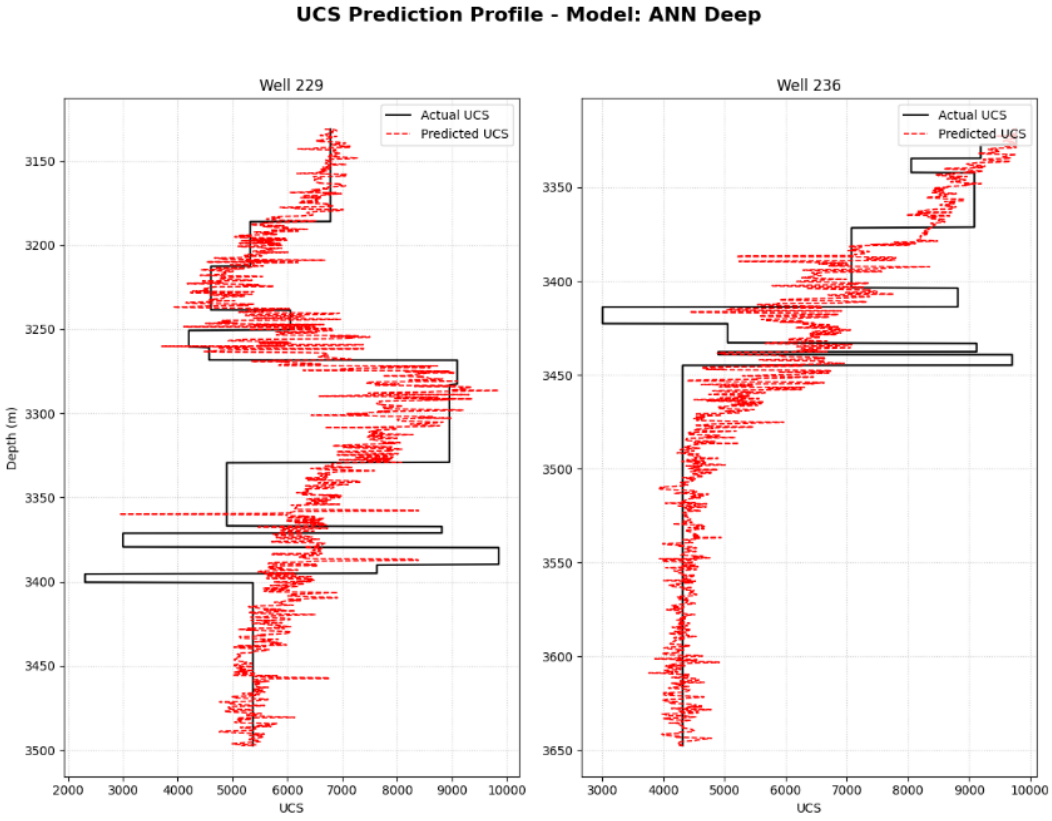

Fig. 8. UCS prediction profile Vs actual - model: ANN

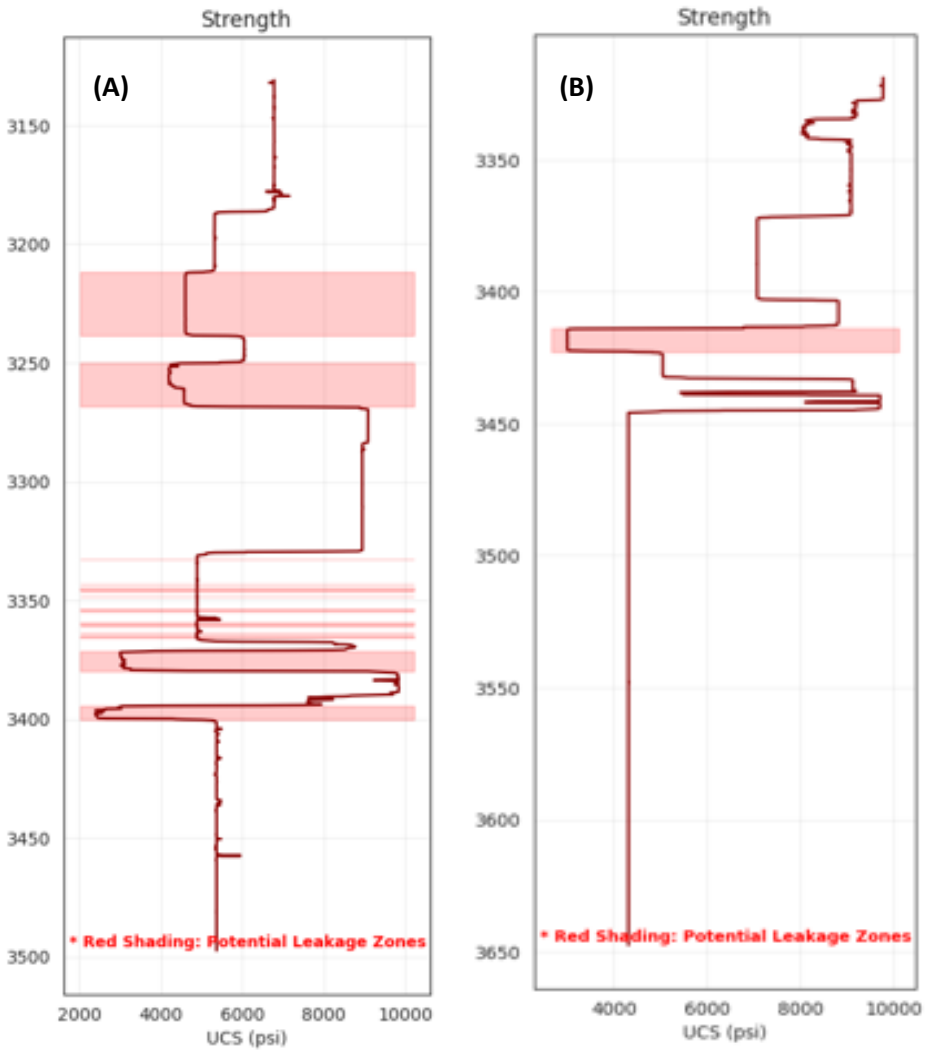

Fig. 9. Integrated geomechanical log from gradient boosting model: A) well229 and B) well 236

Fig. 10 shows Safe Injection-Window of reservoir. The green shaded portion indicates the crucial operational range between pore-pressure and fracture-pressure to assure the integrity of the formation during CO_2 injection. The orange dots point out the 'Weak UCS Zones' which cause a drastic narrowing of this window, thus reducing the injection pressures at those depths to prevent mechanical failure. Besides, a distinct stratigraphic difference is noted between the sand and shale formations, where the Zubair sand gives a larger injection margin in contrast to the shale layers which impose stricter operational constraints.

UCS which the GradientBoosting model produces with high precision develop better 'Safe Injection Pressure' estimates than existing empirical methods. The Zubair formation contains mechanical properties which traditional correlations fail to detect, thus leading to incorrect estimates of its fracture gradient. The study used ML-predicted UCS data to define operational limits which existed in the 'Al-Hammar' and 'Shuaiba' domes. The system needs this level of accuracy because it protects the Shuaiba and Nahr Umr Formations double-seal system, which enables CO2 injection optimization while preventing caprock fracture capacity from being exceeded. The study transitioned from a computational study to a geomechanical system that assesses carbon sequestration safety.

6- Summary of results

A comparative analysis of predictive results showed that the Gradient Boosting algorithm had the highest accuracy of predicting Unconfined Compressive Strength (UCS) in all the datasets. Table 7 highlights consistent strong correlations between modeled predictions and actual geomechanical measurements across the challenging blind-well prediction and field-wide prediction periods.

These results imply the GB model is robust and consistently minimized the MAE, which we strongly interpreted, without extra core sampling, as a reliable quantitative tool for estimating rocks' strength in the Zubair Formation.

The Table 7 below presents a summary of the best models chosen for their excellent performance in reducing the Mean Absolute Error (MAE) and increasing the coefficient of determination (R^2) throughout the different phases of the study.

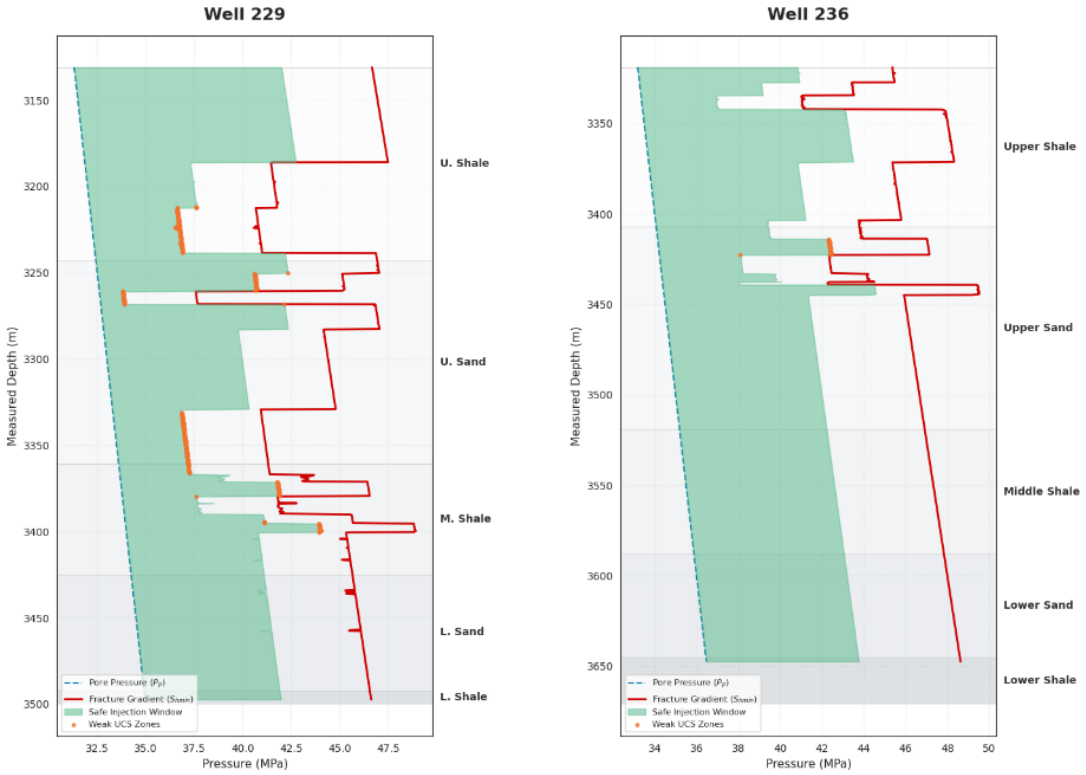

Fig. 10. Illustrates the safe injection window of prediction wells

Table 7. Summary of the best-performing models and their achieved accuracy

Geomechanical Property	Optimum Algorithm	Training Stage (R^2 / MAE)	Blind Well Test (284) (R^2 / MAE)	Field Prediction (236) (R^2 / MAE)	Field Prediction (229) (R^2 / MAE)
UCS	Gradient Boosting	0.9995 / 3.64	0.9854 / 14.82	0.9799 / 19.89	0.9299 / 52.74

7- Conclusion

Based on the obtained results, the following conclusions are summarized:

1. The partial identification of " Physics-Augmented Machine Learning "played a crucial role in the explanation of the machine learning model based on geomechanical laws. Such approach conscripted these models in producing more homogenous predictions, in terms of geomechanical uplift forecasts. Importantly, the proposed UCS profiles lie within a range of physically quantifiable values, thereby reducing the need for tedious and expensive laboratory experimentation.
2. From all six analyzed algorithms, the tree-based ensemble models are strongest in performance. Gradient Boosting Machine (GBM) proved to be the best algorithm for UCS prediction, with a very high R2 value. The GBM model achieved an almost perfect R2 = 0.9995, followed closely by the XGBoost and Random Forest.
3. Ensemble methods based on trees (ExtraTrees, Random Forest, and GBM) have the unique ability to filter geophysical noise while handling stratigraphical heterogeneity fairly efficiently. In contrast, deep ANNs show lower generalization capability, an indication that standard neural architectures require much larger datasets and more sophisticated methods for optimization just to match the performance of the boosting and bagging technique for well-log analysis.
4. Systematic hyperparameter optimization of tree depths and learning rates was crucial in balancing the bias-variance trade-off. By ensuring high predictive performance, the models became robust against overfitting, even within the complex lithologies of the Zubair Formation.

The PIIML framework formulated offers a reliable technique for identifying "Injection Safe Windows" and detecting potential mechanical weakness zones. The study assisted in identifying where these conditions occurred and followed suit by suggesting the optimization of injection pressure, reinforcing the integrity of the caprock for the sequestration of CO_2 in the subsurface.

References

[1] J.-Q. Shi, C. Sinayuc, S. Durucan, and A. Korre, "Assessment of carbon dioxide plume behaviour within the storage reservoir and the lower caprock around the KB-502 injection well at In Salah," *International Journal of Greenhouse Gas Control*, Vol. 7, pp. 115-126, March 2012. https://doi.org/10.1016/j.ijggc.2012.01.002

[2] P. Horsrud, "Estimating Mechanical Properties of Shale from Empirical Correlations," *SPE Drilling & Completion*, Vol. 16, No. 02, pp. 68–73, June 2001. https://doi.org/10.2118/56017-PA

[3] H. Vo Thanh, Q. Yasin, W. J. Al-Mudhafar, and K.-K. Lee, "Knowledge-based machine learning techniques for accurate prediction of CO2 storage performance in underground saline aquifers," *Applied Energy*, Vol. 314, Art. No. 118985, May 2022. https://doi.org/10.1016/j.apenergy.2022.118985

[4] Y. Song, S. Jun, Y. Na, K. Kim, Y. Jang, and J. Wang, "Geomechanical challenges during geological CO2 storage: A review," *Chemical Engineering Journal*, Vol. 452, Art. No. 140968, Jan. 2023. https://doi.org/10.1016/j.cej.2022.140968

[5] S. M. Frailey, "Methods for estimating CO2 storage in saline reservoirs," *Energy Procedia*, Vol. 1, No. 1, pp. 2769–2776, Feb. 2009. https://doi.org/10.1016/j.egypro.2009.02.048

[6] C. Chang, M. D. Zoback, and A. Khaksar, "Empirical relations between rock strength and physical properties in sedimentary rocks," *Journal of Petroleum Science and Engineering*, Vol. 51, No. 3–4, pp. 223–237, May 2006. https://doi.org/10.1016/j.petrol.2006.01.003

[7] M. Issa, M. Issa, A. Abdulkareem, and F. Hadi, "Review of Mechanical Rock Properties Measurement and Estimation Techniques," *International Journal of Analytical Experimental and Finite Element Analysis (IJAEFEA)*, Vol. 9, pp. 43–49, 2022.

[8] W. J. Al-Mudhafar, "Integrating well log interpretations for lithofacies classification and permeability modeling through advanced machine learning algorithms," *Journal of Petroleum Exploration and Production Technology*, Vol. 7, pp. 1023–1033, 2017. https://doi.org/10.1007/s13202-017-0360-0

[9] R. Zhong, M. Tsang, G. Makusha, B. Yang, and Z. Chen, "Improving rock mechanical properties estimation using machine learning," in *Proc. Resource Operators Conf. (ROC)*, University of Wollongong, 2025. https://doi.org/10.71747/uow-r3gk326m.c.7651142.v1

[10] G. E. Karniadakis et al., "Physics-informed machine learning," *Nature Reviews Physics*, Vol. 3, No. 6, pp. 422–441, June 2021. https://doi.org/10.1038/s42254-021-00314-5

[11] A. Latrach, M. L. Malki, M. Morales, M. Mehana, and M. Rabiei, "A critical review of physics-informed machine learning applications in subsurface energy systems," *Geoenergy*, Vol. 3, Art. No. 212938, 2024. https://doi.org/10.1016/j.geoen.2024.212938

[12] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: Unbiased boosting with categorical features," in *Advances in Neural Information Processing Systems*, Vol. 31, 2018, pp. 6639–6649. https://doi.org/10.48550/arXiv.1706.09516

[13] B. Yuan et al., "Physics-informed machine learning in geotechnical engineering: a direction paper," *Geomechanics and Geoengineering*, Vol. 20, No. 5, pp. 1128–1159, 2025. https://doi.org/10.1080/17486025.2025.2502029

[14] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "LightGBM: A highly efficient gradient boosting decision tree," in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., Vol. 30, pp. 3149–3157, 2017.

[15] A. K. Khassaf, Z. M. Al-hameed, N. R. Al-Mohammedawi, W. J. Al-Mudhafar, D. A. Wood, M. A. Abbas, O. Ameur-Zaimeche, and A. A. Alsubaih, "Physics-Informed Machine Learning for Enhanced Permeability Prediction in Heterogeneous Carbonate Reservoirs," in *Proceeding. Offshore Technology Conference*, Houston, TX, USA, May 2025, Paper No. OTC-35892-MS. https://doi.org/10.4043/35892-MS

[16] N. Al-Kanaani, "Geomechanical Properties Characterization and Stability Analysis of the Zubair Formation in the Zubair Oilfield Using Well Logs and Borehole Data," *Iraqi Geological Journal*, Vol. 58, No. 2E, Art. No. 6, 2025. https://doi.org/10.46717/igj.2025.58.2E.6

[17] A. Peter, X. Jin, X. Fan, K. I. Eshiet, Y. Sheng, and D. Yang, "Effect of CO2 Phase on Pore Geometry of Saline Reservoir Rock," *Rock Mechanics and Rock Engineering*, Vol. 55, No. 4, pp. 1907–1930, Apr. 2022. https://doi.org/10.1007/s00603-021-02658-x

[18] M. A. Abbas, W. J. Al-Mudhafar, A. Anees, and D. A. Wood, "Integrating petrophysical data into efficient iterative cluster analysis for electrofacies identification in clastic reservoirs," *Energy Geoscience*, Vol. 5, No. 1, Art. No. 100341, Jan. 2024. https://doi.org/10.1016/j.engeos.2024.100341

[19] E. Fjær, R. M. Holt, P. Horsrud, A. M. Raaen, and R. Risnes, *Petroleum related rock mechanics*, 2nd ed., Vol. 53. Amsterdam, Netherlands: Elsevier, 2008. https://doi.org/10.1016/s0376-7361(07)53002-5

[20] Y. Rimal, N. Sharma, and A. Alsadoon, "The accuracy of machine learning models relies on hyperparameter tuning: student result classification using random forest, randomized search, grid search, bayesian, genetic, and optuna algorithms," *Multimedia Tools and Applications*, Vol. 83, pp. 74349–74364, 2024. https://doi.org/10.1007/s11042-024-18426-2

[21] S. Sen, A. A. Radwan, M. Leila, A. Abdelmaksoud, and M. Ali, "Geomechanical assessment of the Cenozoic stratigraphy of the Tui area, Taranaki Basin, offshore New Zealand – Implications for geological storage in the Paleocene and Eocene sandstones," *International Journal of Greenhouse Gas Control*, Vol. 136, Art. No. 104194, 2024. https://doi.org/10.1016/j.ijggc.2024.104194

[22] A. A. Basheer and M. N. Hajmeer, "Artificial neural networks: fundamentals, computing, design, and application," *Journal of Microbiological Methods*, Vol. 43, No. 1, pp. 3–31, Dec. 2000. https://doi.org/10.1016/s0167-7012(00)00201-3

[23] P. Cosenza, M. Ghoreychi, G. de Marsily, G. Vasseur, and S. Violette, "Theoretical prediction of poroelastic properties of argillaceous rocks from in situ specific storage coefficient," *Water Resources Research*, Vol. 38, No. 11, pp. 1256–1266, Nov. 2002. https://doi.org/10.1029/2001WR001201

[24] M. Ghafoori, A. Rastegarnia, and G. R. Lashkaripour, "Estimation of static parameters based on dynamical and physical properties in limestone rocks," *Journal of African Earth Sciences*, Vol. 137, pp. 22–31, Jan. 2018. https://doi.org/10.1016/j.jafrearsci.2017.09.008

[25] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery, 2016, pp. 785–794. https://doi.org/10.1145/2939672.2939785

[26] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Ann. Statist.*, Vol. 29, No. 5, pp. 1189–1232, Oct. 2001. https://doi.org/10.1214/aos/1013203451

[27] L. Breiman, "Random Forests," *Machine Learning*, Vol. 45, No. 1, pp. 5–32, Oct. 2001. https://doi.org/10.1023/A:1010933404324

[28] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, Vol. 521, No. 7553, pp. 436–444, May 2015. https://doi.org/10.1038/nature14539

[29] E. Fjær, "Relations between static and dynamic moduli of sedimentary rocks," *Geophysical Prospecting*, Vol. 67, No. 1, pp. 128–139, Jan. 2019. https://doi.org/10.1111/1365-2478.12711

[30] Schlumberger Limited, "2015 Annual Report," Houston, TX, USA, Rep. SLB-AR15, 2015.

[31] J. P. Castagna, M. L. Batzle, and R. L. Eastwood, "Relationships between compressional-wave and shear-wave velocities in clastic silicate rocks," *Geophysics*, Vol. 50, No. 4, pp. 571–581, Apr. 1985. https://doi.org/10.1190/1.1441933

[32] A. R. Najibi, M. Ghafoori, G. R. Lashkaripour, and M. R. Asef, "Reservoir geomechanical modeling: In-situ stress, pore pressure, and mud design," *Journal of Petroleum Science and Engineering*,Vol. 151, pp. 31–39, Mar. 2017. https://doi.org/10.1016/j.petrol.2017.01.045

[33] J. H. Friedman, "Stochastic gradient boosting," *Computational Statistics & Data Analysis*, Vol. 38, No. 4, pp. 367–378, 2002. https://doi.org/10.1016/S0167-9473(01)00065-2

التنبؤ بالمقاومة الانضغاطية أحادية المحور للصخور لتخزين ثنائي أكسيد الكربون في التكوينات تحت السطحية باستخدام تعلم الآلة المستند إلى الفيزياء: تقييم مقارن للأداء

فاطمة فلاح مطشر [١،*]، حسن عبد الهادي عبد الحسين [١]

١ قسم هندسة النفط، كلية الهندسة، جامعة بغداد، بغداد، العراق

الخلاصة

تقدم هذه الورقة إطار عمل متطوراً لهندسة الميزات المعززة بالفيزياء للتنبؤ بخاصية جيوميكانيكية حرجة، وهي المقاومة الانضغاطية أحادية المحور (UCS) ، في مكمن الزبير بحقل الزبير لتقليل مخاطر تخزين ثنائي أكسيد الكربون. تم الحصول على النتائج الجيوميكانيكية وسير عمل النمذجة عبر نظام برمجيات متكامل تم تطويره بلغة بايثون(Python) ، بينما أُجريت جميع العمليات الحسابية وتدريب النماذج في بيئة Deepnote للحوسبة السحابية التفاعلية؛ مما أتاح دمج جميع المكتبات اللازمة وقدرات المعالجة عالية الأداء بسلاسة. ومن خلال مقارنة ست خوارزميات تجميعية وهي:

Categorical Boosting (CatBoost), XGBoost (Extreme Gradient Boosting), RF(Random Forest Regressor), ExtraTrees (Extremely Randomized Trees), GBM (Gradient Boosting Machine), and Deep ANN (Deep Artificial Neural Networks),

حدد البحث النماذج الأكثر موثوقية، حيث برز نموذج Gradient Boosting بتحقيقه أفضل مؤشرات التنبؤ بـ UCS بمقاييس (R^2=0.9995, MAE=3.64) ولا توفر هذه النماذج تقييماً دقيقاً وموثوقاً فحسب، بل تُعد أيضاً أداة قابلة للتطوير لتقييم الاستقرار تحت السطحي، وسلامة الاحتواء، والمراقبة الفورية للمكمن.

الكلمات الدالة: تعلم الآلة المستند إلى الفيزياء، تعلم الآلة، المقاومة الانضغاطية أحادية المحور، حقل الزبير، الخاصية الجيوميكانيكية.