



**ARTICLE**

# Geomechanical Characterization of Volcanic Pyroclast Using Machine Learning

Miguel A. Millán<sup>1,\*</sup>, Rubén Galindo<sup>2</sup> and Fausto Molina-Gómez<sup>1</sup>

<sup>1</sup>ETS Arquitectura, Avda. Juan de Herrera n°4, Universidad Politécnica de Madrid, Madrid, Spain

<sup>2</sup>ETSI Caminos, C. y P., C/Profesor Aranguren s/n, Universidad Politécnica de Madrid, Madrid, Spain

\*Corresponding Author: Miguel A. Millán. Email: miguelangel.millan@upm.es

Received: 04 February 2026; Accepted: 18 May 2026; Published: 30 June 2026

**ABSTRACT:** Low-density volcanic rocks have specific geomechanical properties that require complex laboratory tests and characterization that are not usually available in common geotechnical studies. A pyroclastic rock behaves at sufficiently “low” stress levels as if it were a conventional rock under the action of an external load, but when subjected to higher stresses, the bonds between its particles can break, leading to a sudden decrease in its volume and the reorganization of its particles, thus forming a more compact structure than the initial one. This process is known as “mechanical collapse” and involves a drastic change in the properties, which is critical for engineering purposes. A compilation and analysis of many tests performed in our laboratory was conducted, considering several inputs such as origin rock block, lithotype class, dry density, particle density, porosity, failure stresses in Cambridge variables ( $p$  and  $q$ ), deformation at failure (%), and isotropic collapse stress. The collapse stress is not available for all tests, then a reasonable prediction is made using the parabolic model of collapse. The objective of this research is to develop a machine learning procedure using this dataset to predict the isotropic collapse stress, which is challenging to measure in standard laboratory settings. This model will use a set of variables that can be easily obtained in the laboratory. Based on this evaluation, a technician can determine whether a complementary lab test is necessary or if failure is unavoidable within the given range of stresses. Two procedures were tested, XGBoost and artificial neural networks, both showing an outstanding prediction behavior.

**KEYWORDS:** Pyroclast; volcanic rock; machine learning; artificial neural networks; XGBoost; collapse

## 1 Introduction

Pyroclasts are formed by explosive fragmentation during volcanic eruptions, regardless of the subsequent transport and deposition mechanisms. Depending on these processes, they can generate fall deposits, pyroclastic flows, or pyroclastic density currents. The formation process itself imparts unique geomechanical characteristics that clearly distinguish pyroclasts from conventional rocks, differences that are particularly pronounced when pyroclasts are deposited from stable eruptive columns [1].

The eruptive column consists of a mixture of gases and pyroclastic fragments and initially rises due to the thrust generated by gas expansion upon exiting the volcanic conduit. This initial thrust is reinforced by the incorporation of ambient air immediately after emission from the crater, which dilutes the gas-ash mixture and gradually reduces its density. When the density of the mixture becomes lower than that of the surrounding atmosphere, a self-sustained convective regime is activated, allowing the column to remain vertical and stable for minutes to hours. Under these conditions, the column does not collapse onto the volcano flanks, and pyroclasts are dispersed and deposited from the column as fall deposits. However, column collapse occurs when the erupted mixture’s density exceeds the ambient atmospheric density, leading to

the generation of pyroclastic density currents driven by gravitational instability [2]. Conversely, when an excessive amount of solid material is emitted over a short period, the mixture becomes too dense, the column loses stability, and collapse occurs, giving rise to pyroclastic flows.

As a result of these eruptive processes, density becomes a key parameter in characterizing pyroclastic materials. In particular, fall pyroclasts exhibit extremely low densities due to their high internal porosity associated with their mode of formation and deposition [3]. This condition results in behavior that is markedly different from that observed in conventional rocks and soils.

This low-density state is ultimately reflected in the surface arrangement of the deposits, where volcanic materials consolidate to form macroporous deposits of very low density. The geomechanical behavior of these materials is highly distinctive, poorly understood, and difficult to predict, posing a significant challenge from both geotechnical and geological engineering perspectives.

It is especially noteworthy that, due to this low density, rock structures derived from fall pyroclasts are susceptible to mechanical collapse through destructuration. This process involves the breaking of bonds and cementation that confer rigidity to the rock structure, abruptly transforming it into soil-like material. This transformation is accompanied by large deformations and volumetric changes, associated with the high internal porosity of the material. Such abrupt changes in behavior, from rock to soil, substantially modify the geomechanical properties of the material and can severely compromise the safety and functionality of infrastructures such as roads, tunnels, dams, bridges, and other civil works, which are not designed to withstand the limit states induced by such large deformations and settlements.

Despite its importance, the threshold stress or pressure at which this mechanical collapse occurs is not adequately defined or characterized. In practice, its estimation is often based on local approximations or simplified methodologies, which introduce high levels of uncertainty and limit the predictive capacity regarding the behavior of these materials. However, despite this general understanding of the collapse phenomenon, the collapse stress associated with this structural breakdown remains poorly constrained. In particular, experimental procedures for identifying collapse under isotropic compression are not formally established within the recommended methods and suggested practices of the International Society for Rock Mechanics (ISRM), which contributes to the limited standardization and interpretation of such tests.

Therefore, it is necessary to systematically investigate the stress conditions responsible for the mechanical collapse of fall pyroclastic materials. In this context, the present study combines an extensive experimental laboratory campaign on pyroclastic materials obtained from blocks collected from macroporous outcrops in the Canary Islands with machine learning techniques aimed at predicting the collapse stress from measurable mechanical parameters. This approach seeks to improve the understanding of the geomechanical behaviour of these materials and contribute to a more reliable assessment of their response to applied loads.

The combination of very high porosity, low bulk density, and delicate bonding typical of fall pyroclasts in the Canary Islands makes their macro-porous rock frameworks unusually prone to abrupt destructuration under modest confinement, i.e., a collapse transition from rock-like to soil-like behavior once a critical stress threshold is exceeded. Field and laboratory syntheses for volcanic terrains in the archipelago consistently report low densities and low strength across several volcanic lithologies, underscoring the engineering relevance of such collapse phenomena in local infrastructures [4,5]. This sensitivity is further conditioned by diagenetic processes (e.g., zeolitization of pumiceous tuffs), which alter bonding and, therefore, the stress level at which fabric breakdown initiates [6]. Recent experimental work tailored to pyroclasts from the Canaries has also formalized procedures to capture structural collapse in isotropic compression, reinforcing the need to quantify a reliable “collapse stress” for design [5].

In parallel with these geologic and geomechanical insights, supervised machine learning (ML) has matured into a robust framework for mapping measurable physical or petrographic indices to strength metrics in rocks and soils. In intact rocks, gradient-boosting models, particularly XGBoost, and artificial neural networks (ANNs) have repeatedly delivered state-of-the-art predictions of uniaxial compressive strength (UCS) and elastic stiffness using inputs such as porosity, P-wave velocity, Schmidt rebound number, density, and point-load index [7,8]. Boosting models tuned with metaheuristics and interpreted with SHAP values have further improved accuracy and provided transparent rankings of feature influence across diverse lithologies [9,10]. Recent reviews confirm that ANNs and tree-based ensembles (including XGBoost) now dominate AI-based rock strength estimation, frequently achieving  $R^2 > 0.9$  when suitable descriptors are available [11,12]. Notably for volcanic rocks, ANNs have modeled degradation of ignimbrites/tuffs under freeze-thaw and predicted post-weathering UCS from minimal inputs, while other ANN frameworks retrieved igneous-rock strength parameters (UCS, Brazilian tensile strength, abrasion) directly from petrographic composition, evidence that ML can learn structure-property relations in vesicular/pumiceous matrices akin to those of pyroclasts [13,14].

Beyond rocks, ML has proven effective for geotechnical properties of soils, capturing nonlinear effects of index variables on shear strength and compressibility-problems analogous to collapse susceptibility. Boosting and ANN models have predicted soil shear strength from routine indices with high fidelity and provided physically meaningful rankings of variable importance via SHAP analysis [15,16]. Moreover, ML models have been successfully trained to predict the collapse potential of collapsible soils (loess) from oedometer data and index properties, offering a conceptual and methodological precedent for learning a collapse threshold from laboratory compression paths [17,18]. Although loess and fall pyroclasts differ in origin, both materials share metastable, highly porous fabrics whose destructuration is triggered under stress and/or wetting, hence benefiting from data-driven, multi-parameter inference [17]. Moisture-strength couplings reported for fine pyroclastic materials further motivate explainable ML in this context [19].

A complementary approach is the Physics-Informed Neural Networks (PINNs), which train a neural model by enforcing the governing equations, initial and boundary conditions, and data when available. This approach enables physically consistent forward and inverse solutions under sparse/noisy measurements [20,21]. As recent advances, regarding tunnelling-induced deformation, Zhang et al. [22] embed static equilibrium and Mohr-Coulomb elastoplasticity within a PINN to reproduce excavation-driven settlement fields with adaptive sampling near plastic zones. In layered ground, Pan et al. [23,24] propose a multi-physics PINN that links tunnel convergence, face position, and stratigraphy and deep tunnel seepage. More broadly, Luo et al. [25] address steady and free surface seepage via PINNs. In the unsaturated regime, Li et al. [26] couple a PINN with Hydrus 1D to solve Richards' equation for infiltration, while Bandai and Ghezzehei [27] use domain decomposition for layered soils and Depina et al. [28] perform inverse estimation each enforcing the Richards PDE and associated constitutive relations. For consolidation, Wang et al. [29] use PINNs to recover spatiotemporal pore pressure fields and consolidation parameters (Terzaghi/Biot type PDEs), and Hong and Kim [30] predict staged loading settlements directly from field monitoring; Lu and Mei [31] report agreement with FEM in a 2D railway case study. Regarding stability of slopes, Zhang et al. [32] present a physics-guided PINN framework grounded in plasticity and limit analysis (Mohr-Coulomb) for 3D slope stability without labeled data, while Moeineddin et al. [33] encode thermo-poro-mechanical shear band physics to model catastrophic creeping landslides.

Motivated by these advances, we develop ML models to predict the collapse stress associated with the onset of destructuration in fall pyroclastic rocks from an extensive laboratory campaign on block samples from macro-porous outcrops from the Canary Islands, measuring candidate predictors that are either rapidly obtainable or non-destructive. By explicitly targeting the collapse stress, rather than generic strength proxies,

our study bridges a gap between volcanology and geotechnical engineering: it converts a difficult-to-define threshold in highly porous pyroclastic rocks into a predictable function of measurable descriptors, thereby reducing design uncertainty for infrastructures founded on or interacting with such deposits [4,34].

The objective of this research is to develop an artificial intelligence tool that can estimate the collapse pressure of pyroclastic rocks without requiring complex tests in specialized laboratories. This involves creating a machine learning procedure based on a comprehensive set of laboratory tests conducted at the UPM laboratory. The model will take into account key parameters involved in the process and will either directly measure the collapse pressure or estimate it using established models when direct measurement is not possible.

## 2 Pyroclastic Rock Datasets

The detailed definition, hypotheses adopted, and methods employed to solve the proposed problem are outlined in the following subsections.

### 2.1 Laboratory Tests

The materials investigated in this article are pyroclasts originating in the Canary Islands (Spain), specifically collected in La Palma, Tenerife, and El Hierro. The sampling collection addressed the carving of *in situ* blocks of approximately  $30 \times 30 \times 30 \text{ cm}^3$ , which subsequently were transported to the laboratory in Madrid [35]. To reduce the possibility of sample alterations during transport, the following handling procedures were adopted: covering of blocks with wax and construction of wooden boxes with foam walls.

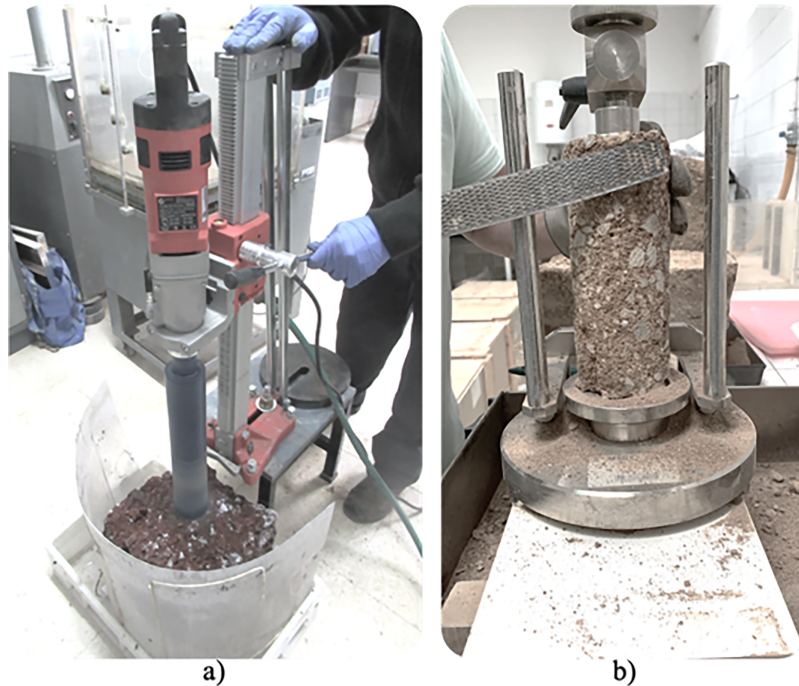
The lithotypes of the blocks used herein cover basaltic and sialic pyroclasts. These lithotypes were categorised according to the classification recommended by the Canarian Government [36], elaborated by Hernández-Gutiérrez and Rodríguez-Lozada. Table 1 shows the classification of lithotypes. The samples examined correspond to the lithotypes: lapilli, scoria, pumice, basaltic ashes, and sialic ashes. In addition, the blocks were separated into welded and lithified materials.

**Table 1:** Lithotypes of pyroclastic materials.

General Type	Lithotype	Structure	Code
Basaltic pyroclasts	Lapilli (LP)	Loose (S)	LPS
		Welded (T)	LPT
	Scoria (Es)	Loose (S)	ESS
		Welded (T)	ESW
	Basaltic ashes (CB)	Loose (S)	CBS
		Welded (T)	CBT
Sialic pyroclasts	Pumice (PZ)	Loose (S)	PZS
		Welded (T)	PZT
	Sialic ashes (CS)	Loose (S)	CSS
		Welded (T)	CST

After identifying the welded and lithified materials, cylindrical samples were prepared for element testing. Due to the high and uneven porosity of Canarian pyroclasts, two distinct procedures for sample preparation were employed [5]. For the welded (structured) materials, blocks were cored using a mechanical

drill designed for rocks. In contrast, the lithified (unstructured) materials were trimmed with a sharp saw. Despite careful preparation, the high porosity of the pyroclasts made the materials sensitive, resulting in various irregularities along the specimens. Consequently, some specimens were unsuitable for mechanical testing, but they were still used for physical classification, including assessments of porosity and specific gravity of solid particles. Fig. 1 illustrates the sample preparation procedures.



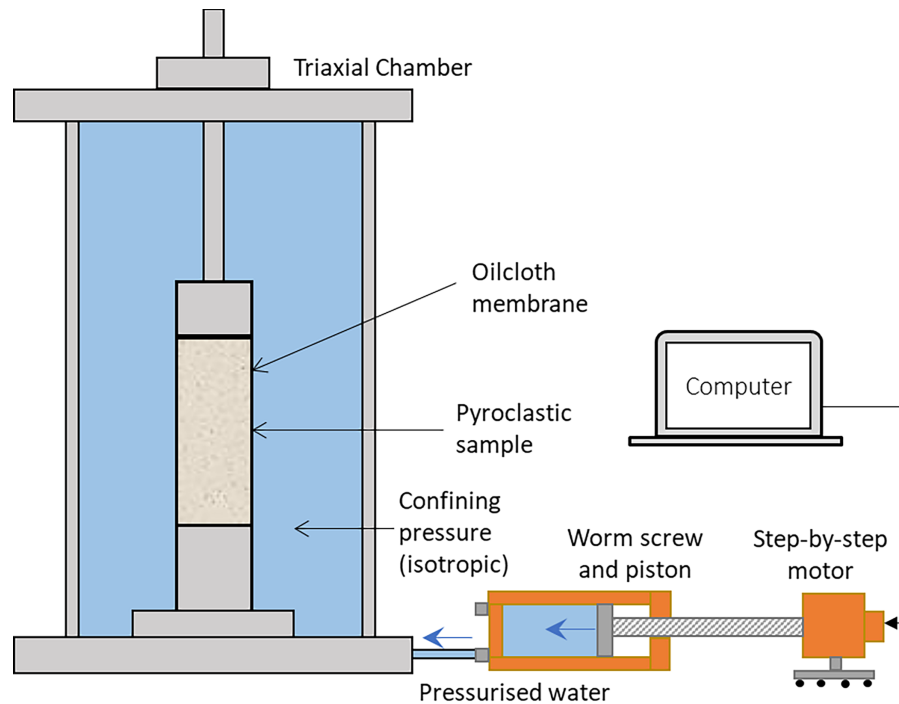
**Figure 1:** Preparation procedures of cylindrical samples (Molina-Gómez et al. 2025 [5]): (a) drilling of welded pyroclasts; (b) trimming of lithified pyroclasts.

The tests conducted to characterise the shear strength of the Canarian pyroclasts included unconfined, triaxial, and isotropic compression tests. These assessments helped establish a failure criterion proposed in Serrano et al. [37] that describes the shear strength of pyroclasts with high porosity. All tests were conducted in dry conditions. For this purpose, the samples were dried in an oven for 24 h before mechanical testing.

The unconfined compression tests were carried out under a stress control rate of 200 kPa/s. The triaxial tests were performed with axial strains up to 30% under various confinement pressures ( $\sigma_3$ ), a minimum of three, to characterise the shear strength of Canarian pyroclasts under different stress paths. To prevent punctures due to the irregularities of porous samples, oilcloth membranes—more durable than traditional latex membranes—were utilised [35]. The isotropic compression tests were designed to estimate the key parameter of the failure criterion known as the collapse pressure ( $P_c$ ), which is the main focus of this research. During these tests, the confining pressure was increased isotropically until a sudden change in volumetric strain ( $\epsilon_v$ ) was observed. This variation in  $\epsilon_v$  indicates the beginning of bond loss between particles, leading to structural collapse and the creation of new material [38].

For the assessment of  $P_c$ , an upgraded triaxial testing apparatus was utilised. This equipment features a metallic triaxial chamber and a series of automatic pressure and volume controllers capable of simultaneously increasing and measuring parameters to estimate both  $\sigma_3$  and  $\epsilon_v$ . The controllers have a capacity of 250 cm<sup>3</sup> and can exert a maximum pressure of 3500 kPa. The test automation of this system is managed by a stepper motor, which drives a worm screw connected to a piston. This piston pushes demineralised and deaerated

water (habitually demineralised and deaerated) into the cylinder, thereby pressurising the triaxial chamber. The controllers are also linked to a computer, which reports pressure measured by an external pore pressure transducer and the volume changes estimated by counting the number of motor steps. Fig. 2 illustrates this innovative configuration and its operating principle.



**Figure 2:** Configuration for isotropic compression tests.

As part of the extensive experimental campaign carried out for this research, a total of 175 mechanical tests were conducted to generate the dataset used for training the machine learning methods. These tests were distributed in 48 unconfined compression tests, 89 triaxial tests, and 25 isotropic compression tests. Additional test results were obtained but later excluded due to inconsistencies such as equipment non-compliance (e.g., membrane punctures) or anomalous values that deviated significantly from the overall trends observed in the dataset.

The selected and validated results contributed to the characterisation of key mechanical parameters, including the mean effective stress and deviatoric stress at failure ( $p_f$  and  $q_f$ , respectively) and  $P_c$ . In addition to the mechanical behaviour, fundamental physical properties were measured for all lithotypes involved in the study. These include porosity ( $n$ ), dry unit weight ( $\gamma_d$ ), and specific gravity of solid particles ( $G_s$ ) of all lithotypes. These results collectively formed a robust and comprehensive database that facilitated the development and validation of a reliable ANN model for predicting the shear strength parameters of welded and lithified pyroclastic materials under various stress path conditions.

## 2.2 Collapse Pressure Definition

The possibility of collapse is a critical aspect of macroporous materials, as it governs their geomechanical behavior and the safety of structures founded on them. Therefore, it is necessary to define a geomechanical parameter that characterizes mechanical collapse and the destructure of the material, a phenomenon specific to these deposits that does not occur in conventional rocks.

In conventional soils and rocks, failure under isotropic loading does not typically occur, and strength is primarily evaluated using shear criteria, such as the Mohr-Coulomb or Hoek-Brown criteria. In contrast, pyroclastic materials exhibit an additional failure mode: collapse. This unique behavior requires a specific parameter to quantify it and predict rock failure under these conditions.

In this context, collapse pressure ( $P_c$ ) is defined as the isotropic load that induces material failure. In laboratory experiments, this parameter is determined by progressively increasing the isotropic stress of the sample until destructuration occurs, accompanied by an abrupt volumetric change. The value of the isotropic stress at the point of failure is assigned as the collapse pressure, providing a quantitative measure of the collapse threshold for these macroporous materials.

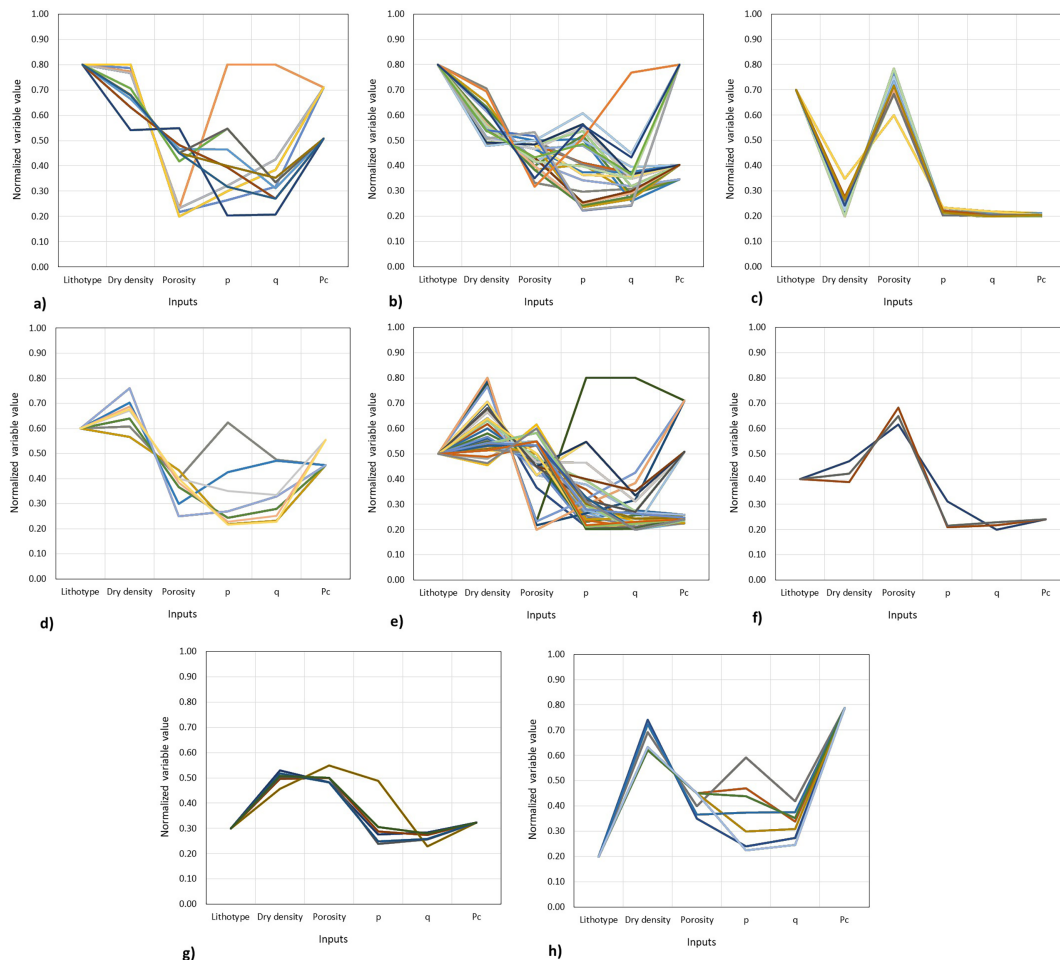
Collapse pressure can be estimated in the laboratory using specific tests with isotropic paths in a triaxial chamber. However, these tests are not routinely performed in standard laboratory practice. Their execution is more demanding, as they require accurate measurement and interpretation of volumetric strain evolution under hydrostatic loading. The identification of the collapse pressure relies on detecting subtle stiffness changes associated with the transition from a structured rock to a soil-like material. While this transition can be relatively clear in highly homogeneous macroporous materials, it becomes significantly more difficult to identify in natural rocks, where pore structure is typically heterogeneous. In such cases, the interpretation of collapse tests is not straightforward and may introduce considerable uncertainty, particularly due to: (1) the difficulty in precisely capturing volumetric strain; (2) the absence of a sharp or well-defined collapse point; (3) the limited experimental experience available compared to conventional tests. Additionally, laboratory determination is often challenging because very high values cannot be reached in conventional chambers, while low values are difficult to obtain in rock cells. Therefore, when specific collapse pressure data are unavailable, the approximation proposed by Serrano et al. [37] based on conventional triaxial tests is commonly used.

### 2.3 Dataset Organization

As described above, a homogeneous set of data is obtained, with every row including origin rock block, lithotype class, dry density, particle density, porosity, failure stresses in Cambridge variables ( $p$  and  $q$ ), deformation at failure (%), and isotropic collapse stress.

Each row of data is not entirely independent from the others, as a specific block generates several specimens, each of which undergoes a predefined series of tests. Therefore, all specimens from the same block are related, and this relationship is crucial for accurately determining the collapse pressure, especially if the value was not obtained directly from testing. Understanding this relationship is also essential for building an effective neural network. Fig. 3 shows the relation (in normalized variables) between input variables corresponding to the different tests grouped by lithotype. As can be seen, the amplitude of the variables is restricted to certain intervals for each group of tests, limiting the ML procedure's possible inputs to be between those combinations of values.

The input features refer to those variables measured in the lab tests and used as input parameters for the machine learning procedure. The parameters obtained in the lab are physical characteristics of the material (such as lithotype, dry density, and porosity), while others are related to its strength (in particular, one specific failure load pair  $p$  and  $q$ ). Each set of parameters can be associated with one specific strength test on one specimen (for example, an unconfined compression test or triaxial test), and their values always come from laboratory measurements. The adopted input for the final dataset comprises two parameter sets obtained from the same bedrock.



**Figure 3:** Graphical representation of the relationships between input variables for the different lithotypes. (a) Ash; (b) Weakly welded ash; (c) Slag; (d) Lapilli; (e) Altered lapilli (f) Pumice; (g) Altered pumice; (h) Heavily lithified pumice.

The target variable is only one variable, the collapse pressure, and should have nearly the same value for a specific set of specimens coming from the same bedrock (assuming certain homogeneity). This value is obtained from a lab test, but due to its difficulty, it is not available in many cases. Not having the collapse pressure for many inputs would have made it impossible to define a machine learning procedure; then, the authors adopted the well-established procedure of filling up the gaps using a well-known physical behavior of the rock. In this case, the number of tests available shows a tendency coherent with the cap shape of the failure criterion; consequently, a parabolical assumption for the missing cases seems reasonable. Consequently, some collapse values were estimated using data from a range of specimens sourced from the same block. This dependency will also influence the predictions made by the machine learning procedures.

When designing the machine learning procedure, two different options were considered.

The first option used the test values of a single specimen as inputs, resulting in a much simpler procedure. In theory, it is possible to predict collapse pressure (Pc) from the test results of a single specimen, because the relevant information is contained in the measured parameters. However, due to the dispersion of these parameters, they do not uniquely capture the rock's failure behavior. Two specimens from the same family can report different Pc values. The tests carried out with this first approach were not successful, so an alternative procedure was adopted.

The second option is based on the idea that two specimens from the same block can complement each other and provide more consistent information to the machine learning model, partially compensating for the dispersion and contradictions that laboratory results often exhibit. An additional advantage is that the number of possible combinations of sibling specimens (including combinations where one specimen is duplicated) is greater than the number of individual specimens, providing a much more comprehensive dataset for training.

In practice, the second option produced significantly better results and was therefore selected for this research.

The final inputs and the single output of the machine learning procedure are detailed in [Table 2](#).

**Table 2:** Inputs and outputs. Variable code names.

Input 1	Input 2	Input 3	Input 4	Input 5	Input 6	Input 7	Input 8	Input 9	Input 10	Output
Lithotype 1	Dry density 1	Porosity 1	p1	q1	Lithotype 2	Dry density 2	Porosity 2	p2	q2	Pc

To overcome this limitation, the input for the net is defined to include two specimen datasets from the same rock block. This approach simplifies and improves the characterization of the block pyroclast and the definition of the pressure collapse.

Since rock mechanical characterization inherently requires multiple specimens and stress paths to define the material response, the use of several conventional tests as input is consistent with standard practice and does not represent an additional burden introduced by the model. The input variables are derived from uniaxial compression and triaxial tests, which are standard, widely available, cost-effective, and straightforward to interpret, allowing the prediction of isotropic collapse without the need to perform collapse tests, which are not routinely performed, and their interpretation—based on volumetric strain evolution—is often uncertain, particularly in natural macroporous rocks with heterogeneous pore structures.

The final number of available input sets and the range of variation of the input parameters is shown in [Table 3](#).

**Table 3:** Datasets and range of variation of the input parameters.

Lithotype	N Tests 163	N Pairs 1041	Dry Density (g/m <sup>3</sup> ) Min/Max	Porosity % Min/Max	P (MPa) Min/Max	q (MPa) Min/Max	Pc (MPa) Min/Max
Ash	7	155	4.68/12.8	0.54/23.6	0.32/5.03	0.05/4.6	0/19.58
Weakly welded ash	6	37	8.53/9.62	0.62/0.66	0.49/3.71	0.62/1.78	4.18/4.18
Slag	3	10	7.5/8.75	0.7/0.74	0.12/1.43	0/0.6	1.43/1.43
Lapilli	63	365	8.51/13.7	0.45/0.7	0.02/7.7	0/12.6	0.87/17.0
Altered lapilli	28	29	9.8/13.18	25/27.41	0.476/0.64	0.2/5.43	0.6/5.8
Pumice	20	155	4.68/6.92	0.69/0.8	0.03/0.43	0/0.4	0.09/0.50
Altered pumice	25	224	8.93/12.3	0.53/0.63	0.47/4.67	1.28/4.1	4.95/6.83
Heavily lithified pumice	11	66	9.8/13.70	0.45/0.66	0.05/7.7	0.16/12.6	10.3/17.0

### 3 Machine Learning Methods

Two artificial intelligence tools are utilized to solve this problem, assessing their relative efficiency and overall performance to determine the best one.

### 3.1 XGBoost

#### 3.1.1 Model Structure Definition and Training

XGBoost (Extreme Gradient Boosting) is a gradient-boosted ensemble of decision trees trained sequentially. The model is built additively: at each boosting iteration, one new tree is added to update the current predictions. Each new tree is fitted to reduce the remaining errors, and the final prediction is the sum of the outputs of all trees. Split selection is guided by first- and second-order derivatives of the loss (gradient and Hessian), which determine both the utility of candidate splits and the corresponding leaf weights under regularization. Typical engineering and scientific applications include reliability prediction, structural health monitoring, materials property estimation, quality control, risk modeling, and learning-to-rank [39–41].

In this study, the dataset contains  $N = 1041$  observations and  $m = 10$  numerical explanatory variables. Although tree models do not require feature scaling, the same initial scaling used for the other methods is also applied here for consistency. Missing values (NaN), if present, are not imputed; instead, XGBoost assigns a default branch direction for missing values at each split, exploiting sparsity-aware training [39].

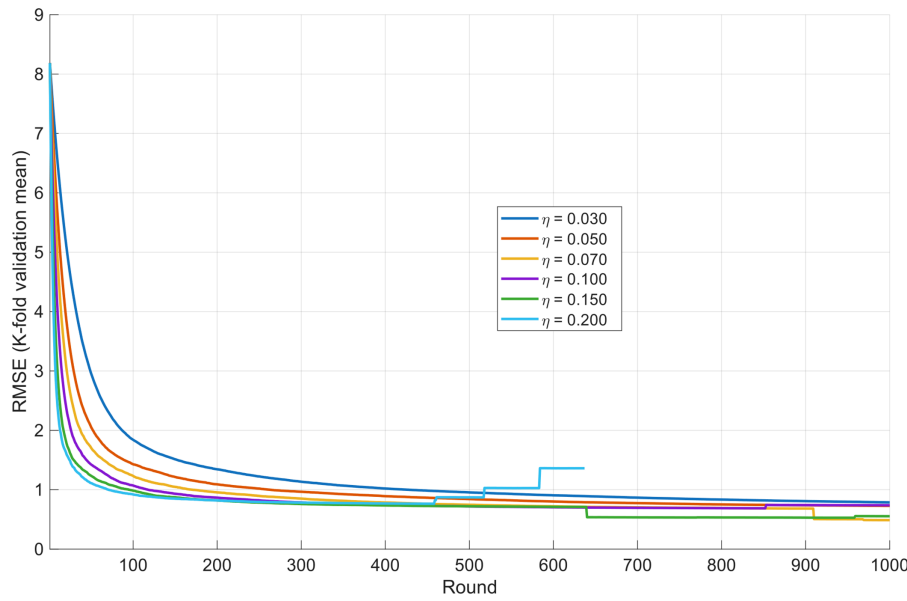
#### 3.1.2 Hyperparameter Tuning

The main hyperparameters that affect model performance are the learning rate (which controls how much each new tree influences the predictions), the maximum number of rounds (the number of iterations), the tree depth (which determines the complexity of each tree; deeper trees capture more interactions but may overfit), and regularization (shrinks leaf values to avoid overfitting).

A series of learning rate ( $\eta$ ) candidates is considered in successive iterations (0.03, 0.05, 0.07, 0.10, 0.15, 0.20). A smaller  $\eta$  typically requires more boosting rounds but yields smoother validation curves and greater resistance to overfitting; a larger  $\eta$  converges faster but risks earlier overfitting. The maximum number of rounds is set conservatively high ( $n_{\text{rounds}} = 1000$ ) to allow early stopping to locate the optimum. Depth and regularization are set to reasonable defaults ( $\text{max\_depth} = 4$ ;  $\text{subsample} = 0.8$ ;  $\text{colsample\_bytree} = 0.8$ ;  $L2 \lambda \approx 1.0$ ; optional  $L1 \alpha \approx 0.0$ ) unless domain constraints suggest otherwise [40].

Model selection is performed with K-fold cross-validation ( $K = 5$ ). For each candidate learning rate ( $\eta$ ), we train up to a high ceiling of boosting rounds ( $n_{\text{rounds}} = 1000$ ) with early stopping based on the chosen validation metric. Early stopping patience is set to 20–50 rounds (30 in this case); the best iteration is the one that achieves the minimum validation error (for lower-is-better metrics) or the maximum score (for higher-is-better metrics). This procedure determines both the optimal  $\eta$  and the effective number of boosting rounds in a statistically robust manner [39]. The validation curves for  $\eta$  obtained in this selection process are illustrated in Fig. 4.

Primary metrics reported are RMSE, MAE, MAPE (percentage), and  $R^2$ . For tuning, a single metric is selected (e.g., RMSE). In the internal scoring, lower-is-better metrics (RMSE/MAE/MAPE) are maximized by negation ( $\text{score} = -\text{metric}$ ), whereas higher-is-better metrics ( $R^2$ ) are used directly ( $\text{score} = R^2$ ). Table 4 presents the results of the iterative optimization process, yielding the selected  $\eta$  and the best iteration (mean  $\pm$  SD across folds).



**Figure 4:** Validation curves by  $\eta$  (RMSE, lower is better)—Full-data.

**Table 4:** Optimization process of statistical results of the optimal learning rate.

Learning Rate ( $\eta$ )	Mean Best Round	STD Best Round	Mean Best Score	STD Best Score
0.03	999.8	0.44721	-0.7867	0.3489
0.05	996.8	3.0332	-0.72836	0.3752
0.07	961.2	51.766	-0.67927	0.40074
0.1	963	78.291	-0.68318	0.36816
0.15	825	153.71	-0.70197	0.39046
0.2	502	78.272	-0.76036	0.34338

Once the optimal  $\eta$  and the effective number of rounds are selected, a final model is trained with all available data. A separate holdout set (20% of the data) is used to evaluate out-of-sample performance with the chosen metric, RMSE. For the optimal  $\eta$ , the results are:

- Outer 5-fold: TEST RMSE = 0.804685 ( $\pm 0.223820$ )
- Full-data CV selected: learning rate  $\eta = 0.070$  | rounds = 961
- Interpretability holdout (20%): RMSE = 0.121195
- Additional parameters: n\_estimators = 500 (max number of trees) with early\_stopping\_rounds = 30 (stop if validation doesn't improve); max\_depth = 4 (maximum tree depth); min\_child\_weight = 1 (minimum leaf/Hessian weight to allow a split);  $\gamma = 0.0$  (minimum loss reduction required for a split); subsample = 0.80 (row subsampling per tree); colsample\_bytree = 0.80 (feature subsampling per tree); Regularization: L2:  $\lambda \approx 1.0$ ; optional L1:  $\alpha \approx 0.0$ .

When analyzing individual errors, it's important to understand that their significance is relative to their magnitude within the context of the overall data range. For instance, a small deviation in values near zero can result in a significant error value, even though it may be negligible in practical terms when compared to the entire range. To address this issue, we define a central error that is relative to the mean value of the range:

$$e = \frac{y_{target} - y_{output}}{y_{target} \cdot \mu(y_{target})} \quad (1)$$

where  $e$  represents the individual centered error,  $y_{target}$  and  $y_{output}$  the reference and predicted values of the collapse pressure, and  $\mu(y_{target})$  the mean value of the whole range.

Fig. 5 presents the prediction errors. Fig. 5a shows the individual errors for each input, and Fig. 5b shows the histogram of errors.

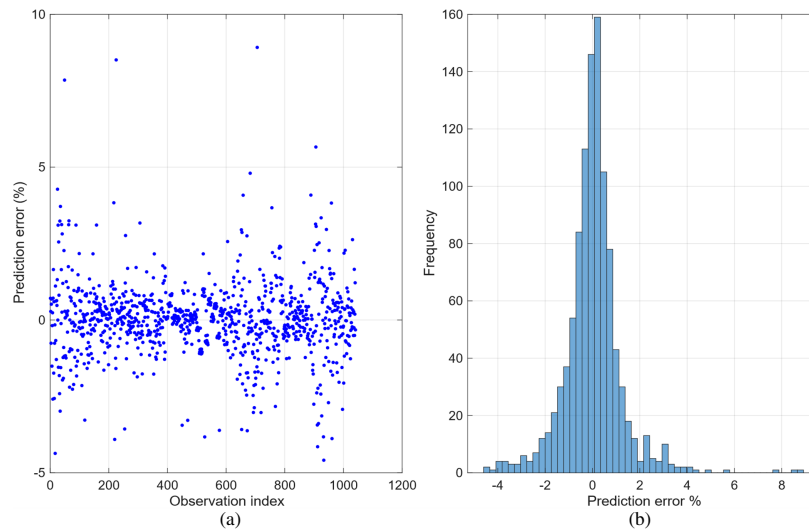


Figure 5: (a) Prediction errors per observation, (b) Residual distribution.

Fig. 6 compares the predictions and the original outputs, showing a very good coincidence. Fig. 6a displays a superposition of both series of values, with very small differences. Fig. 6b presents the regression line between prediction and outputs.

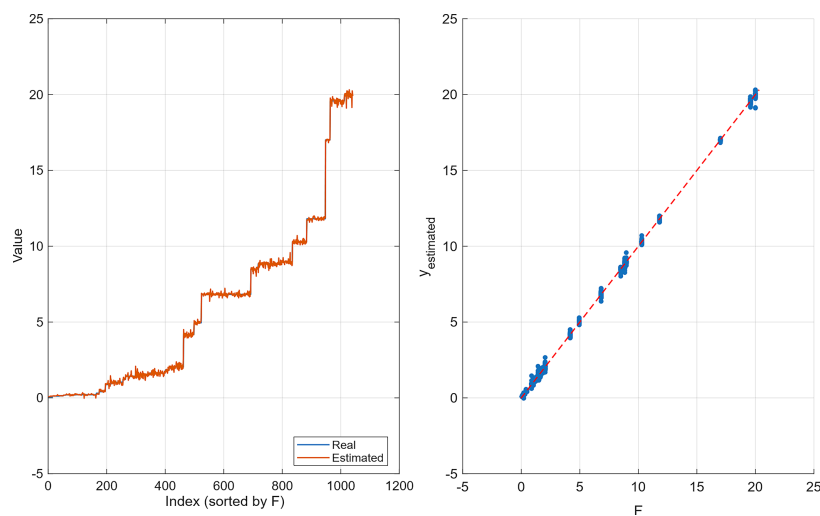


Figure 6: (a) Overlay: Real vs. Estimated, (b) Scatter with  $y = x$ .

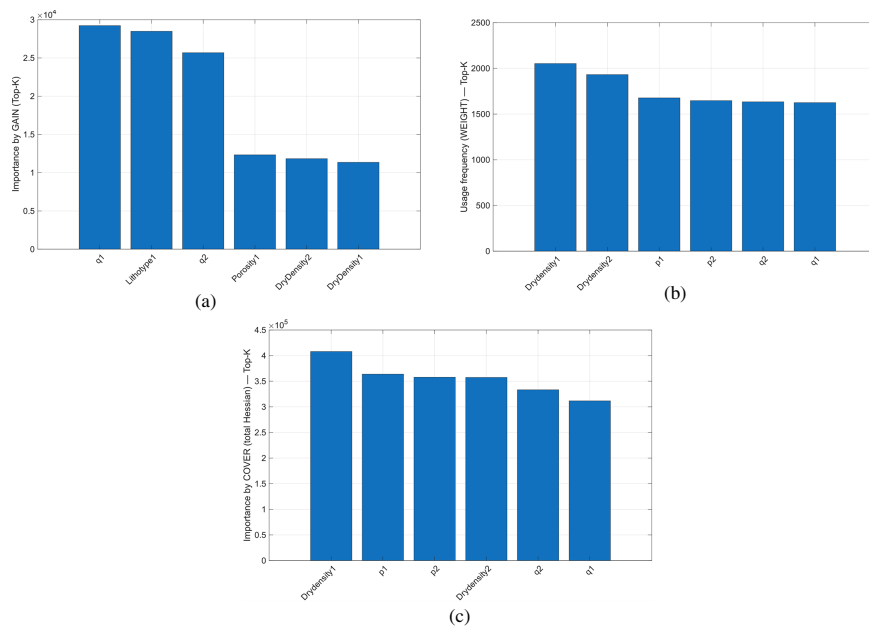
### 3.1.3 Model Performance

Analysis of feature importance allows for the characterization of model performance. Two complementary families of importance are provided: (i) intrinsic tree-based importance (gain, weight/frequency, cover) and (ii) model-agnostic permutation importance.

Fig. 7 shows the three intrinsic graphics of importance. XGBoost computes three intrinsic importance scores for each feature (j) by aggregating statistics from every split in the trained ensemble where that feature is used [39]. “Feature” refers to every input, and “split” refers to the separation that the model performed on any of the trees in the ensemble during training.

In brief, the intrinsic performance indicators show the following results:

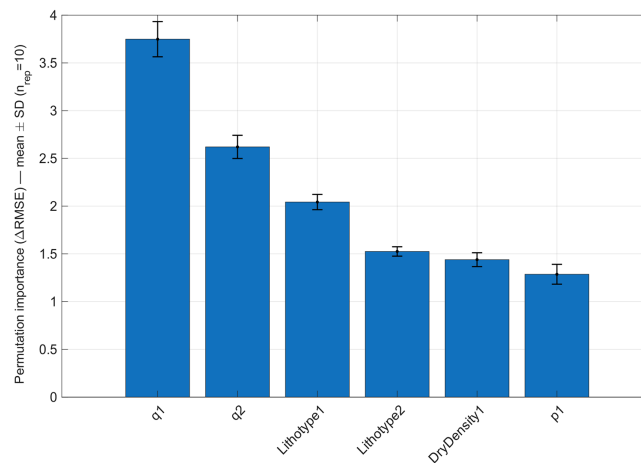
- Gain measures the total improvement in the objective attributable to splits using a feature (j). In other words, it indicates how much the model’s regularized objective improved when a split used feature (j). A feature with high gain tends to produce large reductions in error at the nodes where it is used. It is the most “effect size” oriented metric. Fig. 7a shows the global ranking of features that most reduce loss on the collapse pressure. In order, the most influential features are q1, Lithotype1, and q2, corresponding to the deviatoric stress in specimens 1 and 2, and the rock lithotype (identical for both specimens since they belong to the same family).
- The ratio Weight/Frequency or usage frequency counts how often feature (j) appears in splits across all trees. It captures prevalence rather than effect size. Fig. 7b presents that the features that show more often are particle density 1 and 2, q1 and p1, and q2 and p2, with little difference between them. They do not represent a higher impact per split because their gain is not that relevant, indicating that these features may be used often, yet have small gains each time.
- Cover measures how many training instances are affected by splits on feature (j). In XGBoost, “cover” is computed from the sum of second-order weights (Hessians) over the instances routed through nodes where the feature is used. Fig. 7c shows that the more affected features are particle density 1 and 2, p1 and p2, and q1 and q2, as in Fig. 7b, but with different values.



**Figure 7:** Intrinsic tree-based importance graphics. (a) Gain, (b) Weight/frequency, (c) Cover.

Regarding the model-agnostic permutation importance, permutation importance (PI) quantifies how much the model's generalization score drops when the values of a single feature are randomly shuffled in a validation set, thereby breaking any association between that feature and the target. If the score degrades strongly, the model relied on that feature; if the score stays the same, the feature was not important for that model. This method is model agnostic, meaning it does not depend on the type of model or its internal structure, and can be repeated many times to estimate variability: it shuffles a column and measures the drop in performance; repeating this several times yields a distribution and error bars.

Fig. 8 reports permutation feature importance computed on a held-out validation set. For each feature, we shuffle its values (10 independent permutations), re-evaluate the model, and measure the increase in RMSE relative to the baseline. The bar height is the mean  $\Delta$ RMSE across repeats; the error bars show mean  $\pm$  SD across repeats, enabling robust comparison and highlighting correlated-feature effects that may dampen individual permutations [42,43]. Because RMSE increases when the feature is corrupted, higher bars imply greater reliance on that feature for accurate predictions.



**Figure 8:** Permutation importance represented by the increment of RMSE, including the mean value and the standard deviation with repeats and error bars.

This procedure is recommended for estimating how much our model reliably relies on each variable and for quantifying the variance of the importance under data resampling and random shuffles. These permutations may represent heterogeneity and uncertainty of the physical characteristics of the volcanic material and help to validate the stability of the importance analysis, considering the model's total dependency on one variable.

In this model, q1 is the most influential ( $\Delta$ RMSE  $\approx$  3.2), followed by q2 ( $\approx$  2.4). Both are very stable (net separation and small SD), and the third one, Lithotype 1, is quite stable (clearly above the tail types).

Features Lithotype 2, Dry-density 1, and p1 show more modest effects ( $\approx$  1.3–1.5). The relatively narrow error bars for most features indicate stable importance estimates across permutations, but the SD bars overlap; therefore, we refrain from claiming a stable order among them.

### 3.2 Artificial Neural Network

#### 3.2.1 Artificial Neural Network Architecture

A feed-forward artificial neural network consists of several interconnected layers of neurons, with the outputs of each layer forwarded to the next layer. The first one is the input layer (with as many neurons

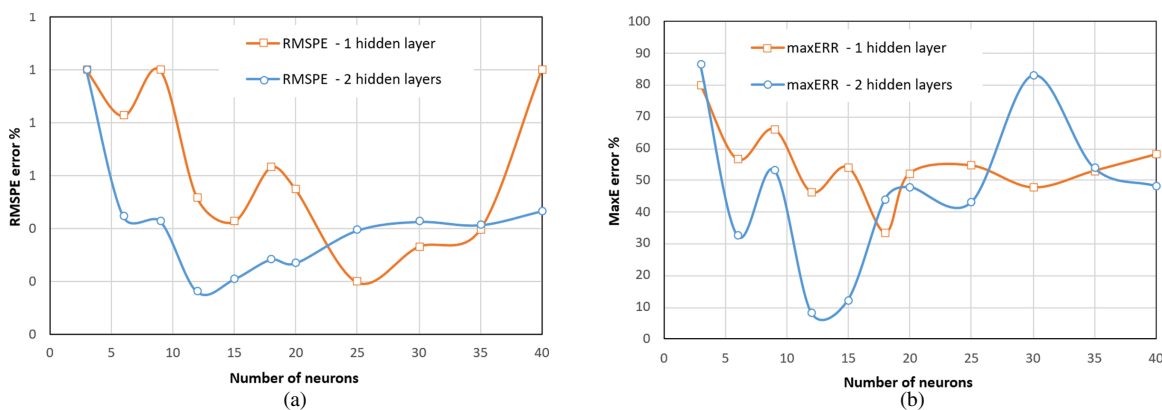
as input variables), followed by several hidden layers (each with a preselected number of neurons) and, at the end, the output layer (with as many neurons as output variables). At each node, the signal is processed by multiplying the input by a specific weight, adding a bias or offset, and applying a nonlinear activation function. A training process obtains the optimum values of weights and biases.

Usual activation functions are the hyperbolic tangent (tanh), the sigmoid function, or the ReLU function. The first one is used in this research.

Defining the ANN architecture extends from the hyperparameter definition to the detailing of weights and biases, all adjusted to obtain a network with minimum complexity and minimum error. The different steps are detailed in the following.

### 3.2.2 Hyperparameter Tuning

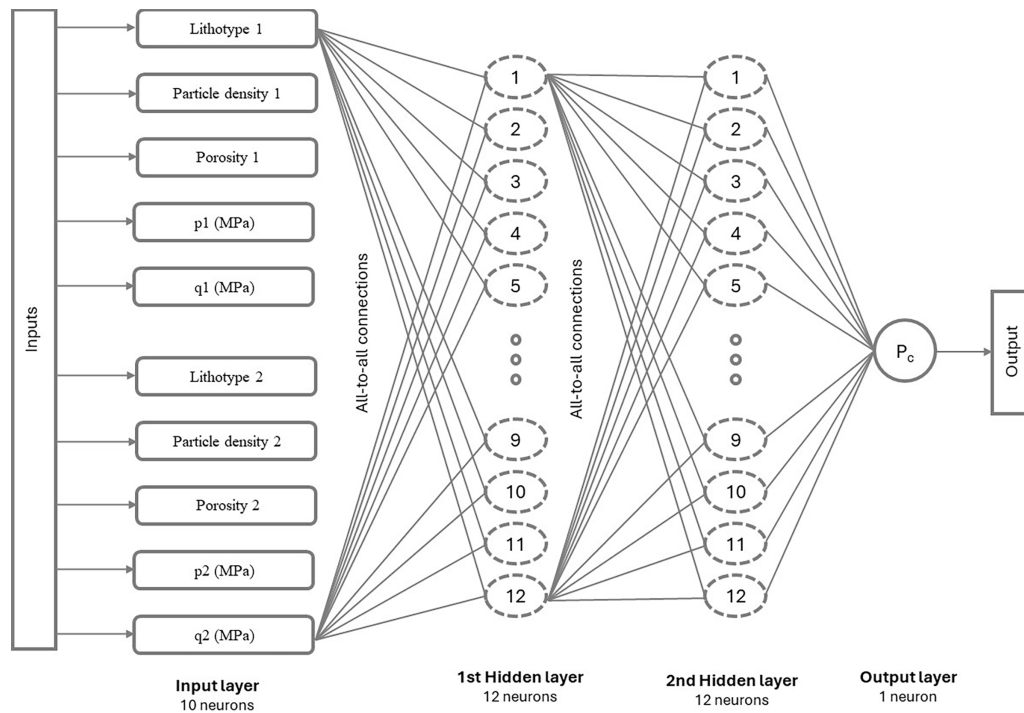
Hyperparameter tuning selects the ANN design choices that are fixed before training (e.g., number of hidden layers, neurons per layer, transfer functions, and training settings), as opposed to weights and biases learned during training. Here, the main hyperparameters explored are the number of hidden layers and the number of neurons per hidden layer, evaluated for both one-hidden-layer and two-hidden-layer configurations by progressively adjusting the neuron count. Since the two-hidden-layer architecture achieved excellent performance, deeper architectures were not pursued. Model selection is based on four error criteria: correlation coefficient (R), root mean square percentage error (RMSPE), mean absolute percentage error (MAPE), and maximum individual error (Max\_E), which provide an interpretable performance assessment (percentage-based errors) across trials. In addition, the transfer function and training algorithm were screened: log-sigmoidal or hyperbolic tangent sigmoid functions were tested for the input/hidden layers with a linear output option, and the final choice was the hyperbolic tangent sigmoid transfer function for the hidden layers and output layer. Training was performed using the Levenberg–Marquardt optimization scheme with mean square error (MSE) as the performance function, using the following settings: maximum steps = 1000, minimum performance gradient =  $10^{-7}$ , maximum momentum =  $10^{10}$ , and 250 validation checks without improvement to limit overfitting. Representative tuning outcomes are summarized in Fig. 9, where Fig. 9a reports the RMSPE trend and Fig. 9b reports the Max\_E trend for the one-hidden-layer and two-hidden-layer networks.



**Figure 9:** Performance indicators of ANN models for each output variable using a network with different numbers of hidden layers and nodes: (a) RMSPE; (b) MaxE.

As expected, increasing the number of neurons improves performance, but not indefinitely. After 25 neurons for a one-hidden-layer case and after 12 neurons for the two-hidden-layer case, errors increase.

Observing Fig. 9, the performance obtained with 2 hidden layers and 12 neurons outperforms all the other configurations. Then, the two-hidden-layer ANN with 12 neurons is selected as the optimum architecture and is used for the detailed training process. Fig. 10 shows the ANN configuration.



**Figure 10:** Graphical representation of the ANN architecture, with two hidden layers and 12 neurons in each of them.

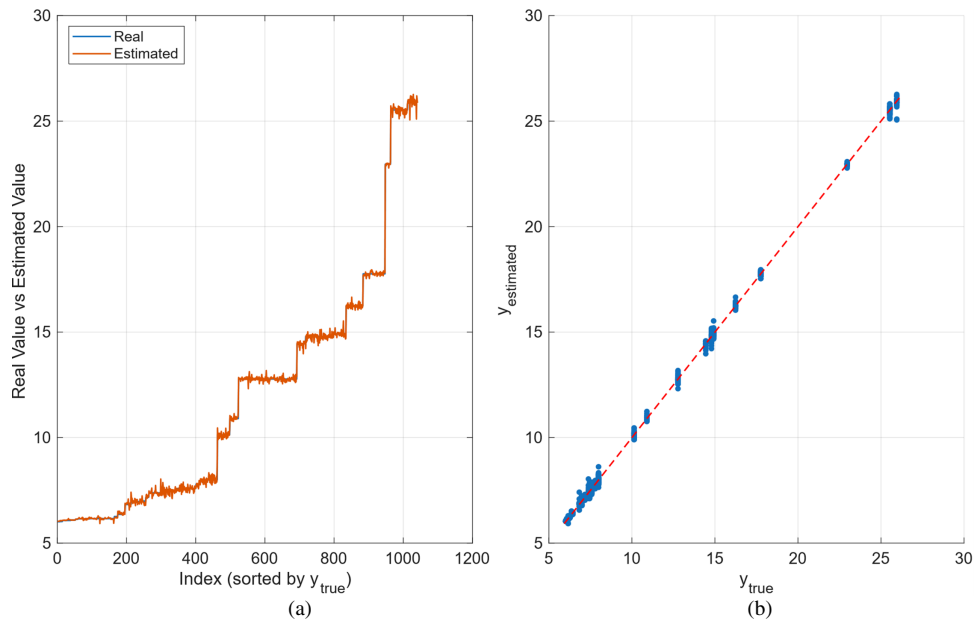
### 3.2.3 Training Process

Once the network configuration is selected, the training process requires an initial data splitting to randomly separate the input dataset into three subsets: training, validation, and test. In this research, a proportion of 70%-15%-15% is chosen for the subsets. The training and validation data are effectively used for the machine learning procedure, checking the trained network at each step with the validation data, and stopping the process if the error increases, as a countermeasure against overfitting. After the ANN achieves its final configuration, the test subset is used to check the network with unseen data and control its generalization capacity.

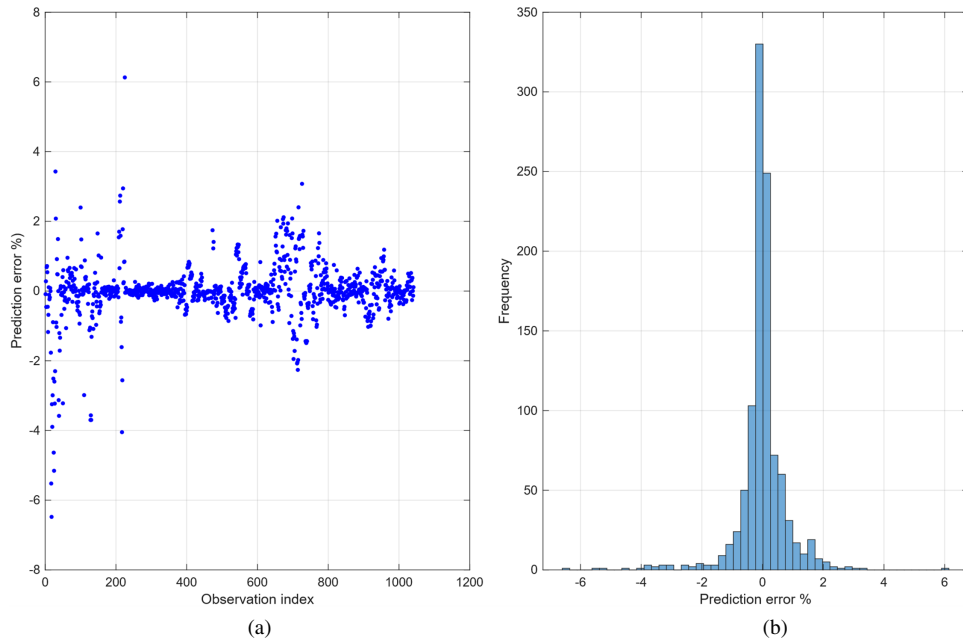
To ensure this generalization independently of the specific chosen subset, a cross-validation K-fold procedure (10-fold in this case) is also implemented. This procedure splits the input dataset successively into 10 configurations (70%-15%-15%) where the subsets change randomly, allowing the ANN to be confronted with different subsets of unseen data and evaluating the final errors. The best group of weights and biases is chosen for the definitive network, and the final errors are evaluated with this ANN and 100% of the data.

Fig. 11a presents the comparison between the value of the collapse pressure of outputs (using the ANN) and targets (original values) for the full input set, represented in scaled values. Fig. 11b shows the individual errors for each input.

The regression value R and the histogram of errors are presented in Fig. 12, where the exceptional performance of the selected configuration is manifest, reaching an R value of 0.99992 and limiting the individual error of prediction to around 6%.



**Figure 11:** Outputs and targets comparison for the full input set (1041 inputs). **(a)** Overlay: Real vs. Estimated (lines), **(b)** Scatter with  $y = x$ .



**Figure 12:** Performance of the optimal network. **(a)** Prediction errors per observation, **(b)** Residual distribution.

## 4 Results and Discussion

### 4.1 Experiment Platform

The experiment platform used in the analysis, including its environment and tools, is summarized below.

Software Environment:

- Main support MATLAB R2025a by MathWorks Inc. [44], employing the MATLAB Machine Learning Toolbox and different custom functions developed to operate.

#### Hardware Specifications:

- A standard computer was used for running the numerical models and building the network: Intel(R) Core (TM) i7-10510U CPU 1.80 GHz (8CPU), 2.30 GHz, 8.00 GB RAM. Graphics Processing Unit: Intel UHD Graphics.
- Operating system Windows 11 Pro.

#### Data Description:

- Source of data: laboratory results from tests performed in the ETSI Caminos, C. y P., UPM (Madrid, Spain).
- Dataset's size: 1041 elements.

#### XGBoost model configuration:

- Model: Gradient-boosted decision trees with 10 input features and 1 continuous target. Objective: to minimize squared loss. Primary metric: RMSE.
- Key hyperparameters: learning rate ( $\eta$ ) = 0.070 (contribution of each new tree); n\_estimators = 500 (max number of trees) with early\_stopping\_rounds = 30 (stop if validation doesn't improve); max\_depth = 4 (maximum tree depth); min\_child\_weight = 1 (minimum leaf/Hessian weight to allow a split);  $\gamma$  = 0.0 (minimum loss reduction required for a split); subsample = 0.80 (row subsampling per tree); colsample\_bytree = 0.80 (feature subsampling per tree); Regularization: L2:  $\lambda \approx 1.0$ ; optional L1:  $\alpha \approx 0.0$ .
- Validation & testing: 5-fold cross-validation (or time-based/grouped CV when appropriate) with early stopping on RMSE; 20% held-out test set. Missing values are handled natively during split finding.

#### Neural Network Configuration:

- Type of network: feedforward.
- Structure: input layer (10 neurons), 2 hidden layers (12 neurons), output layer (1 neuron).
- Activation functions used in all layers: Hyperbolic tangent sigmoid.
- Training algorithm: backpropagation.
- Training: 70%-15%-15% partition and 10-fold cross-validation method.
- Hyperparameters: maximum number of epochs: 1000, minimum performance gradient:  $10^{-10}$ , validation checks without improvement to stop training to avoid overfitting: 250.

#### Experimental Procedure:

- Data scaling and normalization: log scaling to some inputs, normalization to [0.2, 0.8].
- Model performance indicators: R, RMSPE, MAPE, and MaxE (maximum individual error).

## 4.2 Selection of the Machine Learning Procedure

The main results obtained using both procedures for the whole input set are presented in [Table 5](#).

Both methods show outstanding results, with slightly better global indicators for XGBoost and a slightly lower individual error with the neural network.

From those results, the neural network is selected for the following robustness analysis, though either could be equally adequate. The explanatory analysis of the variables obtained from the neural network will be compared with the results already presented for the XGBoost procedure.

**Table 5:** Comparison between error indicators.

	<b>XGBoost</b>	<b>Neural Network</b>
R	0.9997	0.9999
RMSPE (%)	12.44	11.34
MAPE (%)	5.88	4.61
Max Individual Error (%)	12.94	6.48

### 4.3 Robustness Analysis

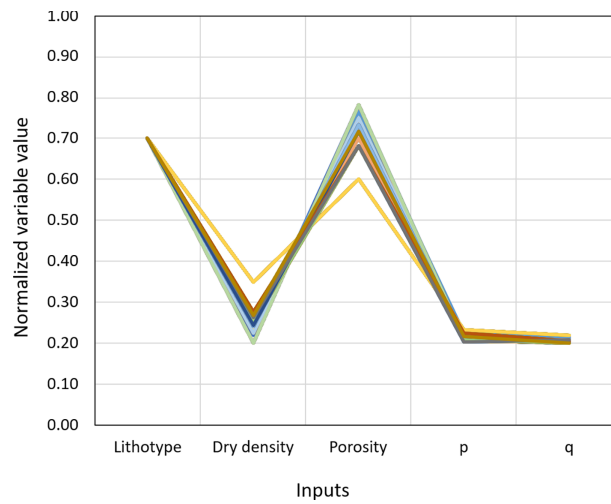
To verify the consistency of the network's outputs across different datasets and its accordance with the expected physical behavior of the system, a coherence sensitivity analysis was performed based on the methodology proposed by Shahin et al. [45]. This analysis confirmed the model's robustness and its capacity to generalize, thereby supporting the reliability of its predictions.

The approach consisted of generating a synthetic set of input data by systematically varying selected parameters over their admissible ranges. Each parameter was modified independently, while the remaining variables were fixed at constant reference values. This process aimed to verify whether the resulting outputs followed a logical trend in line with theoretical expectations and prior empirical knowledge, without producing unexpected or inconsistent behavior.

In this research, due to the specific data organization, including two related lab tests as input for the ANN, selecting a parameter to be changed while maintaining all the rest fixed is not an easy task. The data in each of the two lab tests in each input are related, since they represent results for the same lithotype using different pairs of stress ( $p$ ,  $q$ ). Additionally, variation in density and porosity for two tests on the same material should be of secondary importance. As an alternative to this approach, considering that the objective in this section is to analyze the tendency of predictions when a variable changes, only data from a single test is considered (duplicated to configure a standard input), and the changes are applied to those single-test variables.

Besides, changing the input parameters randomly and independently may generate unrealistic combinations. For example, a given lithotype, such as pumice, should have high porosity in all cases, and it is not compatible with low porosity and very high resistance. A coherent variation of the parameter should be considered that does not necessarily include the full input range, but only that subrange where the combination properties are realistic. The previous statement is illustrated in Fig. 13, which presents all the relationships between variables for the tested pumice lithotype and is expressed in normalized variables (to be able to be shown together). As can be seen, for the lithotype 0.7 (pumice), only low dry densities are compatible, as well as high porosities. Consistent with these parameters, the values of  $p$  and  $q$  at failure are small.

Having all that in mind, the most interesting analysis of parameter variation is the influence of the unconfined compressive strength (UCS) on the porosity ( $n$ ). This variation is analyzed for the previously considered pumice material (similar studies can be performed with other lithotypes, but are not included for the sake of brevity).



**Figure 13:** Values of inputs compatible with the lithotype pumice.

A base dataset is considered, and the values of  $p_1$  and  $q_1$  are adapted to include UCS coherent values. The reference Cambridge variables can be expressed as:

$$p = \frac{\sigma_1 + 2\sigma_3}{3} \quad (2)$$

$$q = \sigma_1 - \sigma_3 \quad (3)$$

where  $p$  represents the mean effective stress,  $q$  denotes the deviatoric stress, and  $\sigma_1$ ,  $\sigma_2$ , and  $\sigma_3$  are the principal, secondary, and tertiary stresses, respectively (being  $\sigma_2 = \sigma_3$  in this case).

The UCS relationship corresponds to the following:

$$\sigma_2 = \sigma_3 = 0 \quad (4)$$

$$p_{UCS} = \frac{\sigma_1}{3} \quad (5)$$

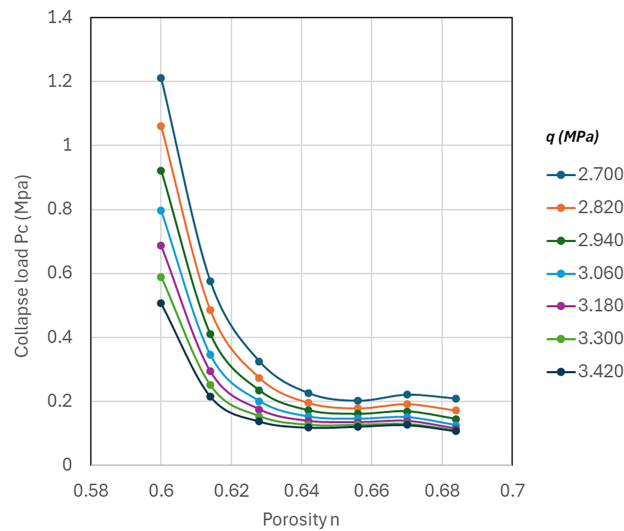
$$q_{UCS} = \sigma_1 \quad (6)$$

The reference values are detailed in [Table 6](#), and the results are shown in [Fig. 14](#).

The results are consistent with the expected behavior of the volcanic material. An increase in porosity produces a rock structure with more pores and weaker, and consequently, with a lower collapse load.

**Table 6:** Input data for the test.

Lithotype1	$\gamma_{d1}$ (kN/m <sup>3</sup> )	Porosity1	$p_1$ (MPa)	$q_1$ (MPa)	Lithotype2	$\gamma_{d2}$ (kN/m <sup>3</sup> )	Porosity2	$p_2$ (MPa)	$q_2$ (MPa)
4	8.68 to 10.984	0.60 to 0.69	0.9 to 1.140	$3 \cdot p_1$	4	10.93	0.602	1.24	1.33



**Figure 14:** Variation of collapse load for different UCS (MPa) and porosity values. Pumice Lithotype.

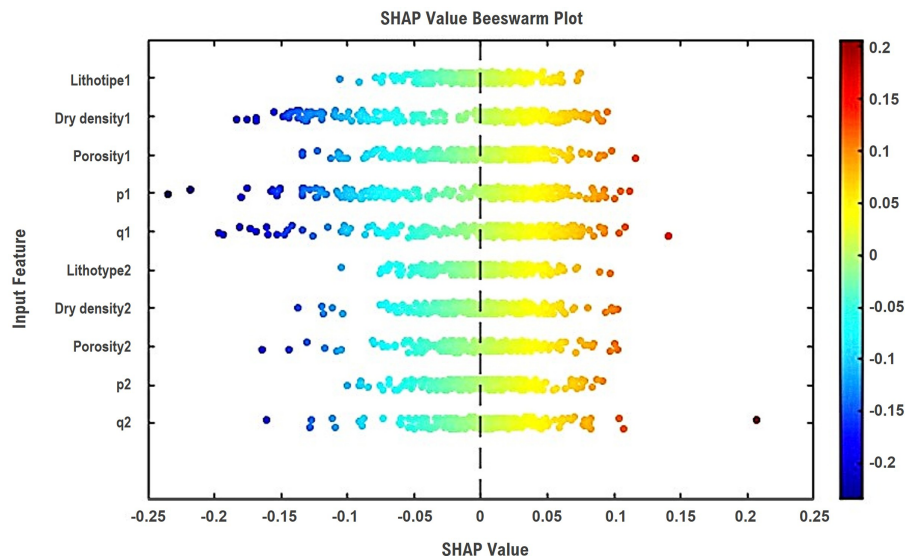
#### 4.4 Predictions Explanation: Conditional SHAP with PCA and Permutations for Dependent Input Features

To explain individual predictions, we use Shapley-value attribution, where we measure how important each feature is for one prediction by adding features one at a time and recording how the prediction changes; we then average this change over many different feature-orderings [46]. This principle is implemented in the SHAP framework for model interpretability [47]. However, standard SHAP commonly assumes feature independence; when inputs are correlated, this assumption can generate unrealistic synthetic samples and misleading attributions. This is relevant for the pyroclastic dataset, where several inputs are mutually constrained (Fig. 3). To avoid implausible feature combinations, we adopt conditional SHAP, which estimates contributions while conditioning on the observed values of the other features, thereby respecting the joint structure of the data [48].

In practice, the conditional sampling step is implemented using Principal Component Analysis (PCA): inputs are mapped to an orthogonal component space that captures the dominant variance directions, and conditional samples are generated by selecting nearby observations in this transformed space and reconstructing the unknown features. This provides a practical approximation of conditional distributions without fitting an explicit probabilistic dependency model, while still preserving the main correlation structure of the inputs [48]. Shapley values are then approximated by Monte Carlo permutations: instead of evaluating all the feature orderings (since the number of permutations grows factorially), we sample a fixed number of random permutations per instance; along each permutation, features are introduced sequentially, and the incremental change in prediction is accumulated, then averaged across permutations to obtain the SHAP estimate [46,47]. The number of permutations trades accuracy for cost; 30–50 permutations are typically sufficient for exploratory analysis, whereas  $\geq 100$  yields more stable estimates when high confidence is required [48]. Overall, combining conditional SHAP with PCA-based conditional sampling and permutation-based approximation produces explanations that remain consistent with physically feasible input relationships, reducing spurious attribution variability and improving repeatability compared with independence-based SHAP on correlated features.

To study the SHAP values across a dataset, we use a Beeswarm plot, which visualizes the distribution of local contributions across all observations (Fig. 15). This is not a snapshot of a specific model since it uses

random permutations of the input features, based on the Principal Component Analysis of the model, to account for the dependencies across the input features. Each dot represents the SHAP value of a particular feature for a single data instance. The horizontal position of each dot indicates the feature's contribution to the model's output. A small vertical jitter is applied to spread overlapping points.



**Figure 15:** Beeswarm plot of the SHAP values of the dataset.

A wide cloud of points indicates a significant influence (global impact) of the variable because it produces a big horizontal dispersion of the prediction. Shift to positive values (right side) indicates that the variation increases the output, while a negative shift (left side) reduces it. If the variable is really important, the graphic shows consistent contributions, not a reduced number of points giving a false average.

Fig. 15 presents the SHAP values for the model developed in this study. The primary variables (Lithotype-1, Dry density-1, Porosity-1, and the stress-related variables p1 and q1) exhibit wide and clearly shifted distributions, indicating a strong and consistent global influence on the model predictions. Their SHAP values are broadly dispersed and predominantly displaced from zero, suggesting that these variables systematically contribute to increasing or decreasing the predicted response. In particular, the inclusion of p1 and q1 highlights the importance of the stress state in controlling the model output, reinforcing their role as key governing variables.

In contrast, the secondary test variables (Lithotype-2, Dry density-2, Porosity-2, and the stress-related variables p2 and q2) display much narrower distributions, with most SHAP values concentrated near zero. This indicates a limited global contribution to the model output. However, the presence of a small number of outliers (points with relatively large positive or negative SHAP values) suggests that these variables can have a significant impact in specific regions of the input space. This behavior is indicative of localized nonlinear effects or interactions, where the influence of these variables becomes relevant only under certain conditions.

Overall, this pattern reveals a clear hierarchy in feature importance: the primary variables, including the initial stress descriptors p1 and q1, dominate the global behavior of the model, while the secondary variables act mainly as modifiers that introduce localized adjustments. This effect is consistent with the use of conditional SHAP values, where dependencies between variables lead to a redistribution of contributions among correlated inputs, resulting in more conservative but physically meaningful attributions.

When comparing the previously analyzed Permutation Importance (mean  $\pm$  SD over 10 repeats) obtained from the XGBoost model with the SHAP beeswarm derived from the ANN model, a consistent qualitative pattern can be observed, although some differences are expected due to the use of different model architectures. Permutation Importance identifies  $q_1$  and  $q_2$  as the most influential and stable predictors, clearly separated from the rest, with Lithotype 1 as the next most relevant feature. The mid-to-lower importance group (Lithotype 2, Dry density 1, and  $p_1$ ) shows overlapping error bars, indicating an unstable ranking among these variables.

This general trend is consistent with the SHAP beeswarm representation, where the  $q$ -related variables exhibit a relatively wide spread of SHAP values, including some extreme contributions. These large contributions, even if occurring in a subset of samples, can strongly affect the model output. As a result, when these variables are permuted, the model loses critical information in those high-impact regions, leading to a significant increase in prediction error and, consequently, high permutation importance values.

In contrast, variables such as Lithotype and Dry density display broader but less extreme SHAP distributions, indicating a more moderate and consistent influence across the dataset, which aligns with their intermediate permutation importance. Taken together, the dominance of  $q_1$  and  $q_2$  appears to be robust across models and interpretability methods, while the relative ranking among mid-tier variables remains less stable and more sensitive to the analysis approach.

Summarily, conditional SHAP values and permutation importance provide a consistent and complementary interpretation of the model. Both methods identify the stress-related variables, particularly  $q_1$  and  $q_2$ , as the main drivers of the predictions, while lithology and density-related features show a more moderate and distributed influence. The conditional SHAP approach yields physically consistent attributions by accounting for feature dependencies, whereas permutation importance captures the global sensitivity of the model, especially in the presence of localized high-impact effects. The agreement between both methods supports the robustness of the main feature hierarchy, while the variability observed in mid-tier variables indicates a more context-dependent influence.

## 5 Conclusions

This research presents a machine learning approximation for the estimated value of the collapse pressure of different pyroclastic rocks, which is not easily evaluated in regular laboratory testing campaigns.

The model allows for quick and relatively accurate predictions based only on two different specimen testings, including only five easy-to-obtain parameters from each one: lithotype class, dry density, porosity, and failure stresses in Cambridge variables ( $p$  and  $q$ ). Though other parameters were included in the initial study (particle density or deformation at failure), they were not influential.

Two different approaches were used: an XGBoost model and a deep neural network. Both showed a very good prediction accuracy, with the individual errors being lower with the ANN. Based on this prediction, a technician can determine whether a complementary lab test is necessary to determine the collapse pressure or if failure is unavoidable within the given range of stresses.

Despite the previously cited advantages of the machine learning algorithms, some limitations should be highlighted. (1) The database used to define the algorithm is limited to the specimen lithotypes and results available in the test campaign, and consequently, the prediction is not applicable outside these limits. (2) Some of the collapse loads were not available from the test results, and these gaps were covered by estimating the collapse values assuming a reasonable parabolic behavior, consistent with the available data. This strategy is needed for the machine learning procedure to be built, but it includes some behavioral assumptions that may be avoided if more data is available in the future.

Future research is needed to enlarge the available dataset, including more lithotypes and ranges of intermediate and collapse loads, to extend the range of validity of the machine learning procedure.

**Acknowledgement:** Not applicable.

**Funding Statement:** This publication is part of the project PID2022-139202OB-I00, Neural Networks and Optimization Techniques for the Design and Safe Maintenance of Transportation Infrastructures: Volcanic Rock Geotechnics and Slope Stability (IA-Pyroslope), funded by the Spanish State Research Agency of the Ministry of Science, Innovation and Universities of Spain and the European Regional Development Fund, MCIN/AEI/10.13039/501100011033/FEDER, EU.

**Author Contributions:** The authors confirm contribution to the paper as follows: Conceptualization, Miguel A. Millán, Rubén Galindo; methodology, Miguel A. Millán, Rubén Galindo; software, Miguel A. Millán; validation, Rubén Galindo; formal analysis, Miguel A. Millán; investigation, Fausto Molina-Gómez; resources, Fausto Molina-Gómez; data curation, Fausto Molina-Gómez; writing—original draft preparation, Miguel A. Millán; writing—review and editing, Rubén Galindo; visualization, Miguel A. Millán; supervision, Rubén Galindo; project administration, Miguel A. Millán, Rubén Galindo; funding acquisition, Miguel A. Millán, Rubén Galindo. All authors reviewed and approved the final version of the manuscript.

**Availability of Data and Materials:** Data available upon reasonable request from the corresponding author.

**Ethics Approval:** Not applicable.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Trolese M, Cerminara M, Esposti Ongaro T, Giordano G. The footprint of column collapse regimes on pyroclastic flow temperatures and plume heights. *Nat Commun.* 2019;10(1):2476. doi:10.1038/s41467-019-10337-3.
2. Damiano E, de Cristofaro M, Brunzo A, Carrieri G, Iavazzo L, Netti N, et al. The mechanical characterization of pyroclastic deposits for landslide early warning systems. *Geosciences.* 2023;13(10):291. doi:10.3390/geosciences13100291.
3. Thiele ST, Kereszturi G, Heap MJ, de Lima Ribeiro A, Kamath AV, Kidd M, et al. Hyperspectral mapping of density, porosity, stiffness, and strength in hydrothermally altered volcanic rocks. *Solid Earth.* 2025;16(11):1249–67. doi:10.5194/se-16-1249-2025.
4. González de Vallejo LI, Hijazo T, Ferrer M. Engineering geological properties of the volcanic rocks and soils of the Canary Islands. *Soils Rocks.* 2008;31(1):3–13. doi:10.28927/sr.311003.
5. Molina-Gómez F, Millán MÁ, Galindo-Aires RÁ. Experimental assessment of structural collapse in pyroclasts. *Civ Environ Eng Rep.* 2025;35(2):15–23. doi:10.59440/ceer/201332.
6. Garcia Hernandez JE, Notario del Pino JS, Gonzalez Martin MM, Reguera FH, Rodriguez Losada JA. Zeolites in pyroclastic deposits in southeastern Tenerife (Canary islands). *Clays Clay Miner.* 1993;41(5):521–6. doi:10.1346/ccmn.1993.0410501.
7. Li C, Zhou J, Dias D, Du K, Khandelwal M. Comparative evaluation of empirical approaches and artificial intelligence techniques for predicting uniaxial compressive strength of rock. *Geosciences.* 2023;13(10):294. doi:10.3390/geosciences13100294.
8. Xu B, Tan Y, Sun W, Ma T, Liu H, Wang D. Study on the prediction of the uniaxial compressive strength of rock based on the SSA-XGBoost model. *Sustainability.* 2023;15(6):5201. doi:10.3390/su15065201.
9. Shahani NM, Qin X, Xin W, Li J, Aizitiliwumaier T, Ma X, et al. Hybrid PSO with tree-based models for predicting uniaxial compressive strength and elastic modulus of rock samples. *Front Earth Sci.* 2024;12:1337823. doi:10.3389/feart.2024.1337823.
10. Xie H, Lin P, Kang J, Zhai C, Du Y. Prediction method of rock uniaxial compressive strength based on feature optimization and SSA-XGBoost. *Sustainability.* 2024;16(19):8460. doi:10.3390/su16198460.

11. Akosah S, Gratchev I, Gidigasu SSR. A systematic literature review on the application of artificial intelligence techniques for rock strength estimation. *Neural Comput Appl.* 2025;37(25):20721–53. doi:10.1007/s00521-025-11517-7.
12. Xie Y, Li X, Min Z. Comparison of machine learning models for rock UCS prediction using measurement while drilling data. *Sci Rep.* 2025;15(1):8434. doi:10.1038/s41598-025-93111-4.
13. Hussain J. Leveraging artificial neural networks to predict igneous rock strength parameters from petrological contents. *Acta Montan Slovaca.* 2025;30(2):470. doi:10.46544/ams.v30i2.16.
14. Varol OO. Predicting uniaxial compressive strength of tuff after accelerated freeze-thaw testing: comparative analysis of regression models and artificial neural networks. *J Mt Sci.* 2024;21(10):3521–35. doi:10.1007/s11629-024-8729-2.
15. Ahmad M, Al Zubi M, Almujiabah H, Sabri Sabri MM, Mustafvi JB, Haq S, et al. Improved prediction of soil shear strength using machine learning algorithms: interpretability analysis using SHapley Additive exPlanations. *Front Earth Sci.* 2025;13:1542291. doi:10.3389/feart.2025.1542291.
16. Momeni E, He B, Abdi Y, Jahed Armaghani D. Novel hybrid XGBoost model to forecast soil shear strength based on some soil index tests. *Comput Model Eng Sci.* 2023;136(3):2527–50. doi:10.32604/cmescs.2023.026531.
17. Motameni S, Rostami F, Farzai S, Soroush A. A comparative analysis of machine learning models for predicting loess collapse potential. *Geotech Geol Eng.* 2024;42(2):881–94. doi:10.1007/s10706-023-02593-4.
18. Zhang W, Guo J, Li Z, Cheng R, Ning C, Niu H, et al. Prediction of loess collapsibility coefficient using Bayesian optimized random forest model. *Sci Rep.* 2025;15(1):25281. doi:10.1038/s41598-025-11121-8.
19. Walding N, Williams R, Dowe N, Rowley P, Thomas M, Osman S, et al. The influence of moisture on ash strength: implications for understanding volcanic stratigraphy. *Bull Volcanol.* 2025;87(6):39. doi:10.1007/s00445-025-01821-4.
20. Yuan B, Choo CS, Yeo LY, Wang Y, Yang Z, Guan Q, et al. Physics-informed machine learning in geotechnical engineering: a direction paper. *Geomech Geoenviron.* 2025;20(5):1128–59. doi:10.1080/17486025.2025.2502029.
21. Farea A, Yli-Harja O, Emmert-Streib F. Understanding physics-informed neural networks: techniques, applications, trends, and challenges. *AI.* 2024;5(3):1534–57. doi:10.3390/ai5030074.
22. Zhang Z, Dias D, Anitescu C, Pan Q, Rabczuk T. PINN for elastic-plastic, mesh-free modelling of tunnelling-induced deformation. *Acta Geotech.* 2026;21(1):349–70. doi:10.1007/s11440-025-02788-4.
23. Pan Q, Wu H, Zhang Z, Song K. Tunneling-induced settlement in composite strata via multi-physics-informed neural network. *Rock Soil Mech.* 2024;45(2):539–51. (In Chinese).
24. Pan Y, Zhang Y, Lu Q, Xia P, Qi J, Luo Q. Enhanced physics-informed neural networks for deep tunnel seepage field prediction: a Bayesian optimization approach. *Water.* 2025;17(17):2621. doi:10.3390/w17172621.
25. Luo T, Feng Y, Huang Q, Zhang Z, Yan M, Yang Z, et al. A novel solution for seepage problems using physics-informed neural networks. *arXiv:2310.17331.* 2023.
26. Li Y, Sun Q, Fu Y, Wei J. Solving the Richards infiltration equation by coupling physics-informed neural networks with Hydrus-1D. *Sci Rep.* 2025;15(1):18649. doi:10.1038/s41598-025-02978-w.
27. Bandai T, Ghezzehei TA. Forward and inverse modeling of water flow in unsaturated soils with discontinuous hydraulic conductivities using physics-informed neural networks with domain decomposition. *Hydrol Earth Syst Sci.* 2022;26(16):4469–95. doi:10.5194/hess-26-4469-2022.
28. Depina I, Jain S, Mar Valsson S, Gotovac H. Application of physics-informed neural networks to inverse problems in unsaturated groundwater flow. *Georisk Assess Manag Risk Eng Syst Geohazards.* 2022;16(1):21–36. doi:10.1080/17499518.2021.1971251.
29. Wang Y, Shi C, Shi J, Lu H. Forward and inverse 2D consolidation using PINNs. *Acta Geotech.* 2024;19:8051–69.
30. Hong S, Kim SR. Physics-informed neural networks for back-analysis and consolidation settlement prediction using field measurements. *Acta Geotech.* 2025;24(4):1–34. doi:10.1007/s11440-025-02888-1.
31. Lu Y, Mei G. A deep learning approach for predicting two-dimensional soil consolidation using physics-informed neural networks (PINN). *Mathematics.* 2022;10(16):2949. doi:10.3390/math10162949.
32. Zhang Z, Wang B, Li Z, Ye X, Sun Z, Dias D. Physics-guided neural network-based framework for 3D modeling of slope stability. *Comput Geotech.* 2024;176(1):106801. doi:10.1016/j.compgeo.2024.106801.

33. Moeineddin A, Seguí C, Dueber S, Fuentes R. Physics-informed neural networks applied to catastrophic creeping landslides. *Landslides*. 2023;20(9):1853–63. doi:10.1007/s10346-023-02072-0.
34. Jones TJ, Beckett F, Bernard B, Breard ECP, Dioguardi F, Dufek J, et al. Physical properties of pyroclastic density currents: relevance, challenges and future directions. *Front Earth Sci*. 2023;11:1218645. doi:10.3389/feart.2023.1218645.
35. Conde M. Caracterización geotécnica de materiales volcánicos de baja densidad (Geotechnical characterization of low-density volcanic materials) [master's thesis]. Madrid, Spain: Universidad Complutense de Madrid; 2013.
36. Canarias GD. GETCAN-011. Guía para la planificación y realización de estudios geotécnicos para la edificación en la Comunidad Autónoma de Canarias (Guide for the planning and execution of geotechnical studies for building construction in the Autonomous Community of the Canary Islands). Las Palmas de Gran Canaria, Spain: Gobierno de Canarias; 2011. 114 p.
37. Serrano A, Perucho A, Conde M. Yield criterion for low-density volcanic pyroclasts. *Int J Rock Mech Min Sci*. 2016;86(1):194–203. doi:10.1016/j.ijrmms.2016.04.014.
38. Serrano A, Galindo R, Perucho Á. Ultimate bearing capacity of low-density volcanic pyroclasts: application to shallow foundations. *Rock Mech Rock Eng*. 2021;54(4):1647–70. doi:10.1007/s00603-020-02341-7.
39. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: ACM; 2016. p. 785–94. doi:10.1145/2939672.2939785.
40. Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Stat*. 2001;29(5):1189–232. doi:10.1214/aos/1013203451.
41. Hastie T, Tibshirani R, Friedman J. *The elements of statistical learning*. New York, NY, USA: Springer; 2009. doi:10.1007/978-0-387-84858-7.
42. Breiman L. Random forests. *Mach Learn*. 2001;45(1):5–32. doi:10.1023/A:1010933404324.
43. Fisher A, Rudin C, Dominici F. All models are wrong, but many are useful: learning a variable's importance by studying an entire class of prediction models simultaneously. *J Mach Learn Res*. 2019;20:1–81.
44. MathWorks Inc. MATLAB R2025a [software]. Natick, MA, USA: The MathWorks Inc.; 2025.
45. Shahin M, Jaksa M, Maier H. State of the art of artificial neural networks in geotechnical engineering. *Electron J Geotech Eng*. 2008;8:1–26.
46. Shapley LS. A value for n-person games. In: Kuhn HW, Tucker AW, editors. *Contributions to the theory of games*. Vol. 2. Princeton, NJ, USA: Princeton University Press; 1953. p. 307–17.
47. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*; 2017 Dec 4–9; Red Hook, NY, USA.
48. Aas K, Jullum M, Løland A. Explaining individual predictions when features are dependent: more accurate approximations to Shapley values. arXiv:1903.10464. 2019.