Kharazmi University

Journal of Engineering Geology

Journal home page https://jeg.khu.ac.ir

OPEN ACCESS

Online ISSN 2981-1600

Uncertainty quantification of ucs prediction in sedimentary rocks using petrographic features: a machine learning–monte carlo approach

Seyyed Mahmoud Fatemi Aghda[1]✉, Asieh Hamidi[2], Fatemeh Amiri[3]

1. Professor, Applied Geology Group, Department of Applied Geology, University of Kharazmi, Tehran, Iran. E-mail: fatemi@khu.ac.ir
2. Postdoctoral Researcher, Department of Civil and Mining Engineering, University of Toronto, Canada. E-mail: asieh.hamidi@utoronto.ca
3. M.Sc, Applied Geology Group, Department of Applied Geology, University of Kharazmi, Tehran, Iran. E-mail: fatemeh.amiri@khu.ac.ir

Article Info

Article type:
Research Article

Article history:
Received 18 September 2025
Accepted 17 November 2025

Keywords:
Uniaxial compressive strength, Machine learning, Petrographic features, Uncertainty quantification, Monte Carlo simulation, Rock engineering, Iranian coastal rocks.

ABSTRACT

The evaluation of mechanical strength, particularly uniaxial compressive strength (UCS) of rocks, is critical for the design and performance prediction of surface and underground structures, impacting project costs and safety. Traditional UCS testing methods are destructive, time-consuming, and costly, while indirect methods often lack reliability due to rock heterogeneity. This study addresses these limitations by developing advanced machine learning frameworks that integrate petrographic features with conventional rock properties to predict UCS and quantify uncertainties. A comprehensive dataset from sedimentary rocks along Iran's southern coasts (Persian Gulf and Gulf of Oman) was used, including mechanical properties (UCS, Brazilian tensile strength, point load index, porosity, ultrasonic pulse velocity), durability indices (Los Angeles abrasion, slake durability, aggregate impact value), and petrographic characteristics derived from thin-section analysis. Three complementary approaches were applied: (1) hybrid Neural Network-Gradient Boosting regression (ANN-GBR), (2) AutoML-optimized Random Forest, and (3) Monte Carlo simulation for uncertainty quantification. Results showed that the AutoML-optimized Random Forest model demonstrated exceptional predictive performance with $R^2 = 0.9884$, RMSE = 0.5732 MPa, and MAPE = 3.6%, significantly outperforming traditional empirical methods. The ANN-GBR hybrid approach achieved $R^2 = 0.9412$ with RMSE = 1.385 MPa. Monte Carlo simulations provided robust probabilistic assessments with 95% confidence intervals and systematic bias identification. Feature importance analysis revealed that soundness parameters and mineralogical composition were the most influential predictors. The developed framework offers significant practical benefits, including reduced laboratory testing costs, faster prediction for quality control, and enhanced risk assessment through uncertainty quantification, providing a robust and cost-effective approach for rock strength assessment.

Introduction

The Uniaxial (unconfined) compressive strength (UCS) represents a fundamental mechanical behaviour of rock materials and plays a critical role in the design and performance of large-scale infrastructure projects (e.g., dams, tunnels, breakwaters, etc.). UCS, as a comprehensive measure of the strength, is extensively applied in the engineering rock mass classification schemes (RMR, Q, etc.), and analysis of structural stability for geotechnical engineering design

Cite this article: Fatemi Aghda, S. M., Hamidi, A., Amiri, F., (2025). Uncertainty quantification of ucs prediction in sedimentary rocks using petrographic features: a machine learning–monte carlo approach. Journal of Engineering Geology, 19 (5), 648-676. https//doi.org/10.22034/JEG.2025.19.5.1020471

Publisher: Kharazmi University.
DOI: https//doi.org/10.22034/JEG.2025.19.5.1020471

[DOR: 20.1001.1.22286837.1404.19.5.2.3]

[DOI: 10.22034/JEG.2025.19.5.1020471]

applications (Siegesmund and Török, 2014; Xue et al., 2021; Kong et al., 2023; Alizadeh et al., 2023; Xie et al., 2025a). Practical experiments mainly measure UCS according to the international guidelines established by ISRM and ASTM (ISRM, 1978; Standard, 2008). However, the direct measurement methods are destructive, time- and cost-consuming, and require high-quality rock core samples. In soft and weak rocks, due to the fragility of the specimens, conducting tests is highly challenging or impossible; while in hard rocks, testing becomes difficult and time-intensive (Aliyu et al., 2019; Hemmati et al., 2020; Azadmehr et al., 2024). Due to the importance of UCS in engineering projects and the limitations of direct assessment, indirect methods such as the point load test (PLT), Brazilian tensile strength (BTS) test, Schmidt Hammer, and ultrasonic wave velocity measurements are index tests for evaluation (Aladejare et al., 2024; Narimani and Vásárhelyi, 2025). Although index tests are useful for estimating UCS, they are not always reliable due to inherent heterogeneity of rock materials (Kong et al., 2024). Recent studies have introduced the use of petrographic characteristics as an effective, rapid, and low-cost approach for predicting mechanical parameters of rock. Petrographic characteristics, such as mineral composition, texture, and cement types strongly influence the mechanical behavior and UCS of rocks. These features, identifiable through petrographic analysis, have been shown in numerous studies to correlate significantly with the mechanical properties of rocks. (Petrounis et al., 2018; Fereidooni, 2022). Anisotropic rocks display varying mechanical behavior under stress due to structural discontinuities and fabric heterogeneities, including cleavage, bedding, schistosity, grain size distribution, and porosity. This anisotropy significantly affects rock strength and deformation behavior. (Askaripour et al., 2022; Daoud et al., 2025). Sandstone, as one of the most common sedimentary rocks, plays a vital role in engineering applications due to its abundance and distinct physical properties. Its mechanical behavior is largely governed by variations in texture and internal structure, making petrographic analysis essential for understanding its strength, deformation behavior, and failure mechanisms (Kamenev et al., 2021; Khajevand, 2023; Guo et al., 2024); thus, understanding of the complex and nonlinear behavior of rocks originating from geological processes and their internal structure is essential for reliable estimates in rock engineering design (Hamidi et al., 2024).

1.1. Petrographic Control on Rock Strength Behavior

The mechanical behavior of sedimentary rocks is fundamentally controlled by three primary petrographic characteristics: texture, mineralogical composition, and microstructural properties. Textural maturity in sandstones directly influences UCS through grain arrangement and matrix content. Immature textures, characterized by angular grains, poor sorting, and >5% clay matrix, exhibit reduced strength due to ineffective grain contacts and weak matrix materials. Conversely, mature textures with well-sorted, rounded grains and minimal clay matrix demonstrate enhanced load distribution and higher compressive strength (Folk, 1954). Porosity and sorting relationships critically affect mechanical performance through their control on particle contact efficiency. While increased porosity typically correlates with reduced UCS due to diminished particle cohesion, well-sorted aggregates can maintain structural integrity through uniform load distribution despite higher void ratios. Poorly sorted rocks, although potentially exhibiting lower porosity through grain size variation, suffer from stress concentration effects that compromise overall strength (Zhao et al., 2023; Jia., 2025). Mineralogical composition creates heterogeneous mechanical responses due to varying mineral properties. Hard minerals such as quartz enhance load-bearing capacity and structural integrity, while soft minerals like

calcite introduce weak zones susceptible to brittle failure under stress concentration. The presence of fossils and intraclasts further compromises matrix continuity, creating preferential failure pathways that reduce overall rock strength. This mineralogical heterogeneity necessitates comprehensive petrographic analysis for accurate strength prediction in engineering applications (Khan et al., 2023; Stopka, 2024). Customary approaches such as empirical and statistical methods fall short in accurately prediction estimates and quantifying the uncertainties thus unreliable engineering decision-making. The inability to properly quantify such uncertainties may lead to instability in rock structures and, ultimately, failure in engineering projects (Pérez-Díaz et al., 2020; Yeh et al., 2021). Recent advancements in ML- and AI-based techniques offer powerful tools (Wang et al., 2023) for accurately modeling nonlinear relationships between parameters relevant to petrographic features and UCS of rocks. Table 1 summarizes a brief finding from these studies, emphasizing that such AI based approaches not only improve prediction accuracy, but they also enable effective uncertainty quantification, a key step toward more reliable, informed decision-making in geotechnical engineering design.

Table 1. Summary of recent research on predicting UCS (as the output) using ML techniqueS

References	Input for predictions	Applied models
Singh et al. (2001)	Mineralogical composition, Grain size, Aspect ratio, Form factor, Area weighting, Orientation of foliation planes (planes of weakness)	ANN
Gokceoglu. (2002)	Black andesite, Pink andesite, Tuff	Multivariate regression techniques, Fuzzy logic and neural network approaches
Gokceoglu and Zorlu. (2004)	Unit weight, Iv by weight, Apparent porosity, Vp, Rn, Block punch index, PLT, Tensile strength	Multiple regression, FIS
Zorlu et al. (2008)	Quartz content, Packing density, Concavo–convex type grain contact	Multiple regression, ANN
Çanakcı et al. (2009)	Ultrasound pulse velocity, Iv, ρ, Saturated density, Bulk density	ANN, GEP
Dehghan et al. (2010)	Vp, PLT, Rn, porosity	ANN
Jahanbakhshi et al. (2011)	Vp, Density, porosity	ANN
Ceryan et al. (2012)	Total porosity, Effective porosity, Slake durability index, Vp in dry samples and in the solid part of samples	GRNNs, Feed-forward back-propagation algorithm-based neural networks
Yesiloglu-Gultekin et al. (2013)	Quartz, Plagioclase, Orthoclase	Five Nonlinear Multiple Regressions, ANFIS
Ali et al. (2014)	Microfabric properties including Grain size, Shape factor, Quartz content.	ANN, FIS, MR
Liu et al. (2015)	Calcite, Clay, Quartz, Opaque minerals, Biotite, Specific density, Dry unit weight, Total porosity, Effective porosity, Slake durability index (fourth cycle), Vp	ELM, SVM, ANN

[DOR: 20.1001.1.22286837.1404.19.5.2.3]
[DOI: 10.22034/JEG.2025.19.5.1020471]

Madhubabu et al. (2016)	Porosity, Density, Vp, υ, PLT	MLRA, ANN
Sharma et al. (2017)	Ultrasonic P-wave velocity, Density, Slake durability index	ANFIS, MLR, ANN
Heidari et al. (2018)	Porosity, Density, Vp, υ , PLT	MLR, ANN
Asheghi et al. (2019)	Density, Porosity, Vp, PLT, Iv	GFFN incorporated with ICA
Barzegar et al. (2020)	Vp, Rn, Porosity, PLT	RF, M5 Model Tree, MARS, ANN-committee
Saedi and Mohammadi.(2021)	Mineral area, Orientation, Concavity, Index of interlocking, Feldspar content, Saturation	ANN
Yang et al. (2022)	Diameter, Grain size, Bulk density, Vp, Strain rate, SCS.	ELM, RF, SVR, PSO-SVR
Afolagboye et al. (2023)	ACV, AIV, LAAV, Rn, PLT	RF, RVM, SVM, ANN
Amiri et al. (2024)	ρ, Iv, Vp, PLT, BTS	DNN
Sun et al. (2024)	PLT, Vp, porosity	AOA-XGB
Hussain et al. (2024)	Calcite, Dolomite, Clay, Pyrite, Quartz, Feldspar, Large Porosities, Fossils, Lithics	RF, GB, MLP, CB
Xie et al. (2025b)	Thrust Force, Torque, Drilling Speed, Rotational Speed.	MLP, SVR, CNN, RT, LSTM
Swamy et al. (2025)	Number of ball, Grinding media weight, Grind duration, Mill volume fraction occupied by sample charge, Mill volume fraction by ball charge, Interstitial filling ratio, Charge ratio, Mill filling, Representative particle sizes at which 10%, 50% and 90% of the particles by weight are finer	MLR, KNN , SVR, RFR

Table notes: ANN: artificial neural networks. FIS: fuzzy inference systems. GEP: gene expression programming. GRNNs: generalized regression neural networks. ANFIS: adaptive neuro-fuzzy inference system. MR : multivariate regression. ELM: extreme learning machine, SVM: support vector machines. MLRA: multiple linear regression analysis. MLR: multiple linear regression. GFFN: generalized feedforward neural network. ICA: imperialist competitive algorithm. MARS: multivariate adaptive regression splines. RF: random forest. PSO-SVR: aparticle swarm optimization- support vector regression. RVM: relevance vector machine. DNN: deep neural networks. AOA-XGB: archimedes optimization algorithm - extreme gradient boosting. CB: categorical boosting. GB: gradient boosting. MLP: multi-layer perceptron. SVR: support vector regression. CNN: convolutional neural networks. RT: random trees. LSTM: long short-term memory networks. RFR: random forest regression. KNNR: k-nearest neighbor regression. Vp: p-wave velocity. PLT: point load test. Rn: schmidt hammer rebound number. υ: Poisson's ratio. SCS: Static compressive strength. ACV: Aggregate crushing value. AIV: aggregate impact value. LAAV: los angeles aggregate value. ρ: dry density. BTS: brazilian tensile strength. Iv: water absorption.

In this context, modeling is essential for analyzing the relationships between petrographic characteristics and rock mechanical strength (Morgenroth et al., 2019). Given the nonlinear and heterogeneous nature of rock materials, traditional methods face limitations. Therefore, this paper develops a robust predictive framework to address the challenges of accurately assessing UCS, using petrographic characteristics as inputs for advanced machine learning (ML) and artificial intelligence (AI) algorithms.

2. Geology of the study

The southern coasts of Iran, where materials for the breakwaters were sourced, lie within three main geological zones: the Folded Zagros (Bushehr and Hormozgan) and the Makran (Sistan and Baluchestan). The coastal zone of Bushehr lies within the Folded Zagros belt, characterized by NW–SE trending folds and Tertiary formations. The dominant

[DOR: 20.1001.1.22286837.1404.19.5.2.3]
[DOI: 10.22034/JEG.2025.19.5.1020471]

[DOR: 20.1001.1.22286837.1404.19.5.2.3]

[DOI: 10.22034/JEG.2025.19.5.1020471]

lithostratigraphic units belong to the Fars Group (Gachsaran, Mishan, and Aghajari formations of Miocene age) and the Bakhtiari Formation (Miocene–Pliocene). Regional structures such as folds, faults, and drainage networks align with the Zagros structural trend. Hormozgan coasts are strongly influenced by diapirism, with numerous Infracambrian salt domes bringing Hormoz Formation deposits and associated igneous rocks (intrusive and extrusive) to the surface. These diapiric structures have played a key role in island formation (e.g., Qeshm, Hormoz, Greater and Lesser Tunbs). Along the coastal outcrops, the Fars Group formations and the Bakhtiari Formation are widespread, with limited occurrences of younger local units (e.g., Qeshm Formation). East of Minab, the coast transitions into the Makran zone. The coastal belt of Sistan and Baluchestan belongs to the Makran zone, extending eastward along the Oman Sea. This zone is tectonically active, with open anticlines, synclines, and faulted cliffs (mainly E–W trending). The stratigraphy comprises Tertiary–Quaternary units: marls, silty marls, bioclastic limestones, sandstones, and interbedded breccias and conglomerates, with widespread Quaternary alluvium. Coastal geomorphology includes uplifted marine terraces, mud volcanoes, and erosional badlands, particularly around Chabahar. The lithological diversity across these provinces provides a range of sedimentary and igneous rock types that have been utilized for the construction of coastal breakwaters in the region. The study area data is presented in Figure 1.

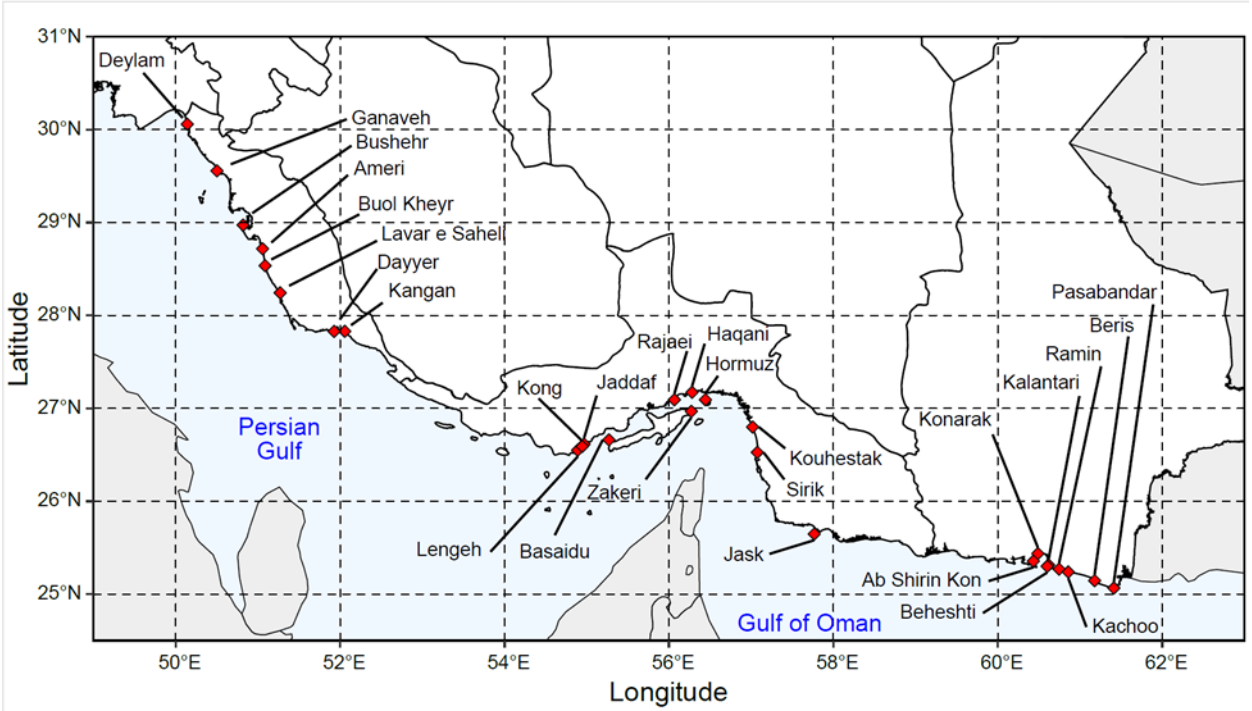

Fig. 1. The distribution map of studied breakwaters across southern coastline, Iran. (Hamidi et al., 2024)

3. Methodology

In this research, a subset of experimental data obtained through the sampling and laboratory measurements conducted by (Hamidi, 2024; Hamidi et al., 2024) have been used, including sedimentary rock samples from the southern coastlines. The following sections describe further information about data acquisitions and properties, and the employed methods for

analysis purposes. Figure 2 is a developed flowchart illustrating data analyses milestone.

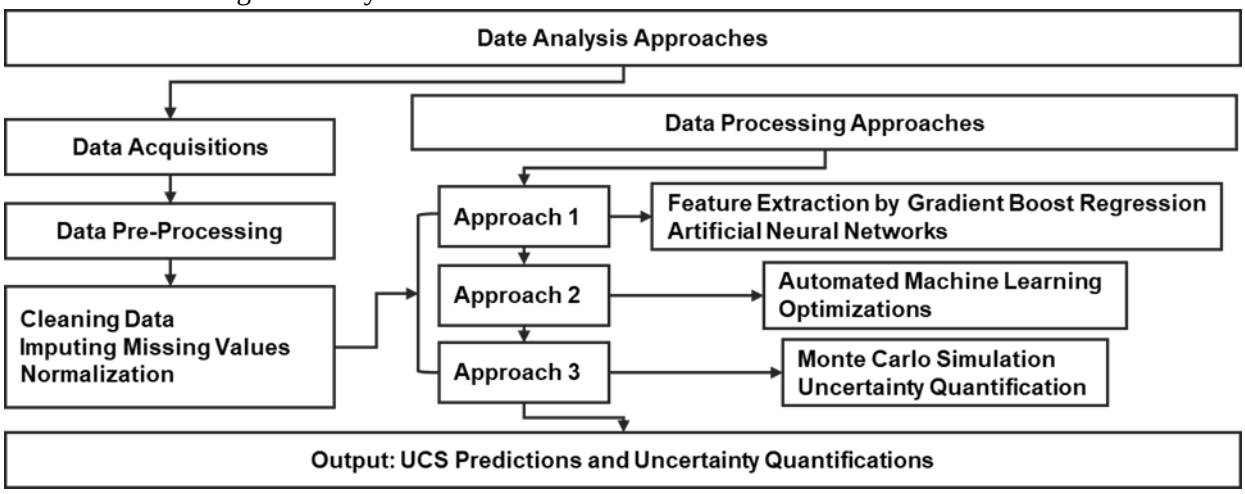

Fig. 2. Flowchart of the applied methods through data analysis.

Data Acquisition and Description

The dataset employed in this study encompasses a comprehensive range of rock mechanical and physical properties obtained from sedimentary rock samples (Hamidi et al., 2024). The statistical summary of the sedimentary rock dataset is presented in Table 2.

Table 2. Statistical summary of the sedimentary rock dataset

No.	Parameter	Units	Count	Max	Min	Mean	Std Dev	Missing values
1	UCS	MPa	29	30.210	0.889	10.303	5.480	7
2	PLT	MPa	17	3.498	0.981	1.914	0.750	19
3	BTS	MPa	14	5.442	1.536	2.865	1.339	22
4	Vp	m/s	20	6595.745	2653.670	3879.695	1096.165	16
5	E	%	21	20.766	3.429	11.773	4.617	15
6	Γ	g/cm³	21	2.291	1.714	1.927	0.187	15
7	SN05	%	20	52.087	0.751	14.039	17.113	16
8	SN10	%	20	61.410	3.870	25.244	18.119	16
9	SN15	%	21	81.360	4.289	35.089	21.345	15
10	ID05	%	36	96.135	73.962	84.957	7.296	0
11	ID10	%	36	93.543	66.666	78.553	8.780	0
12	ID15	%	36	91.862	62.417	71.951	9.991	0
13	LAA	%	36	77.910	69.200	75.019	4.096	0
14	AIV	%	11	65.664	34.567	47.241	10.705	25
15	ACV	%	11	81.155	36.139	64.059	19.345	25

Table notes: UCS: unconfined compressive strength. PLT: point load strength. BTS: brazilian tensile strength. AIV: aggregate impact value. ACV: aggregate crushing value. LAA: los angeles abrasion. e: porosity. Vp: P- wave ultrasonic velocity; γ: bulk density; SN05, SN10, SN15: soundness at 5, 10, and 15 cycles, respectively; ID05, ID10, ID15: durability at 5, 10, and 15 cycles, respectively.

[DOR: 20.1001.1.22286837.1404.19.5.2.3]
[DOI: 10.22034/JEG.2025.19.5.1020471]

[DOR: 20.1001.1.22286837.1404.19.5.2.3]

[DOI: 10.22034/JEG.2025.19.5.1020471]

Data Analysis Framework

The primary objective of this study is to predict the uncertainty of UCS of rocks using advanced data-driven approaches, specifically machine learning (ML) and artificial neural networks (ANNs). Three complementary frameworks were implemented: (i) machine learning models, (ii) artificial neural networks with feature fusion, and (iii) gradient boosting regression with uncertainty quantification via Monte Carlo simulation (MCS). A total of 36 rock samples were collected and tested; this number forms the basis of the final dataset (equal to the UCS count). Due to physical limitations and sample fragility, not all mechanical and durability tests could be conducted on every specimen. Consequently, the number of available measurements varies among variables, as reported in Table 2. Missing values were handled using the Iterative Imputer with a Random Forest estimator, ensuring consistent data integration across all variables. A complete list of all samples, variable values, and missing entries is provided in Supplementary Table 3.

Table 3. Summary of all samples and variables used in modeling

Sample ID	UCS (MPa)	PLT (Mpa)	BTS (Mpa)	Vp (m/s)	E (%)	Γ (g/cm³)	SN05 (%)	SN10 (%)	SN15 (%)	ID05 (%)	ID10 (%)	ID15 (%)	LAA (%)	AIV (%)	ACV (%)	Texture1	Texture2	Mineral1	Mineral2
1	30.21	N/A	N/A	N/A	N/A	N/A	1.92	3.87	4.29	91.17	89.73	88.21	N/A	N/A	N/A	25.62	10	63.44	24.66
2	14.28	3.14	5.44	N/A	3.47	2.20	0.75	6.08	10.14	N/A	N/A	N/A	N/A	N/A	N/A	15	52.71	46.77	10
3	14.28	3.13	4.95	4565/66	3.44	2.18	0.75	6.08	10.14	N/A	N/A	N/A	N/A	N/A	N/A	15	53.89	48.92	10
4	14.28	N/A	N/A	4251/82	3.43	2.29	0.75	6.08	10.14	N/A	N/A	N/A	N/A	N/A	N/A	15	45.7	45.9	10
5	21.53	N/A	N/A	4649/000	N/A	N/A	0.75	6.08	10.14	N/A	N/A	N/A	N/A	N/A	N/A	15	54.11	51.51	10
6	14.28	N/A	N/A	N/A	N/A	N/A	0.75	6.08	10.14	N/A	N/A	N/A	N/A	N/A	N/A	15	47.56	52.77	10
7	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	34.26	10	20	10
8	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	72.20	65.66	76.06	30.98	10	50.43	24.88
9	9.31	N/A	N/A	4120/97	N/A	N/A	8.16	27.42	37.47	N/A	N/A	N/A	N/A	N/A	N/A	28.98	10	61.57	10
10	9.50	N/A	N/A	N/A	N/A	N/A	8.16	27.42	37.47	N/A	N/A	N/A	N/A	N/A	N/A	22.34	10	52.14	10
11	9.50	2.31	2.23	N/A	N/A	N/A	8.16	27.42	37.47	N/A	N/A	N/A	N/A	N/A	N/A	22.34	10	50.62	10
12	9.50	1.45	1.54	N/A	8.97	2.00	8.16	27.42	37.47	N/A	N/A	N/A	N/A	N/A	N/A	20.87	10	55.85	10
13	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	32.99	10	47.82	30.94
14	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	96.13	93.54	91.86	N/A	N/A	N/A	29.02	10	61.04	29.56
15	5.12	1.26	1.73	6595/74	20.77	2.12	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	30.62	10	46.49	33.31
16	5.12	N/A	N/A	2653/67	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	20.31	10	64.74	27.08
17	5.12	0.98	N/A	3051/95	15.71	1.85	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	34.55	10	60.44	21.79
18	5.12	N/A	N/A	2985/00	15.45	1.85	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	32.49	10	48.97	30.7
19	0.89	3.50	N/A	N/A	N/A	N/A	N/A	N/A	81.36	N/A	N/A	N/A	N/A	N/A	N/A	23.19	10	45.11	31.41
20	10.77	N/A	N/A	N/A	15.37	1.73	10.95	25.01	39.21	86.00	78.04	68.05	77.91	54.10	81.15	22.73	10	61.31	10
21	10.77	N/A	N/A	N/A	15.43	1.74	10.95	25.01	39.21	86.00	78.04	68.05	77.91	54.10	81.15	22.75	10	59.14	10
22	10.77	N/A	N/A	3829/46	N/A	N/A	10.95	25.01	39.21	86.00	78.04	68.05	77.91	54.10	81.15	24.56	10	59.58	10
23	10.77	1.23	1.99	3134/00	15.68	1.72	10.95	25.01	39.21	86.00	78.04	68.05	77.91	54.10	81.15	27.87	10	60.43	10
24	10.77	1.44	1.78	2869/67	15.86	1.73	10.95	25.01	39.21	86.00	78.04	68.05	77.91	54.10	81.15	26.48	10	46.48	10

25	10.77	1.20	2.70	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	24.37	10	20	10
26	10.77	N/A	5.25	6214/29	11.08	1.75	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	29.18	10	20	10
27	10.77	1.97	2.34	3334/00	13.35	1.83	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	22.09	10	20	10
28	5.48	1.96	2.56	N/A	11.77	1.91	15.73	25.82	29.34	90.14	84.60	77.91	69.20	34.57	36.14	24.38	10	52.17	10
29	5.48	2.07	2.98	3946/00	15.19	1.90	15.73	25.82	29.34	90.14	84.60	77.91	69.20	34.57	36.14	25.5	10	47.32	10
30	N/A	N/A	N/A	3291/00	7.74	2.15	52.09	61.41	65.30	73.96	66.67	62.42	N/A	38.11	50.18	26.84	10	20	10
31	N/A	1.71	N/A	3692/77	9.60	2.11	52.09	61.41	65.30	73.96	66.67	62.42	N/A	38.11	50.18	31.78	10	20	10
32	N/A	1.20	N/A	3940/40	9.60	2.08	52.09	61.41	65.30	73.96	66.67	62.42	N/A	38.11	50.18	23	10	20	10
33	8.87	1.83	1.97	4457/63	12.89	1.71	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	15	45.98	62.26	28.42
34	8.87	2.14	2.76	3169/01	10.94	1.81	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	15	59.23	57.47	31.56
35	8.87	N/A	N/A	2832/86	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	15	59.48	51.62	27.41
36	8.87	N/A	N/A	N/A	11.49	1.80	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	15	57.13	46.27	27.84

Table note: N/A indicates unmeasured values. Missing data were imputed using Iterative Imputer with a Random Forest estimator as described in Section 3.2.1.

[DOR: 20.1001.1.22286837.1404.19.5.2.3]
[DOI: 10.22034/JEG.2025.19.5.1020471]

[DOR: 20.1001.1.22286837.1404.19.5.2.3]

[DOI: 10.22034/JEG.2025.19.5.1020471]

Machine Learning Framework

ML, a subfield of AI, enables algorithms to learn patterns directly from data without explicit programming. In engineering contexts where relationships between variables are highly nonlinear, ML provides efficient, accurate, and cost-effective alternatives to traditional empirical and laboratory methods for UCS estimation (Atici et al., 2011; Rezaei et al., 2014; Hasanipanah et al., 2015; Mistry et al., 2024).

To establish a robust ML workflow, AutoML tools within PyCaret were employed to automate data preprocessing, model training, hyperparameter optimization, and performance benchmarking. This minimizes human bias, ensures consistency, and improves computational efficiency. The final modeling dataset consisted of 36 samples, where missing data were imputed using the Iterative Imputer (Random Forest estimator) as part of the preprocessing pipeline.

All preprocessing steps, including normalization, feature scaling, and imputation, were implemented within the PyCaret AutoML framework using built-in data pipelines. These operations were performed independently inside each cross-validation fold, ensuring no information leakage between training and validation subsets and preventing optimistic bias in model performance.

Model training and evaluation were performed using 10-fold cross-validation within the PyCaret AutoML pipeline. Each fold applied preprocessing (imputation, normalization, and scaling) independently, ensuring no data leakage between training and validation subsets. Given the limited dataset (36 samples), k-fold cross-validation was adopted instead of holding out an independent test set to maintain statistical robustness.

Data Preparation for ML approach

Two datasets were integrated: (i) **Laboratory dataset:** quantitative measurements of rock properties (Table 2); and (ii) **Petrographic dataset:** quantified descriptive features derived from thin section analysis, representing the presence of specific minerals, textures, and microstructural features.

The preprocessing petrographic dataset in this paper includes the following steps:

- Removal of non-informative columns (e.g., sample identifiers).
- Missing value imputation using Iterative Imputer with Random Forest as the estimator.
- Feature standardization to normalize the input scales.
- Data integration by combining laboratory and petrographic features.

Random Forest (RF) Algorithm and Implementation in PyCaret

RF, an ensemble algorithm, constructs multiple decision trees using bootstrap aggregation and random feature selection, with the final regression prediction obtained by averaging individual tree outputs. This method is particularly suitable for high-dimensional, nonlinear datasets due to its ability to reduce variance and mitigate overfitting (Ngo et al., 2025).

- Base estimators: PyCaret initializes RF with 10 trees (n_estimators = 10), differing from scikit-learn's default of 100.
- Hyperparameter tuning: Conducted via randomized grid search with 10 iterations (n_iter = 10), optimizing R^2 by default for regression.
- Safeguard: The choose_better=True parameter ensures that the tuned model replaces the base model only if performance improves.

Artificial Neural Networks

ANNs are computational models inspired by the human brain, widely applied for capturing complex nonlinear relationships in data (Azadeh et al., 2008; Rubo et al., 2019; Pradeep and Samui, 2022). Their strength lies in learning hidden patterns between multi-source inputs and target variables without explicit functional assumptions, making them suitable for UCS prediction (Sabri et al., 2024; Onyelowe et al., 2025).

[DOR: 20.1001.1.22286837.1404.19.5.2.3]

[DOI: 10.22034/JEG.2025.19.5.1020471]

Architecture

A multi-input feedforward ANN was developed to jointly process laboratory and petrographic features:

- Laboratory data branch: Input → Dense (32) with ReLU activation
- Petrographic data branch: Input → Dense (16) with ReLU activation
- Combined representation: Concatenation → Dense (16) with ReLU activation → Linear output

This ANN served as a feature extractor to generate latent representations, later utilized in ensemble learning.

Gradient Boosting Regression

GBR is an ensemble technique that sequentially combines weak learners (typically shallow decision trees) to iteratively minimize residual errors, thereby improving predictive accuracy. Unlike RF, which averages parallel trees, GBR builds trees sequentially, allowing fine-grained adjustment of predictions (Wijesundara et al., 2025).

Optimization:

The ANN-derived feature representations were passed to a GBR model, optimized using grid search cross-validation across:

- Number of estimators: [100, 200]
- Maximum depth: [3, 5]
- Learning rate: [0.01, 0.1]
- Subsample ratio: [0.8, 1.0]

Uncertainty Quantification via Monte Carlo Simulation (MCS)

To explicitly address input variability and parameter uncertainty, Monte Carlo simulations were applied. MCS generates probabilistic UCS distributions by sampling input variables repeatedly, thereby producing robust confidence bounds. This approach, widely used in geotechnical engineering (Hamidi and Harrison, 2025), enhances reliability compared to deterministic methods by integrating inherent variability into model outputs.

The implementation involved systematic noise injection, generating multiple noisy versions of the dataset by adding Gaussian noise proportional to each feature's standard deviation. For each feature x, noisy realizations were created using:

$X_{noisy} = x + N(0, \alpha \cdot \sigma_x)$

where $\sigma_x$is the standard deviation of $x$across all samples, $\alpha = 0.1$scales the noise magnitude (representing typical laboratory measurement uncertainty), and $N(0, \alpha \cdot \sigma_x)$is a Gaussian random sample with zero mean and standard deviation $\alpha \cdot \sigma_x$. Noise was applied independently to each feature for every sample, so each parameter, laboratory measurements and petrographic characteristics, received perturbations according to its variability. For each test sample, 100 noisy realizations were produced, and mean predictions, standard deviations, and 95% confidence intervals ($\pm 1.96\sigma$) were computed to provide robust uncertainty estimates.

The Evaluation of Model Convergence and Stability with Increased Iterations and Sensitivity Analysis

In this analysis, the number of simulation iterations with Gaussian noise was increased from 100 to 1000 to further investigate the model's convergence and stability. The results showed that even with the increased number of iterations, the model remained stable with minimal changes in the R^2 values, indicating its robustness. Additionally, a sensitivity analysis was conducted for different values of α (0.05, 0.1, 0.2) to assess whether the results remained consistent across varying levels of regularization. Furthermore, the distribution of predictions was examined, and the normality assumption was verified using the Shapiro-Wilk test, (Table 4) and Q-Q plots, (Figure 3). In this context, increasing the amount of noise led to better fulfillment of the normality condition by the model. In cases where the sample size was small, the introduction of noise initially caused more significant random fluctuations, as the

Shapiro-Wilk test is highly sensitive in small samples and reports even small deviations from normality. However, with the increase in noise and the number of iterations, the model's performance improved, resulting in more stable results, confirming that the distribution of predictions was normal. In other words, as the noise increased, the normality test showed higher p-values, indicating convergence, model stability, and normality. These comprehensive checks, including the increase in iteration count, sensitivity analysis, and noise adjustments, ensure that the model maintains high stability and reliability under different conditions.

Table 4. Normality test of predictions (Shapiro-Wilk) with varying numbers of simulations and levels of noisenoise.

n_simulations	noise_level= 0.1	noise_level= 0.2
100	0.1841	0.0882
1000	0.1841	0.0882

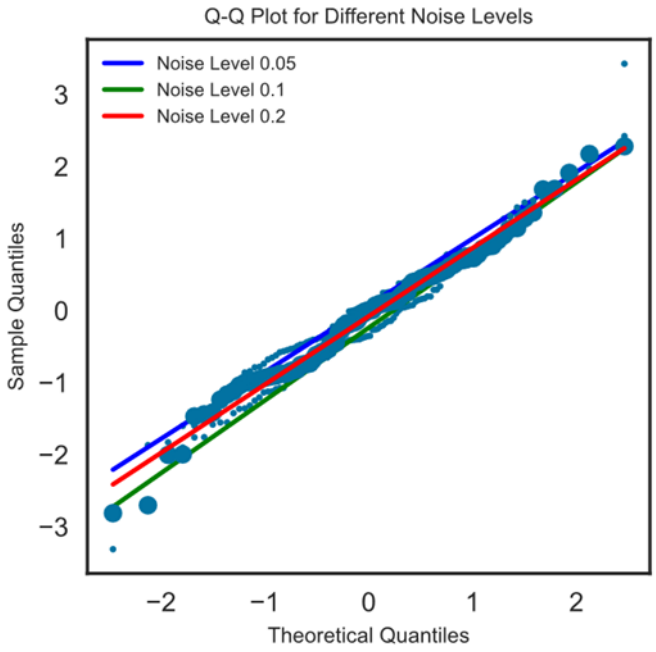

Fig. 3. Q-Q plots for monte carlo simulations with 100 and 1000 iterations at different noise levels

To evaluate the convergence of the model, 100 simulation iterations were performed, where Gaussian noise was added to the input features in each iteration to simulate the uncertainty of experimental measurements. The goal of this process was to examine whether the model converges to a stable value for the performance metric (R^2 in this case) and whether it demonstrates sufficient stability across multiple iterations. Convergence refers to reaching a steady state where there are no significant changes in the evaluation metrics after multiple iterations. In this study, R^2 was chosen as the performance metric to assess convergence. The results showed that after a few initial iterations, the R^2 value quickly stabilized and the variations minimized, indicating model stability. The convergence plot in Figure 4 confirms these results. The largest change in R^2 was less than 0.001, which signifies rapid and effective convergence of the model. This analysis confirms that the chosen number of iterations (100 iterations) was sufficient to stabilize the model's performance. Moreover, the convergence and stability process is particularly important in scenarios with noisy and uncertain

[DOR: 20.1001.1.22286837.1404.19.5.2.3]
[DOI: 10.22034/JEG.2025.19.5.1020471]

data, as it demonstrates that the model has learned efficiently and consistently from the data without overfitting or underfitting.

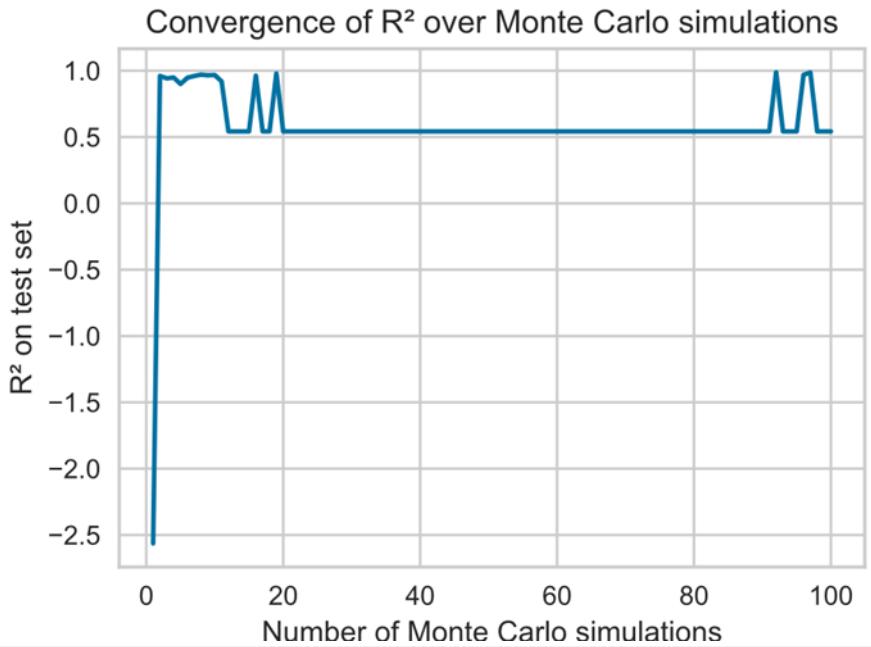

Fig. 4. Convergence plot of the model for the Monte Carlo Simulation with 100 iterations and 0.1 level noise.

Model Performance Evaluations

Model performance was evaluated using six complementary statistical metrics:

- **Mean Absolute Error (MAE):** average absolute error magnitude.
- **Mean Squared Error (MSE):** average squared prediction error, penalizing larger deviations.
- **Root Mean Squared Error (RMSE):** square root of MSE, representing standard prediction error.
- **Coefficient of Determination (R^2):** variance proportion explained by the model (0–1 scale).
- **Root Mean Squared Logarithmic Error (RMSLE):** error metric suitable for skewed or exponential data.
- **Mean Absolute Percentage Error (MAPE):** relative prediction error expressed as a percentage.

In the concept of frequentist statistics, the smaller error values (MAE, MSE, RMSE, RMSLE, MAPE) and higher R^2 indicate better predictive performance.

Results and Discussions

Dataset Characteristics

The Unconfined Compressive Strength (UCS), defined as the sole target variable in this study, is analyzed through a histogram that visualizes its distribution pattern and provides descriptive statistical insights into central tendency, variability, and data concentration (Figure 5). The histogram constructed from the uniaxial compressive strength data of limestone sandstones from southern Iran reveals a clear depiction of the distribution pattern of these mechanical properties. The data demonstrates a concentration of values predominantly within the 8–12 MPa range, indicating that a significant portion of the samples exhibit medium-strength characteristics. Notably, the presence of values both below 2 MPa and up to approximately 30 MPa suggests a relatively broad spectrum of compressive strengths, potentially reflecting structural or compositional heterogeneity within the samples. The overall shape of the histogram exhibits positive skewness, which implies that

[DOR: 20.1001.1.22286837.1404.19.5.2.3]
[DOI: 10.22034/JEG.2025.19.5.1020471]

while most samples have moderate strengths, a few exhibits higher compressive strengths (Armaghani et al., 2016).

Kernel Density Estimation Analysis

Kernel Density Estimation (KDE), a non-parametric method for estimating the probability density function of continuous random variables, was applied to complement the histogram analysis. In contrast to histograms, which rely on discrete binning and may be sensitive to bin width and placement, KDE generates a smooth, continuous curve that better represents the underlying distribution of data. The KDE curve revealed a dominant peak around 10–12 MPa, confirming that most samples exhibit medium-strength behavior. This method also detected finer distributional features such as skewness, mild multimodality, and local fluctuations in density that are often overlooked in traditional histograms. KDE thus serves as a powerful tool for gaining deeper statistical insight into the mechanical variability and potential heterogeneity of geological materials (Chen, 2017; Mokhtari, 2022).

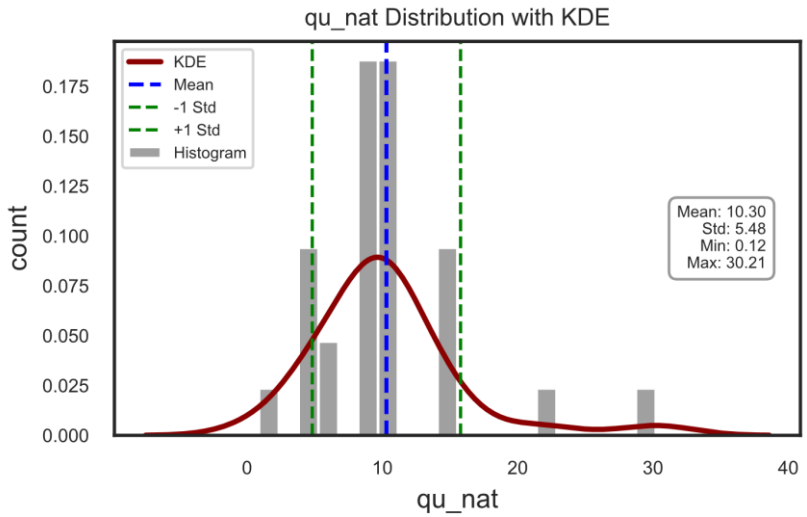

Fig. 5. Histogram of UCS as the target variable.

Petrographic Characterization

Petrographic analyses were conducted in accordance with ASTM C1721 (Hamidi, 2024) to identify the sedimentary and chemical composition of the rock aggregates. Thin sections were prepared and examined using a polarizing optical microscope to determine petrographic characteristics such as mineralogical composition, cement type, and textural features. A summary of the petrographic analysis, highlighting the key mineralogical and textural characteristics of the rock samples, is provided in Table 5.

Table 5. The summary of employed petrographical features.

1. No.	2. Parameter	3. Units	4. Count	5. Max	6. Min	7. Mean	8. Std Dev
9. 1	10. Texture1	11. %	12. 36	13. 34.55	14. 15.00	15. 23.75	16. 6.32
17. 2	18. Texture2	19. %	20. 36	21. 59.48	22. 10.00	23. 20.72	24. 19.00
25. 3	26. Mineral1	27. %	28. 36	29. 64.74	30. 20.00	31. 47.18	32. 14.69
33. 4	34. Mineral2	35. %	36. 36	37. 33.31	38. 10.00	39. 16.65	40. 9.18

Table note: Texture1: immature clastic texture. Texture2: mature clastic texture
Mineral1: Quartz mineral. Mineral2: Chert mineral.

[DOR: 20.1001.1.22286837.1404.19.5.2.3]
[DOI: 10.22034/JEG.2025.19.5.1020471]

[DOR: 20.1001.1.22286837.1404.19.5.2.3]

[DOI: 10.22034/JEG.2025.19.5.1020471]

AI-Based Models, Results and Performance

Neural Network-Gradient Boosting and Feature Extraction

The UCS prediction results using the Gradient Boosting algorithm are presented in Table 6. The hybrid Neural Network–Gradient Boosting model (ANN–GBR) demonstrated strong predictive performance with an RMSE of 1.385 MPa and an R^2 value of 0.9412.

Table 6. Gradient Boosting model performance in predicting UCS

41.	42.	43. GB
44. UCS	45. RMSE	46. 1.385
47.	48. R2	49. 0.9412

Gradient Boosting Algorithm For UCS Prediction

Figure 6 illustrates the relationship between the predicted UCS by the GB model and the actual UCS values, indicating the model performance. Each red point in the plot represents a data sample where the predicted UCS value is plotted against its corresponding actual value. In an ideal scenario, if the predictions were perfectly accurate, all the points would lie along the blue reference line ($y = x$), indicating a one-to-one match between the predictions and observations. The predictions and actual values are tightly clustered around the reference line, indicating a high degree of agreement between the predictions and the actual UCS values, which emphasizes the high accuracy and precision of the model. The majority of these points fall within the light blue area, representing a ±10% error margin, a range considered precise and acceptable for engineering predictions. From a quantitative perspective, the high coefficient of determination ($R^2 = 0.941$) significantly indicates that 94.1% of the variance in the UCS data can be explained by the GB model. This value is notably high, especially in an engineering database, which often contain nonlinear complexities, demonstrating the robustness and efficiency of the model. Additionally, the RMSE = 1.3857 shows that, on average, the model's predictions deviate from the actual UCS values by only 1.39 MPa. This level of error is considered low and acceptable, particularly in rock mechanics, where data typically exhibit complex variations and high scatter. The results statistically highlight the effectiveness of the hybrid Neural Network–Gradient Boosting model (ANN–GBR) in generating accurate and reliable UCS predictions.

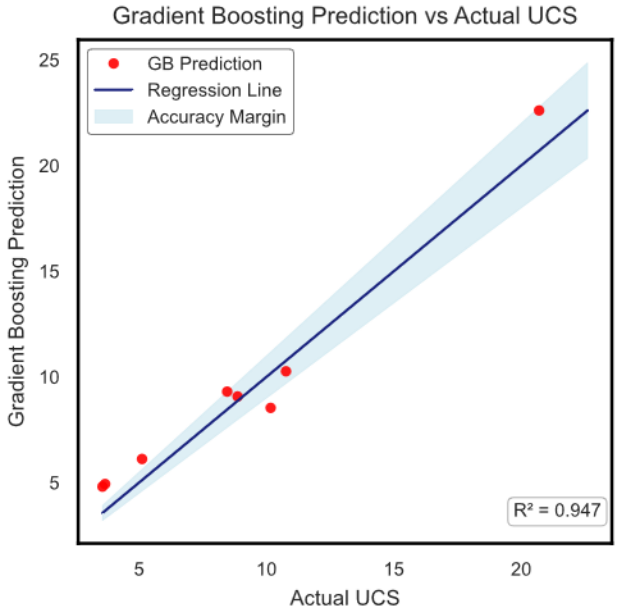

Fig. 6. UCS prediction plot with Gradient Boosting

Automated ML and Optimization

Automated Machine Learning (AutoML) was employed to reduce the cost and expertise required for model development by systematically exploring algorithms and hyperparameter configurations. Using the PyCaret library, preprocessing (scaling, encoding, transformation), model evaluation, and hyperparameter tuning were performed across multiple regression algorithms. Random Forest (RF) emerged as the top-performing model, consistent with its suitability for capturing the nonlinear complexities of geotechnical and rock engineering data (Aladejare et al., 2024).

Random Forest UCS Prediction

The PyCaret-optimized Random Forest (RF) model achieved a mean $R^2 = 0.988 \pm 0.015$, RMSE = 0.57 ± 0.09 MPa, and MAE = 0.39 ± 0.07 MPa across 10 cross-validation folds. These results indicate consistently high predictive accuracy with minimal variation between folds. The ±5% and ±10% prediction bands shown in Figure 7 further confirm model stability despite the small sample size. In summary, the AutoML-driven RF model provides highly accurate and reliable UCS predictions, making it a practical tool for rock engineering applications. Detailed results are presented in Table 7.

[DOR: 20.1001.1.22286837.1404.19.5.2.3]
[DOI: 10.22034/JEG.2025.19.5.1020471]

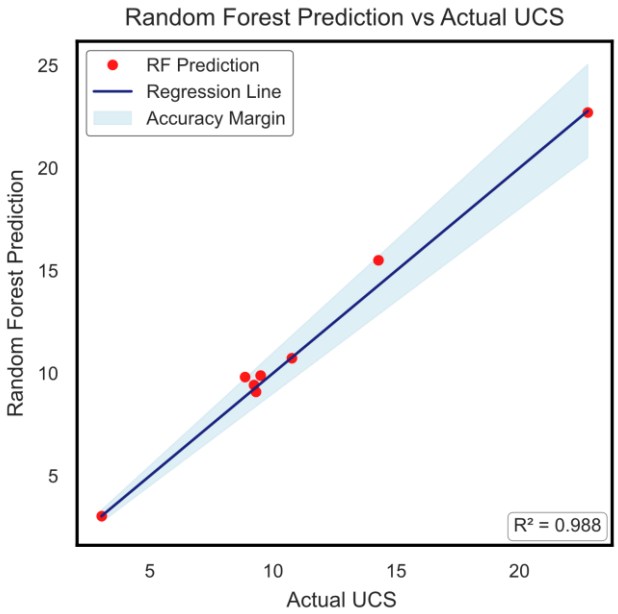

Fig. 7. UCS prediction plot with Random Forest

Table 7. Random Forest model performance in predicting UCS

50.	51.	52. RF
53. UCS	54. MAE	55. 0.3862
56.	57. MSE	58. 0.3285
59.	60. RMSE	61. 0.5732
62.	63. R2	64. 0.9884
65.	66. RMSLE	67. 0.0453
68.	69. MAPE	70. 0.036

Uncertainty Quantifications Through Monte Carlo Simulation

Uncertainty Analysis

Monte Carlo Simulation was employed to quantify prediction uncertainty in UCS, providing a probabilistic modeling approach that accounts for variability in geomaterial properties and input parameter fluctuations. This method serves as an effective tool for risk analysis and enhancing design reliability in geotechnical engineering applications (Fattahi et al., 2013; Wang, 2013; Wang and Coa, 2014). The implementation involved systematic noise injection where multiple noisy versions of the dataset were generated by adding Gaussian noise proportional to each feature's standard deviation. For For all of the laboratory and petrographic features, the formula $x\_noisy = x + N(0, \alpha \cdot \sigma\_x)$ was applied with $\alpha = 0.1$. The best-performing AutoML model was then retrained on these multiple noisy datasets to generate predictive ensembles. For each test sample, 100 noisy realizations were produced, with mean predictions, standard deviations, and 95% confidence intervals ($\pm 1.96\sigma$) computed to provide robust uncertainty estimates incorporating both measurement errors and model variability. Model performance was assessed using R^2, RMSE, and MAE, complemented by comprehensive uncertainty analysis through prediction standard deviations and confidence intervals. Data visualization proved essential for transforming complex datasets into understandable formats, enabling identification of patterns, trends, and relationships not easily detectable in raw data, thereby facilitating informed decision-making for both technical and non-technical audiences.

[DOR: 20.1001.1.22286837.1404.19.5.2.3]
[DOI: 10.22034/JEG.2025.19.5.1020471]

Visualization of true versus predicted UCS with uncertainty bands (Figure 8) demonstrated strong model performance, with the predicted UCS values (red line) closely matching true UCS values (blue line). The ±1σ uncertainty band (light red) was narrower in regions of high prediction confidence and wider where greater variability was observed. The 95% confidence interval (light purple) encompassed predictions in 95% of cases. Observed discrepancies were attributed to inherent model uncertainty and fluctuations in input parameters, with uncertainty bands clearly indicating zones of higher prediction risk. Overall, the Monte Carlo-based uncertainty quantification demonstrates model robustness while identifying areas requiring improved prediction reliability, thereby enhancing the interpretability and practical applicability of UCS predictions for geotechnical engineering applications (Shinde and Shivthare, 2024).

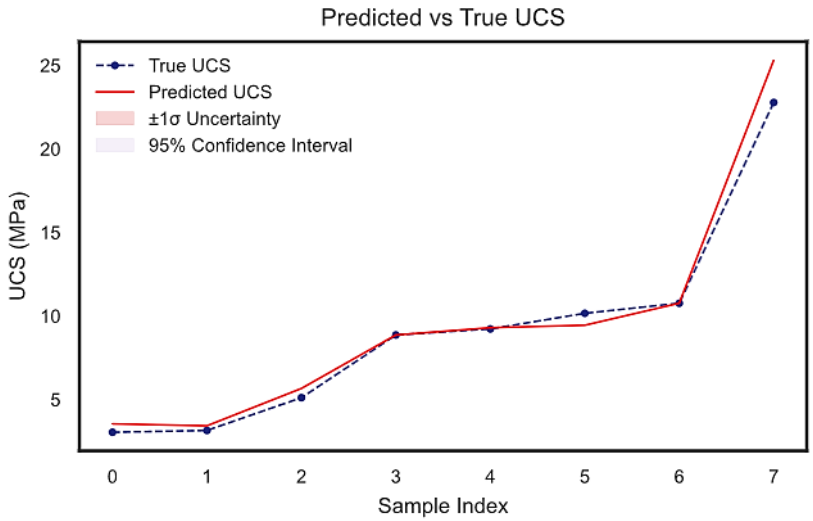

Fig. 8. True vs predicted with uncertainty

Residual analysis to identify systematic biases

Residual analysis was performed to evaluate the Monte Carlo simulation results by comparing actual UCS values with predicted ones (Figure 9). Residuals, defined as the difference between actual and predicted UCS, provide critical insights into model performance and potential systematic biases. Ideally, residuals should be randomly distributed around zero, reflecting unbiased predictions. However, Figure 9 reveals discernible patterns in certain UCS ranges, indicating over- or under-prediction. The 95% confidence intervals in Figure 10 represent the expected prediction range. Narrow intervals correspond to higher accuracy and lower uncertainty, while wider intervals indicate greater uncertainty, particularly in regions of high variability or limited data. Similarly, the ±1σ uncertainty bands highlight zones of reduced predictive reliability. Broader bands suggest lower accuracy, often associated with complex geological conditions or insufficiently captured geotechnical parameters. Overall, Figure 10 underscores areas where systematic bias or elevated uncertainty persists, emphasizing the need for model recalibration, additional data integration, and enhanced feature representation to improve UCS prediction accuracy. The distribution of prediction uncertainty (Figure11) demonstrates approximately normal behavior, supporting the use of ±1.96σ for 95% confidence intervals.

[DOR: 20.1001.1.22286837.1404.19.5.2.3]
[DOI: 10.22034/JEG.2025.19.5.1020471]

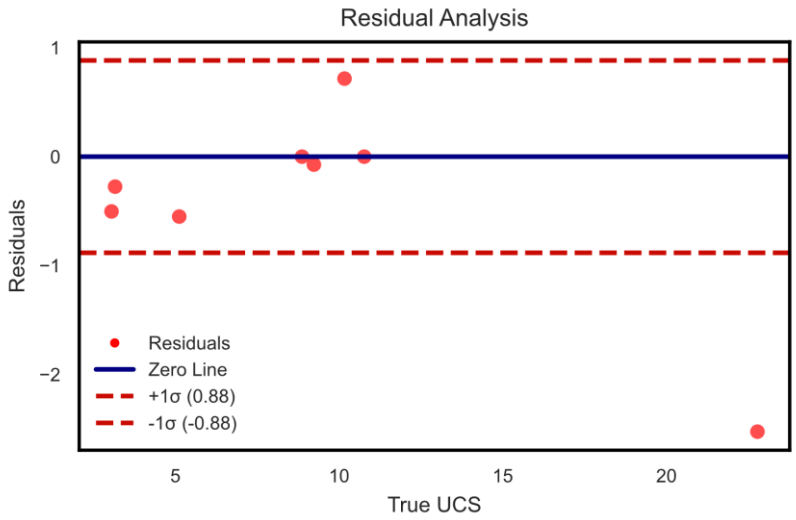

Fig. 9. Residual analysis to identify systematic biases

Feature importance for model interpretability

The feature importance of visualization highlights the relative contribution of input variables to the model's predictions (Figure10). Among them, soundness15 emerges as the most influential feature for predicting UCS, underscoring its role in material resistance to failure. Closely following, id15 and mineral2 also show strong influence, emphasizing the importance of sample identification and mineralogical properties in model accuracy. In contrast, features such as texture1 and aiv have limited impact, suggesting minimal relevance to predictive performance. This ranking not only illustrates the model's reliance on critical attributes but also provides guidance for refining feature selection and improving predictive robustness.

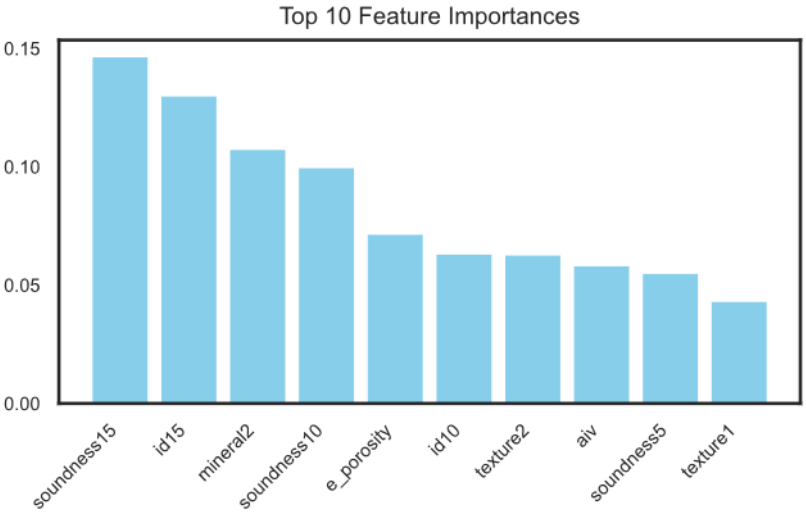

Fig. 10. Feature importance visualization for model interpretability

Distribution comparison between true and Predicted Values

The distribution comparison of true versus predicted values provides insight into the model's predictive fidelity (Figure 11). High similarity between the two distributions indicates effective pattern capture, whereas pronounced differences may reflect overfitting or underfitting. Deviations or outliers in the predictions can further signal systematic errors

requiring adjustment. When combined with uncertainty analysis, this comparison enhances understanding of model reliability and highlights potential areas for refinement in predictive methodology.

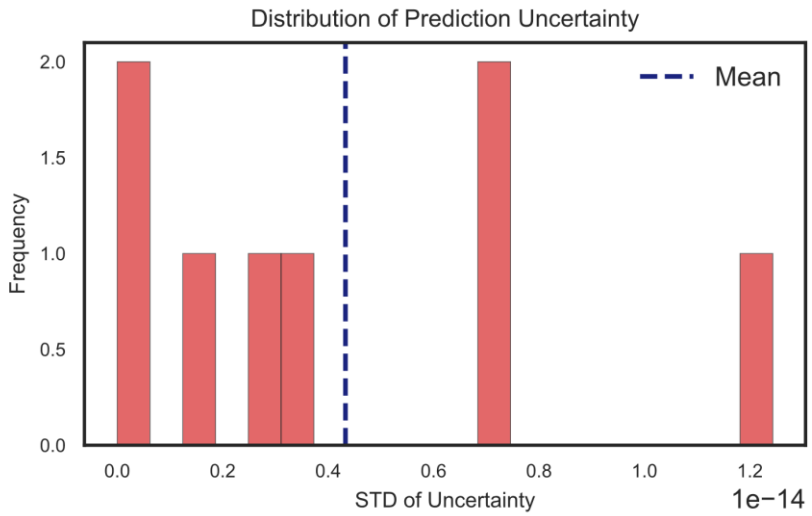

Fig. 11. Distribution comparison between true and predicted values

Correlation analysis between features and target variable

The correlation heatmap illustrates the relationships between features and the target variable, with coefficients ranging from –0.2 to 0.8 (Figure 12). Positive correlations (red hues) indicate direct relationships, while negative correlations (blue hues) suggest inverse dependencies; values near zero denote weak or negligible associations. This analysis identifies features most strongly related to the target, offering opportunities to improve model performance, while also indicating variables of limited

predictive utility that may be reconsidered in future modeling efforts.

[DOR: 20.1001.1.22286837.1404.19.5.2.3]
[DOI: 10.22034/JEG.2025.19.5.1020471]

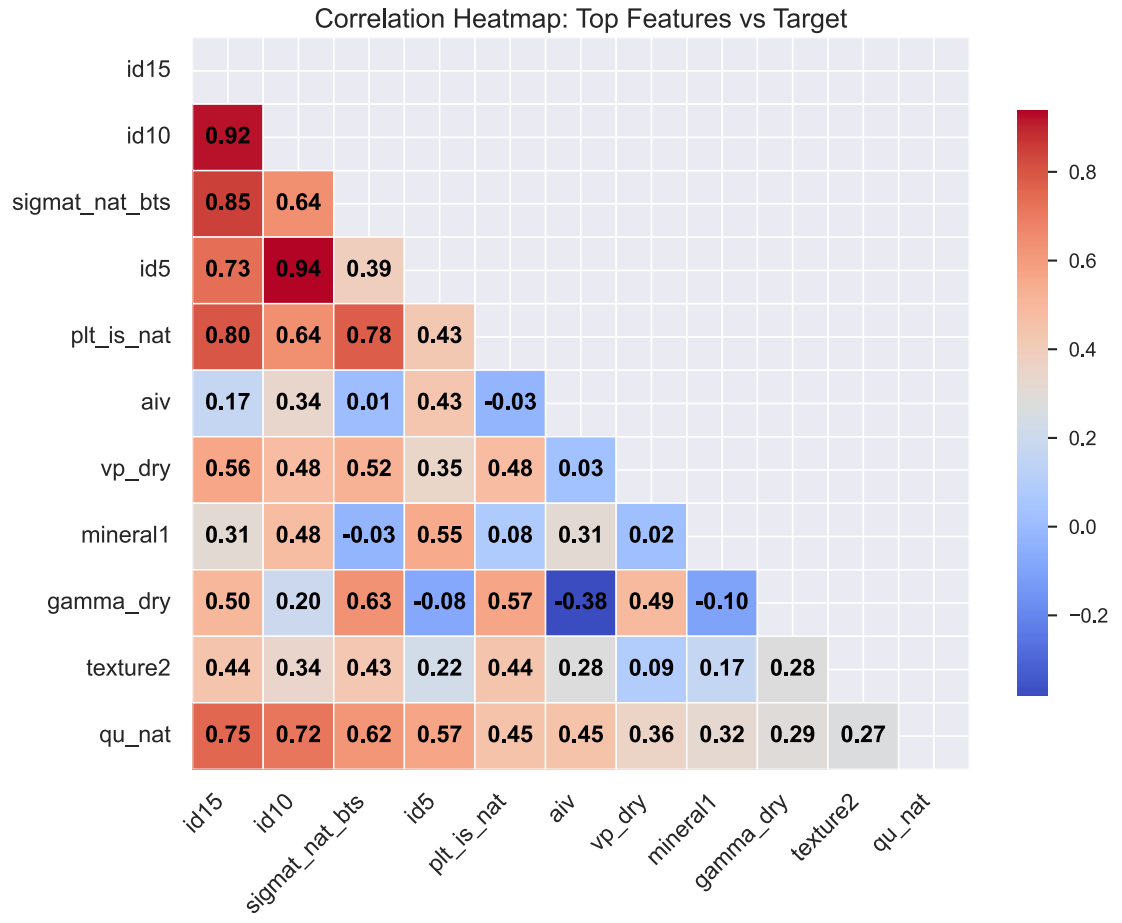

Fig. 12. Correlation analysis between features and target variable

Practical Implementation and Engineering Application

The developed models provide notable advantages for geotechnical engineering by reducing laboratory testing costs while maintaining high predictive accuracy ($R^2 > 0.98$) and low error margins (RMSE < 0.6 MPa). Incorporating uncertainty quantification enables probabilistic design and improved risk assessment, while feature importance analysis supports material characterization and decision-making. Moreover, the ability to generate rapid predictions makes the models suitable for real-time quality control, particularly during preliminary design and quality assurance phases. The framework was implemented in Python, employing TensorFlow/Keras for neural networks, Scikit-learn for classical machine learning and evaluation, and PyCaret for automated pipeline optimization. Data handling was performed with NumPy and Pandas, while visualization used Matplotlib and Seaborn. Its modular architecture allows seamless integration of new feature engineering, alternative models, or uncertainty quantification methods, ensuring flexibility and adaptability for future engineering applications.

Conclusion

This study successfully developed and validated comprehensive machine learning frameworks for predicting the UCS of sedimentary rocks using integrated petrographic and conventional rock property data from Iran's southern coastlines. The research demonstrates significant advances in both predictive accuracy and uncertainty quantification for rock

[DOR: 20.1001.1.22286837.1404.19.5.2.3]
[DOI: 10.22034/JEG.2025.19.5.1020471]

[DOR: 20.1001.1.22286837.1404.19.5.2.3]

[DOI: 10.22034/JEG.2025.19.5.1020471]

engineering applications. The AutoML-optimized Random Forest model achieved outstanding performance metrics (R^2 = 0.9884, RMSE = 0.5732 MPa, MAPE = 3.6%), representing a substantial improvement over traditional empirical correlations and statistical methods. The hybrid Neural Network-Gradient Boosting approach provided robust alternative predictions (R^2 = 0.9412, RMSE = 1.385 MPa), demonstrating the effectiveness of ensemble learning techniques for complex geotechnical problems. The implementation of Monte Carlo simulation methodology successfully addressed a critical gap in conventional rock engineering practice by providing probabilistic UCS distributions, 95% confidence intervals, and systematic uncertainty quantification. Feature importance analysis revealed that soundness parameters (soundness15) and mineralogical composition (mineral2) are the most influential predictors of UCS, validating the fundamental role of petrographic characteristics in controlling mechanical behavior. This finding bridges the gap between microscopic rock fabric and macroscopic engineering properties, providing scientific justification for incorporating detailed petrographic analysis in rock characterization protocols. The correlation analysis and residual evaluation further demonstrated the models' ability to capture complex nonlinear relationships between multiple input parameters and rock strength. The developed framework addresses major limitations in current rock engineering practice by reducing dependence on expensive and time-consuming laboratory testing while maintaining high prediction reliability. The rapid prediction capabilities enable real-time quality control during construction phases, while uncertainty quantification supports risk-informed decision making in engineering design. The modular Python implementation ensures seamless integration into existing engineering workflows, making these advanced techniques accessible to practicing engineers. The success of petrographic feature integration demonstrates the value of incorporating geological knowledge into data-driven engineering solutions. This approach is particularly relevant for coastal protection structures, where rock durability and strength are critical for long-term performance under harsh marine environments. The uncertainty quantification framework provides a foundation for probabilistic design approaches that can improve safety factors and optimize material selection processes. The research establishes a foundation for expanding these methodologies to other rock types and geological settings. Future work should focus on developing standardized petrographic quantification protocols, integrating additional microstructural parameters, and extending uncertainty analysis to include spatial variability and time-dependent degradation effects. The framework's modular design facilitates incorporation of emerging machine learning techniques and adaptation to specific regional geological conditions. Engineering practitioners are encouraged to adopt integrated petrographic-geotechnical characterization protocols, particularly for critical infrastructure projects. The demonstrated accuracy and reliability of these machine learning approaches support their implementation in preliminary design phases, material selection processes, and quality assurance programs. The uncertainty quantification capabilities should be leveraged for risk assessment and probabilistic design approaches in geotechnical engineering applications.

This research contributes to the evolution of rock engineering from purely empirical approaches toward scientifically grounded, data-driven methodologies that enhance both accuracy and reliability in engineering decision-making processes.

Acknowledgments

The authors would like to thank Prof. John Harrison, professor of Engineering Rock Mechanics at the University of Toronto,

acknowledging the technical contributions he made to this paper.

References

Afolagboye, L. O., Ajayi, D. E., & 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.

Aladejare, A. E., Idowu, K. A., & Ozoji, T. (2024). Reliability of Monte Carlo simulation approach for estimating uniaxial compressive strength of intact rock. Earth Science Informatics, 17(3), 2043-2053.

Ali, E., Guang, W., & Ibrahim, A. (2014). Empirical relations between compressive strength and microfabric properties of amphibolites using multivariate regression, fuzzy inference and neural networks: A comparative study. Engineering Geology, 183, 230-240.

Alizadeh, R., Marji, M. F., Abdollahipour, A., & Sagand, M. P. (2023). Numerical simulation of fatigue crack propagation in heterogeneous geomaterials under varied loads using displacement discontinuity method. Journal of Rock Mechanics and Geotechnical Engineering, 15(3), 702-716.

Aliyu, M. M., Shang, J., Murphy, W., Lawrence, J. A., Collier, R., Kong, F., & Zhao, Z. (2019). Assessing the uniaxial compressive strength of extremely hard cryptocrystalline flint. International Journal of Rock Mechanics and Mining Sciences, 113, 310-321.

Amiri, M., Amiri, M., Karrari, S. S., & Moradi, S. (2024). Predicting uniaxial compressive strength of different rocks using principal component analysis and deep neural network. Journal of Geomine, 2(2), 78-89.

Armaghani, D. J., Amin, M. F. M., Yagiz, S., Faradonbeh, R. S., & Abdullah, R. A. (2016). Prediction of the uniaxial compressive strength of sandstone using various modeling techniques. International Journal of Rock Mechanics and Mining Sciences, 85, 174-186.

Asheghi, R., Abbaszadeh Shahri, A., & Khorsand Zak, M. (2019). Prediction of uniaxial compressive strength of different quarried rocks using metaheuristic algorithm. Arabian Journal for Science and Engineering, 44(10), 8645-8659.

Askaripour, M., Saeidi, A., Mercier-Langevin, P., & Rouleau, A. (2022). A review of relationship between texture characteristic and mechanical properties of rock. Geotechnics, 2(1), 262-296.

Atici, U. (2011). Prediction of the strength of mineral admixture concrete using multivariable regression analysis and an artificial neural network. Expert Systems with applications, 38(8), 9609-9618.

Azadeh, A., Ghaderi, S. F., & Sohrabkhani, S. (2008). A simulated-based neural network algorithm for forecasting electrical energy consumption in Iran. Energy policy, 36(7), 2637-2644.

Azadmehr, A., Irji, A., Danesh, M., & Saffarian, M. (2024). Application of intelligent and statistical methods in estimating the uniaxial compressive strength of sandstones. Iranian Journal of Engineering Geology, 17(3).

Barzegar, R., Sattarpour, M., Deo, R., Fijani, E., & Adamowski, J. (2020). An ensemble tree-based machine learning model for predicting the uniaxial compressive strength of travertine rocks. Neural Computing and Applications, 32(13), 9065-9080.

Çanakcı, H., Baykasoğlu, A., & Güllü, H. (2009). Prediction of compressive and tensile strength of Gaziantep basalts via neural networks and gene expression programming. Neural Computing and Applications, 18(8), 1031-1041.

Ceryan, N., Okkan, U., & 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), 1055-1072.

Chen, Y. C. (2017). A tutorial on kernel density estimation and recent advances. Biostatistics & Epidemiology, 1(1), 161-187.

Daoud, A. M., Abdelkader, M. M., Sediek, K. N., Elsharief, A. M., Rashed, M. A., Elamein, A. M., ... & Rózsa, P. (2025). Relationship between petrography and geomechanical properties of the sandstone: A case study from Wadi Halfa, North Sudan. International Review of Applied Sciences and Engineering, 16(2), 193-209.

Dehghan, S., Sattari, G. H., Chelgani, S. C., & Aliabadi, M. A. (2010). Prediction of uniaxial compressive strength and modulus of elasticity for Travertine samples using regression and artificial neural networks. Mining Science and Technology (China), 20(1), 41-46.

[DOR: 20.1001.1.22286837.1404.19.5.2.3]

[DOI: 10.22034/JEG.2025.19.5.1020471]

Fattahi, H., Shojaee, S., Farsangi, M. A. E., & Mansouri, H. (2013). Hybrid Monte Carlo simulation and ANFIS-subtractive clustering method for reliability analysis of the excavation damaged zone in underground spaces. Computers and Geotechnics, 54, 210-221

Fereidooni, D. (2022). Importance of the mineralogical and textural characteristics in the mechanical properties of rocks. Arabian Journal of Geosciences, 15(7), 637.

Folk, R. L. (1954). The distinction between grain size and mineral composition in sedimentary-rock nomenclature. The Journal of Geology, 62(4), 344-359.

Gokceoglu, C. (2002). A fuzzy triangular chart to predict the uniaxial compressive strength of the Ankara agglomerates from their petrographic composition. Engineering Geology, 66(1-2), 39-51.

Gokceoglu, C., & Zorlu, K. (2004). A fuzzy model to predict the uniaxial compressive strength and the modulus of elasticity of a problematic rock. Engineering Applications of Artificial Intelligence, 17(1), 61-72.

Guo, S., Wei, Q., Qi, S., Xue, L., Zheng, B., Wang, H., Li, J., Song, S., Liang, N., & Zou, Y. (2024). Research progress on the geomechanical properties of block-in-matrix rocks. Materials, 17(5), 117 Multidisciplinary Digital Publishing Institute (MDPI).

Hamidi, A., Harrison, J. (2025). Uncertainty analysis of rock properties using Monte Carlo machine learning (ML). 9th Int. Symp. for Geotechnical Safety and Risk (ISGSR2025), Oslo, Norway.

Hamidi, A., Harrison, J., Hafezi Moghads, N., Rahmani, I., & Ghafoori, M. (2024). Predicting long-term durability of rock material for breakwater design with machine learning algorithms. Scientific Quarterly Journal of Iranian Association of Engineering Geology, 17(2), 75-93.

Hamidi A. (2024). New contributions to predicting engineering properties of rock material in coastal protection structures using machine learning algorithms. Doctoral Diss., Ferdowsi University of Mashhad, Iran. (In Persain)

Hasanipanah, M., Monjezi, M., Shahnazar, A., Armaghani, D. J., & Farazmand, A. (2015). Feasibility of indirect determination of blast induced ground vibration based on support vector machine. Measurement, 75, 289-297.

Heidari, M., Mohseni, H., & Jalali, S. H. (2018). Prediction of uniaxial compressive strength of some sedimentary rocks by fuzzy and regression models. Geotechnical and Geological Engineering, 36(1), 401-412.

Hemmati, A., Ghafoori, M., Moomivand, H., & Lashkaripour, G. R. (2020). The effect of mineralogy and textural characteristics on the strength of crystalline igneous rocks using image-based textural quantification. Engineering Geology, 266, 105467.

Hussain, J., Zafar, T., Fu, X., Ali, N., Chen, J., Frontalini, F., ... & Koumoutsakou, O. (2024). Petrological controls on the engineering properties of carbonate aggregates through a machine learning approach. Scientific Reports, 14(1), 31948.

ISRM, I (1978). Suggested methods for determining tensile strength of rock materials. International Journal of Rock Mechanics and Mining Sciences and Geomechanics Abstracts, 15(3), 99-103.

Jahanbakhshi, R., Keshavarzi, R., & Azinfar, M. J. (2011). Intelligent prediction of uniaxial compressive strength for sandstone. In ARMA US Rock Mechanics/Geomechanics Symposium (pp. ARMA-11). ARMA.

Jia, T., Zhang, L., Yan, Y., Zhang, R., Zeng, Q., Liu, P., & Yang, B. (2025). Compaction and pore preservation mechanisms of deep-buried sandstone reservoirs under the influence of sedimentary grain size characteristics: Insights from DEM simulations. Marine and Petroleum Geology, 173, 107286.

Kamenev, P. A., Bogomolov, L. M., Usoltseva, O. M., Tsoi, P. A., & Semenov, V. N. (2021). Geomechanical parameters of sedimentary rocks of Southern Sakhalin. In IOP Conference Series: Earth and Environmental Science (Vol. 946, No. 1, p. 012013). IOP Publishing Ltd.

Khajevand, R. (2023). Estimating geotechnical properties of sedimentary rocks based on physical parameters and ultrasonic P-wave velocity using statistical methods and soft computing approaches. Iranian Journal of Science and Technology, Transactions of Civil Engineering, 47(6), 3785-3809.

Khan, M. A., Khan, T., Ali, A., Bello, A. M., & Radwan, A. E. (2023). Role of depositional and

diagenetic controls on reservoir quality of complex heterogenous tidal sandstone reservoirs: An example from the Lower Goru formation, Middle Indus Basin, Southwest Pakistan. Marine and Petroleum Geology, 154, 106337.

Kong, F., Xue, Y., Shang, J., Zhu, C., Han, M., Qu, Z., & Yang, K. (2024). Predicting uniaxial compressive strength of building stone based on index tests: Correlations, validity, reliability, and unification. Construction and Building Materials, 438, 137227.

Kong, F., Xue, Y., Gong, H., Jiang, X., Song, Q., Fu, Y., & Fu, K. (2023). The formation mechanism of dynamic water and mud inrush geohazard triggered by deep-buried tunnel crossing active fault: Insights from the geomechanical model test. Tunnelling and Underground Space Technology, 142, 105437.

Liu, Z., Shao, J., Xu, W., & Wu, Q. (2015). Indirect estimation of unconfined compressive strength of carbonate rocks using extreme learning machine. Acta Geotechnica, 10(5), 651-663.

Madhubabu, N., Singh, P. K., Kainthola, A., Mahanta, B., Tripathy, A., & Singh, T. N. (2016). Prediction of compressive strength and elastic modulus of carbonate rocks. Measurement, 88, 202-213.

Mistry, H. K., Mavani, C., Goswami, A., & Patel, R. (2024). Artificial intelligence for networking. Educational Administration: Theory and Practice, 30(7), 813-821.

Mokhtari, M. (2022). Predicting the Young's modulus and uniaxial compressive strength of a typical limestone using the principal component regression and particle swarm optimization. Journal of Engineering Geology, 16(1), 95.

Morgenroth, J., Khan, U. T., & Perras, M. A. (2019). An overview of opportunities for machine learning methods in underground rock engineering design. Geosciences, 9(12), 504.

Narimani, S., & Vásárhelyi, B. (2025). Leveraging machine learning for precision prediction of geomechanical properties of granitic rocks: a comparative analysis of MLR, ANN, and ANFIS models. Earth Science Informatics, 18(1), 1-27.

Ngo, T. P., Vu, H. N., & Bui, Q. B. (2025). Application of machine learning models for the optimisation of compressive strength and water resistance of geopolymer stabilised compacted earth. Case Studies in Construction Materials, 22, e04203.

Pérez-Díaz, L., Alcalde, J., & Bond, C. E. (2020). Introduction: Handling uncertainty in the geosciences: Identification, mitigation and communication. Solid Earth, 11(3), 889-897.

Petrounias, P., Giannakopoulou, P. P., Rogkala, A., Stamatis, P. M., Lampropoulou, P., Tsikouras, B., & Hatzipanagiotou, K. (2018). The effect of petrographic characteristics and physico-mechanical properties of aggregates on the quality of concrete. Minerals, 8(12), 577.

Pradeep, T., & Samui, P. (2022). Prediction of rock strain using hybrid approach of ANN and optimization algorithms. Geotechnical and Geological Engineering, 40(9), 4617-4643.

Onyelowe, K. C., Ebid, A. M., Hanandeh, S., & Kamchoom, V. (2025). Evaluating the slope behavior for geophysical flow prediction with advanced machine learning combinations. Scientific Reports, 15(1), 6531.

Rezaei, M., Majdi, A., & Monjezi, M. (2014). An intelligent approach to predict unconfined compressive strength of rock surrounding access tunnels in longwall coal mining. Neural Computing and Applications, 24(1), 233-241.

Rubo, R. A., de Carvalho Carneiro, C., Michelon, M. F., & dos Santos Gioria, R. (2019). Digital petrography: Mineralogy and porosity identification using machine learning algorithms in petrographic thin section images. Journal of Petroleum Science and Engineering, 183, 106382.

Sabri, M. S., Jaiswal, A., Verma, A. K., & Singh, T. N. (2024). Advanced machine learning approaches for uniaxial compressive strength prediction of Indian rocks using petrographic properties. Multiscale and Multidisciplinary Modeling, Experiments and Design, 7(6), 5265-5286.

Saedi, B., & Mohammadi, S. D. (2021). Prediction of uniaxial compressive strength and elastic modulus of migmatites by microstructural characteristics using artificial neural networks. Rock Mechanics and Rock Engineering, 54(11), 5617-5637.

Siegesmund, S., & Török, Á. (2014). Stone in architecture: Properties, durability (5th ed.). Springer.

Singh, V. K., Singh, D., & Singh, T. N. (2001). Prediction of strength properties of some schistose rocks from petrographic properties

[DOR: 20.1001.1.22286837.1404.19.5.2.3]

[DOI: 10.22034/JEG.2025.19.5.1020471]

using artificial neural networks. International Journal of Rock Mechanics and Mining Sciences, 38(2), 269-284.

Sharma, L. K., Vishal, V., & Singh, T. N. (2017). Developing novel models using neural networks and fuzzy systems for the prediction of strength of rocks from key geomechanical properties. Measurement, 102, 158-169.

Shinde, B. G., & Shivthare, S. (2024). Impact of data visualization in data analysis to improve the efficiency of machine learning models. Journal of Advanced Zoology, 45, 107-112.

Standard, A. S. T. M. (2008). Standard test method for splitting tensile strength of intact rock core specimens. ASTM D3967-08. ASTM International, West Conshohocken, USA.

Stopka, G., Gieleta, R., Panowicz, R., Wałach, D., & Kaczmarczyk, G. P. (2024). High strain rate response of sandstones with different porosity under dynamic loading using split hopkinson pressure bar (SHPB). Applied Sciences, 14(12), 5347.

Swamy, S. V., Kunar, B. M., Chandar, K. R., Alwetaishi, M., Krishnan, S., & Reddy, S. (2025). Prediction of uniaxial compressive strength of limestone from ball mill grinding characteristics using supervised machine learning techniques. Scientific Reports, 15(1), 28395.

Sun, J., Zhang, R., Zhang, A., Wang, X., Wang, J., Ren, L., ... & Zhang, Z. (2024). Rock strength prediction based on machine learning: A study from prediction model to mechanism explanation. Measurement, 238, 115373.

Wang, Y. (2013). MCS-based probabilistic design of embedded sheet pile walls. Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards, 7(3), 151-162.

Wang, Y., & Cao, Z. (2014). Practical reliability analysis and design by Monte Carlo Simulation in spreadsheet. Risk and reliability in geotechnical engineering, 301.

Wang, Y., Hasanipanah, M., Rashid, A. S. A., Le, B. N., & Ulrikh, D. V. (2023). Advanced tree-based techniques for predicting unconfined compressive strength of rock material employing non-destructive and petrographic tests. Materials, 16(10), 3731.

Wijesundara, S., Wijesundara, K., & Bandara, S. (2025). Machine learning approach for predicting the compressive strength of ultra-high performance fiber reinforced concrete (UHPFRC). In Structures (Vol. 75, p. 108704). Elsevier.

Xie, W. Q., Liu, X. L., Zhang, X. P., Liu, Q. S., & Wang, E. Z. (2025a). A review of test methods for uniaxial compressive strength of rocks: Theory, apparatus and data processing. Journal of Rock Mechanics and Geotechnical Engineering, 17(3), 1889-1905.

Xie, Y., Li, X., & Min, Z. (2025b). Comparison of machine learning models for rock UCS prediction using measurement while drilling data. Scientific Reports, 15(1), 8434.

Xue, Y., Kong, F., Li, S., Zhang, Q., Qiu, D., Su, M., & Li, Z. (2021). China starts the world's hardest "sky-high road" project: challenges and countermeasures for Sichuan-Tibet railway. Innovation, 2(2), 100105.

Yang, Z., Wu, Y., Zhou, Y., Tang, H., & Fu, S. (2022). Assessment of machine learning models for the prediction of rate-dependent compressive strength of rocks. Minerals, 12(6), 731.

Yeh, C. H., Dong, J. J., Khoshnevisan, S., Juang, C. H., Huang, W. C., & Lu, Y. C. (2021). The role of the geological uncertainty in a geotechnical design–A retrospective view of Freeway No. 3 Landslide in Northern Taiwan. Engineering Geology, 291, 106233.

Yesiloglu-Gultekin, N., Sezer, E. A., Gokceoglu, C., & Bayhan, H. (2013). An application of adaptive neuro fuzzy inference system for estimating the uniaxial compressive strength of certain granitic rocks from their mineral contents. Expert Systems with Applications, 40(3), 921-928.

Zhao, Y., He, X., Jiang, L., Wang, Z., Ning, J., & Sainoki, A. (2023). Influence analysis of complex crack geometric parameters on mechanical properties of soft rock. International Journal of Coal Science & Technology, 10(1), 78.

Zorlu, K., Gokceoglu, C., Ocakoglu, F., Nefeslioglu, H. A., & Acikalin, S. J. E. G. (2008). Prediction of uniaxial compressive strength of sandstones using petrography-based models. Engineering Geology, 96(3-4), 141-158.

[DOR: 20.1001.1.22286837.1404.19.5.2.3]

[DOI: 10.22034/JEG.2025.19.5.1020471]

نشریه زمین شناسی مهندسی
Journal home page https://jeg.khu.ac.ir
دانشگاه خوارزمی
دسترسی آزاد
شاپا الکترونیکی ۲۹۸۱-۱۶۰۰

کمی‌سازی عدم‌قطعیت پیش‌بینی UCS در سنگ‌های رسوبی با استفاده از ویژگی‌های پتروگرافی: رویکرد یادگیری ماشین_شبیه‌سازی مونت کارلو

سید محمود فاطمی عقدا[۱] ✉ ، آسیه حمیدی [۲]، فاطمه امیری [۳]

۱. استاد، گروه زمین شناسی کاربردی، دانشکده علوم زمین، دانشگاه خوارزمی، تهران، ایران، رایانامه: fatemi@Khu.ac.ir

۲. پژوهشگر پسادکتری، دانشکده عمران و معدن، دانشگاه تورنتو، کانادا، رایانامه: asieh.hamidi@utoronto.ca

۳.کارشناشی ارشد،گروه زمین شناسی کاربردی ، دانشکده علوم زمین، دانشگاه خوارزمی، تهران، رایانامه: fatemeh.amiri@Khu.ac.ir

اطلاعات مقاله

نوع مقاله: مقاله پژوهشی

تاریخ دریافت: ۱۴۰۴/۰۶/۲۷

تاریخ پذیرش: ۱۴۰۴/۰۸/۲۶

کلیدواژه‌ها:
مقاومت فشاری تک محوری، ماشین لرنینگ، ویژگی‌های پتروگرافی، کمی سازی عدم قطعیت، شبیه سازی مونت کارلو، مهندسی سنگ، سنگ‌های ساحلی ایران.

چکیده

ارزیابی مقاومت مکانیکی، به‌ویژه مقاومت فشاری تک‌محوری (Uonixal Compersive Strength) سنگ‌ها، برای طراحی و پیش‌بینی عملکرد سازه‌های سطحی و زیرزمینی حیاتی است و تأثیر زیادی بر هزینه‌ها و ایمنی پروژه‌ها دارد. روش‌های آزمایشگاهی سنتی برای ارزیابی UCS مخرب، زمان‌بر و پرهزینه هستند، در حالی که روش‌های غیرمستقیم اغلب به دلیل ناهمگنی سنگ‌ها از دقت و قابلیت اطمینان کافی برخوردار نیستند. این مطالعه با توسعه چارچوب‌های پیشرفته یادگیری ماشین که ویژگی‌های پتروگرافی را با ویژگی‌های سنتی سنگ‌ها ترکیب می‌کنند، این محدودیت‌ها را برطرف کرده و به پیش‌بینی UCS و کمی سازی عدم‌قطعیت‌ها پرداخته است. داده‌های جامع از سنگ‌های رسوبی سواحل جنوبی ایران (خلیج فارس و دریای عمان) استفاده شده است که شامل ویژگی‌های مکانیکی (UCS، مقاومت کششی برزیلی، شاخص بار نقطه‌ای، تخلخل، سرعت پالس فراصوت)، شاخص‌های دوام (سایش لس آنجلس، دوام ترک‌خوردگی، ارزش تأثیر مصالح) و ویژگی‌های پتروگرافی استخراج شده از تحلیل مقاطع نازک می‌با شد. سه رویکرد مکمل به‌کار گرفته شده‌اند: (۱) رگرسیون ترکیبی شبکه عصبی–افزایش گرادیان (ANN-GBR)، (۲) جنگل تصادفی بهینه شده با AutoML، و (۳) شبیه سازی مونت کارلو برای کمی سازی عدم‌قطعیت. . نتایج نشان داد که مدل جنگل تصادفی بهینه شده با AutoML عملکرد پیش‌بینی استثنائی با $R^2 = 0.9884$، RMSE = 0.5732 MPa و MAPE = 3.6% داشت که به‌طور چشمگیری از روش‌های تجربی سنتی بهتر عمل کرده است. رویکرد ترکیبی ANN-GBR موفق به کسب $R^2 = 0.9412$ و RMSE = 1.385 MPa شده است. شبیه‌سازی‌های مونت کارلو ارزیابی‌های احتمالاتی مقاوم با فواصل اطمینان ۹۵٪ و شناسایی بایاس سیستماتیک را ارائه داده است. تحلیل اهمیت ویژگی‌ها نشان داد که ویژگی های سلامت سدیم و ترکیب کانی شناسی مهم‌ترین پیش‌بینی‌کننده‌ها هستند. چارچوب توسعه یافته مزایای عملی قابل‌توجهی از جمله کاهش هزینه‌های آزمایشگاهی، پیش‌بینی سریع‌تر برای کنترل کیفیت و ارزیابی بهبود یافته ریسک از طریق کمی‌سازی عدم‌قطعیت‌ها را فراهم می‌آورد و رویکردی مقاوم و مقرون‌به‌صرفه برای ارزیابی مقاومت سنگ‌ها ارائه می‌دهد.

مقدمه

مقاومت فشاری تک‌محوری (UCS) به عنوان یک ویژگی کلیدی در ارزیابی مکانیکی سنگ‌ها، نقش حیاتی در طراحی پروژه‌های مهندسی سطحی و زیر سطحی زمین ایفا می‌کند (Xue et al., 2021, Kong et al., 2023, Alizadeh et al., 2023, Xie et al., 2025). آزمایش‌های مستقیم UCS به‌دلیل زمان‌بر بودن و هزینه‌های بالا، به ویژه در

استناد: فاطمی عقدا، س. م.، حمیدی، آ.، امیری، ف. (۱۴۰۴). کمی‌سازی عدم‌قطعیت پیش‌بینی UCS در سنگ‌های رسوبی با استفاده از ویژگی‌های پتروگرافی: رویکرد یادگیری ماشین_شبیه‌سازی مونت کارلو. مجله زمین شناسی مهندسی، ۱۹ (۵)، ۶۴۸-۶۷۶ https://doi.org/10.22034/JEG.2025.19.5.1020471

ناشر: دانشگاه خوارزمی.

سنگ‌های نرم و سخت، معمولاً چالش‌برانگیز هستند. به‌همین‌دلیل، روش‌های غیرمستقیم مانند آزمایش بار نقطه‌ای، چکش اشمیت و سرعت امواج فراصوت به‌عنوان آزمایش‌های شاخص برای ارزیابی UCS استفاده می‌شوند در حالی که این روش‌ها به دلیل ناهمگنی ذاتی مصالح سنگی همیشه دقیق و قابل اطمینان نیستند (Aliyu et al., 2019, Kong et al., 2023, Aladejare et al., 2024). در این راستا، ویژگی‌های پتروگرافی مانند ترکیب معدنی، بافت و نوع سیمان، تأثیر زیادی بر UCS دارند و می‌توانند از طریق تحلیل‌های پتروگرافی سریع و کم هزینه شناسایی شوند (Petrounis et al., 2018). به‌ویژه، بافت‌های بالغ و مواد معدنی سخت مانند کوارتز می‌توانند موجب افزایش UCS شوند، در حالی که مواد نرم مانند کلسیت، مقاومت را کاهش می‌دهند (Folk, 1954, Khan et al., 2023). پیشرفت‌های اخیر در استفاده از یادگیری ماشین (ML) و هوش مصنوعی (AI) به‌عنوان ابزارهای قدرتمند برای مدل‌سازی روابط پیچیده بین ویژگی‌های پتروگرافی و UCS نشان داده‌اند که این روش‌ها می‌توانند دقت پیش‌بینی را به‌طور قابل‌توجهی افزایش دهند و عدم قطعیت‌ها را کمی‌سازی کنند که موجب تصمیم‌گیری‌های دقیق‌تر در طراحی‌های مهندسی می‌شود (Wang et al., 2023).

روش کار

این تحقیق به پیش‌بینی UCS سنگ‌های رسوبی سواحل جنوبی با استفاده از روش‌های داده محور پیشرفته ماشین لرنینگ و شبکه عصبی مصنوعی پرداخته است. داده‌ها از دو مجموعه‌داده آزمایشگاهی (ویژگی‌های مهندسی سنگ‌ها) و پتروگرافی (ویژگی‌های میکروساختاری سنگ‌ها) به دست آمده‌اند (حمیدی و همکاران، ۲۰۲۴). سه چارچوب اصلی برای تحلیل داده‌ها توسعه داده شده است:

مدل‌های یادگیری ماشین: برای مدل‌سازی روابط غیرخطی پیچیده بین ویژگی‌های سنگ‌ها از الگوریتم‌های MLاستفاده شده است.

شبکه عصبی مصنوعی: یک مدل ANN با چندورودی برای ترکیب ویژگی‌های آزمایشگاهی و پتروگرافی به‌کار رفته که روابط غیرخطی را به‌طور دقیق شبیه‌سازی می‌کند.

رگرسیون تقویت‌شده با الگوریتم افزایش گرادیان: این مدل برای بهبود پیش‌بینی‌های UCS و کاهش خطاهای باقیمانده به‌کار رفته است.

پیش‌پردازش داده‌ها: داده‌ها پس از حذف ستون‌های غیرمفید، جایگزینی مقادیر گم‌شده با استفاده از تکمیل‌کننده تکراری (Iterative Imputer) و استفاده از الگوریتم جنگل تصادفی به‌عنوان تخمین‌گر، و استانداردسازی ویژگی‌ها، برای مدل‌سازی آماده شدند.

مدل‌سازی و بهینه سازی:

جنگل تصادفی: این الگوریتم تجمیعی با استفاده از درخت‌های تصمیمی برای پیش‌بینی UCS به‌کار رفته است. این روش به دلیل توانایی در کاهش واریانس و جلوگیری از بیش‌برازش برای داده‌های غیرخطی و ابعاد بالا مناسب است. و با استفاد از ماشین خودکار(AutoML) با Paycaret تنظیمات پارامترهای مدل برای یافتن بهترین پارامترها بهینه سازی شده است.

شبکه عصبی مصنوعی: معماری ANN شامل دو شاخه موازی برای پردازش ورودی‌های آزمایشگاهی و پتروگرافی است که با لایه‌های متراکم و تابع فعال‌سازی ReLU به‌طور مؤثر ویژگی‌ها مهم در ارتباط با هدف را استخراج می‌کند.

رگرسیون تقویت‌شده با گرادیان: مدل GBR با استفاده از جستجوی شبکه‌ای و اعتبارسنجی متقابل برای بهینه‌سازی

پارامترهایی چون تعداد برآوردگرها و نرخ یادگیری به‌کار گرفته شده است.

کمی‌سازی عدم قطعیت: برای مدل‌سازی و ارزیابی عدم قطعیت‌های مرتبط با ورودی‌ها، شبیه‌سازی مونت کارلو (MCS)انجام شد. این روش به تولید توزیع‌های احتمالاتی UCSکمک می‌کند و قابلیت اطمینان مدل را در برابر تغییرات ورودی‌ها افزایش می‌دهد.

ارزیابی عملکرد: عملکرد مدل‌ها با استفاده از معیارهای آماری نظیر MAE، MSE، RMSE، R^2، RMSLE و MAPE ارزیابی شد. مقادیر پایین‌تر خطاها و R^2 بالاتر نشان‌دهنده دقت پیش‌بینی بالای مدل‌ها هستند.

بحث و نتایج

در این مطالعه UCS به عنوان متغیر هدف تحلیل شده است. داده‌ها از سنگ‌های آهکی ماسه‌سنگی جنوب ایران جمع‌آوری شده‌اند و از طریق هیستوگرام و تحلیل برآورد چگالی هسته‌ای (KDE) مورد بررسی قرار گرفته‌اند. هیستوگرام با تمرکز بیشتر مقادیر در Mpa ۱۲-۸ نشان‌دهنده توزیع کجی مثبت است، KDE به طور دقیق‌تر الگوی توزیع را نشان داده است و ویژگی‌هایی مانند کجی و نوسانات محلی را شناسایی می‌کند. پیش‌بینی UCS با استفاده از مدل‌های یادگیری ماشین، از جمله ANN– GBR و RF انجام شد. تحلیل‌ها نشان داد که هر دو مدل در پیش‌بینی UCS عملکرد قابل‌قبولی داشتند، با مدل ANN–GBRکه RMSE برابر MPa ۱.۳۹ و R^2 برابر ۰.۹۴۱۲ را به دست آورد و RF که موفق به کسب R^2 برابر ۰.۹۸۸۴ و RMSE برابر MPa ۰.۵۷۳۲ شد. برای ارزیابی دقت پیش‌بینی‌ها، شبیه‌سازی مونت کارلو برای کمی‌سازی عدم قطعیت در مدل‌ها به‌کار گرفته شد. این شبیه‌سازی به طور مؤثری تغییرپذیری داده‌ها و نویز سیستماتیک را در نظر گرفت و باندهای عدم قطعیت ۹۵٪ را برای پیش‌بینی‌ها محاسبه کرد که نشان داد مدل‌ها در بیشتر موارد در ناحیه های بالا ، پیش‌بینی‌های دقیقی انجام داده‌اند. نمودارتحلیل باقی‌مانده‌ها برای شناسایی سوگیری‌های سیستماتیک، به‌ویژه در برخی از بازه‌ها UCS، نشان داد که نیاز به بهبود پیش‌بینی‌ها در این نواحی وجود دارد. و تحلیل اهمیت ویژگی‌ها نشان داد که ویژگی‌های soundness15 و mineral2بیشترین تأثیر را در پیش‌بینی UCS دارند در حالی که ویژگی‌هایی مانند texture1 و aiv نقش کمتری ایفا می‌کنند.

نتیجه‌گیری

این مطالعه چارچوب‌های یادگیری ماشینی پیشرفته‌ای را برای پیش‌بینی UCS سنگ‌های رسوبی با استفاده از داده‌های پتروگرافی و ویژگی‌های مهندسی سنگ‌ها در سواحل جنوبی ایران توسعه و اعتبارسنجی کرد. نتایج نشان می‌دهند که مدل جنگل تصادفی بهینه‌شده با AutoML، با پیش بینی R^2 برابر ۰.۹۸۸، RMSE برابر MPa ۰.۵۷۳۲، MAPE برابر MPa ۳.۶ عملکرد بسیار بهتری نسبت به روش‌های سنتی و تجربی دارد. همچنین، مدل شبکه عصبی- تقویت گرادیان توانسته است، پیش‌بینی مقاومی با R^2 برابر ۰.۹۴۱۲، RMSE برابر Mpa ۱.۳۸۵ ارائه دهد. این تحقیق با استفاده از شبیه‌سازی مونت کارلو موفق به کمی‌سازی عدم قطعیت و تعیین توزیع‌های احتمالاتی برای UCS شد و ابزاری مفید برای تصمیم‌گیری‌های مهندسی فراهم آورد. تحلیل اهمیت ویژگی‌ها نشان داد که پارامترهای استحکام و ترکیب معدنی بیشترین تاثیر را در پیش‌بینی UCS دارند که نشان‌دهنده اهمیت پتروگرافی در تحلیل‌های مهندسی است. این چارچوب نه‌تنها وابستگی به آزمایش‌های گران‌قیمت را کاهش می‌دهد، بلکه قابلیت‌های پیش‌بینی سریع UCS، امکان بهبود کنترل کیفیت در زمان واقعی و تصمیم‌گیری مبتنی بر ریسک را فراهم می‌آورد. این رویکرد

[DOR: 20.1001.1.22286837.1404.19.5.2.3]
[DOI: 10.22034/JEG.2025.19.5.10204471]

می‌تواند به‌ویژه در پروژه‌های زیرساختی حیاتی مانند سازه‌های حفاظت ساحلی کاربرد گسترده‌ای داشته باشد. این مطالعه به مهندسی سنگ کمک می‌کند تا از روش‌های تجربی به سمت روش‌های علمی مبتنی بر داده حرکت کند و دقت و اطمینان در فرآیندهای طراحی و انتخاب مصالح را در پروژه های مهندسی افزایش دهد.

[DOR: 20.1001.1.22286837.1404.19.5.2.3]
[DOI: 10.22034/JEG.2025.19.5.1020471]