geotechnics

MDPI

Article

Predicting the CERCHAR Abrasiveness Index (CAI) from Geotechnical and Geophysical Rock Properties Using Machine Learning (ML) Regression Models

Pushpinder Singh Rana, John Molgaard and Stephen Douglas Butt *

The Department of Process Engineering, Memorial University of Newfoundland and Labrador, St. John's, NL A1C 5S7, Canada; psr162@mun.ca (P.S.R.); jmolgaard@mun.ca (J.M.)
* Correspondence: sdbutt@mun.ca

Abstract

The CERCHAR Abrasiveness Index (CAI) is widely used to quantify rock abrasiveness, yet its contact-based scratch procedure is highly sensitive to operational factors such as scratch length, surface preparation, stylus hardness, confining pressure, scratch velocity, and observer variability. The requirement for a prepared, polished surface and styli of specific Rockwell hardness further complicates testing, making CAI difficult to standardize across laboratories and challenging to compile into consistent databases. This study evaluates whether CAI can be reliably predicted from geotechnical and geophysical properties typically obtained through in situ testing or routine core characterization. A dataset of 175 igneous, sedimentary, and metamorphic rocks was assembled from publicly available CAI databases and laboratory sources. A rigorous validation pipeline was implemented, including Leave-One-Out Cross-Validation, K-Fold and Repeated K-Fold validation, and hyperparameter tuning using Nested K-Fold across seven regression models. A four-predictor model using Schmidt hammer rebound hardness, Mohs hardness number, longitudinal wave velocity, and grain size is proposed. Across all tested seeds and rock types, the Random Forest model showed the most stable performance, achieving K-Fold R^2 values of 0.94–0.98 with low RMSE and MAE. Because these predictors can be measured directly on outcrops, blocks, or tunnel faces, the approach enables a non-destructive, field-deployable CAI estimation without the need for prepared samples or specialized laboratory infrastructure.

Keywords: CERCHAR abrasion index; unconfined compressive strength; grain size; ultrasonic velocities; rebound hardness of rocks; machine learning regression modeling

Check for updates

Academic Editors: Meghdad Bagheri and Arya Assadi Langroudi

Received: 28 July 2026
Revised: 25 September 2026
Accepted: 2 October 2026
Published: 9 October 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Wear is a fundamental material response in any system involving relative motion between contacting surfaces. In rock excavation and drilling, wear leads to direct losses through tool and equipment degradation and indirect losses through Non-Productive Time (NPT) associated with maintenance, repair and component replacement. Rock abrasiveness is a primary driver of wear. Mouritz and Hutchings [1] demonstrated this by quantifying wear rates for sandstone and limestone against different steel types, showing that sandstone produced wear rates 140–200 times higher than limestone. Their work also emphasized the role of relative hardness: abrasion becomes effective when the hardness ratio (H_a/H_m) exceeds approximately 1.2 (where H_a is the hardness of the tool cutter and H_m is the hardness

of the rock). A variety of methods have been developed to characterize rock abrasiveness. Nilsen et al. [2] summarized widely used approaches, including Mohs hardness, petrographic analysis, X-ray diffraction, and mechanical abrasion tests such as the Norwegian University of Science and Technology (NTNU) test and the Laboratoire Central des Ponts et Chaussées (LCPC) procedures. The LCPC test evaluates wear on a rotating impeller using 4–6.3 mm aggregates, whereas the NTNU test uses slurry-based abrasion. Among these abrasion tests, the CERCHAR Abrasiveness Index (CAI) remains the most widely adopted method because it simulates two-body abrasion conditions relevant to cutting tools. The CAI test traces its origins to early scratch-based work by Valantin [3], later standardized by Cerchar and Hoenig [4] through specifications on stylus material and applied load. Although CAI is widely used, its results are sensitive to stylus hardness, surface condition and operational parameters and measurements often vary across laboratories. These limitations have motivated extensive research into petrographic and mechanical controls on CAI. Figure 1 shows the basic principle of the CAI test procedure. It is common to use styli with Rockwell hardness type C (HRC) 40 or HRC 55, with a tolerance of (±) 1 HRC. HRC 40-rated hardened styli are used for softer rocks like weaker limestones and HRC 55 hardness-rated styli are preferred for harder and more abrasive rocks like granites, quartzites and tight sandstones.

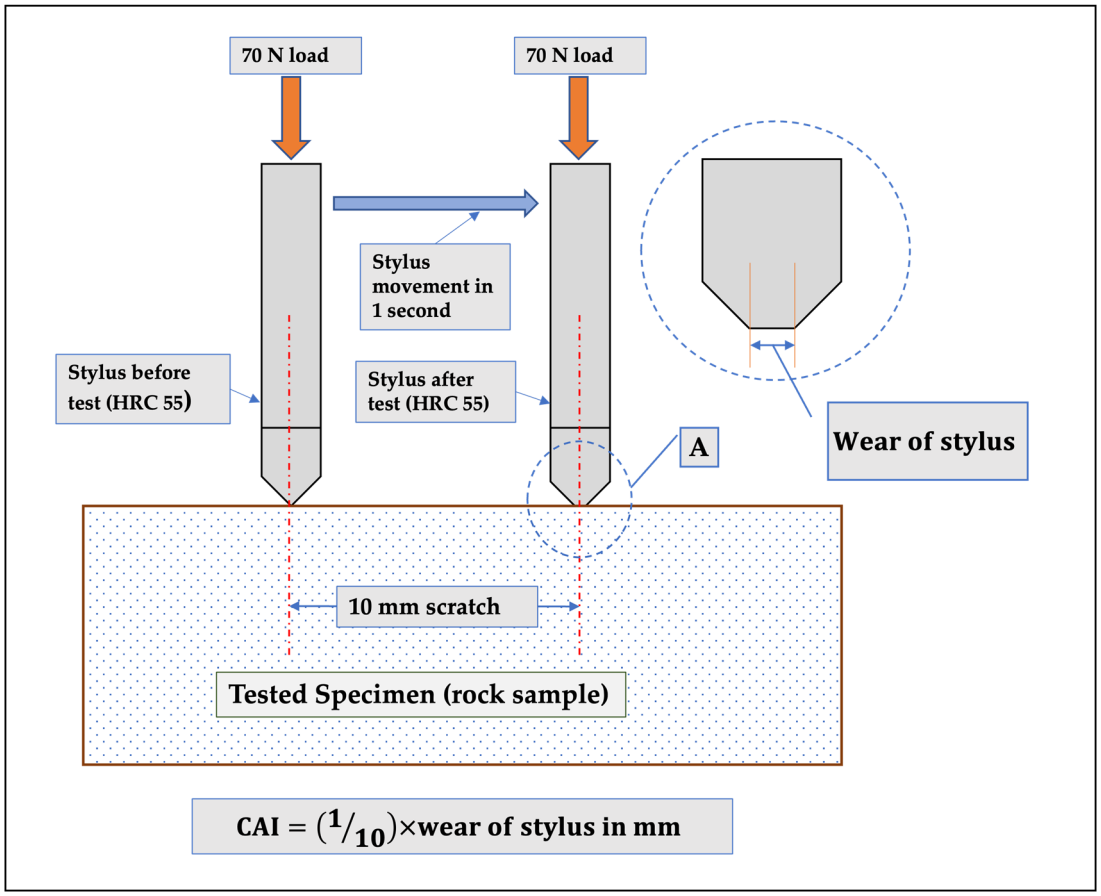


Figure 1. CERCHAR abrasion test principle, detail view 'A' shows the magnified image of the stylus tip wear (for West-type equipment; scratch velocity should be at least 1 mm s^{-1}).

Mechanical and petrographic indicators have long been linked to wear behavior. West et al. [5] grouped abrasiveness measures into petrological indicators (e.g., Mohs hardness, Rosiwal hardness, quartz content) and mechanical tests (e.g., hacksaw test, Tabor test, CAI). Yilmaz et al. [6] reported linear relationships between Specific Wear

Rate (SWR) and composite hardness measures such as Mohs hardness, Vickers hardness, and Rosiwal hardness. Empirical wear prediction using Equivalent Quartz Content (EQC), Rock Abrasivity Index (RAI), and CAI is common practice [7–10].

Despite its widespread use, CAI exhibits substantial variability. Michalakopoulos et al. [11] showed that stylus hardness strongly influences CAI, with linear correlations between CAI measured using 40 HRC and 55 HRC styli, while Gharahbagh et al. [12] reported similar trends on rough surfaces. Stanford and Hagan [13] found that CAI remains stable when stylus hardness is held within a narrow range (50–55 HRC), but systematic variation from 10 to 60 HRC produces linear increases in CAI. Suana and Peters [7] linked mineralogy, quartz equivalence, and grain size to abrasiveness, and Plinninger and Thuro [8] demonstrated that broken surfaces yield higher CAI values than sawn surfaces, highlighting the influence of surface preparation. Thuro and Plinninger [14] showed that most stylus wear occurs within the first 2 mm of the scratch, underscoring the importance of scratch length and initial contact conditions. Additional studies have explored geological and mechanical influences, including grain size and foliation [15], confining pressure [16,17], and scratch velocity [18], all of which affect measured CAI. The effect of reusing the styli, along with the effects of scratch length and confining pressure, was studied and reported for natural sedimentary rocks and concrete rock analogs by Rana et al. [19]; improvements in the after-test styli wear flat measurement technique were also proposed. Recent work by Zhang et al. [20] further highlighted the influence of scratch velocity on CAI variability. Using a newly developed CAI apparatus with controlled and constant scratch speed, Zhang demonstrated that stabilizing scratch velocity significantly reduces operator-dependent scatter and improves inter-laboratory repeatability. This reinforces the need for tighter control of operational parameters in CAI testing. Collectively, these findings indicate that CAI is highly sensitive to both operational parameters and sample preparation. The requirement for machined, polished surfaces and controlled stylus hardness, combined with the influence of confining stress, scratch length, and surface condition, makes it difficult to obtain consistent results across different laboratories, equipment types, and operators. As a result, CAI databases often exhibit significant scatter, and empirical correlations between CAI and rock properties tend to be geology-specific and lack generalizability. Mineralogical correlations have also been widely investigated. West et al. [10] observed that quartz content in coal measures correlates with CAI, though predictive strength remained limited. Thuro et al. [21] identified quartz-rich gravels as major contributors to LCPC abrasiveness, with grain size and angularity playing significant roles. Cheshomi et al. [22] and Hosseini et al. [23] established relationships between CAI and strength parameters, reporting increases in CAI with rising quartz content, longitudinal wave velocity, Brazilian tensile strength, and UCS. Hosseini et al. also proposed a correlation between CAI and Schmidt hammer rebound number for tuff samples, suggesting potential for non-destructive hardness-based prediction. Additional correlations involving Schmidt rebound hardness, V_p, UCS, and BTS have been explored by Yarali [24], Altindag [25], and Capik and Yilmaz [26].

Machine learning (ML) approaches have recently gained traction in rock mechanics due to their ability to handle multicollinearity and integrate large datasets. Ghademejad and Esmaeili [27] applied ML to relate CAI, Leeb hardness, and XRD results for 159 rock types. Heydarian et al. [28], Choi and Ko [29], Wani et al. [30], and Tripathy et al. [31] used ML to investigate mechanical and mineralogical relationships in sedimentary rocks. Zhang et al. [32] explored CAI prediction using drilling parameters and rock properties. Beyond CAI-focused studies, machine learning has seen increasing adoption in broader geotechnical and rock-engineering applications, demonstrating its suitability for complex, nonlinear prediction tasks. Recent work has applied hybrid and ensemble ML models to practical engineering problems, including seismic performance assessment of under-

ground structures using a hybrid MFO-XGBoost framework, lateral capacity estimation of partially rock-socketed monopiles through combined finite-element and ML approaches, and empirical prediction of soil-nail bond strength in weathered soils using gene-expression programming. These studies illustrate the growing maturity of ML techniques in real-world geomechanics and support their potential for developing reliable, transferable models for rock abrasiveness prediction [33–35].

Existing empirical correlations between CAI and rock properties—such as UCS, EQC, V_p, BTS, or Schmidt rebound hardness—are typically geology-specific, linear, and lack generalizability across rock types. Few studies develop a robust, multi-geology CAI prediction framework validated across multiple seeds, multiple cross-validation schemes, and a large multi-source dataset. This study evaluates whether CAI can be reliably predicted from geotechnical and geophysical properties that are routinely obtained through in situ testing or standard core characterization. A dataset of 175 igneous, sedimentary, and metamorphic rocks was compiled from publicly available CAI databases and laboratory sources. A rigorous validation pipeline was implemented, including Leave-One-Out Cross-Validation, K-Fold and Repeated K-Fold validation, and hyperparameter tuning using Nested K-Fold across seven regression models and multiple random seeds. A four-predictor model using Schmidt hammer rebound hardness, Mohs hardness number, longitudinal wave velocity (V_p), and grain size (GS) is proposed. These predictors are non-destructive, field-deployable, and eliminate the need for prepared CAI samples, offering a practical alternative for estimating rock abrasiveness in drilling and excavation applications.

2. Materials and Methods

2.1. Dataset Compilation

2.1.1. Main Dataset Compilation

Two datasets are used in this study. The first dataset is called "Main dataset", and the other dataset is called "Dataset 1". The main dataset contains 175 rock samples representing a wide spectrum of igneous, sedimentary and metamorphic lithologies. Each rock class exhibits characteristic mineralogical and geo-mechanical signatures that influence its abrasiveness and mechanical behavior. The dataset integrates samples from the open-source Norwegian Database (NSD) and individual sources referenced in the paper [36–43]. Information is not available about the equipment used for collecting the main dataset. The main dataset includes 68 individual igneous rock specimens, including granite, granodiorite, diorite, gabbro, basalt, andesite, rhyolite, peridotite, syenite, and tuff. They exhibit high UCS values ranging from 150 to 270 MPa, reflecting strong interlocking crystalline textures. Longitudinal wave velocity is towards the higher end of the spectrum for main dataset ranging between 5000 and 6600 ms^{-1}. CAI varies from 3 to 6, driven by quartz, feldspar and mafic minerals. Tuffs are the softest igneous rocks, showing a CAI between 2.8 and 3.1. Quartz content varies widely, with granites (28–35%), rhyolites (25–27%), quartz-poor basalts (5–8%), and ultramafic peridotites (3–5%). Grain size varies from fine-grained basalts (0.4–0.5 mm) to coarse granites (3–4 mm). The sedimentary portion of the dataset consists of 54 specimens, including limestone, dolomite, travertine, shale, sandstone, quartz sandstone, and marble. These rocks exhibit low to moderate UCS values, typically between 70 and 150 MPa for carbonates and 110 to 150 MPa for sandstones, reflecting variations in cementation, porosity and depositional environment. Longitudinal wave velocities range from 3800 to 5200 ms^{-1}, consistent with the higher porosity and layered structures characteristic of sedimentary formations. CAI values span from 1.5 to 4.8, strongly influenced by quartz content and grain angularity. Carbonate rocks such as limestone, dolomite, travertine, and marble contain only 0–3% quartz and therefore show very low abrasiveness, whereas sandstones with 25–30% quartz exhibit moderate CAI values. Quartz sandstones,

containing 70–80% quartz, are the most abrasive sedimentary rocks, with CAI values approaching those of igneous lithologies. Grain size ranges from 0.4 to 3.9 mm, with coarser sandstones generally showing higher abrasiveness. The dataset includes 53 metamorphic rock specimens, comprising gneiss, schist, quartzite, amphibolite, phyllite, and migmatite. These rocks display high UCS values ranging from 140 to 245 MPa, reflecting recrystallized mineral fabrics and strong interlocking textures formed under elevated temperature and pressure. Longitudinal wave velocities lie between 5000 and 6250 ms^{-1}, indicative of dense, low-porosity metamorphic structures. CAI values are consistently high, typically between 4.5 and 6.5, with quartzites occupying the uppermost range due to their nearly pure quartz composition. Quartz content varies significantly across metamorphic types, from 92 to 99% in quartzites to 35–43% in gneiss and 34–39% in schist, influencing their abrasiveness and mechanical behavior. Grain size ranges from very fine (0.2 mm) in quartzites to coarser textures in gneiss and schist. Quartzites are the hardest and most abrasive rocks in the dataset, while gneiss and schist exhibit anisotropic mechanical behavior but consistently high abrasiveness due to quartz-rich banding and mica-bearing foliation. Amphibolite shows high hardness and velocity, reflecting hornblende–plagioclase mineralogy and dense metamorphic textures.

The main dataset integrates 175 rock specimens sourced from various sources, each contributing distinct lithological groups and measurement completeness. NSD provides the broadest coverage, including igneous (granite, basalt, gabbro, andesite, rhyolite, peridotite, tuff), sedimentary (limestone, dolomite, travertine, shale, sandstone, marble), and metamorphic rocks (gneiss, schist, quartzite, amphibolite, phyllite, migmatite), with nearly complete measurements across CAI, UCS, V_p, density, grain size, and hardness. Both Plinninger [36], Plinninger and Thuro [37] contribute well-characterized igneous and metamorphic rocks with full mineralogical and mechanical properties, while Gokceoglu's samples [38] similarly contain complete geomechanical fields. In contrast, Kahraman's dataset [39] exhibits systematic missing values in Schmidt rebound hardness (H), Shore hardness type D (R), and grain size (GS), particularly for basalts, granites, gneiss, and quartzites, requiring imputation. Thuro's samples [40] show missing Schmidt rebound values (H) and grain size entries, and Suana's [43] quartz sandstones are complete except for minor variability in secondary parameters. Yarali's sedimentary rocks [41] and Yavuz [42] are fully populated with no missing fields.

Mohs hardness number is estimated using the available information on mineral content and their respective scratch hardness.

$$H_{rock} = \sum_{i=1}^{n} F_i \times H_i \quad (1)$$

where F_i is the Volume/Weight fraction of the mineral and H_i is the Mohs hardness of the mineral. For calculating dynamic elastic moduli, the following equations are used. Wave velocity was calculated as:

$$V = \frac{L}{t} \quad (2)$$

where V is velocity (ms^{-1}), L is path length, and t is signal travel time.

Dynamic elastic properties (Shear modulus, G; Bulk modulus, K; Young's modulus, E; Poisson's ratio, υ) were calculated using the measured V_p and V_s, and bulk density ρ for Dataset 1.

$$G = \rho V_s^2 \quad (3)$$

$$K = \rho\left(V_p^2 - \frac{4}{3}V_s^2\right) \quad (4)$$

$$E = \frac{9KG}{3K + G} \tag{5}$$

$$\upsilon = \frac{3K - E}{6K} \tag{6}$$

For the main dataset, shear wave-related information and Poisson's ratio were estimated using the following formulas:

Christensen [44]:

$$\upsilon = \frac{1}{2}\left(1 - \frac{2}{\left(V_p / V_s\right)^2 - 1}\right) \tag{7}$$

This model is best for igneous and metamorphic rocks.

Brocher [45]:

$$\upsilon = 0.32 - 0.33\left(\frac{V_p}{V_s}\right) \tag{8}$$

This model is good for mixed-crystal rock.

Barton [46]:

$$\upsilon = \frac{E}{2G} - 1 \tag{9}$$

This model is good for metamorphic and crystalline rocks.

Han [47]:

$$\upsilon = f\left(V_p, \varnothing, C_{clay}\right) \tag{10}$$

where Φ is the porosity of the rock and C_{clay} is the clay content of the rock measured in fraction or percentage. This model is good for sedimentary rocks.

Rafavich [48]:

$$\upsilon = \frac{1}{2}\left(1 - \frac{2V^2}{{V_p}^2 - {V_s}^2}\right) \tag{11}$$

This model is excellent for carbonates.

Castagna's Murdock [49]:

$$V_s = \left(0.862 \times V_p\right) - 1.172 \tag{12}$$

$$\upsilon = \frac{1}{2}\left(\frac{\left(V_p / V_s\right)^2 - 2}{\left(V_p / V_s\right)^2 - 1}\right) \tag{13}$$

These multiple models allowed cross-validation of reconstructed elastic parameters, and the resulting moduli were compared against published ranges for each lithology to ensure consistency with expected mechanical behavior. This approach provided a defensible reconstruction of dynamic elastic properties across diverse rock types using a combination of measured ultrasonic data, mineralogical information, and established empirical relationships.

2.1.2. Dataset 1

The smaller laboratory dataset consists of 12 rock specimens sourced from [50,51], representing a focused selection of lithologies, including limestone, fine-grained and coarse-grained gabbro, various sandstone types, granite variants, quartzite, and low-, medium-, and high-strength concrete analogs. These rocks span a wide mechanical range, with UCS values from 20 MPa in low-strength Rock Like Material (L-RLM) to 180 MPa in quartzite, reflecting differences in mineralogy, cementation and metamorphic overprint. Longitudinal wave velocities vary between 3560 and 6335 ms^{-1}, while shear velocities range from 1975 to

3660 ms^{-1}, indicating contrasting densities and elastic frameworks. CAI values range from 0.68 in weakly cemented sandstones to 4.53 in quartzite, consistent with increasing quartz content and crystalline interlocking. Elastic moduli also show strong variability, with shear modulus spanning 9–39 GPa and Young's modulus ranging from 23 to 94 GPa. Mohs hardness increases systematically from softer carbonates (~3) to quartz-rich rocks (~7). Overall, Dataset 1 captures a compact but geologically diverse set of rocks. Figure 2 shows CERCHAR test equipment used to measure CAI for Dataset 1, along with the styli used, wear scratch measurement and the styli wear measurement setup.

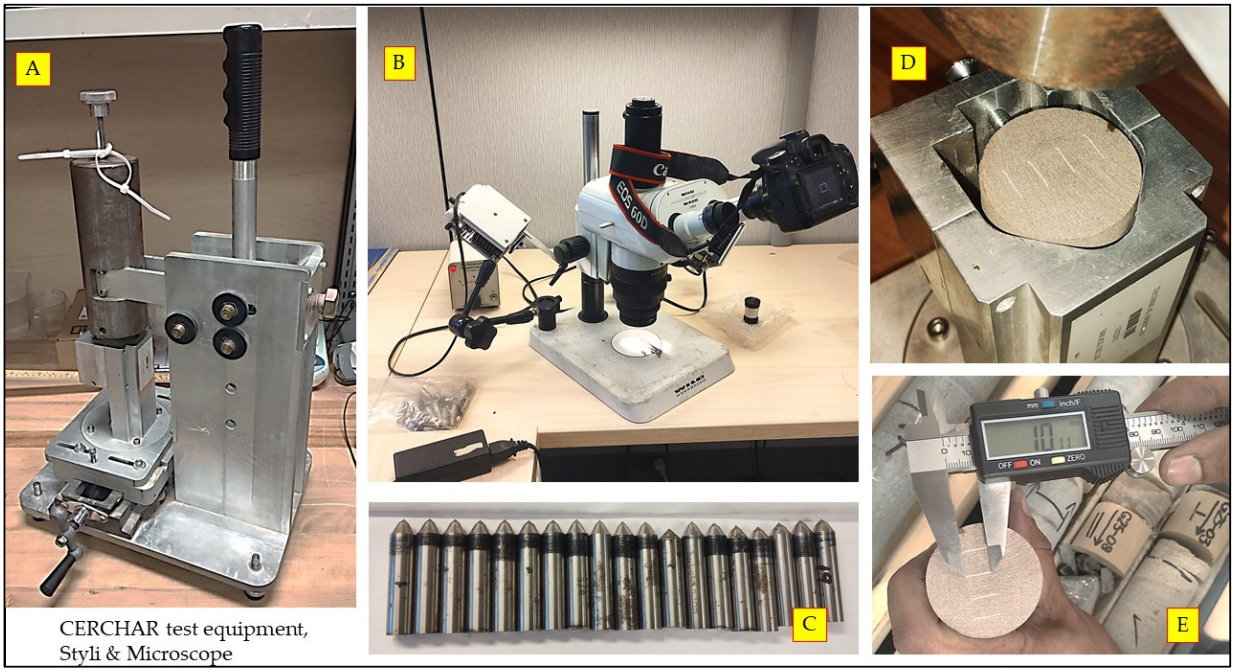
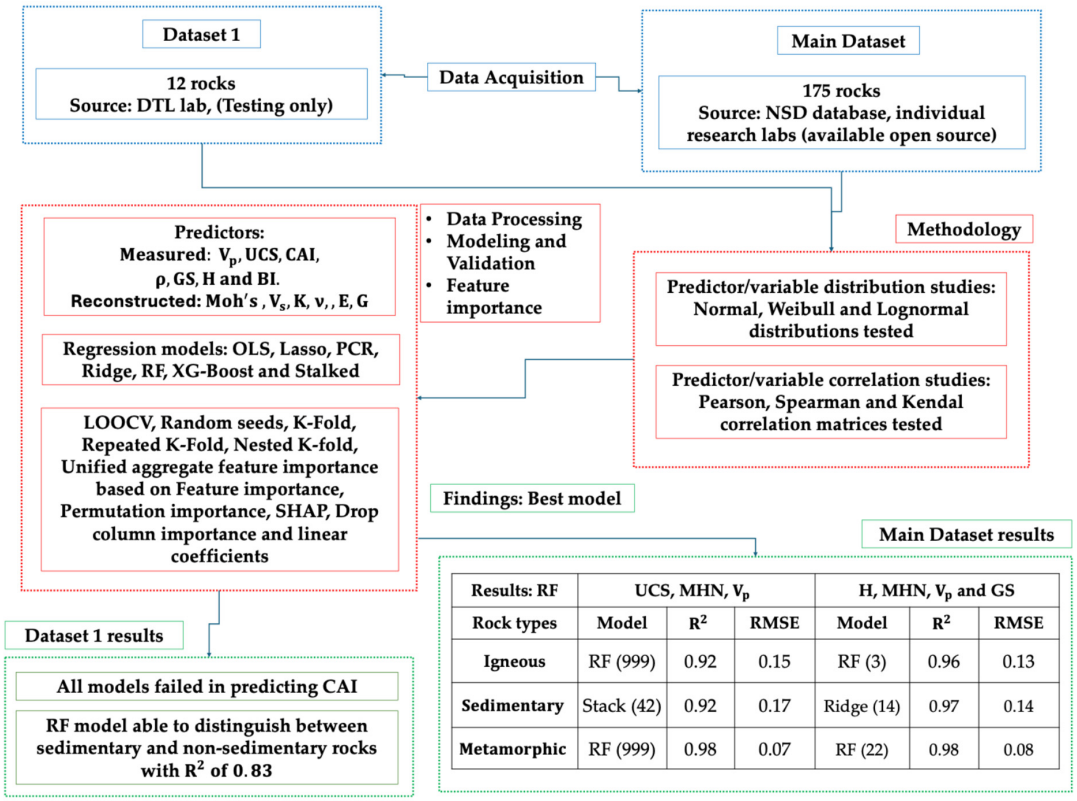


Figure 2. (**A**) CERCHAR test equipment used to measure CAI for Dataset 1; (**B**) setup to measure wear flat of stylus; (**C**) prepared styli (HRC 55); (**D**) red sandstone sample with test scratch marks; (**E**) scratch length measurement in mm.

For measuring UCS for Dataset 1, an Instron 6800 series universal testing machine was used. For ultrasonic velocity measurements, Evident V-153 (Evident Corporation, Waltham, MA, USA) type transducers were used in conjunction with an oscilloscope (Tektronix TD 1002, Tektronix Inc., Beaverton, OR, USA) and pulser receiver unit (Olympus 5070 PR, Olympus Corporation of the Americas, Waltham, MA, USA) was used. Standard lab equipment was used and related ASTM standards were followed [52–55]. Although ASTM D2845 [55] for ultrasonic pulse velocity in rocks has been formally withdrawn, its methodology remains widely used and technically robust; therefore, the same measurement principles involving measuring direct travel time over a known rock sample length. Figure 3 shows the flowchart followed for conducting the current research.

The workflow begins with two independent datasets: a small laboratory dataset containing 12 rocks tested at the DTL lab, and a larger main dataset of 175 rocks compiled from NSD and open-source laboratory studies. Both datasets include directly measured predictors such as V_p, UCS, CAI, density, grain size, Shore hardness, and brittleness index, along with reconstructed properties including Mohs hardness, V_s, bulk modulus, Poisson's ratio, Young's modulus, and shear modulus. After acquisition, the data undergo distribution analysis (normal, Weibull, and lognormal) and correlation studies using Pearson, Spearman, and Kendall matrices. A suite of regression and machine learning models—OLS, Lasso, PCR, Ridge, Random Forest, XGBoost, and stacked ensembles—is trained and validated us-

ing LOOCV, random seeds, K-Fold, repeated K-Fold, and nested K-Fold strategies. Unified feature importance analysis is performed through permutation importance, SHAP values, drop-column tests, and linear coefficients. The workflow concludes with model evaluation for each rock type, highlighting the best-performing models and their predictive accuracy.



Results: RF	UCS, MHN, V_p			H, MHN, V_p and GS		
Rock types	Model	R^2	RMSE	Model	R^2	RMSE
Igneous	RF (999)	0.92	0.15	RF (3)	0.96	0.13
Sedimentary	Stack (42)	0.92	0.17	Ridge (14)	0.97	0.14
Metamorphic	RF (999)	0.98	0.07	RF (22)	0.98	0.08

Figure 3. Flowchart showing the outline of the process followed for the presented work, arrows denote the sequential relationships and information transfer between components of the methodology.

2.2. Predictor Selection

2.2.1. Main Dataset Predictor Selection

Predictor selection was carried out using a combination of statistical correlation analysis, geological reasoning, and model-based importance evaluation applied to both datasets. The initial pool of variables included all measured and reconstructed properties: UCS, density, longitudinal and shear wave velocities, grain size, Schmidt rebound hardness, brittleness index, mineral percentages, elastic moduli, Poisson's ratio, and Mohs hardness. Distribution studies showed that several predictors exhibited strong geological grouping, with quartz-rich rocks consistently clustering at higher CAI values and carbonates forming the lowest-abrasiveness end of the spectrum. Correlation matrices (Pearson, Spearman, and Kendall) revealed stable, repeatable relationships between CAI and a subset of predictors, particularly Schmidt rebound hardness, Mohs hardness, V_p, and grain size, while variables such as density, Poisson's ratio, and bulk modulus showed weak or geology-dependent correlations. Predictors with unstable behavior, high missingness, or inconsistent geological trends, such as Shore hardness, brittleness index, and some reconstructed elastic parameters, were excluded from the final reduced model. The resulting predictor set balances geological relevance, measurement practicality, and statistical robustness, forming a compact group of variables capable of capturing the mineralogical and mechanical controls on rock abrasiveness across diverse lithologies. Figure 4 shows the Normal, Weibull and Lognormal distribution plots for individual predictors for the main dataset.

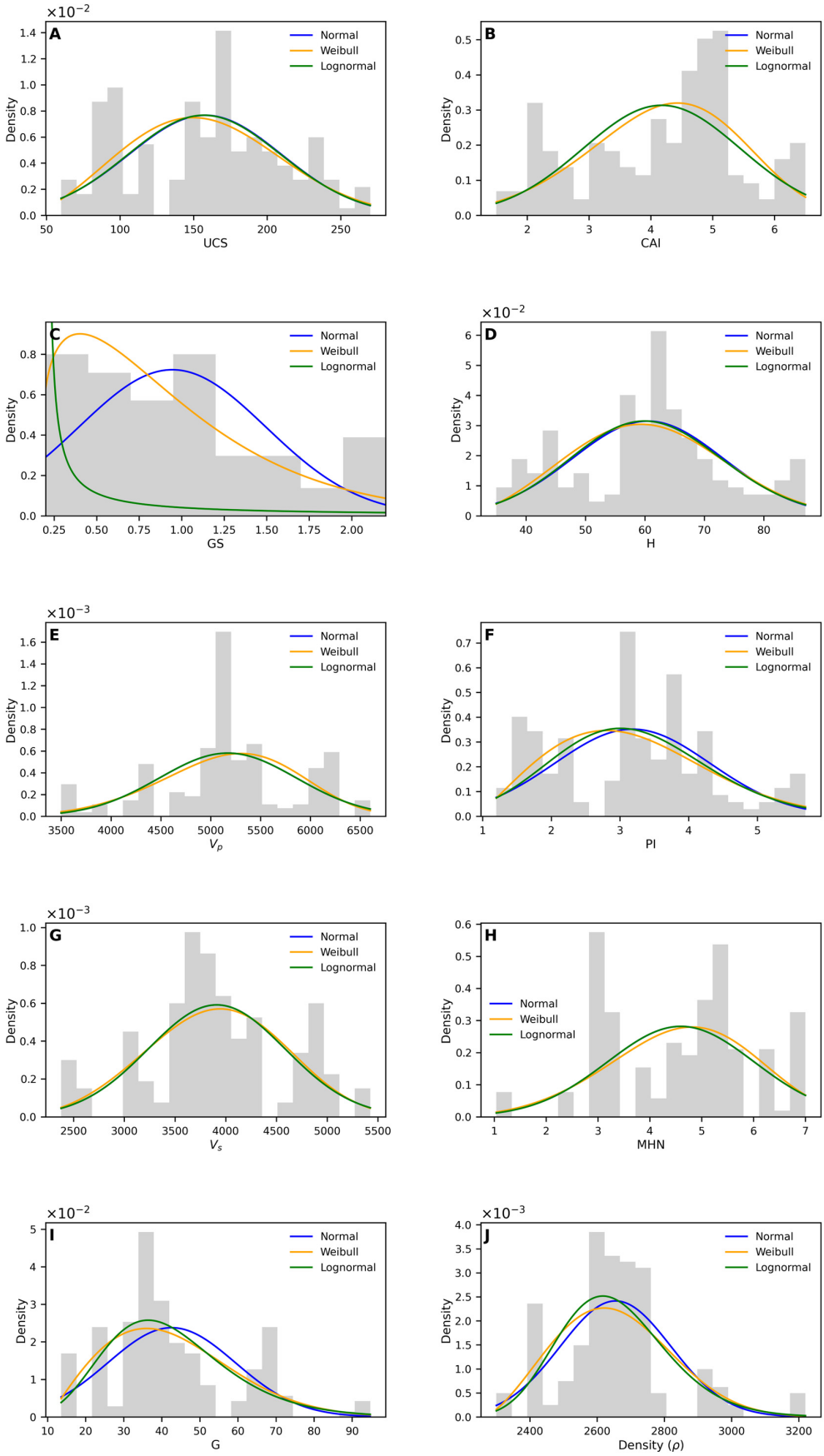


Figure 4. Normal, Weibull and Lognormal distributions for the main dataset for (**A**) UCS, (**B**) CAI, (**C**) density, (**D**) Schmidt rebound hardness, (**E**) longitudinal wave velocity, (**F**) petrographic index, (**G**) shear wave velocity, (**H**) Mohs hardness number, (**I**) Dynamic shear modulus and (**J**) density.

Predictor selection for the main dataset was guided by a combination of correlation structure, distribution behavior, and model-based importance analysis. The Spearman correlation heatmap (Figure 5) shows that hardness-related variables (Schmidt rebound hardness H, Mohs hardness, and elastic moduli) exhibit strong positive correlations with CAI, while density, Poisson's ratio, and bulk modulus display weak or geology-dependent relationships. Quartz content correlates positively with CAI across igneous, sedimentary, and metamorphic rocks, whereas calcite content shows a clear negative trend, reinforcing the mineralogical control on abrasiveness. Distribution-fit analyses further informed predictor stability: variables such as UCS, V_p, MHN, V_s, and shear modulus (G) follow lognormal or Weibull-type distributions rather than normal behavior, indicating skewed geological variability that favors predictors with consistent right-tailed distributions. In contrast, parameters like grain size and petrographic index show broader, less stable distributions, reducing their reliability as primary predictors. The combined interpretation of correlation strength and distribution robustness highlighted a compact group of variables: Schmidt rebound hardness, Mohs hardness, longitudinal wave velocity, and grain size—as the most geologically meaningful and statistically stable predictors. These variables consistently captured the mineralogical and mechanical controls on CAI across all lithologies, while predictors with weak correlations, multimodal distributions, or high missingness were excluded from the final reduced model. Visual diagnostics: trend plots, pairwise relationships, and response patterns, were used to confirm that each selected predictor exhibited a consistent and interpretable influence on the target variable.

2.2.2. Dataset 1 Predictor Selection

The distribution-fit analyses show that UCS, CAI, V_p, V_s, density, Mohs hardness, and shear modulus (G) all follow distinctly skewed lognormal or Weibull-type distributions rather than normal behavior, reflecting the natural variability of strength and elastic properties across carbonates, sandstones, granites, and quartzites. These skewed distributions indicate that the dataset captures genuine geological contrasts but lacks the breadth needed for stable multivariate modeling. Although hardness-related parameters and wave velocities display tight right-tailed distributions that make them reliable indicators of abrasiveness, several predictors, particularly density and elastic moduli in weaker sandstones and recrystallized limestones, show broad or unstable behavior, reducing their discriminatory power. Because Dataset 1 contains only 12 samples, its limited lithological diversity and small sample size make it unsuitable for training regression models without risking overfitting or spurious correlations. Instead, it provides a controlled environment for testing model behavior, evaluating predictor sensitivity, and validating geological trends observed in the larger main dataset. For these reasons, Dataset 1 was used exclusively as a test-only dataset, allowing the main dataset to drive model training while Dataset 1 served as an independent check on predictor robustness and model generalization. Figure 6 shows the Normal, Weibull and Lognormal distribution plots for individual predictors for Dataset 1.

2.3. Data Preprocessing

Missing values in the main dataset were handled using a targeted imputation strategy designed to preserve geological trends while avoiding artificial bias. Variables with systematic gaps (primarily Schmidt rebound hardness, Shore hardness number, and grain size) in samples from Kahraman [39] and Thuro [40] were imputed using lithology-specific medians, ensuring that each rock type retained its characteristic mechanical behavior. For continuous predictors with small, random gaps, distribution-aware imputation was applied, using values drawn from the fitted lognormal or Weibull distributions identified

during the distribution-fit analysis. This approach maintained the natural skew present in strength and velocity parameters. No global mean imputation was used, as it would distort mineralogical contrasts between carbonates, sandstones, and crystalline rocks. Overall, the imputation process ensured a complete predictor matrix while preserving the geological integrity of each lithology group. In addition, post-imputation diagnostics were performed to verify that the filled values did not introduce artificial clustering or shift the central tendencies of key mechanical variables. Density, UCS, and P-wave velocity distributions were compared before and after imputation to confirm that lithology-dependent trends remained intact. Correlation matrices were also re-evaluated to ensure that imputed values did not inflate or suppress relationships among predictors. Finally, all imputed fields were flagged within the preprocessing pipeline, allowing downstream model evaluation to explicitly track the influence of imputed versus measured data.

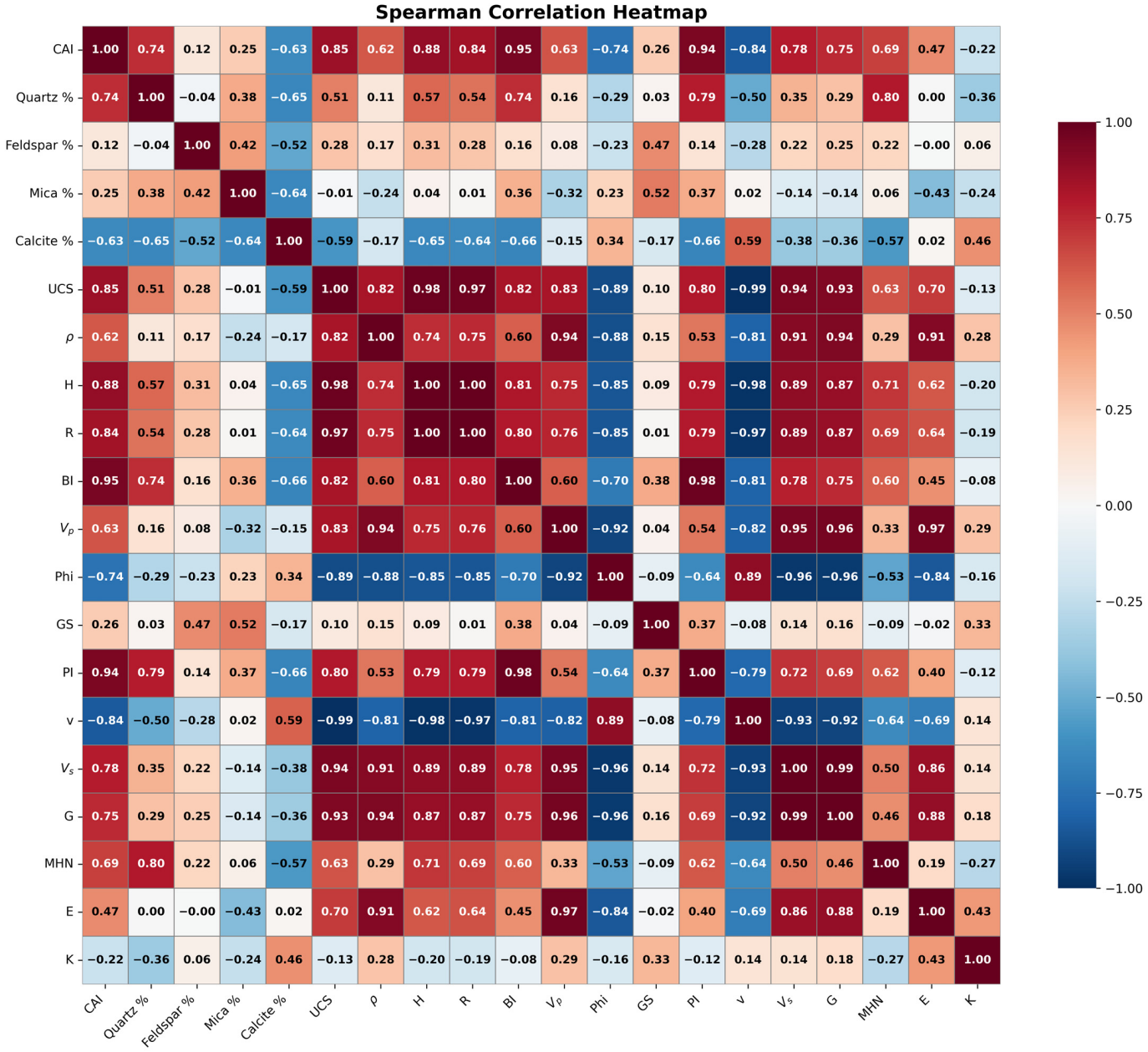


Figure 5. Spearman correlation heatmap for main dataset.

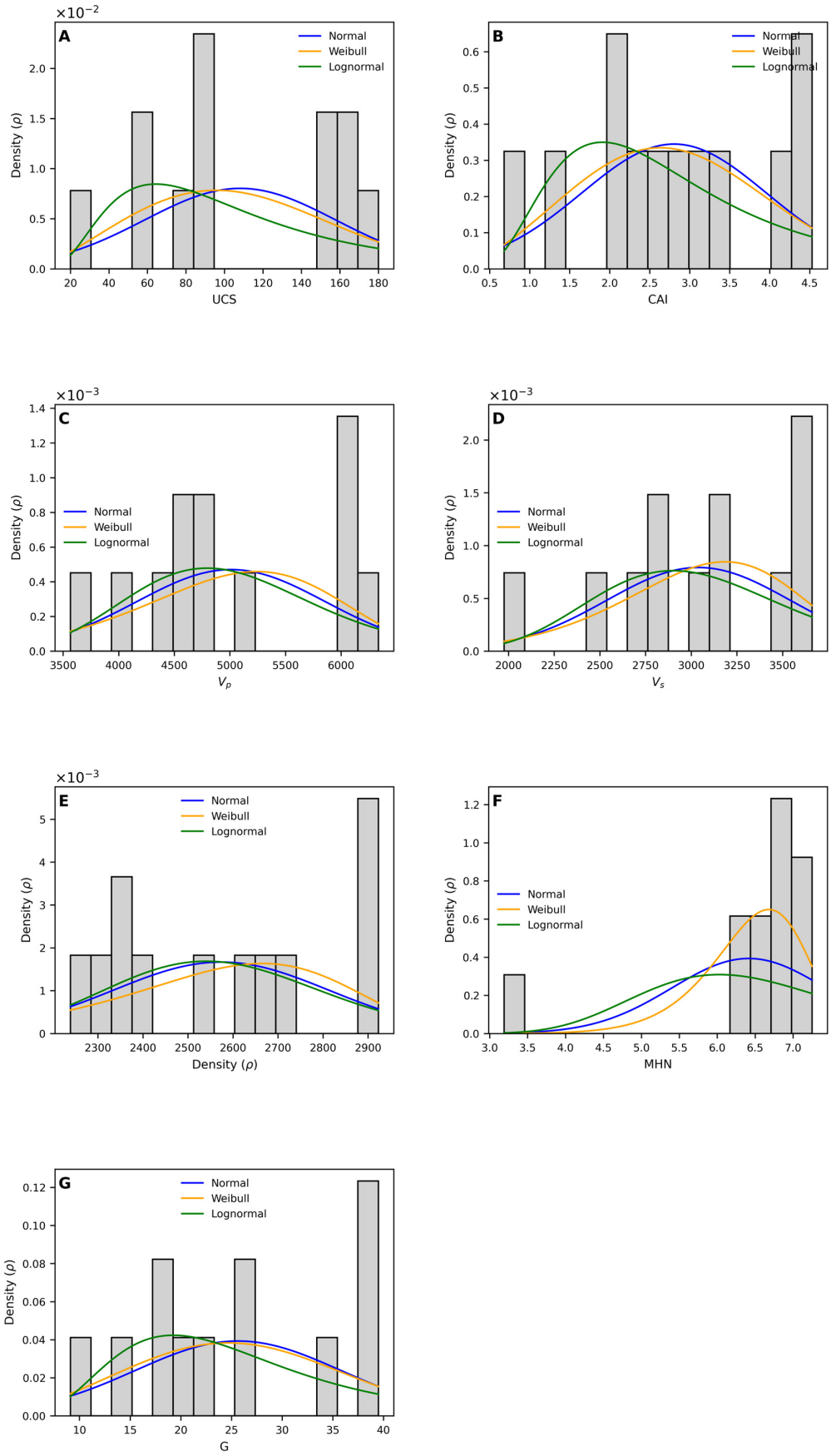


Figure 6. Normal, Weibull and Lognormal distributions for Dataset 1 for (**A**) UCS, (**B**) CAI, (**C**) longitudinal wave velocity, (**D**) shear wave velocity, (**E**) density, (**F**) Mohs hardness number and (**G**) dynamic shear modulus of elasticity.

2.4. Regression Models

A suite of regression models was implemented to evaluate the predictive relationships between rock properties and the CERCHAR Abrasivity Index (CAI), ranging from simple linear formulations to advanced ensemble methods. Linear models, including Ordinary Least Squares, Ridge, Lasso, and Principal Component Regression, were used to assess baseline linearity, quantify coefficient stability, and examine how multicollinearity among mineralogical and mechanical predictors influenced model behavior. These were complemented by non-linear models such as Random Forest and XGBoost, which capture complex interactions between hardness, wave velocities, grain size, and mineral composition without requiring distributional assumptions. A stacked hybrid model combining linear and tree-based learners was also developed to integrate complementary strengths of both model families. All models were trained exclusively on the main dataset, while Dataset 1 served as an independent test set to evaluate generalization and prevent overfitting. This multi-model approach allowed robust comparison of linear versus non-linear predictor influence and ensured that the final framework captured both geological trends and statistical structure present across diverse lithologies.

2.5. Performance Metrics

Model performance was evaluated using standard classification accuracy, which measures the proportion of correctly predicted rock classes relative to the total number of samples. Accuracy is an appropriate metric for the main dataset because the test set is small and the class distribution is balanced, allowing direct comparison across models without the need for weighted adjustments. Higher accuracy values indicate that a model successfully captures the underlying relationships between mechanical properties and lithological classes, while lower values reflect limited generalization or sensitivity to the dataset's skewed distributions. Together, these metrics provide a straightforward benchmark for assessing how effectively each algorithm distinguishes rock types based on their measured properties.

2.6. Validation Strategy

Model validation was designed to ensure reliable generalization across the diverse lithologies represented in the main dataset while preventing information leakage into the independent test set. The main dataset was split using stratified K-Fold cross-validation, preserving the proportional representation of igneous, sedimentary, and metamorphic rocks within each fold. This approach allowed the models to be evaluated on multiple unseen subsets while maintaining geological balance. All preprocessing steps, including imputation, scaling, and encoding, were fitted exclusively on the training folds and applied to the validation folds to avoid bias. Dataset 1 was reserved entirely as an external test set, providing an independent assessment of model performance under laboratory-controlled conditions. This two-tier validation strategy ensured that the models were evaluated both internally through repeated cross-validation and externally through a fully isolated test dataset, offering a robust measure of predictive stability and generalization.

2.7. Model Selection Criteria

Model selection was based on a combination of predictive accuracy, stability across cross-validation folds, and the ability to capture non-linear relationships inherent in geological data. Algorithms were first screened using stratified K-Fold validation on the main dataset, and models showing inconsistent performance or high variance across folds were discarded. Emphasis was placed on models that could handle skewed, lognormal-type distributions and complex interactions between hardness, wave velocities, and mineralogical

parameters. Ensemble and kernel-based methods were prioritized when they demonstrated superior generalization, while simpler linear and instance-based models were retained only when they provided interpretable baselines.

2.8. Implementation Environment

All data processing, statistical analysis, and model development were carried out in a Python-based environment using standard scientific computing libraries. The workflow was implemented in Jupyter Notebook 6.5 running Python 3.10, with NumPy 1.26 and Pandas 2.1 used for data handling, SciPy 1.11 for statistical analysis, and scikit-learn 1.3 for regression, classification, cross-validation, and performance evaluation. Visualization and distribution-fit assessments were performed using Matplotlib 3.8 and Seaborn 0.13. XGBoost 1.7 was executed through its dedicated Python 3.10 API, and all experiments were run on a laptop and workstation with sufficient CPU resources to support ensemble and kernel-based models. This unified environment ensured reproducible preprocessing, consistent model training, and seamless integration of statistical and machine learning components.

3. Results

3.1. Results of Main Dataset

The results are presented in a progressive sequence that reflects increasing model complexity and validation rigor. The analysis begins with LOOCV and source-based validation to establish baseline performance and quantify how individual data sources influence predictive stability. This is followed by a detailed study across the three major lithological groups: igneous, sedimentary and metamorphic, using three predictor configurations (measured predictors, reconstructed predictors and all predictors) and multiple cross-validation schemes, including K-Fold and repeated K-Fold at multiple random seeds and nested K-Fold at a few selected random seeds, to evaluate robustness under varying sampling conditions. The subsequent sections present the outcomes of Script 1, Script 3, and Script 4, each representing increasingly refined modeling pipelines with expanded predictor sets and more advanced evaluation metrics. Finally, the independent Dataset 1 test results are reported, providing an external assessment of generalization and confirming whether trends observed in the main dataset hold under controlled laboratory conditions.

3.1.1. CAI Prediction for Main Dataset LOOCV at Data Source Level

This regression study was performed on the dataset to assess how well the model can predict the CAI for a new dataset that it has not seen before. It comprises nine folds based on nine distinct sources; the main dataset is comprised of these. Table 1 shows that not all the sources rank equally when it comes to their ability to train a model. One of the reasons for this irregularity is the imputation required due to some missing predictors for a few data sources. The following are the predictors used for the regression modeling: UCS, CAI, MHN, BI, PI, Quartz %, Mica %, Calcite %, Feldspar %, density, H, R, V_p, Phi, GS, ν, V_s, G, K and E. Group-wise imputation was opted for based on the unique requirements of the dataset. Table 1 shows the Random Forest regression model results for the main dataset with grouping based on data source type, giving 9 folds in total. The model is trained by keeping a source out of the training dataset and is tested later. The LOOCV results clearly show that CAI predictability depends far more on dataset consistency than dataset completeness. Sources such as Plinninger [36], Plinninger and Thuro [37], Gokceoglu [38], and Yarali [41] achieve extremely high R^2 values because their CAI, mechanical, and petrographic measurements follow uniform testing protocols and exhibit stable inter-property relationships. In contrast, datasets like Yavuz [42] and Suana [43], even though complete,

produce strongly negative R^2 values because their CAI ranges are narrow, their mechanical properties are scaled differently, and their indentation procedures deviate from the dominant protocol, causing the model to extrapolate rather than interpolate. Kahraman [39] and Thuro [40] perform poorly for a different reason; both require substantial imputation due to missing petrographic or mechanical variables, and the imputed values distort the covariance structure that tree-based models rely on, lowering R^2 despite moderate MAE/RMSE. Overall, poor performance arises not from data quantity but from protocol mismatch, measurement heterogeneity, and the need for imputation, all of which break the CAI–mechanical relationships learned from the dominant literature datasets.

Table 1. Regression model results (RF only), main dataset (LOOCV) based on the references/sources of the data.

LOOCV	Source	MAE	RMSE	R^2
0	NSD	0.45	0.73	0.66
1	Plinninger2003	0.09	0.12	0.99
2	Plinninger and Thuro2003	0.07	0.08	0.99
3	Gokceoglu	0.05	0.07	0.99
4	Kahraman	0.87	1.03	0.08
5	Thuro2003	0.62	0.82	0.39
6	Yarali2011	0.12	0.13	0.96
7	Yavuz2008	0.27	0.28	−5.16
8	Suana1982	0.55	0.60	−2.57

Figure 7 shows the agreement-weighted consensus bar chart that integrates feature importance, drop importance, SHAP and permutation importance methods into a single composite importance score, producing a unified ranking of predictors under LOOCV. Features such as BI, Phi, PI, and the lithology indicators emerge with the highest consensus scores, showing that they consistently influence CAI prediction across all importance frameworks. Quartz content, MHN, H, and rebound index also rank highly, reflecting their strong geological connection to abrasiveness. Conversely, variables such as Calcite%, V_p, and Poisson's ratio receive negative or low consensus scores, indicating limited or inverse contribution to CAI prediction under LOOCV. Figure 7 also highlights the effect of multicollinearity on elastic parameters: although UCS, E, G, and K show moderate importance in individual methods, their consensus scores are lower because their influence is not consistently captured across all four frameworks. This reinforces that CAI is more strongly governed by mineralogical and textural attributes than by bulk elastic properties. The clear separation between high-importance mineralogical predictors and low-importance velocity-derived parameters demonstrates that the consensus approach effectively filters out method-specific biases. Additionally, the smooth gradient of bar colors reflects the stability of the ranking, with no abrupt transitions that would indicate disagreement among methods. The negative scores for Calcite% and Poisson's ratio further suggest that these variables may act as suppressors rather than drivers of abrasiveness in the combined dataset. Overall, the consensus plot provides a robust, method-agnostic interpretation of predictor influence, ensuring that the final ranking reflects geological reality rather than artifacts of any single importance technique.

The Spearman agreement matrix (Figure 8) illustrates how consistently different feature importance methods rank predictors when LOOCV is applied across all sources. Random Forest and permutation importance show very strong agreement (~0.96–1.00), indicating that both methods capture similar non-linear relationships between hardness, velocity, and mineralogical variables. Drop-column importance shows moderate positive

correlation with these two methods, reflecting its sensitivity to model structure and predictor interactions. In contrast, SHAP importance exhibits negative correlations with all other methods, highlighting that SHAP captures a different aspect of predictor influence, specifically, local, sample-level contributions rather than global model behavior. Together, the matrix demonstrates that while the methods broadly agree on dominant predictors, SHAP provides a complementary perspective that enriches the overall interpretation.

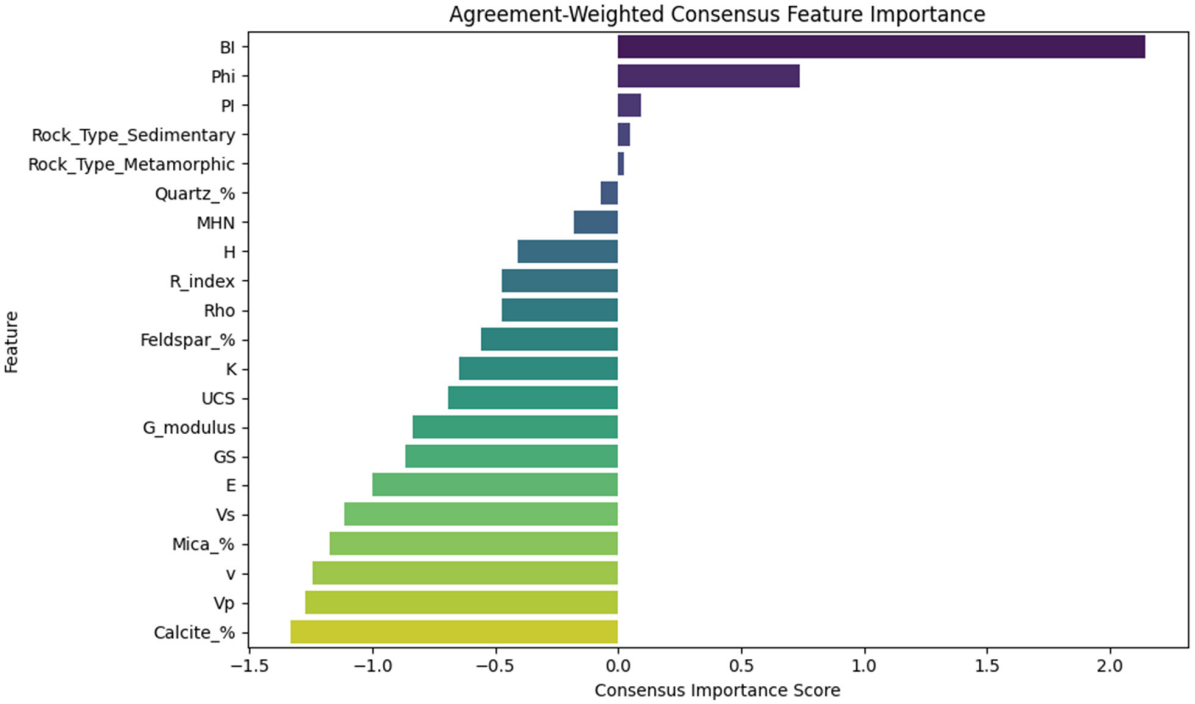


Figure 7. Agreement-weighted consensus feature importance for the main dataset, data-source grouped LOOCV, the color gradient reflects the consensus importance score, where yellow tones correspond to lower importance and darker purple tones indicate higher importance.

3.1.2. Main Dataset Regression Model Results (3 Rock Types and 3 Predictors Set)

The performance trade-off plot (Figure 9) confirms model stability: models cluster tightly in the high-R^2, low-MAE region, with small RMSE values, showing that no model suffers from large error penalties despite differences in complexity. The performance results (Figure 10) show all regression models achieve consistently high predictive accuracy across the different validation schemes, with R^2 values remaining close to unity under K-Fold, repeated K-Fold, and nested K-Fold cross-validation. The bar chart shows minimal variation in mean R^2 across models, indicating that the predictor set is highly informative and that model performance is robust to changes in sampling strategy. The rock-type heatmap (Figure 11) reveals that performance remains strong across igneous, sedimentary, and metamorphic subsets, with ensemble models such as Random Forest, XGBoost, and the stacked model achieving the highest R^2 values, particularly for igneous rocks where nonlinear relationships are more pronounced. Overall, the figures collectively show that the regression framework is stable, generalizes well across lithologies, and maintains high accuracy regardless of the validation scheme applied.

The agreement-based consensus rankings reveal clear differences in which predictors dominate CAI behavior within each lithological group. For igneous rocks (Table 2), grain size (GS) consistently emerges as the strongest predictor across measured and combined sets, with high agreement among RF, permutation, and SHAP importance. The constructed brittleness index (BI) also ranks highly, followed by hardness-related variables such as

MHN and elastic parameters including G, V_s and V_p, though these show lower agreement and weaker contributions. Overall, igneous rocks are primarily governed by textural and brittleness-related predictors.

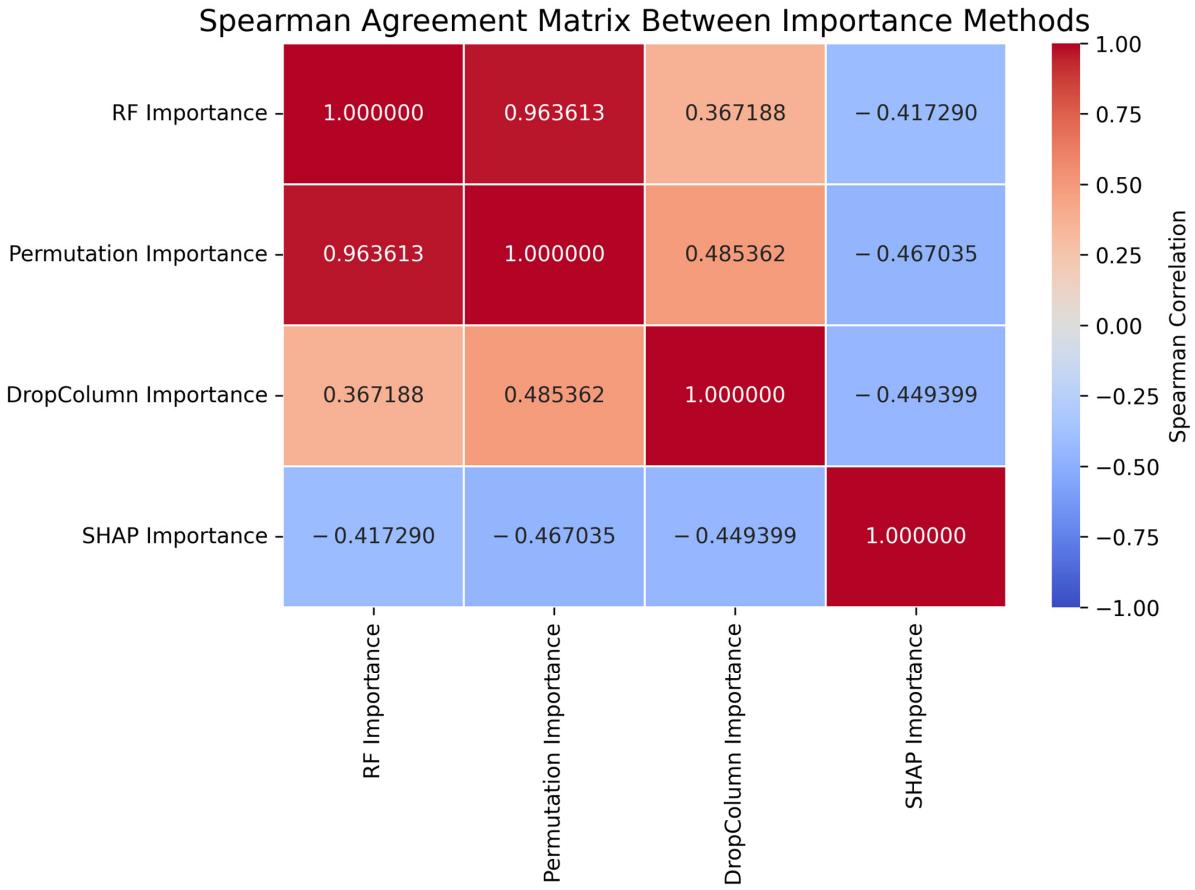


Figure 8. Spearman agreement matrix for RF importance, permutation importance, drop-column importance and SHAP importance for the main dataset, source-based LOOCV.

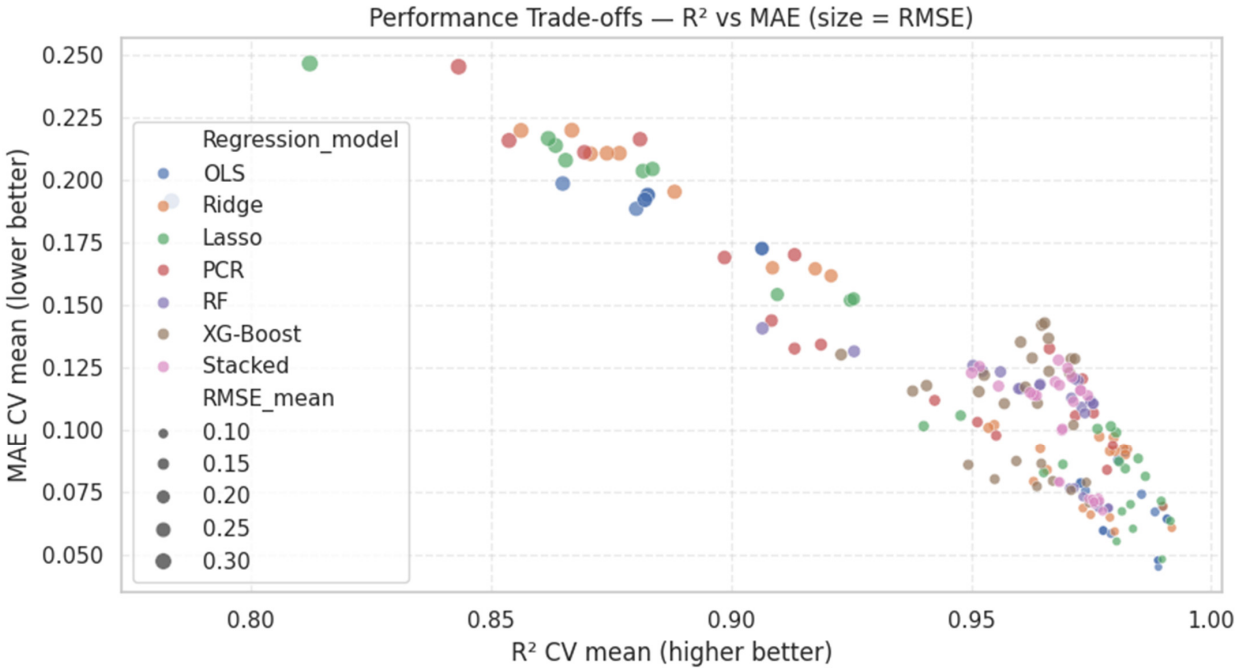


Figure 9. Mean absolute error for regression models (all) on the main dataset.

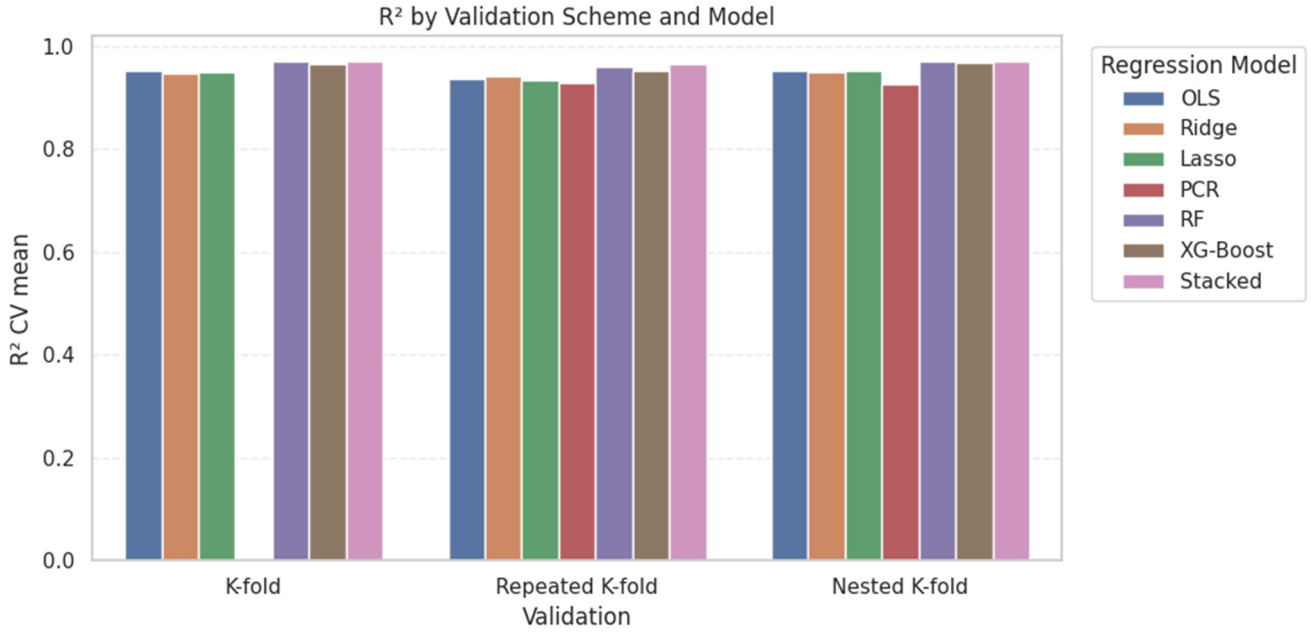


Figure 10. Validation of regression models using K-fold (42, 123 and 999), repeated K-fold (42, 123 and 999) and nested K-fold (42) for the main dataset (random seeds).

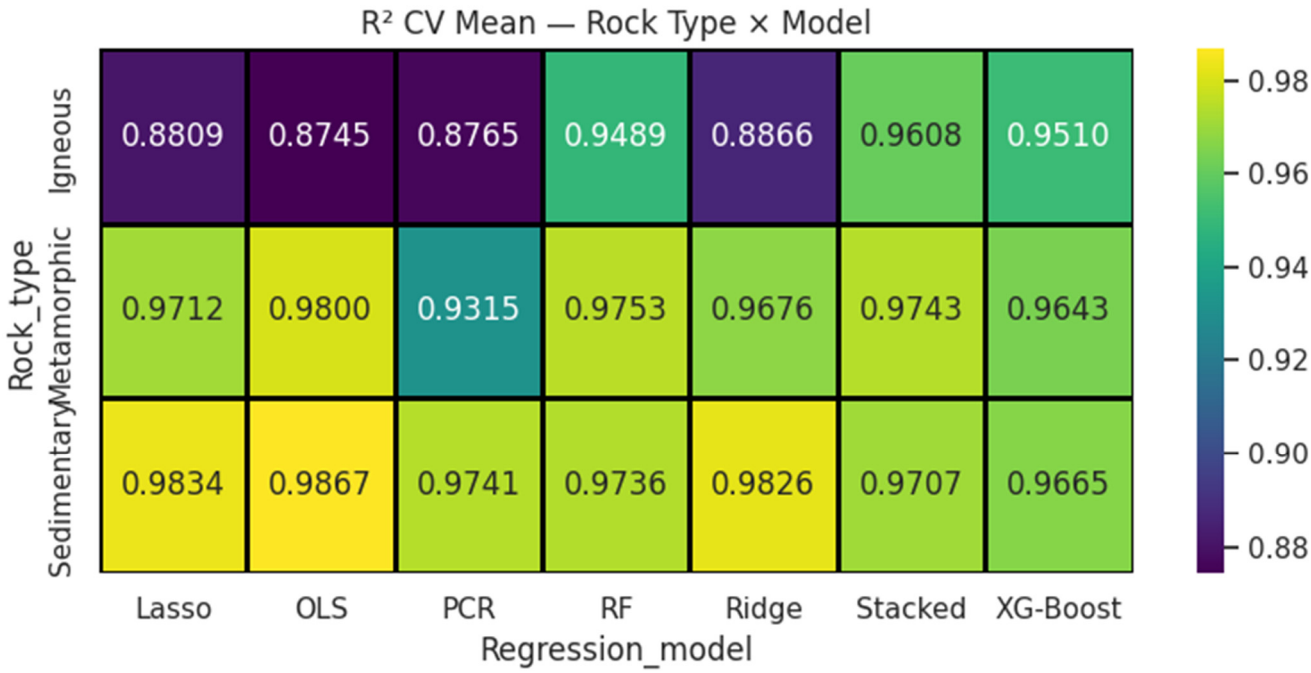


Figure 11. All regression model performance heatmaps against rock type for the main dataset.

In sedimentary rocks (Table 3), the constructed brittleness index again holds the top position, supported by strong agreement across all importance methods. UCS and MHN follow closely, reflecting the dominant role of strength and hardness in controlling abrasiveness in clastic and carbonate lithologies. Quartz content, both measured and combined, also ranks prominently, highlighting mineralogical composition as a key driver. Additional predictors such as Schmidt rebound hardness (H), Poisson's ratio (v), and dynamic Young's elastic modulus (E) contribute moderately.

For metamorphic rocks (Table 4), MHN and UCS share the highest consensus importance, indicating that hardness and strength are the primary controls in foliated and crystalline metamorphic lithologies. Poisson's ratio and quartz content occupy the next tier,

reflecting the influence of mineralogical stiffness and deformation characteristics. Elastic parameters such as G and K, along with mica content and V_s, show lower but still meaningful contributions. The overall pattern suggests that metamorphic rocks rely more heavily on hardness-strength interactions than on textural predictors.

Table 2. Agreement-based consensus rank feature importance for igneous rocks for the main dataset, predictors characterized as measured, reconstructed and all.

Feature	Predictor Set	Agreement	Consensus Rank	RF	Perm	Drop	SHAP
GS	Measured	4.00	1.00	0.64	0.76	0.04	0.52
BI	Constructed	3.10	1.00	0.58	0.72	0.02	0.41
GS	All	2.79	1.25	0.56	0.75	0.00	0.45
MHN	Constructed	1.42	2.00	0.23	0.34	0.01	0.22
BI	All	0.55	2.50	0.11	0.11	0.00	0.08
G	All	0.59	3.00	0.08	0.15	0.00	0.10
G	Constructed	0.41	3.25	0.05	0.11	0.00	0.07
Vp	Measured	0.31	3.25	0.05	0.06	0.00	0.06
UCS	Measured	1.29	3.50	0.21	0.35	0.00	0.25
Vs	Constructed	0.39	3.75	0.05	0.10	0.00	0.07

Igneous rocks feature importance

Table 3. Agreement-based consensus rank feature importance for sedimentary rocks for the main dataset, predictors characterized as measured, reconstructed and all.

Feature	Predictor Set	Agreement	Consensus Rank	RF	Perm	Drop	SHAP
BI	Constructed	2.38	1.00	0.39	0.59	0.01	0.35
UCS	Measured	1.56	1.50	0.25	0.38	0.01	0.24
MHN	Constructed	1.74	2.00	0.33	0.41	0.00	0.30
BI	All	0.84	2.25	0.15	0.19	0.00	0.15
Quartz_%	All	0.84	2.75	0.17	0.19	0.00	0.15
Quartz_%	Measured	1.48	3.00	0.29	0.37	0.00	0.26
H	Measured	1.09	3.75	0.22	0.26	0.00	0.19
v	Constructed	0.89	3.75	0.18	0.21	0.00	0.16
E	Constructed	0.28	4.00	0.03	0.07	0.00	0.05
R	Measured	0.68	4.25	0.14	0.13	0.00	0.13

Sedimentary rocks feature importance

Table 4. Agreement-based consensus rank feature importance for metamorphic rocks for the main dataset, predictors characterized as measured, reconstructed and all.

Feature	Predictor Set	Agreement	Consensus Rank	RF	Perm	Drop	SHAP
UCS	Measured	1.21	1.00	0.28	0.25	0.00	0.17
MHN	Constructed	2.43	1.00	0.56	0.47	0.01	0.34
v	Constructed	0.96	2.00	0.24	0.19	0.00	0.14
UCS	All	0.77	2.50	0.20	0.14	0.00	0.12
Quartz_%	Measured	0.73	3.25	0.18	0.16	0.00	0.11
Quartz_%	All	0.56	3.50	0.14	0.10	0.00	0.09
G	Constructed	0.38	3.75	0.04	0.12	0.00	0.06
K	Constructed	0.34	4.00	0.05	0.08	0.00	0.04
Mica_%	Measured	0.48	4.25	0.12	0.10	0.00	0.07
Vs	Constructed	0.33	4.50	0.05	0.09	0.00	0.06

Metamorphic rocks feature importance

Across all three lithologies, the consensus approach highlights that while the dominant predictors differ by rock type, hardness, strength, and brittleness consistently form the core

set of influential variables, with mineralogical indicators contributing more strongly to sedimentary and metamorphic groups.

3.1.3. Script 1 Results

Script 1 evaluates the predictive performance of UCS, MHN, and V_p across igneous, sedimentary, and metamorphic subsets using K-Fold cross-validation at seeds 42, 123, and 999. The results show consistently high R^2 values for all models, confirming that these three predictors form a stable and highly informative core set across lithologies.

For metamorphic rocks, all models maintain R^2 values near unity with minimal seed-to-seed variation, indicating strong robustness and low sensitivity to sampling randomness. Sedimentary rocks display similarly high and uniform performance, with linear, regularized, and ensemble models all achieving near-identical R^2 values across seeds. Igneous rocks exhibit slightly more variability, particularly for PCR at seed 999, yet Random Forest, OLS, and XGBoost remain highly stable and deliver the strongest performance. Overall, Script 1 demonstrates that UCS, MHN, and V_p provide reliable predictive strength across all rock types, and that model behavior remains consistent across different random seeds, validating the stability of this reduced predictor set. In metamorphic and sedimentary subsets, the tight clustering of R^2 values suggests that the predictors are highly informative and that the models are not overfitting to specific partitions of the data. The slight drop in PCR performance for igneous rocks highlights its sensitivity to multicollinearity and its weaker ability to represent nonlinear relationships compared with ensemble methods. Meanwhile, the consistently strong performance of RF and XGBoost across all lithologies indicates that tree-based models effectively capture the complex interactions among UCS, MHN, and V_p.

The nested cross-validation results (Figure 12) for Script 1, using XGBoost with the reduced predictor set (UCS–MHN–V_p), show clear and consistent performance differences across the three lithological groups. Metamorphic rocks achieve the highest nested R^2, indicating that the reduced predictor set captures their strength–hardness relationships with exceptional stability even under strict, nested validation. Sedimentary rocks also exhibit strong performance, reflecting the strong linear and non-linear associations between UCS, MHN, and V_p in clastic and carbonate formations. Igneous rocks show slightly lower nested R^2 values, consistent with their greater mineralogical variability and more complex elastic behavior.

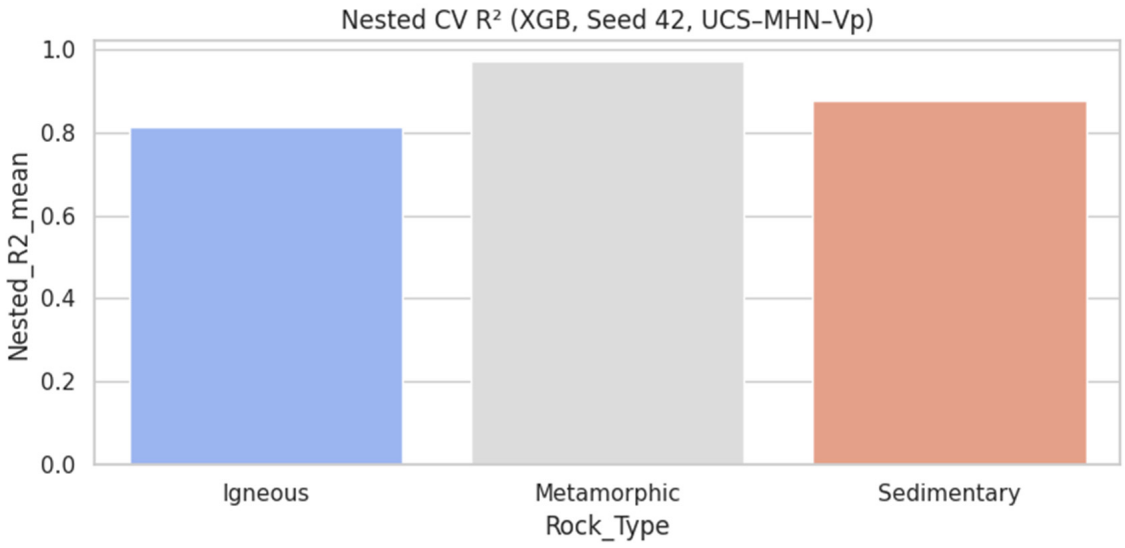


Figure 12. Nested K-Fold predictive performance at random seed 42 for igneous, metamorphic and sedimentary rocks for Script 1 predictors.

Overall, Script 1 demonstrates that the reduced predictor set generalizes well under nested CV, with metamorphic and sedimentary rocks showing the strongest predictive response and igneous rocks displaying moderate but still reliable performance. The unified feature importance analysis for the reduced predictor set, shown in Figure 13, shows strong cross-method agreement on the dominance of UCS, with consistently high importance values across Random Forest, permutation, SHAP, and linear-model coefficients. MHN forms the second-most influential predictor, supported by moderate to high importance across most methods, particularly RF permutation, XGB permutation, and SHAP. V_p contributes the least, with uniformly low values across all methods, reflecting its weaker direct relationship with CAI compared with strength and hardness parameters. Notably, XGBoost gain importance is zero for all three predictors. This occurs because, under the reduced predictor set and LOOCV structure, XGBoost frequently constructs extremely shallow trees, often single-split or pruned trees, where the calculated gain falls below the internal precision threshold and is stored as zero. Despite this, XGBoost permutation and drop-column importance remain non-zero, confirming that the model still uses these predictors meaningfully even when gain is suppressed. Overall, the unified heatmap demonstrates that UCS overwhelmingly drives predictive performance, MHN provides secondary support, and V_p plays a minor role, with consistent agreement across nine different importance frameworks.

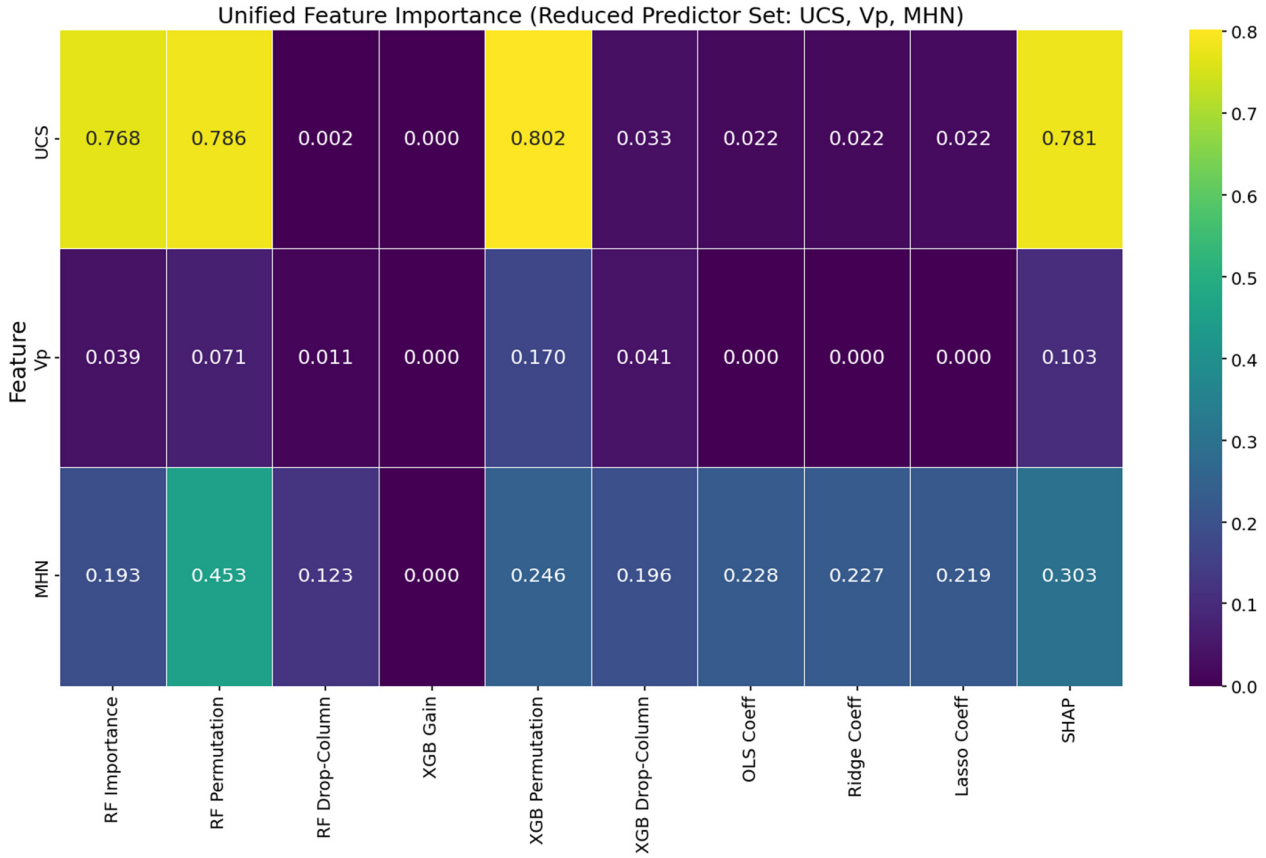


Figure 13. Feature importance for Script 1, reduced predictors.

3.1.4. Script 3 Results

Script 3 evaluates the expanded predictor set—UCS, MHN, V_p, PI, and GS, across igneous, sedimentary, and metamorphic rocks using K-Fold cross-validation at seeds 42, 123, and 999. Igneous rocks exhibit more variation, particularly for PCR, which performs poorly across all seeds, while RF, OLS, Ridge, and XGBoost maintain high R^2 values

and demonstrate strong robustness. Sedimentary rocks show similarly consistent behavior, confirming that the expanded predictor set captures both mechanical and textural controls effectively. Metamorphic rocks remain the most stable group, with all models, including linear, regularized, and ensemble approaches, clustered tightly near $R^2 \approx 1.0$. The results show uniformly high R^2 values for sedimentary and metamorphic rocks, with all models achieving strong and stable performance across seeds, indicating that the addition of petrographic index and grain size enhances predictive strength without introducing variability. Across all three lithologies, the addition of PI and GS enhances the predictive signal without introducing instability or seed-specific fluctuations. Sedimentary and metamorphic rocks show near-perfect agreement across seeds, demonstrating that the expanded predictor set effectively represents both mechanical and petrographic controls. Igneous rocks remain the most challenging group, yet RF, Ridge, and XGBoost consistently deliver high R^2 values despite PCR's weaker performance. The K-Fold results confirm that PI and GS contribute meaningful geological information, improving model robustness while maintaining consistency across seeds.

The performance evaluation for Script 3, using the expanded predictor set, shows consistently strong results across all three rock types (Figure 14). RMSE values remain low for most models, with ensemble approaches such as Random Forest, XGBoost, and the stacked model achieving the smallest errors across sedimentary and metamorphic rocks (Figure 15). Igneous rocks exhibit slightly higher RMSE, particularly for OLS, reflecting their greater mineralogical variability and more complex mechanical behavior. MAE values follow the same trend (Figure 16): XGBoost consistently produces the lowest MAE across all lithologies, while RF shows slightly elevated MAE for igneous rocks. The R^2 results reinforce these patterns, with metamorphic and sedimentary rocks achieving values near unity for all models, indicating excellent predictive stability. Igneous rocks show more variation, with PCR performing notably worse than other models, while RF, Ridge, OLS, and XGBoost maintain high accuracy. Overall, Script 3 demonstrates that the expanded predictor set provides strong predictive performance across lithologies, with ensemble models delivering the most stable and accurate results and igneous rocks showing the greatest sensitivity to model choice.

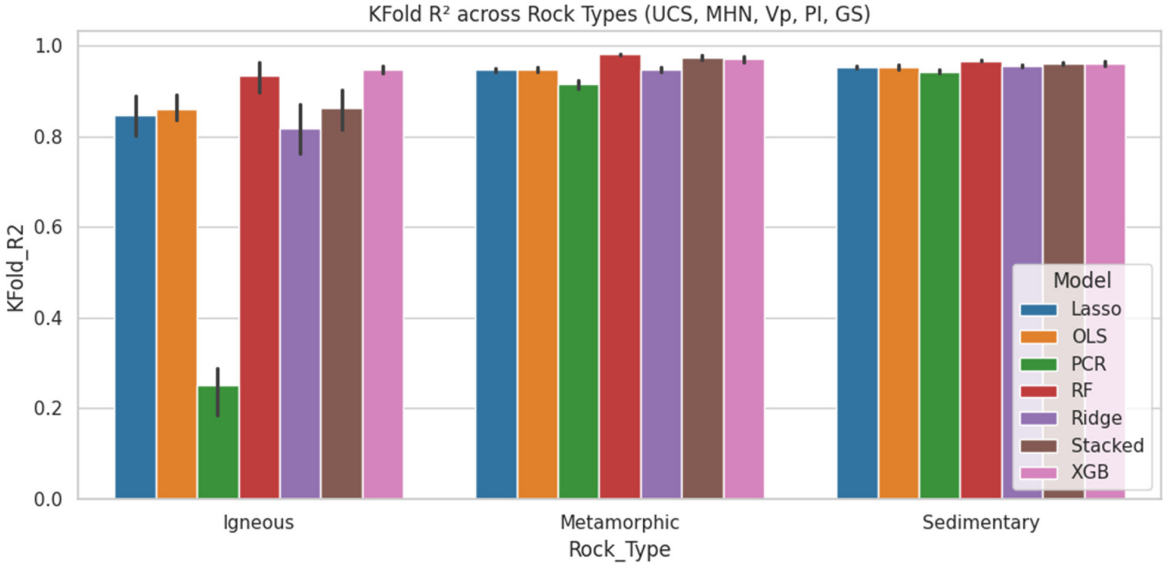


Figure 14. Performance results for all rocks main dataset all regression models for Script 3 (UCS, MHN, V_p, PI and GS). The vertical lines extending from each bar represent the standard deviation of the K-Fold R^2 values for that model and predictor set. Shorter lines represent higher model stability and longer lines represent higher variability.

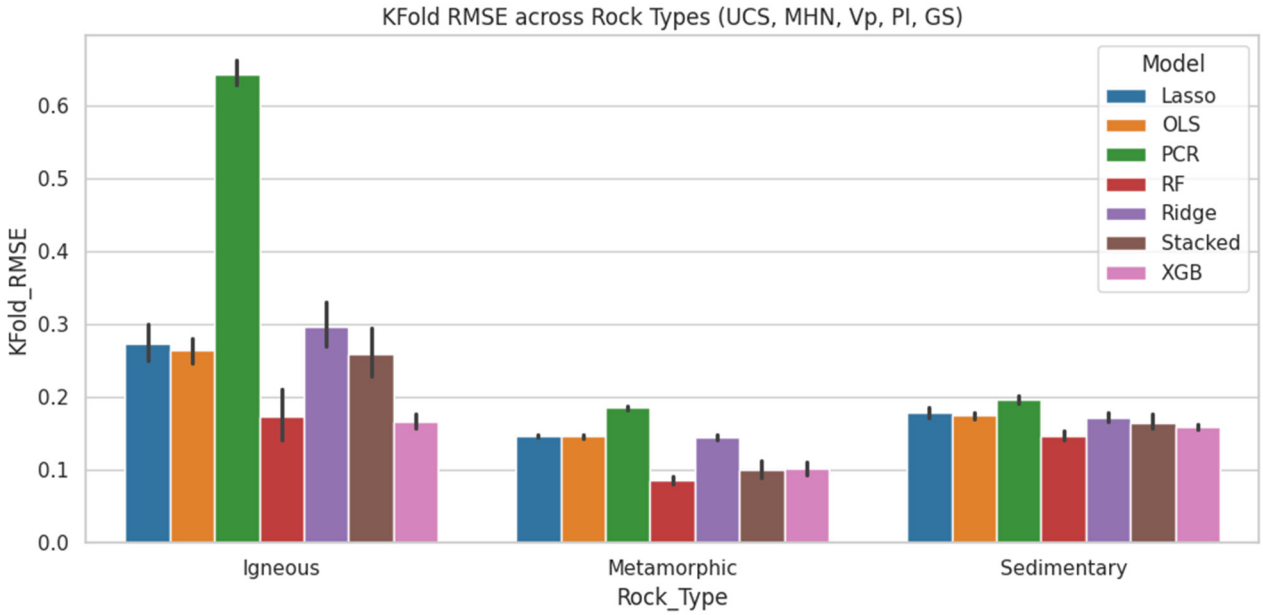


Figure 15. Root mean square error (RMSE) for all rocks main dataset all regression models for Script 3 (UCS, MHN, V_p, PI and GS). The vertical lines extending from each bar represent the standard deviation of the K-Fold RMSE values for that model and predictor set. Shorter lines represent higher model stability and longer lines represent higher variability.

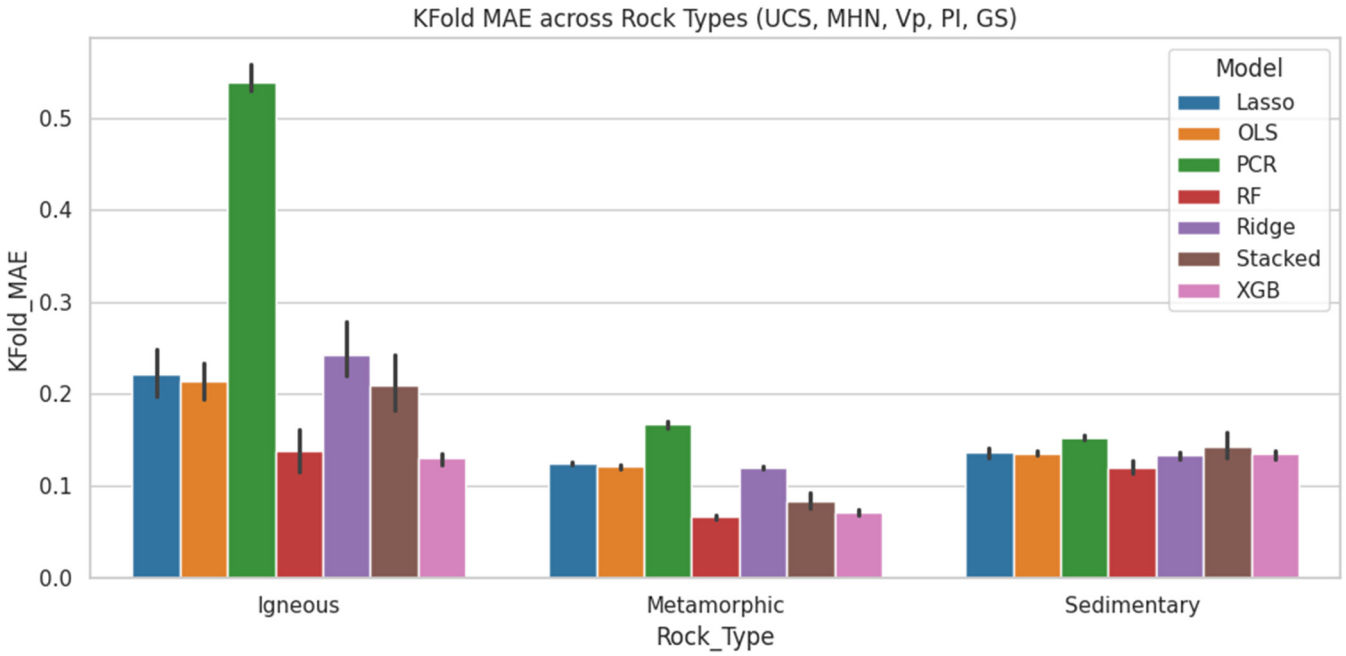


Figure 16. Mean absolute error (MAE) for all rocks main dataset all regression models for Script 3 (UCS, MHN, V_p, PI and GS). The vertical lines extending from each bar represent the standard deviation of the K-Fold MAE values for that model and predictor set. Shorter lines represent higher model stability and longer lines represent higher variability.

The unified feature importance results (Figure 17) for Script 3 show overwhelming cross-method agreement that UCS is the dominant predictor when using the expanded variable set (UCS, MHN, V_p, PI, GS). UCS receives the highest importance values across nearly all frameworks, including Random Forest, permutation, linear-model coefficients, and especially SHAP, where its contribution is an order of magnitude larger than all other features. MHN, V_p, and PI form a secondary tier of predictors, each showing

modest but consistent influence across permutation-based and SHAP methods, while GS contributes minimally.

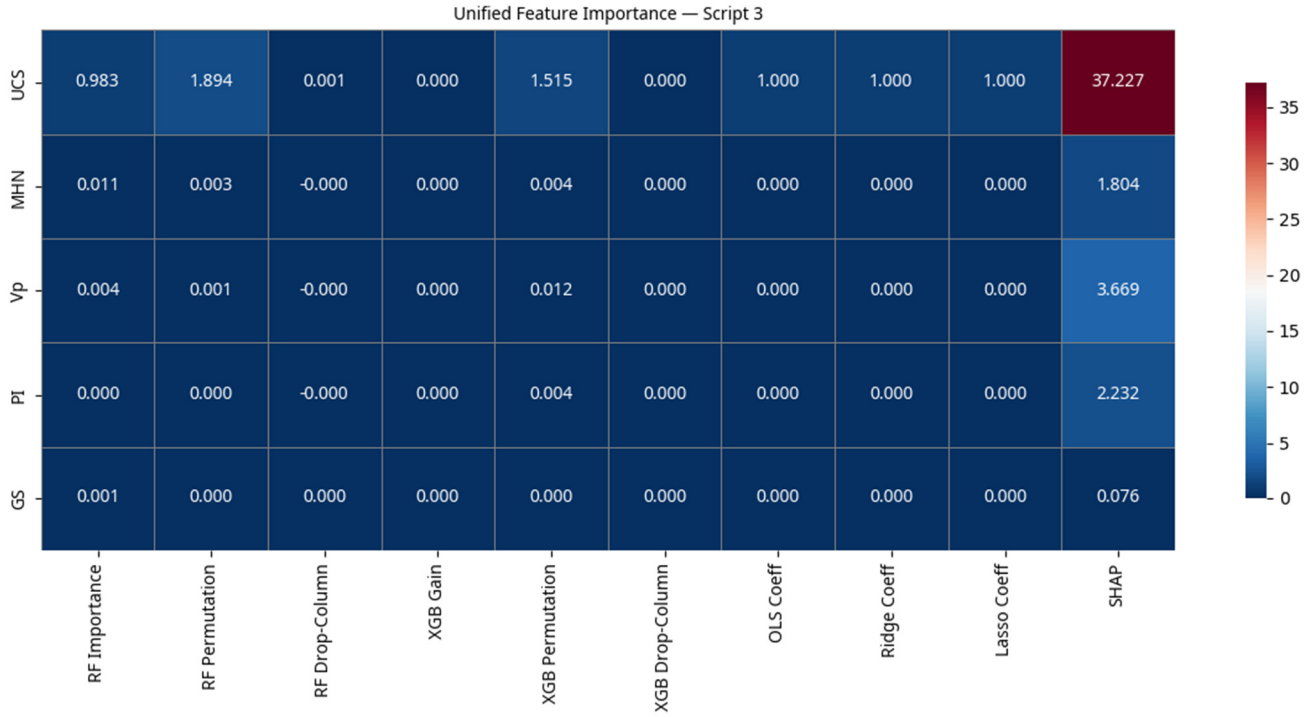


Figure 17. Unified feature importance for Script 3, reduced predictor set.

The comparison across Scripts 1–3 (Figures 18–20) shows a clear and systematic improvement in model performance as the predictor set expands from UCS–MHN–V_p (Script 1) to UCS–MHN–V_p–PI (Script 2) and finally to UCS–MHN–V_p–PI–GS (Script 3). The K-Fold R^2 results (Figure 18) demonstrate that all models benefit from the additional predictors, with ensemble methods (RF, XGB, Stacked) showing the largest gains and achieving values consistently near unity in Script 3.

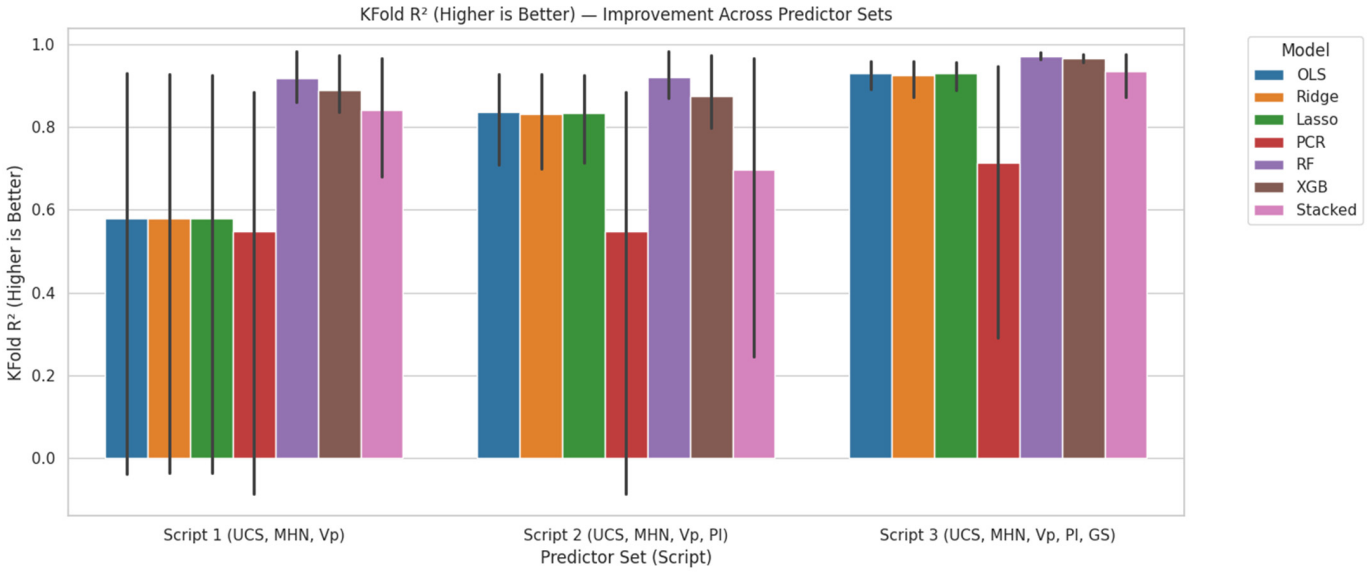


Figure 18. All regression models' performance comparison for three reduced predictor scripts. The vertical lines extending from each bar represent the standard deviation of the K-Fold R^2 values for that model and predictor set. Shorter lines represent higher model stability and longer lines represent higher variability.

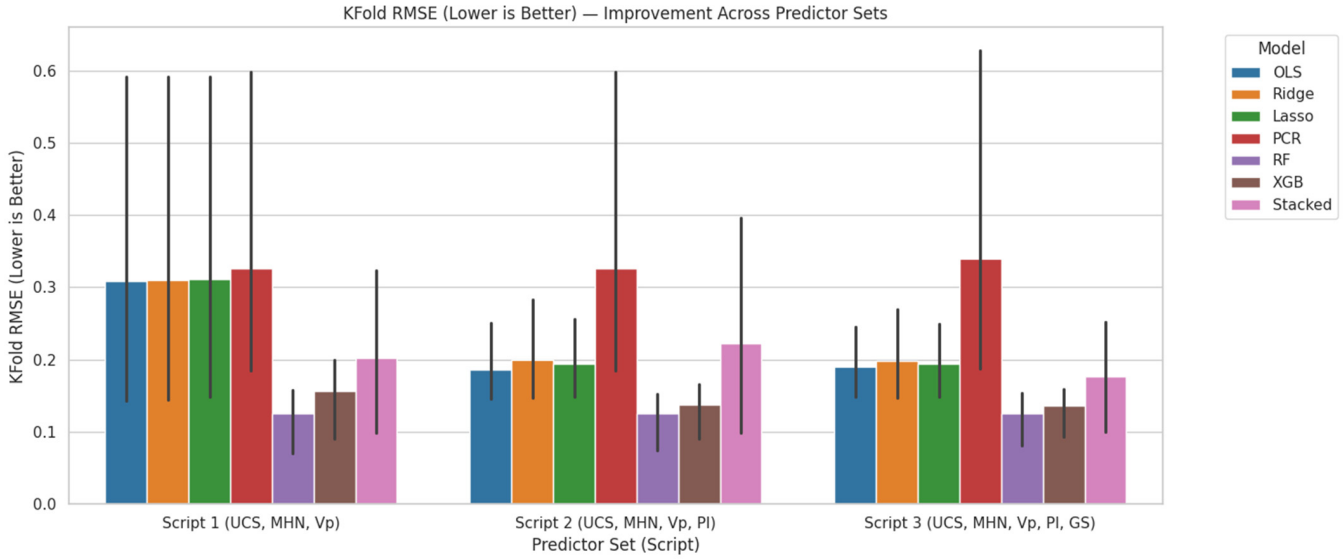


Figure 19. All regression models' RMSE comparison for three reduced predictor scripts. The vertical lines extending from each bar represent the standard deviation of the K-Fold RMSE values for that model and predictor set. Shorter lines represent higher model stability and longer lines represent higher variability.

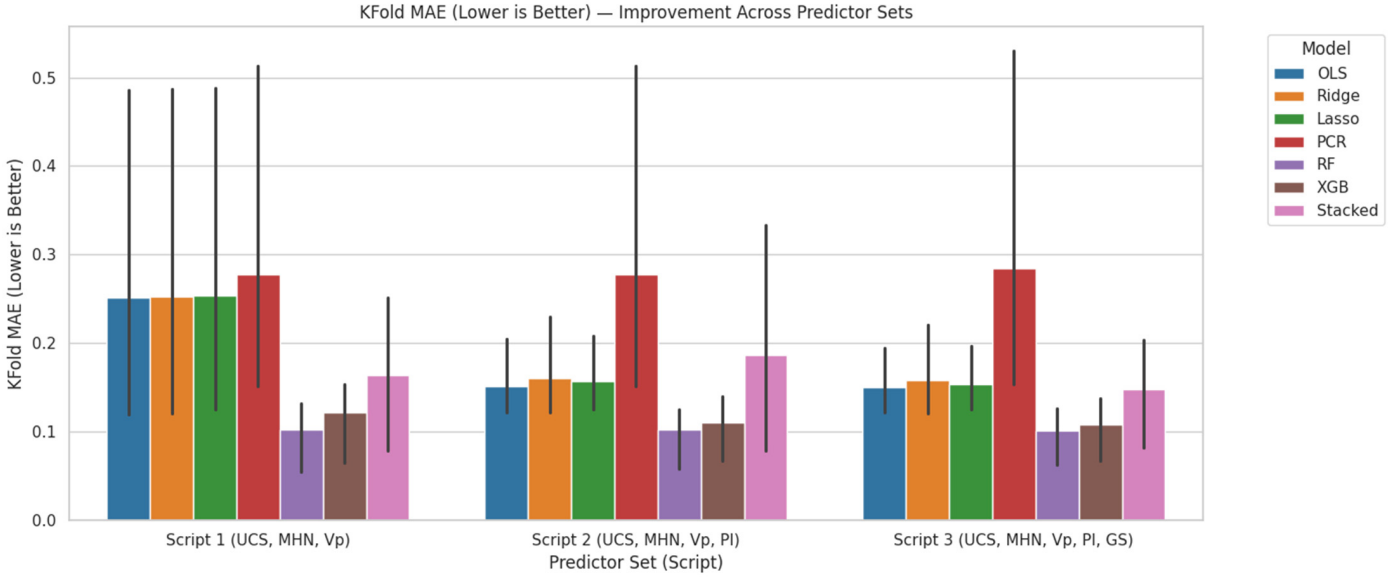


Figure 20. All regression models' MAE comparison for three reduced predictor scripts. The vertical lines extending from each bar represent the standard deviation of the K-Fold MAE values for that model and predictor set. Shorter lines represent higher model stability and longer lines represent higher variability.

Correspondingly, RMSE (Figure 19) decreases across scripts, indicating that the expanded predictor sets reduce residual error and improve the precision of CAI predictions. This trend is most pronounced for igneous rocks, where Script 1 models show higher RMSE and variability, while Script 3 models achieve substantially lower error magnitudes. The MAE results (Figure 20) reinforce this pattern: complex models such as RF, XGB, and the stacked model consistently achieve the lowest MAE values, and their performance improves steadily as PI and GS are added. Linear and regularized models also benefit from the expanded predictor sets, though to a lesser extent. Overall, the three figures collectively show that expanding the predictor set enhances model stability, reduces error,

and increases predictive accuracy across all lithologies, with Script 3 providing the most robust and reliable performance.

The figures also highlight that Script 2 provides a notable intermediate improvement, suggesting that PI alone captures a substantial portion of petrographic variability relevant to abrasiveness. Script 3 further refines this by incorporating grain-size information, which appears particularly beneficial for models sensitive to textural heterogeneity, such as RF and XGB. The reduction in error bars across Scripts 2 and 3 indicates that the expanded predictor sets not only improve mean performance but also reduce fold-to-fold variability, strengthening model reliability. The seed-stability analysis for Script 1, evaluated across a wide range of random seeds, shows that the reduced predictor set (UCS, MHN, V_p) produces highly consistent K-Fold R^2 values for most models and rock types.

The boxplots demonstrate that sedimentary and metamorphic rocks maintain tight R^2 distributions across seeds, indicating strong robustness and minimal sensitivity to sampling randomness (Figure 21). Igneous rocks exhibit greater variability, particularly for PCR and occasionally RF, reflecting the higher geological heterogeneity and more complex mechanical behavior of igneous lithologies. The rock-type stability boxplots reinforce that sedimentary rocks are the most stable group, metamorphic rocks show slightly higher but still controlled variability, and igneous rocks present the widest spread (Figure 21). The seed-stability heatmap (Figure 22) further confirms these trends: linear and regularized models (OLS, Ridge, Lasso) show low standard deviations across seeds, while PCR displays the highest instability, and ensemble models (RF, XGB, Stacked) remain consistently stable with low to moderate variability. Overall, Script 1 demonstrates strong seed stability across models and lithologies, with variability concentrated primarily in igneous rocks and PCR, while ensemble and linear models remain reliably consistent across all tested seeds.

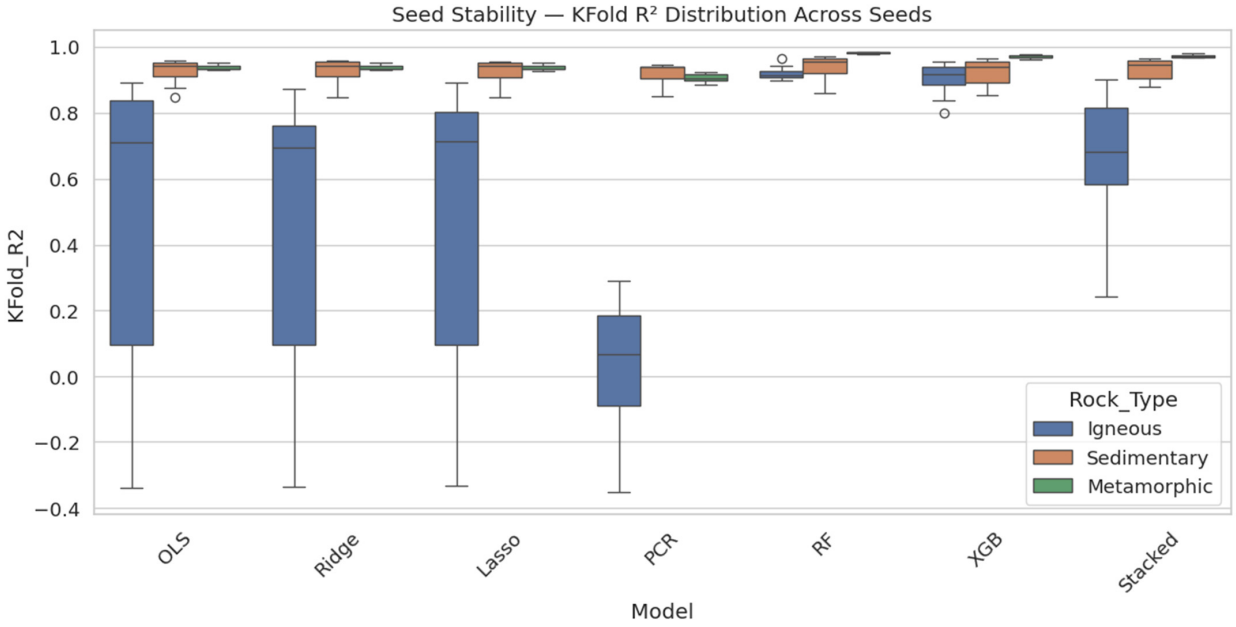


Figure 21. Prediction performance stability across multiple seeds for Script 1. Circles represent the underlying K-Fold R^2 standard deviation values from each fold, highlighting the spread and potential outliers across models.

The seed-stability analysis for Script 3, evaluated across an extended range of random seeds, shows that the expanded predictor set (UCS, MHN, V_p, PI, GS) produces highly stable K-Fold R^2 values for most models and rock types (Figure 23). The heatmap of mean R^2 standard deviations demonstrates that ensemble models—RF, XGB, and the stacked model—maintain consistently low variability across seeds, confirming their robustness

under repeated resampling. Linear and regularized models (OLS, Ridge, Lasso) also show low to moderate variability, indicating that the addition of PI and GS improves their stability relative to Script 1. PCR remains the least stable model, with noticeably higher standard deviations across seeds, reflecting its sensitivity to changes in data partitioning and its reliance on principal-component structure.

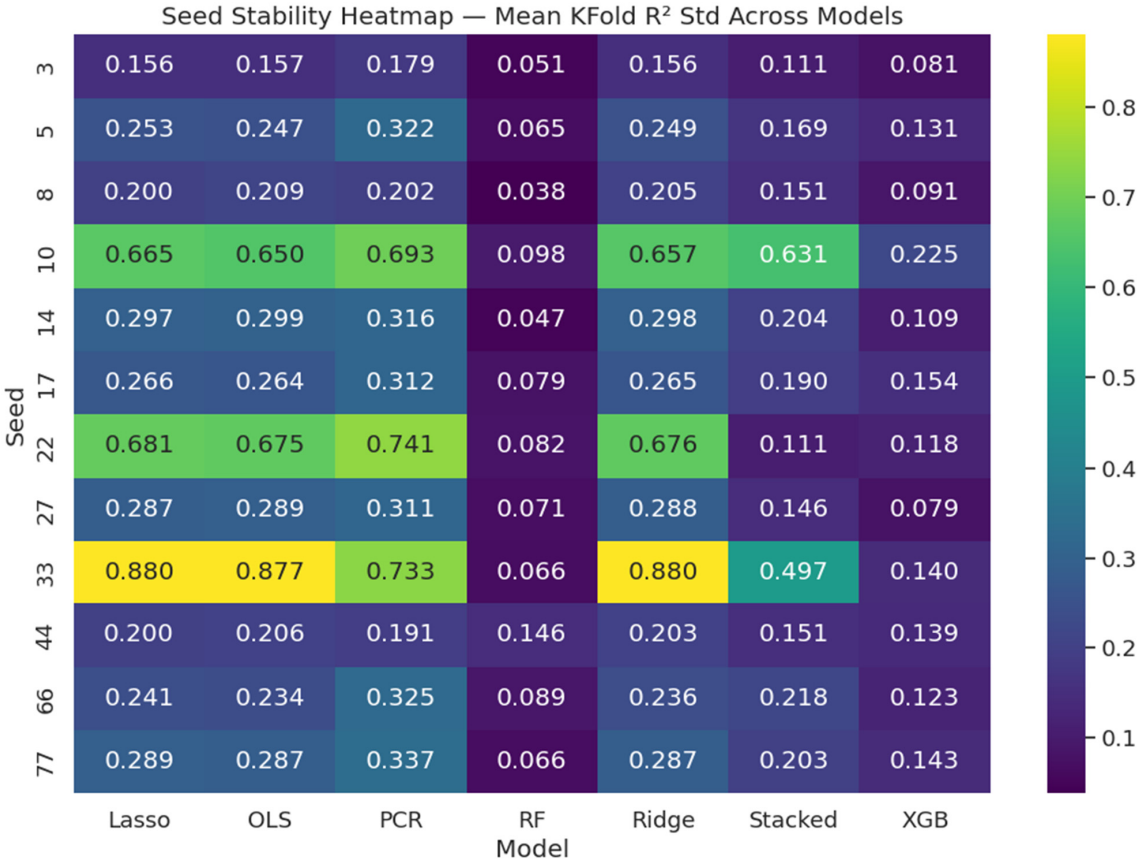


Figure 22. Seed stability heatmap for Script 1.

The rock-type, seed stability boxplots (Figure 24) reinforce these trends: sedimentary and metamorphic rocks exhibit tight distributions with minimal spread, while igneous rocks show the highest variability, particularly for PCR and occasionally RF, consistent with their greater mineralogical heterogeneity. Overall, Script 3 demonstrates strong seed stability across models and lithologies, with ensemble methods showing the greatest resilience to seed variation and igneous rocks presenting the only notable sensitivity.

The nested-CV seed-stability analysis for Script 3 (Figure 25), using XGBoost with the full predictor set (UCS, MHN, V_p, PI, GS), shows exceptionally consistent performance across all tested seeds (42, 187, 432) and rock types. The nested R^2 values remain uniformly high—near unity—for igneous, sedimentary, and metamorphic rocks, indicating that the model's generalization ability is largely insensitive to seed variation. Sedimentary and metamorphic rocks exhibit the tightest clustering of nested R^2 values, reflecting strong and stable relationships between the expanded predictor set and CAI behavior. Igneous rocks show slightly more spread across seeds, consistent with their greater mineralogical and mechanical heterogeneity, yet still maintain high nested R^2 values that confirm reliable predictive performance. The prediction-stability analysis for Script 3, using XGBoost with the full predictor set (UCS, MHN, V_p, PI, GS), shows that the model maintains excellent predictive fidelity across both stable and unstable seeds. For the most stable seed (27), predicted CAI values align almost perfectly with actual values, with all rock types clustering

tightly around the 1:1 line and showing no systematic bias (Figure 26). Even under the least stable seed (14), the predicted–actual relationship remains strong, with only minor dispersion visible, primarily for igneous samples, yet still preserving a clear linear trend and high overall accuracy (Figure 27).

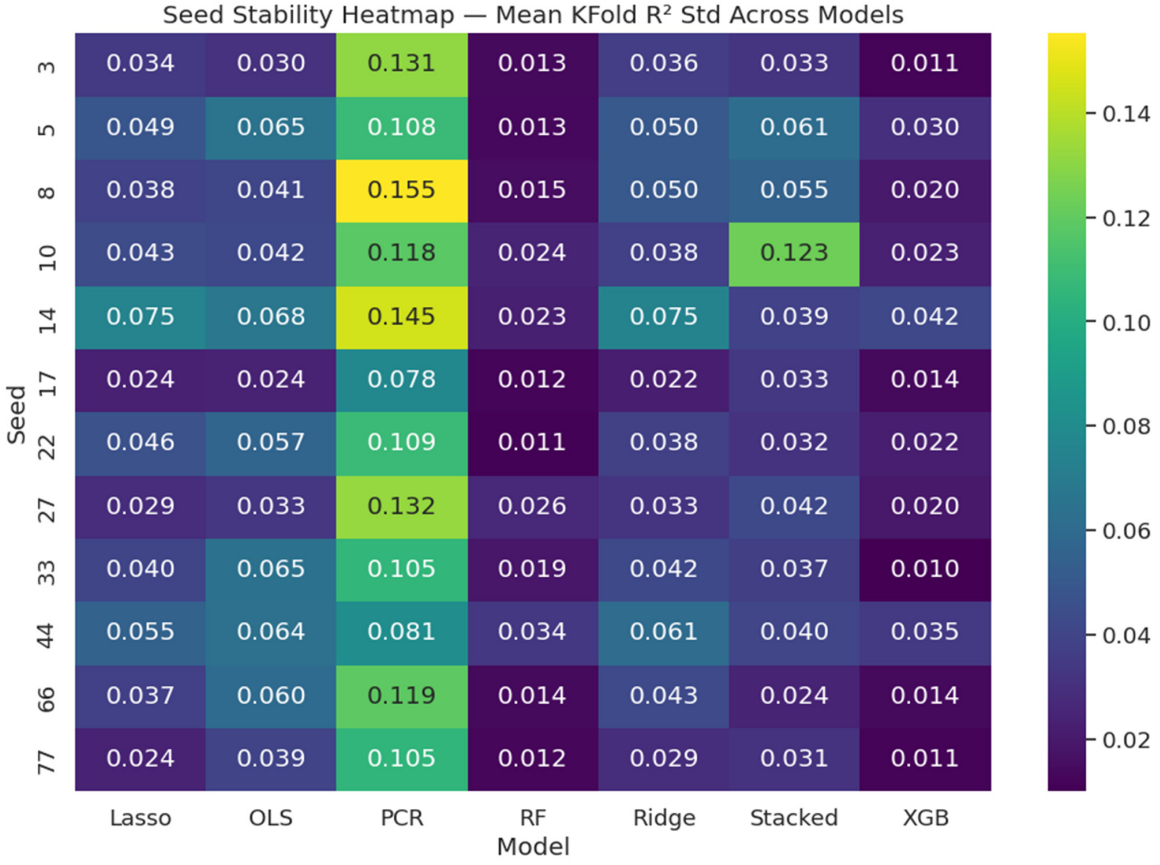


Figure 23. Seed stability heatmap for Script 3.

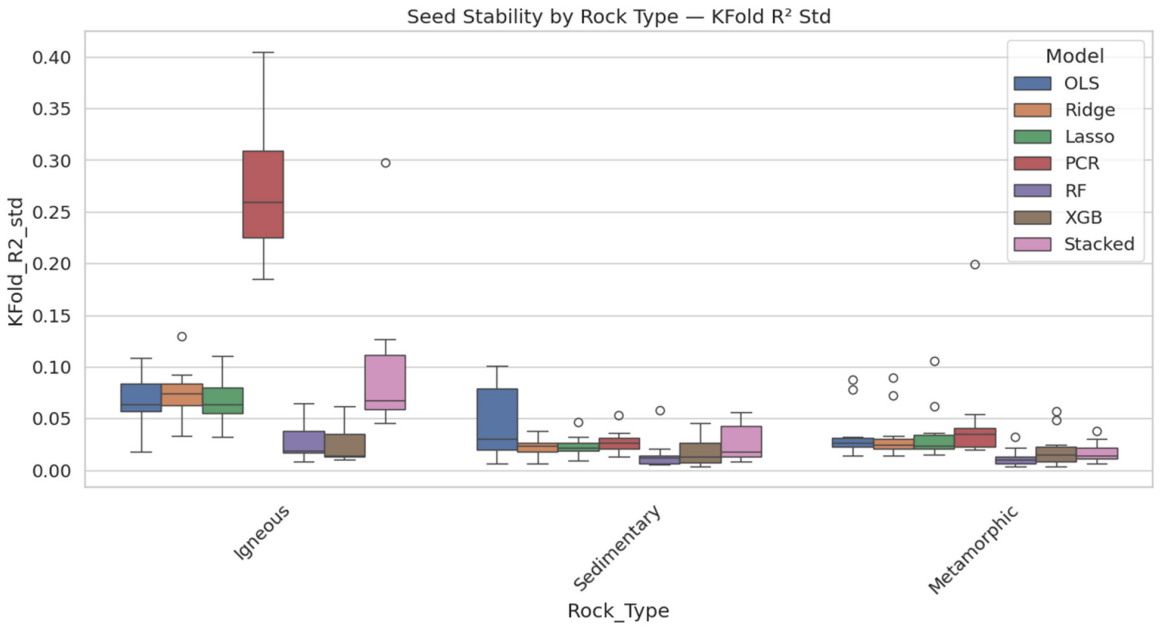


Figure 24. Prediction performance stability across multiple seeds for Script 3, Circles represent the underlying K-Fold R^2 standard deviation values from each fold, highlighting the spread and potential outliers across models.

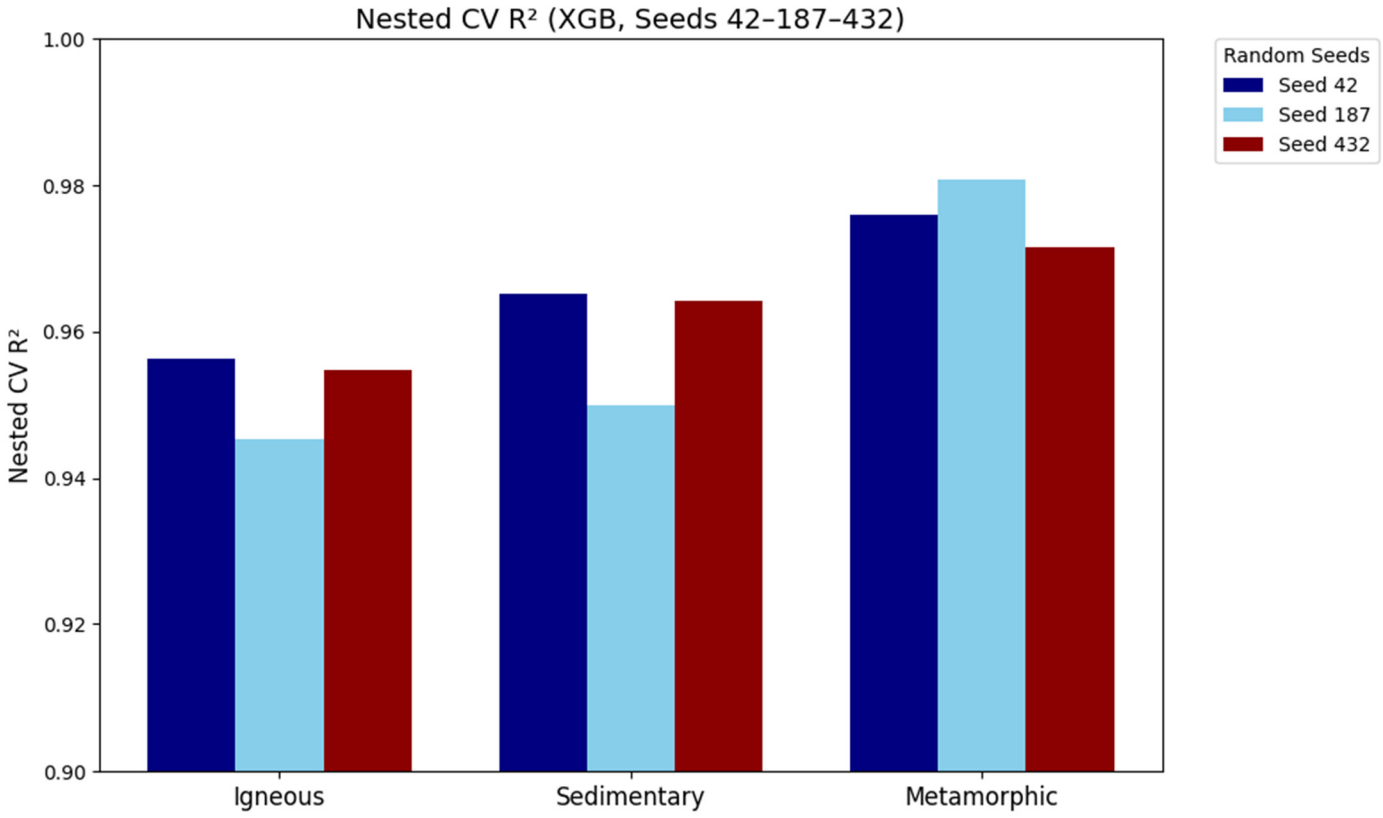


Figure 25. Nested K-Fold CV results for Script 3 tested at random seeds 42, 187 and 432.

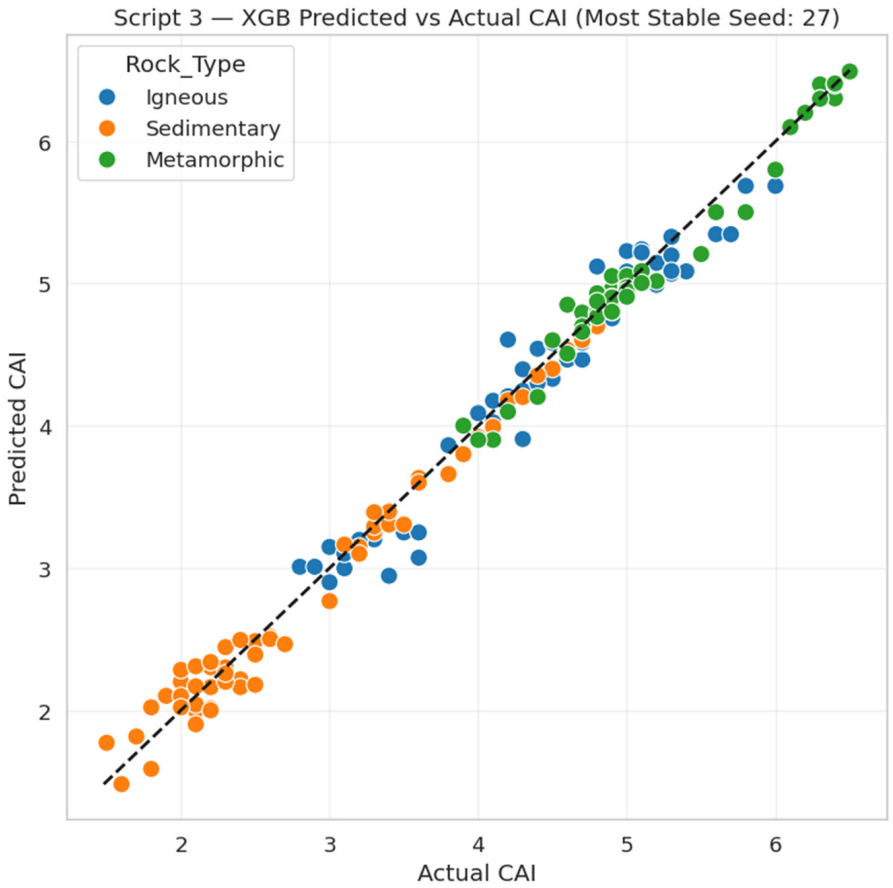


Figure 26. Predicted vs. actual CAI measurements for the most stable seed and Script 3.

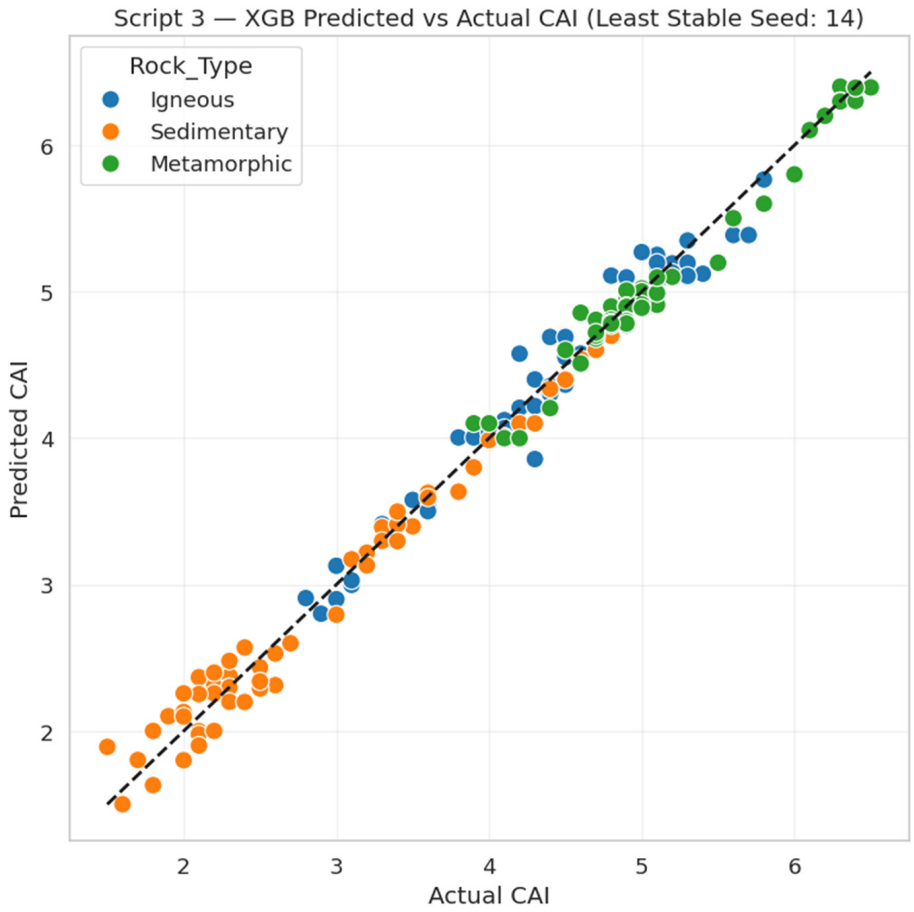


Figure 27. Predicted vs. actual CAI measurements for least stable seed and Script 3.

The 12-seed confidence-interval plot (Figure 28) reinforces these findings: mean predictions remain close to the actual CAI values, and the associated error bars are small across sedimentary and metamorphic rocks, indicating high robustness to seed variation. Igneous rocks exhibit slightly larger uncertainty, consistent with their greater mineralogical heterogeneity, but still maintain narrow confidence intervals relative to the full CAI range.

Together, these results demonstrate that Script 3 delivers highly stable and reliable CAI predictions across seeds and lithologies, with XGBoost showing strong generalization and minimal sensitivity to initialization.

3.1.5. Script 4 Results

Script 4 (H, MHN, V_p, PI, GS) provides a comprehensive view of model performance variability using repeated K-Fold cross-validation (RKFold_R^2), revealing clear differences in both model stability and rock-type behavior. The heatmap of mean RKFold_R^2 across models and rock types integrates these patterns numerically (Figure 29). Random Forest and XGBoost achieve the highest mean R^2 values across all lithologies, with igneous rocks reaching ~0.93–0.94 and metamorphic and sedimentary rocks exceeding ~0.96 and ~0.92, respectively. Linear models cluster around ~0.90 for metamorphic and sedimentary rocks but drop to ~0.45 for igneous rocks, while PCR performs substantially worse for igneous and metamorphic rocks (0.197 and 0.524), though still reasonably for sedimentary rocks (0.899). The stacked model provides strong overall performance, particularly for metamorphic rocks (0.946), reflecting its ability to blend linear and ensemble strengths. Ensemble methods—Random Forest, XGBoost, and the stacked model—produce the highest and most concentrated R^2 distributions, indicating strong predictive consistency across repeated folds. Linear and regularized models (OLS, Ridge, Lasso) exhibit moderate spread with lower cen-

tral tendencies, while PCR shows the widest and lowest-centered distribution, confirming its sensitivity to data partitioning and its weaker ability to capture CAI-relevant structure.

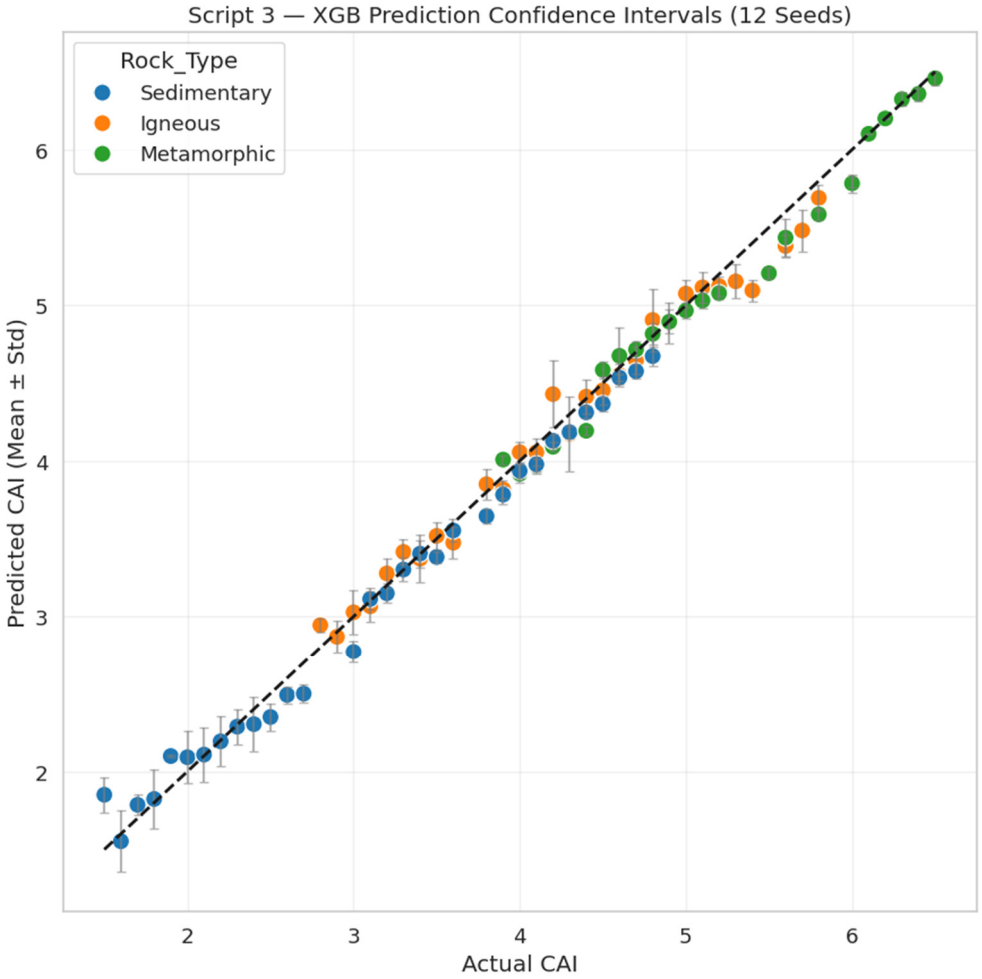


Figure 28. Predicted vs. actual CAI measurements for all seeds with confidence intervals for Script 3.

Sedimentary rocks display a narrow, high-centered distribution tightly clustered near $R^2 \approx 1$, reflecting their strong and predictable relationships with the expanded predictor set. Metamorphic rocks show slightly broader distributions but still maintain high central values, indicating robust performance with modest variability. Igneous rocks exhibit the widest spread, ranging from low to high R^2, consistent with their greater mineralogical heterogeneity and more complex mechanical behavior. Together, these three visualizations show that Script 4 delivers a detailed and stable performance profile: ensemble models consistently outperform others, sedimentary and metamorphic rocks yield highly reliable predictions, and igneous rocks remain the most challenging but still achieve strong results with tree-based models.

The unified feature importance analysis for Script 4 (Figure 30), using the reduced predictor set (H, MHN, V_p, GS), shows strong cross-method agreement on the dominant role of hardness (H) in predicting CAI. H receives the highest importance values across Random Forest importance, permutation, and SHAP, confirming that it is the primary driver of model behavior under repeated K-Fold evaluation. MHN forms the second-most influential predictor, with moderate contributions across permutation, drop-column, and linear-model coefficients, indicating that it provides meaningful secondary structure complementary to hardness. V_p shows consistently low importance across nearly all methods, reflecting its weaker direct relationship with CAI in this reduced predictor

configuration. GS displays an interesting pattern: although its importance is modest in RF and permutation methods, it becomes highly influential under XGB drop-column and linear-model coefficients, suggesting that grain-size effects are captured more strongly by additive and regularized models than by tree-based gain metrics.

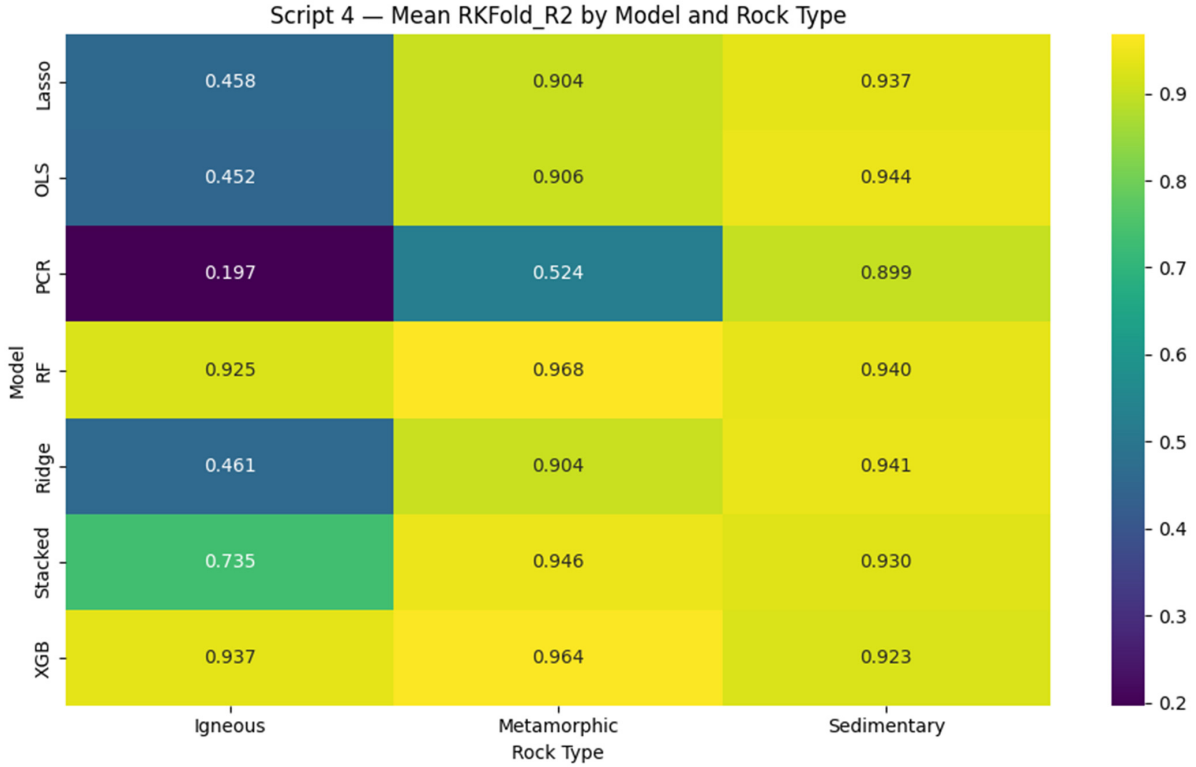


Figure 29. Prediction performance heatmap for Script 4, all regression models and three rock types.

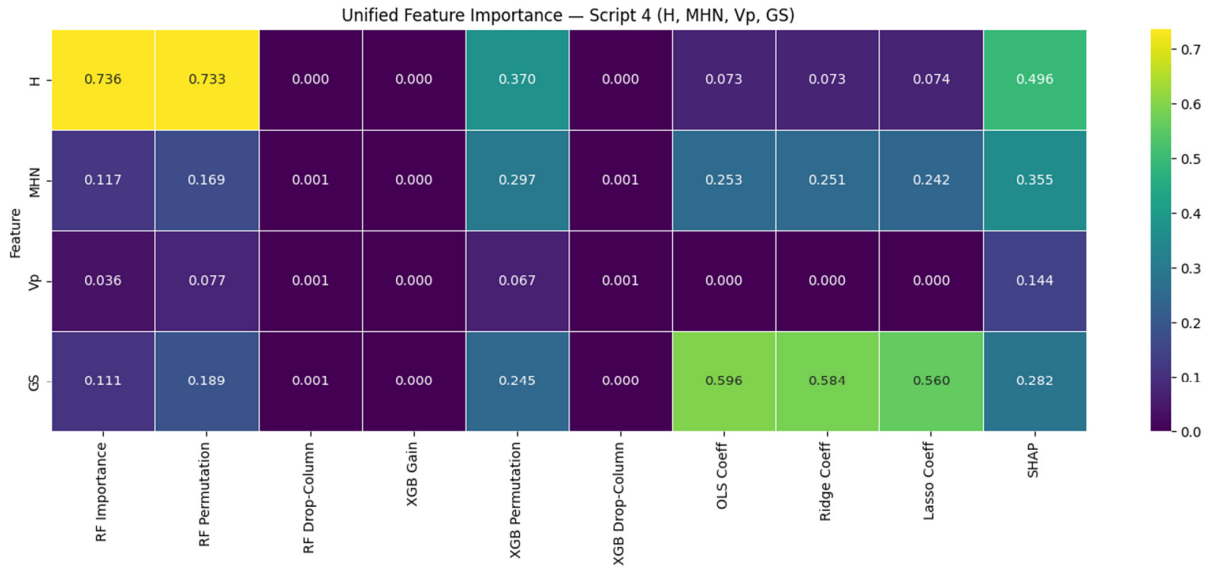


Figure 30. Unified feature importance for script 4.

Overall, Script 4's unified importance results show that hardness overwhelmingly governs predictive performance, MHN provides consistent secondary support, V_p contributes minimally, and GS plays a selective but meaningful role depending on the importance framework. The strong agreement across all methods confirms the stability and interpretability of the Script 4 predictor set.

The seed-stability analysis for Script 4, evaluated across a wide range of random seeds, shows that the repeated K-Fold framework produces highly consistent results for most models, with variability concentrated in only a few seeds and model types. A small subset of seeds, particularly 14, 22, and 33, exhibit slightly wider distributions, reflecting occasional sensitivity in models such as PCR and, to a lesser extent, RF (Figure 31).

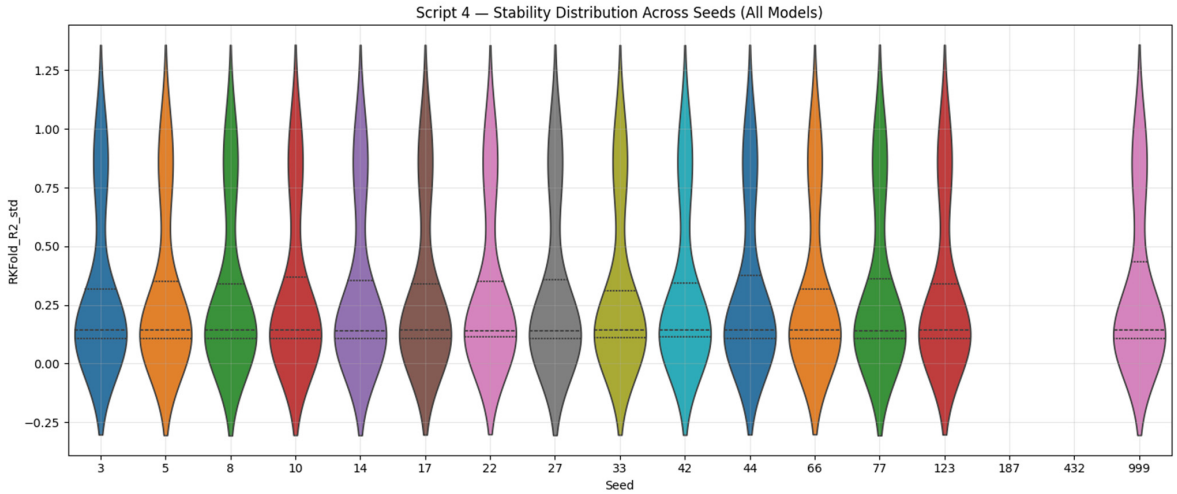


Figure 31. Seed stability distribution across all seeds and all models, dashed lines represent the underlying Repeated K Fold R^2 standard deviation values for each seed, highlighting the spread and stability characteristics across models.

Together, these results show that Script 4 delivers strong seed stability across the full model suite, with XGBoost standing out as the most stable and reproducible model, and only minor seed-dependent fluctuations appearing in weaker learners such as PCR. The narrow, symmetric violin shapes across most seeds indicate that XGB's internal regularization and gradient-boosting structure effectively suppress noise introduced by different train–test splits. Seeds such as 22, 33, and 42 produce almost flat distributions, demonstrating near-perfect reproducibility of R^2 values across folds. Even seeds that show slightly wider spread, such as 14 or 999, remain well within acceptable variability limits, with no evidence of instability or multimodal behavior. These findings reinforce that Script 4's repeated K-Fold framework provides a dependable measure of model robustness, with XGB serving as the benchmark for seed-insensitive predictive behavior.

Extended results and scripts are provided in the Supplementary Materials (Figures S1 and S2; Supplementary File S2).

3.2. CAI Prediction for Data Set 2

The binary classification results for Dataset 1 (Table 5) show substantial variation in model performance when predicting sedimentary versus non-sedimentary rocks. Random Forest and SVM (RBF) achieve the highest accuracy (0.83), indicating that non-linear decision boundaries capture the separation between the two classes more effectively than linear models. Logistic Regression, SVM (Linear), and Decision Tree all reach moderate accuracy (0.58), suggesting that linear separability is limited and simple tree structures struggle to capture the underlying relationships. kNN and Naive Bayes perform the weakest (0.50), reflecting their sensitivity to local neighborhood structure and distributional assumptions, respectively. XGBoost achieves moderate accuracy (0.58), likely due to the small dataset size and limited feature diversity, which restrict the depth and complexity of boosted trees.

Table 5. Binary classification results for Dataset 1 using various regression models.

S.No	Model	Accuracy
1	Random Forest	0.83
2	Decision Tree	0.58
3	kNN (k = 3)	0.50
4	Logistic Regression	0.58
5	SVM (Linear)	0.58
6	SVM (RBF)	0.83
7	Naive Bayes	0.50
8	XGBoost	0.58

The confusion matrix (Figure 32) provides deeper insight into these results. The classifier correctly identifies all non-sedimentary samples (5 correct, 0 misclassified), demonstrating strong specificity and reliable detection of the non-sedimentary class. However, two sedimentary samples are misclassified as non-sedimentary (5 correct, 2 incorrect), indicating reduced sensitivity for sedimentary rocks. This imbalance suggests that sedimentary samples exhibit greater overlap with non-sedimentary features in this dataset, making them harder to distinguish. The overall pattern aligns with the accuracy table: models that capture non-linear structure (RF, SVM-RBF) handle this overlap more effectively, while simpler or distribution-based models struggle. Furthermore, the misclassification trend highlights that sedimentary rocks occupy a broader and more diffuse region in feature space, increasing the likelihood of overlap with mechanically stronger lithologies. This reinforces the need for non-linear classifiers capable of modeling complex boundaries between classes. The confusion matrix also confirms that the dataset exhibits asymmetric class separability, where non-sedimentary rocks form a compact, well-defined cluster, while sedimentary rocks display greater internal variability. As a result, models with limited expressive power tend to bias predictions toward the more tightly clustered non-sedimentary class. Overall, the confusion matrix supports the conclusion that binary separation is feasible but requires sufficiently flexible algorithms to capture the underlying geo-mechanical relationships.

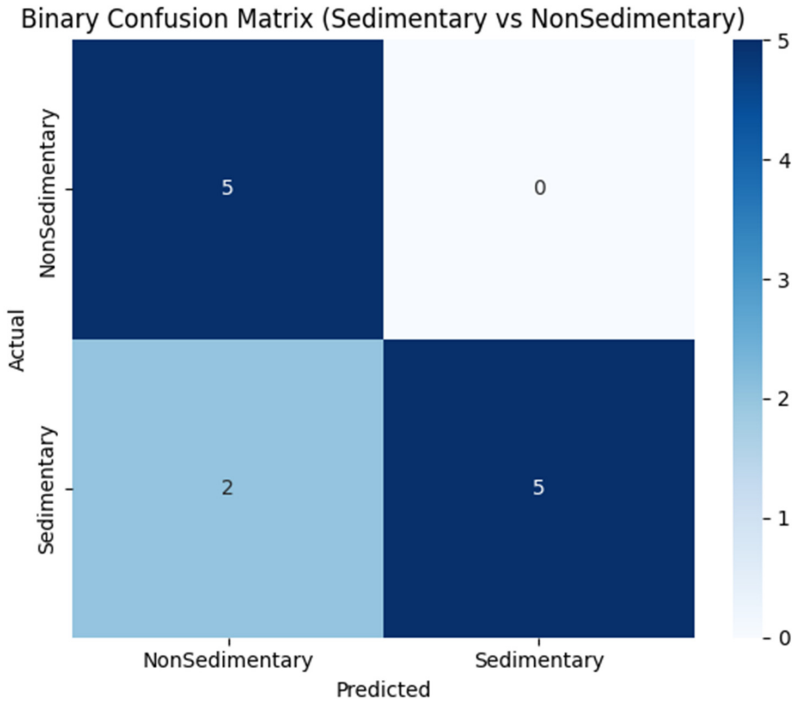


Figure 32. Binary confusion matrix for predicting between sedimentary and non-sedimentary rocks for Dataset 1.

4. Discussion

Predominantly, the study showed lithology-wise predictive behavior of the regression models trained and tested on the main dataset. A consistent pattern across all scripts is seen: predictive stability for sedimentary and metamorphic rocks and comparatively higher variability for igneous rocks. Sedimentary rocks exhibit tightly clustered R^2 distributions and minimal error, reflecting their more uniform mineralogical composition and predictable mechanical behavior. Metamorphic rocks show slightly broader distributions but still maintain high central tendencies, likely due to their foliated textures and well-defined strength–hardness relationships. Igneous rocks, however, present greater heterogeneity in grain size, mineralogy, and microstructural features, which leads to wider spreads in R^2, higher RMSE, and greater seed sensitivity. This lithology-dependent behavior underscores the importance of geological context in CAI prediction and highlights the need for more expressive models when dealing with igneous variability.

Across Scripts 1–3, a clear trend emerges: expanding the predictor set consistently improves model performance, stability, and generalization. The addition of PI in Script 2 and GS in Script 3 enhances predictive strength across all lithologies, with the largest gains observed for igneous rocks, which initially showed the greatest variability. Ensemble models, particularly Random Forest and XGBoost, benefit most from the expanded feature space, achieving near-unity R^2 values and substantially reduced RMSE and MAE. These results indicate that CAI behavior is governed by a combination of mechanical, hardness-related, and textural parameters, and that richer predictor sets allow models to capture these relationships more effectively.

Tree-based ensemble models consistently outperform linear and regularized models across all scripts. Random Forest and XGBoost demonstrate strong robustness to seed variation, high predictive accuracy, and excellent generalization under nested cross-validation. Their ability to capture non-linear interactions is particularly beneficial for igneous rocks, where linear models struggle. PCR is the weakest performer across all scripts, showing high instability, low R^2, and poor sensitivity to geological variability. This is expected, as PCR compresses the predictor space into principal components that may not align with CAI-relevant directions. Linear models (OLS, Ridge, Lasso) perform well for sedimentary and metamorphic rocks but show limitations for igneous samples, reinforcing the need for non-linear modeling in complex lithologies.

Seed-stability analyses across Scripts 1–4 reveal that ensemble models are highly robust to random initialization. XGBoost maintains exceptionally low RKFold_R^2 standard deviations across nearly all seeds, even under repeated K-Fold and nested CV frameworks. The most stable seeds produce near-perfect predicted-vs.-actual alignment, while even the least stable seeds show only minor dispersion. PCR exhibits the highest seed sensitivity, especially for igneous rocks, confirming its structural fragility. The stability patterns demonstrate that Script 3 and Script 4 provide reliable predictive frameworks, with minimal dependence on seed choice and strong reproducibility across resampling.

Unified feature importance analyses across Scripts 3 and 4 reveal consistent geological insights. UCS is the dominant predictor when included, reflecting its direct relationship with CAI and its role as a fundamental mechanical property. Schmidt hardness (H) becomes the primary driver in Script 4, with MHN providing meaningful secondary support. V_p consistently shows low importance, suggesting that velocity alone does not strongly govern CAI behavior. GS plays a selective but significant role, particularly in linear and drop-column frameworks, indicating that grain-size effects are captured more effectively by additive models. The zero values in XGB gain are methodological rather than geological, arising from shallow trees and precision thresholds, and do not imply lack of feature usage.

Overall, the importance patterns align well with known geological controls on rock strength and hardness.

The sedimentary vs. non-sedimentary classification results for Dataset 1, highlight the challenges of distinguishing lithologies using limited features. Non-sedimentary rocks are classified with perfect specificity, while sedimentary rocks show reduced sensitivity, with two misclassifications. Models with non-linear decision boundaries (RF, SVM-RBF) achieve the highest accuracy, indicating that lithological separation requires expressive decision surfaces. Simpler models struggle due to overlapping feature distributions, reinforcing the need for richer predictor sets or additional geological descriptors for reliable lithology classification.

Taken together, this study substantially advances the ability to predict CAI across diverse lithologies and under varying data conditions. By systematically evaluating multiple predictor sets, model families, stability regimes, and geological contexts, the work demonstrates that accurate and reliable CAI estimation is achievable with modern ensemble methods and appropriately selected features. While the inherent complexity of igneous rocks and the limited separability in binary lithology classification (Dataset 1) reveal that some geological challenges remain, the overall findings show that the core problem, i.e., developing a robust, generalizable, and interpretable framework for CAI prediction, can be solved within the scope of the available data. The study provides clear evidence that strength- and hardness-based predictors dominate CAI behavior, that ensemble models offer superior stability and performance, and that careful feature selection meaningfully improves outcomes. These insights not only validate the modeling approach but also offer practical guidance for future CAI estimation workflows, highlighting where predictive confidence is strongest, where geological variability demands more expressive models, and where additional data could further refine performance. In this sense, the work delivers both a functional solution and a deeper understanding of the geological and methodological factors that govern CAI prediction.

A comparison with existing CAI-prediction studies further reinforces the validity of the outcomes obtained in this work. Choi and Ko [29] demonstrated that a reduced predictor subset consisting of brittleness index B1, density, EQC, and UCS can achieve high accuracy (Test R^2 = 0.907), emphasizing that density and quartz-related mineralogical hardness are primary drivers of abrasiveness. Their reliance on tensile-based brittleness indices and petrographic EQC aligns with our finding that strength–hardness parameters dominate CAI behavior, but our results show that hardness-derived brittleness (BI, Phi, MHN, H), petrographic index (PI), and grain size (GS) provide equally strong and often superior predictive capability—particularly for igneous rocks where mineralogical heterogeneity is greatest. The findings of this study are also consistent with previous CAI-prediction research focused exclusively on igneous rocks [30]. The igneous-specific machine learning study in the reference list reported that CAI in crystalline rocks is strongly governed by mineralogical hardness, quartz abundance, and grain-scale heterogeneity, and that nonlinear ML models outperform linear regression when dealing with complex igneous textures. This aligns directly with our observation that igneous rocks exhibit the highest variability in R^2, RMSE, and seed sensitivity across Scripts 1–4, reflecting their diverse mineral assemblages and microstructural anisotropy. The literature further shows that ensemble models—particularly Random Forest and boosting algorithms—provide superior accuracy for igneous datasets due to their ability to capture nonlinear interactions among strength, hardness, and mineralogical parameters.

The predictors identified across existing CAI estimation studies provide a clear scientific foundation for the modeling choices and outcomes of this work. Prior research consistently highlights the importance of strength- and hardness-related parameters, such

as UCS, Schmidt hardness, brittleness indices, and density, as primary controls on abrasivity, with mineralogical descriptors like EQC, quartz content, petrographic indices, and grain-size heterogeneity offering additional explanatory power for lithologies with complex textures [23,29,30]. Soft-computing and machine learning studies further demonstrate that non-linear models outperform linear regression when predicting CAI, particularly in igneous rocks where mineralogical variability and microstructural anisotropy are greatest [31]. Studies incorporating drilling parameters show that operational metrics such as thrust, torque, and penetration rate also correlate strongly with CAI, underscoring the value of integrating both laboratory and in situ predictors for robust abrasivity estimation [32]. These collective insights directly reinforce the findings of the present work: the dominance of strength and hardness predictors in feature importance analyses, the superior performance of ensemble models, the heightened variability observed in igneous rocks, and the substantial gains achieved when expanding the predictor set to include PI and GS. In this context, our results not only align with established CAI-prediction literature but extend it by demonstrating how systematically structured predictor sets and modern ensemble methods can produce highly stable, generalizable, and lithology-aware CAI estimates.

Future work should also extend the CAI-prediction framework to massive sulfide and oxide rock formations, which represent a geologically and mechanically distinct class of materials. Recent electromagnetic and rock-physics characterization of massive sulfides [56] highlights their unique mineralogical composition, high density, strong anisotropy, and complex elastic behavior, factors that directly influence hardness and abrasivity but are not yet represented in current CAI datasets. In this study, Schmidt hammer hardness V_p, V_s, indirect tensile strength (IT), and point-load index (PLI) were collected for massive sulfide and oxide samples, providing an initial foundation for developing CAI-prediction models tailored to these ore-bearing lithologies. Integrating these measurements into future scripts would allow exploration of how sulfide-specific mechanical signatures differ from those of silicate-dominated rocks, and whether hardness-based predictors alone can reliably estimate CAI in mineralogically extreme environments. Given that massive sulfides often exhibit high metallic mineral content, strong acoustic attenuation, and brittle–ductile transitions that do not present in typical igneous or metamorphic rocks, incorporating these datasets will enable testing whether ensemble models remain robust under such conditions or whether hybrid physics-ML approaches are required. Ultimately, leveraging Schmidt hammer data and elastic parameters from massive sulfides offers a promising pathway toward expanding CAI prediction into mining-focused domains, improving abrasivity estimation for drilling, blasting, and geotechnical design in ore-rich environments.

Although this study demonstrates that CAI can be predicted with high accuracy using both laboratory-derived and in situ predictors, several limitations remain. The dataset size is modest, particularly for igneous and metamorphic rocks, which constrains model generalization and contributes to the higher variability observed in these lithologies. Some predictors, such as PI and GS, are not uniformly available across all samples, limiting the consistency of Scripts 2–3 and reducing the ability to explore more complex interactions. Future work should focus on expanding the dataset, especially for igneous rocks, incorporating mineralogical or geochemical descriptors to improve lithology separation, and exploring hybrid models that combine physical constraints with machine learning flexibility. Additional field validation of Script 4 predictors would further strengthen the operational applicability of in situ CAI estimation and integrating real-time drilling parameters may enable continuous CAI prediction during drilling operations. Together, these directions offer a clear pathway for refining and extending the predictive framework established in this study.

5. Conclusions

- The study demonstrates that CAI can be predicted with high reliability across major lithological groups, addressing a long-standing challenge in rock-property estimation where CAI is typically measured through destructive, time-consuming laboratory procedures.
- By systematically evaluating four predictor sets, the work shows that strength- and hardness-based parameters (UCS, H, MHN) are sufficient to achieve near-operational accuracy, meaning that CAI can be estimated without full mechanical testing, reducing laboratory workload and enabling faster decision-making in drilling, mining, and geotechnical workflows.
- Unified feature importance results provide a clear geological interpretation: UCS and hardness dominate CAI behavior, resolving ambiguity in the literature regarding which mechanical or textural parameters most strongly govern CAI. This gives practitioners a defensible basis for selecting minimal predictor sets.
- Script 4 shows that CAI can be predicted using exclusively in situ measurable parameters (H, MHN, V_p and GS), eliminating the need for core preparation and destructive UCS testing. This directly addresses the practical field challenge of estimating CAI when intact cores are unavailable or when rapid, non-destructive assessment is required.
- Ensemble models, especially Random Forest and XGBoost, consistently outperform linear and PCA-based approaches, providing a practical solution to the field problem of non-linear CAI behavior, which conventional regression methods fail to capture.
- Seed-stability analyses confirm that XGBoost maintains extremely low variability across repeated K-Fold and nested CV frameworks, meaning the study solves the reproducibility problem commonly encountered in small-sample geomechanical datasets.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/geotechnics6040105/s1.

Author Contributions: Conceptualizing: P.S.R., J.M. and S.D.B. Methodology: P.S.R. and S.D.B. Software: P.S.R. Validation: P.S.R. Formal analysis: P.S.R. Investigation: P.S.R. Resources: P.S.R. and S.D.B. Data curation: P.S.R. Writing—Original draft preparation: P.S.R. Writing—review and editing: P.S.R. and S.D.B. Visualization: P.S.R., J.M. and S.D.B. Supervision: J.M. and S.D.B. Project administration: S.D.B. Funding acquisition: S.D.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Novamera Inc. through Memorial University of Newfoundland, research contracts No. 20241493 and No. 20220331, Mathematics of Information Technology and Complex Systems (MITACS) Accelerate award IT26846, Digital Research Alliance of Canada, and the National Science and Engineering Research Council (NSERC) Alliance Grant ALLRP 597085-24 for the development and optimization of Surgical Mining Technology.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Acknowledgments: I acknowledge the efforts and contributions of the researchers at the Drilling Technology Labs (DTL), who conducted the rock formation measurements used in this manuscript. Proper references are provided throughout the manuscript wherever necessary.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

CAI	CERCHAR Abrasiveness Index
DTL	Drilling Technology Labs
UCS	Unconfined Compressive Strength
MHN	Mohs Hardness Number
OLS	Ordinary Least Square
PCR	Principal Component Regression
NPT	Non-Productive Time
WC-CO	Tungsten-Cobalt substrate
H_a	Hardness of Tool/cutter
H_m	Hardness of the Rock
NTNU	Norwegian University of Science and Technology
LCPC	Laboratoire Central des Ponts et Chaussées
QC	Quartz content
QE	Quartz Equivalence
RHR	Rosiwal's Hardness Number
VHN	Vickers Hardness Number
V_p	Longitudinal wave velocity
V_s	Shear wave velocity
v	Poisson's Ratio
BTS	Brazilian Tensile Strength
SRH	Schmidt Rebound Hardness
ML	Machine Learning
SWR	Specific Wear Rate
HRC	Rockwell Hardness Number type C
RAI	Rock Abrasiveness Index
MPa	Mega Pascals
L-RLM	Low-strength rock-like material
M-RLM	Medium-strength rock-like material
H-RLM	High-strength rock-like material
LS	Limestone
F.Gr	Fine-grained gabbro
C.Gr	Coarse-grained Gabbro
R-SS.	Red Sandstone
G/B-SS	Gray/Brown Sandstone
W-SS	Wallace Sandstone
Gr (G)	Granite Green
Gr (P)	Granite Pink
Q	Quartz
Rock_Name	Rock Nomenclature
Rock_Type	Type of Rock (Igneous, Sedimentary and Metamorphic)
Source_Reference	CAI Data Source Reference
Quartz_%	Percentage of Quartz measurement for a rock
Feldsfar_%	Percentage of Feldsfar measurement for a rock
Mica_%	Percentage of Mica measurement for a rock
Calcite_%	Percentage of Calcite measurement for a rock
Rho	Density
Phi	Porosity
H	Shore Hardness Number
R	Schmidt Hardness Number
BI	Brittleness Index
GS	Grain Size

PI	Photographic Index
G	Dynamic Shear Modulus
E	Dynamic Young's Modulus of Elasticity
K	Dynamic Bulk Modulus
Source_Type	Type of Source (Data center or Individual Lab results)
NSD	Norwegian Source Database

References

1. Mouritz, A.P.; Hutchings, I.M. *The Abrasive Wear of Rock Drill Bit Materials*; SPE Manuscript 024222; Society of Petroleum Engineers: Richardson, TX, USA, 1991.
2. Nilsen, B.; Dahl, F.; Holzhauser, J.; Raleigh, P. Abrasivity Testing for Rocks and Soils. *Tunn. Tunn. Int.* **2006**, *38*, 47–49.
3. Valantin, A. Test Cerchar Pour la Mesure de la Dureté et de l'Abrasivité des Roches. In *Annexe de l'Exposé Présenté aux Journées de l'Information*; Techniques de Creusement; European Commission: Luxembourg, 1973.
4. Cerchar, J.; Hoenig, M. *Cerchar Abrasivity Index (CAI): Its Definition and Application*; CERCHAR Technical Report; Centre d'Études et Recherches des Charbonnages de France: Verneuil-en-Halatte, France, 1980.
5. West, G. A Review of Rock Abrasiveness testing for Tunneling. In Proceedings of the International Symposium of Weak Rock, Tokyo, Japan, 21–24 September 1981.
6. Yilmaz, N.G. Abrasivity Assessment of Granite building stones in relation to Diamond Tool Wear Rate using Mineralogy based Rock Hardness Indexes. *Rock Mech. Rock Eng.* **2011**, *44*, 725–733. [CrossRef]
7. Suana, M.; Peters, T.J. The CERCHAR Abrasivity Index and its Relation to Rock Mineralogy and Petrography. *Rock Mech.* **1982**, *15*, 1–8. [CrossRef]
8. Plinninger, R.J.; Thuro, K. Wear Prediction in Hardrock Excavation using CERCHAR Abrasiveness Index (CAI). In *EUROCK 2004 & 53rd Geomechanics Colloquium*; Schubert, W., Ed.; VGE: Essen, Germany, 2004.
9. Plinninger, R.J. *Hardrock Abrasivity Investigation Using the Rock Abrasivity Index (RAI)*; Taylor and Francis Group: London, UK, 2010.
10. West, G. A Relation between Abrasiveness and Quartz Content for some Coal Measures sediments. *Geotech. Geol. Eng.* **1986**, *4*, 73–78. [CrossRef]
11. Michalakopoulos, T.N.; Anagnostou, V.G.; Bassanou, M.E.; Panagiotou, G.N. The Influence of Steel Styli Hardness on the CERCHAR Abrasiveness Index Value. *Int. J. Rock Mech. Min. Sci.* **2006**, *43*, 321–327. [CrossRef]
12. Gharahbagh, A.E.; Rostami, J.; Ghasemi, A.R.; Tonon, F. Review of Rock Abrasion Testing. In Proceedings of the 45th US Rock Mechanics/Geomechanics Symposium, San Francisco, CA, USA, 26–29 June 2011; p. ARMA 11-141.
13. Stanford, J.; Hagan, P. An Assessment of the Impact of Stylus Metallurgy on CERCHAR Abrasiveness Index. In *Coal Operators' Conference*; University of Wollongong: Wollongong, Australia; Australasian Institute of Mining and Metallurgy: Melbourne, Australia, 2009; pp. 348–355.
14. Thuro, K.; Plinninger, R.J. Hard rock tunnel boring, cutting, drilling and blasting: Rock parameters for excavatability. In *ISRM Congress*; Southern African Institute of Mining and Metallurgy: Johannesburg, South Africa, 2003.
15. Lassnig, K.; Latal, C.; Klima, K. Impact of Grain Size on the CERCHAR Abrasiveness test. *Geomech. Tunn.* **2008**, *1*, 71–76. [CrossRef]
16. Ruehl, S.; Alber, M. *Initial Stress Conditions Influencing the CERCHAR Abrasiveness Index*; Geological Society of London: London, UK, 2006; p. IAEG-436.
17. Alber, M. Stress dependency of CERCHAR Abrasivity Index (CAI) and its effect on wear of selected rock cutting tools. *Tunn. Undergr. Space Technol.* **2008**, *23*, 351–359. [CrossRef]
18. Hamzaban, M.-T.; Karami, B.; Rostami, J. Effect of Pin Speed on CERCHAR Abrasion Test Results. *J. Test. Eval.* **2019**, *47*, 121–139. [CrossRef]
19. Rana, P.S.; Molgaard, J.; Butt, S.D. CERCHAR Abrasiveness Index test limitations and improvements under bench-top and in-situ conditions. In Proceedings of the 23rd Canadian Rock Mechanics Symposium (RockEng 2025), Montreal, QC, Canada, 17–19 August 2025.
20. Zhang, G.; Thuro, K.; Song, Z.; Dang, W.; Bai, Q. Cerchar Abrasivity test and its applications in rock engineering: A review. *Int. J. Coal Sci. Technol.* **2025**, *12*, 13. [CrossRef]
21. Thuro, K.; Singer, J.; Kasling, H.; Bauer, M. Determining Abrasivity with LCPC Test. In Proceedings of the 1st Canada-U.S. Rock Mechanics Symposium, Vancouver, BC, Canada, 27–31 May 2007; p. ARMA 07-103.
22. Cheshomi, A.; Moradizadeh, M. The Relationship between Strength and Abrasion Characterizations in Granite Building Stone. *Q. J. Eng. Geol. Hydrogeol.* **2021**, *54*, qjegh2020-045. [CrossRef]

23. Hosseini, M.; Fakhri, D.; Behnam, N. Correlation between Cerchar Abrasivity index with Rock Properties in Sandstone and Tuff samples. *J. Miner. Resour. Eng.* **2024**, *9*, 47–60.
24. Yarali, O. Investigation into Relationships between CERCHAR Hardness Index and some Mechanical Properties of Coal measure rocks. *Geotech. Geol. Eng.* **2017**, *35*, 1605–1614. [CrossRef]
25. Altindag, R. Correlation between P-Wave velocity and some Mechanical properties for Sedimentary rocks. *J. South. Afr. Inst. Min. Metall.* **2012**, *112*, 229–237.
26. Capik, M.; Yilmaz, A.O. Correlation between CERCHAR Abrasivity Index, Rock properties and Drill bit lifetime. *Arab. J. Geosci.* **2017**, *10*, 15. [CrossRef]
27. Ghadernejad, S.; Esmaeili, K. Investigating the relationship between Geochemistry, Leeb rebound hardness and Cerchar Abrasiveness Index. *Int. J. Geomech.* **2024**, *24*, 04024280. [CrossRef]
28. Heydarian, P.; Asef, M.R.; Hamidi, J.K.; Talkhablo, M. The relationship between Mechanical properties and mineralogical composition of some sedimentary rocks. *Q. J. Eng. Geol. Hydrogeol.* **2024**, *57*, qjegh2024-069. [CrossRef]
29. Choi, S.W.; Ko, T.Y. Data driven estimation of Cerchar Abrasivity Index using Rock Geomechanical and Mineralogical Characteristics. *Appl. Sci.* **2026**, *16*, 552. [CrossRef]
30. Wani, S.R.; Teshnizi, E.S.; Jalota, S. Correlation between Cerchar Abrasivity Index and Geotechnical Properties of igneous rocks: A comprehensive analysis using Machine Learning algorithms and interpretative analysis. *Measurement* **2026**, *257*, 118989. [CrossRef]
31. Tripathy, A.; Singh, T.N.; Kundu, J. Prediction of Abrasiveness index of some Indian rocks using soft computing methods. *Measurement* **2015**, *68*, 302–309. [CrossRef]
32. Zhang, S.R.; She, L.; Wang, C.; Wang, Y.J.; Cao, R.L.; Li, Y.L.; Cao, K.L. Investigation on the relationship among the Cerchar Abrasivity Index, Drilling parameters and Physical and Mechanical properties of the rock. *Tunn. Undergr. Space Technol.* **2021**, *112*, 103907. [CrossRef]
33. Nguyen, V.; Tran, V.; Nguyen, D.; Sadiq, S.; Park, D. Novel hybrid MFO-XGBoost model for predicting the racking ratio of the rectangular tunnels subjected to seismic loading. *Transp. Geotech.* **2022**, *37*, 100878. [CrossRef]
34. Sadiq, S.; Won, M.; Kim, H.J.; Hong, C. Hybrid finite element and machine learning approach for estimating lateral capacity of partially rock-socketed monopiles. *Occean Eng.* **2025**, *330*, 121286. [CrossRef]
35. Won, M.; Sadiq, S.; Joung, Y.; Kim, H. GEP based empirical models for estimation of soil nail bond strength in weathered soil. *KSCE J. Civ. Eng.* **2025**, *29*, 100115. [CrossRef]
36. Plinninger, R. Hardness and drillability of rocks, Geological Society of London. *Eng. Geol. Spec. Publ.* **2003**, *19*, 43–52. [CrossRef]
37. Plinninger, R.; Thuro, K. Rock mass quality and drillability. In *Proceedings of the 3rd Austrian Swiss German Geomechanics Colloquium*; ÖGG/DGGT/STUVA: Vienna, Austria, 2003; pp. 125–136.
38. Gokceoglu, C.; Zorlu, K. A fuzzy model to predict the uniaxial compressive strength and modulus of elasticity of problematic rocks. *Eng. Appl. Artif. Intell.* **2004**, *17*, 61–72. [CrossRef]
39. Kahraman, S.; Bilgin, N.; Feridunoglu, C. Dominant Rock Properties Affecting the Penetration Rate of Percussive drills. *Int. J. Rock Mech. Min. Sci.* **2003**, *40*, 711–723. [CrossRef]
40. Thuro, K. Drillability prediction: Geological influences on TBM performance. *Geol. Soc. Lond. Eng. Geol. Spec. Publ.* **2003**, *20*, 1–14.
41. Yarali, O.; Soyer, E. The Effect of Mechanical Rock Properties and Brittleness on drillability. *Sci. Res. Essays* **2011**, *6*, 1077–1088.
42. Yavuz, M.A. Selection of plant location in the natural stone industry using the fuzzy multiple attribute decision making method. *J. South. Afr. Inst. Min. Metall.* **2008**, *108*, 641–649.
43. Suana, F.; Peters, J. The relationship between rock properties and penetration rate in rotary drilling. *Int. J. Rock Mech. Min. Sci.* **1982**, *19*, 107–118.
44. Christensen, N.I. Poisson's ratio and crustal seismology. *J. Geophys. Res. Solid Earth* **1996**, *101*, 3139–3156. [CrossRef]
45. Brocher, T.M. Empirical relations between elastic wave speeds and density in the Earth's crust. *Bull. Seismol. Soc. Am.* **2005**, *95*, 2081–2092.
46. Barton, N. Some new Q-value correlations to assist in site characterization and tunnel design. *Int. J. Rock Mech. Min. Sci.* **2002**, *39*, 185–216. [CrossRef]
47. Han, D.-H. Effects of Porosity and Clay Content on Acoustic Properties of Sandstones and Shales. Ph.D. Dissertation, Stanford University, Stanford, CA, USA, 1996.
48. Rafavich, F.; Kendall, C.H.; Todd, T. The relationship between acoustic properties and the petrographic character of carbonate rocks. *Geophysics* **1984**, *49*, 1622–1636. [CrossRef]
49. Castagna, J.P.; Batzle, M.L.; Eastwood, R.L. Relationship between compressional-wave and shear-wave velocities in clastic silicate rocks. *Geophysics* **1985**, *50*, 571–581. [CrossRef]
50. Quan, W.; Zhang, Z.; Rana, P.S.; Futheiz, I.; Alvaar, A.; Heydari, O.; Conhyedass, B.; Shah, S. *DTL Internal Reports and Quarterly Reports on Rock Characterization Between May 2017 and May 2024*; Memorial University of Newfoundland and Labrador: St. John's, NL, Canada, 2024.

51. Zhang, Z. Development and Characterization of Synthetic Rock-Like Materials for Drilling and Geomechanics Experiments. Master's Thesis, Memorial University of Newfoundland and Labrador, St. John's, NL, Canada, 2017. Available online: https://research.library.mun.ca/view/theses_dept/FacEngineering.html (accessed on 15 May 2017).
52. *ASTM D7625-22*; Standard Test Method for Laboratory Determination of Abrasiveness of Rock Using the CERCHAR Method. ASTM International: West Conshohocken, PA, USA, 2022.
53. *ASTM D4543-19*; Standard Practices for Preparing Rock Core as Cylindrical Test Specimen and Verifying Conformance to Dimensional and Shape Tolerances. ASTM International: West Conshohocken, PA, USA, 2019.
54. *ASTM D7012-14e1*; Standard Test Method for Compressive Strength and Elastic Moduli of Intact Rock Core Specimen Under Varying States of Stress and Temperature. ASTM International: West Conshohocken, PA, USA, 2014.
55. *ASTM D2845-08*; Standard Test Method for Laboratory Determination of Pulse Velocities and Ultrasonic Elastic Constants of Rocks. ASTM International: West Conshohocken, PA, USA, 2008.
56. Abbasian, L.; Rana, P.S.; Leitch, A.; Butt, S.D. Electromagnetic and Rock Physics Characterization of Massive Sulfide Rock Formations. *Geosciences* **2025**, *16*, 171. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.