Journal of Geochemistry, Geophysics, Geosystems

Supporting Information for

Volcanic ash classification through Machine Learning

Damià Benet[1,2,3,], Fidel Costa[1], Christina Widiwijayanti[2]

[1]Institut de Physique du Globe de Paris, Université Paris Cité, CNRS, Paris, France.

[2] EOS, Earth Observatory of Singapore, Nanyang Technological University, Singapore.

[3] Asian School of the Environment, Nanyang Technological University, Singapore.

Contents of this file

Introduction

The following information below includes details on the used features and their abbreviations, a series of tables with the accuracies obtained through different experiments to choose amongst the different types of machine learning-based models and to choose the optimal hyperparameters. Other supplementary information includes the results of classification from the One-Versus-One and One-Versus-Rest classification techniques, two figures to summarize classification scores across different types of models, and an example of a confusion matrix.

Table S1. List of measured features, and their abbreviation and calculation. The reader is referred to Benet et al., (*preprint*) for more details and a reference.

Feature	Abbreviation	Equation
Convexity	convexity	P_h/P_p
Rectangularity	rectangularity	$\dfrac{P_p}{2H+2W}$
Elongation	elongation	$\dfrac{D_{MaxFeret}^2}{E_{maj}}$
Roundness	roundness	$\dfrac{4A_p}{\pi D_{MaxFeret}^2}$
Circularity by Dellino and la Volpe (1996)	circ_dellino	$\dfrac{P_p}{2\sqrt{\pi A_p}}$

Circularity by Cioni et al. (2014)	circ_cioni	$\dfrac{4\pi A_p}{P_p{}^2}$
Solidity	solidity	$\dfrac{A_p}{2H + 2W}$
Aspect ratio	aspect_rat	W/H
Compactness	compactness	$\dfrac{A_p}{HW}$
Contrast	contrast	$\displaystyle\sum_{i,j=0}^{levels-1} P_d^\theta (i-j)^2$
Dissimilarity	dissimilarity	$\displaystyle\sum_{i,j=0}^{levels-1} P_d^\theta |i-j|$
Homogeneity	homogeneity	$\displaystyle\sum_{i,j=0}^{levels-1} \dfrac{P_d^\theta(i,j)}{1+(i-j)^2}$
ASM	asm	$\displaystyle\sum_{i,j=0}^{levels-1} P_d^\theta(i,j)^2$
Energy	energy	$\sqrt{ASM}$
Correlation	correlation	$\displaystyle\sum_{i,j=0}^{levels-1} P_d^\theta \left[\dfrac{(i-\mu_i)\,(j-\mu_j)}{\sqrt{(\sigma_i^2)(\sigma_j^2)}}\right]$
Channel[1] mean	channel_mean (e.g., hue_mean)	$\dfrac{1}{N}\displaystyle\sum_{i=i}^{n} x_i$
Channel standard dev	channel_std (e.g., value_std)	$\sqrt{\dfrac{1}{N-1}\displaystyle\sum_{i=1}^{N}(x_i-\overline{x})^2}$
Channel mode	channel_mode (e.g., red_mode)	Computationally found as the most common value in the array by Scipy's stats.mode function

Symbols used: A_p, particle area; P_p, particle perimeter; A_h, hull area; P_h, hull perimeter; W, width of bounding box; H, height of bounding box; $D_{MaxFeret}$, Feret maximum diameter the maximum distance between two parallel lines tangential to the particle outline; E_{maj}, major ellipse axis; levels, pixel intensities from the ROI used for Grey-Level Cooccurrence-Matrix (GLCM) calculation; $P_d^\theta(i,j)$, probability of pixel pairs at a given distance (d) and angle (θ) in GLCM; μ_i, GLCM mean; σ_i^2, standard deviation; N, number of pixels per channel; x_i, pixel value; $\overline{x}$, mean of pixel values.

Table S2: Optimal hyperparameter obtained from the highest cross-validation score for various models.

Hyperparameter	XGB	RF	DTC	KNN	GBC
colsample_bytree	0.47	–	–	–	–

learning_rate	0.01	–	–	–	0.01
max_depth	10	7	7	–	10
n_estimators	45	22	–	–	48
reg_alpha	1	–	–	–	–
reg_lambda	1	–	–	–	–
min_samples_split	–	22	25	–	30
n_neighbors	–	–	–	5	–

Table S3: Evaluation of optimized models. Support indicates the number of particles used for evaluation.

	XGB	RF	DT	KNN	GBC
precision	0.75	0.68	0.65	0.64	0.69
recall	0.75	0.72	0.69	0.68	0.72
F1-score	0.75	0.69	0.65	0.64	0.70
accuracy	0.81	0.75	0.70	0.70	0.77
support	315	315	315	315	315

Table S4: Statistical measures of mean, first and second standard deviations of the distribution of F1-scores by particle type and their aggregated macro F1-score.

	Altered material	Free-crystal	Lithic	Juvenile	Overall
Mean	0.87	0.57	0.73	0.88	0.76
Standard deviation	0.01	0.04	0.02	0.01	0.015
Second standard deviation	0.02	0.09	0.04	0.03	0.03
Particles in train	2310	326	1122	1281	5040
Particles in test	577	81	280	320	1260

Table S5. List of the base hyperparameters for each model provided by their authors. Note, in bold, the name of the model according to the authors.

Hyperparameter	Value
Vision Transformer (ViT-B{16,32})	**Dosovitskiy et al., 2020**
Learning rate	8×10^{-4}
Epochs	7
Residual neural network (R50x{1,2})	**He et al., 2016**
Learning rate	10^{-3}
Epochs	7

Convolutional neural network (ConvNeXt-T/S/B/L/XL	**Liu et al., 2022**
Optimizer	Adam
Learning rate	5×10^{-5}
Epochs	30

Table S6. Accuracies obtained from grid search at varying learning rate and batch size.

Learning rate / Batch size	6e-4	8e-4	1e-5	3e-5
4	86.66	87.32	87.18	86.66
8	86.55	87.79	86.55	86.55
16	86.93	87.50	86.97	87.25
32	86.13	86.99	87.07	87.08
64	86.34	87.21	86.87	87.08

Table S7. Comparison of optimizers' performance based on accuracy.

Optimizer	**Accuracy**
AdamW	87.50%
SGD	81.72%
Adagrad	85.59%

Table S8. *F1-scores* obtained from the OVO and OVR strategies for each particle type, and their unweighted average (i.e., macro), for all particles in the test set (Overall) and across the associated binary classifiers. These measurements have an estimated precision of ±0.03 (see 'Effect of the train and test split' in Section 2.2.6 for its calculation).

	One-vs-One (OVO)							One-vs-Rest (OVR)				
	Overall	*F* vs *A*	*F* vs *L*	*F* vs *J*	*A* vs *L*	*A* vs *J*	*L* vs *J*	Overall	*A* vs Rest	*F* vs Rest	*L* vs Rest	*J* vs Rest
F1-score (macro)	0.75	0.81	0.78	0.9	0.88	0.97	0.84	0.76	0.89	0.74	0.82	0.92
F^1	0.56	0.67	0.64	0.82	–	–	–	0.55	–	0.52	–	–
A^2	0.9	0.95	–	–	0.92	0.98	–	0.88	0.88	–	–	–
L^3	0.71	–	0.92	–	0.86	–	0.84	0.73	–	–	0.73	–
J^4	0.85	–	–	0.96	–	0.97	0.85	0.88	–	–	–	0.88
$Rest^5$								–	0.89	0.97	0.9	0.96

[1]*F*: Free-crystal   [2]*A*: Altered material   [3]*L*: Lithic   [4]*J*: Juvenile

[5]Rest includes all the particles that do not belong to the class of interest (e.g., Lithic vs Non-lithic)

Table S9. *F1-scores* obtained from the ViT classifier of each particle type and their unweighted *F1-score* average (i.e., macro) for all particles in the test set (Overall), and across volcanoes and eruptive styles.

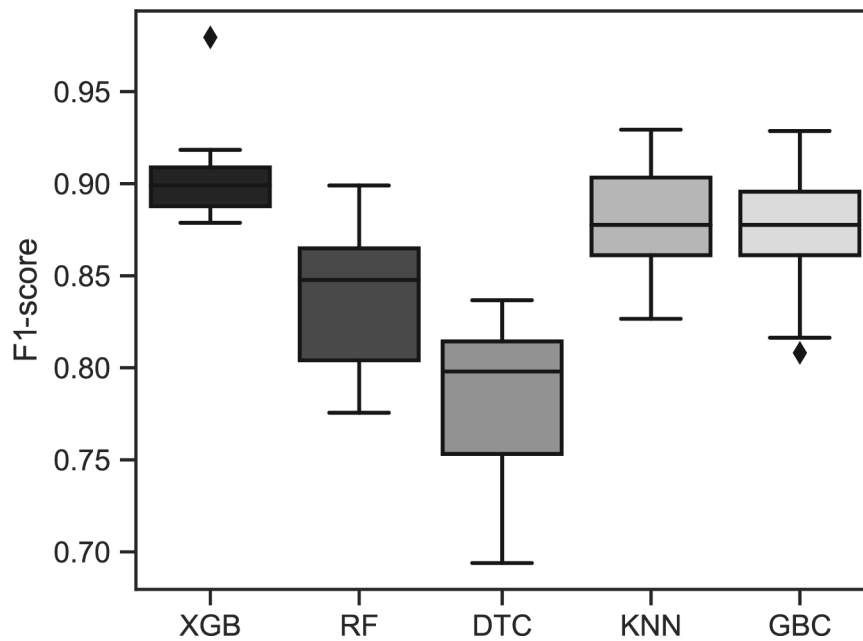|	Overall	Volcano					Eruptive style			
		Soufrière de Guadeloupe	Merapi	Nevados de Chillán	Cumbre Vieja	Kelud	Phreatic	Dome explosion	Lava fountaining	Sub-plinian / Plinian
F1-score	0.93	0.95	0.80	0.85	0.91	0.91	0.95	0.85	0.91	0.95
F[1]	0.91	0.90	0.72	0.95	–	0.92	0.94	0.86	–	0.92
A[2]	0.95	0.99	0.95	0.80	–	0.93	0.99	0.90	–	0.93
L[3]	0.89	0.96	0.75	0.91	0.88	0.77	0.93	0.90	0.88	0.77
J[4]	0.95	–	–	0.72	0.94	1	–	0.73	0.94	1

Figure S1. Whisker plots of the *F1-score* values obtained from 10-fold cross validation (see 'Hyperparameter optimization' in Section 2.2.2 for an explanation of this technique) of Extreme Gradient Boosting (XGB), Random Forest (RF), Decision Tree Classifier (DTC), K-Nearest Neighbor (KNN) and Gradient Boost Classifier (GBC). Performances are measured with the *F1-score* (see 'Model evaluation' in Section 2.2.3 for its calculation).

Each whisker plot shows the median (horizontal line), 25th and 75th percentiles (box upper and lower side). Whisker lengths are at 1.5 times the interquartile ranges, beyond which are the outliers (diamonds).
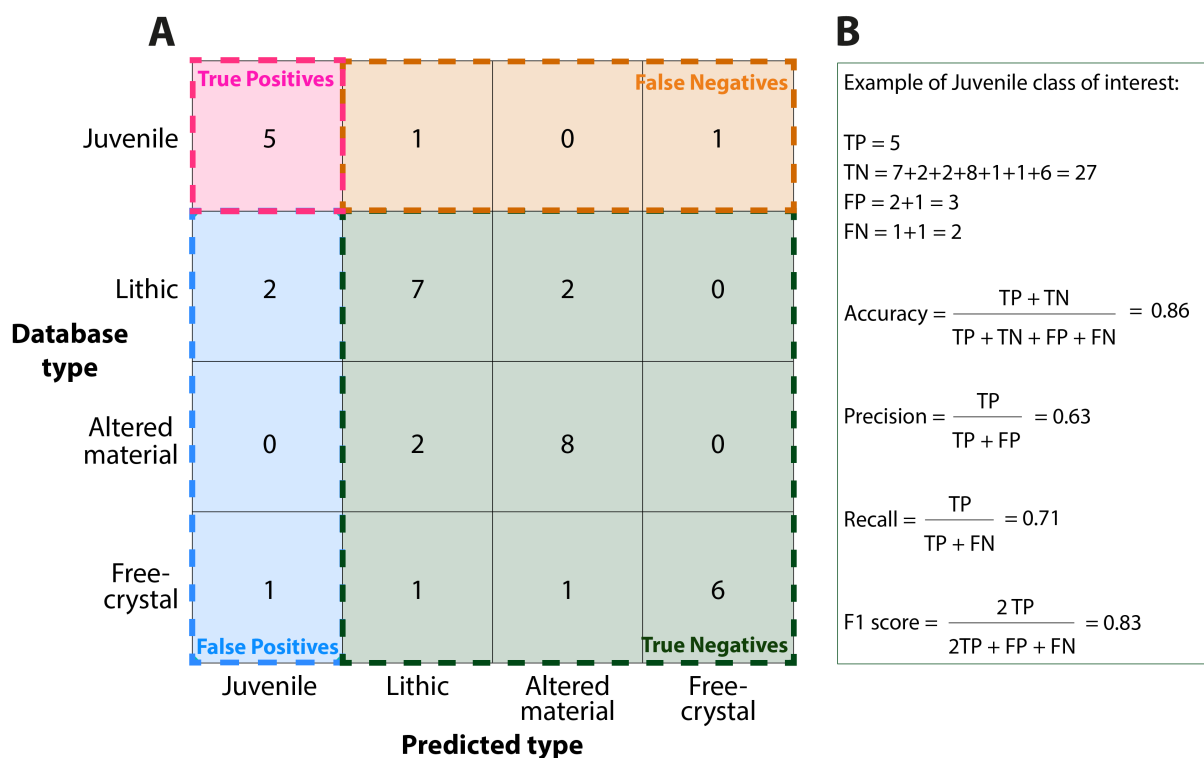

Figure S2. (A) Example of a confusion matrix for a four particle-classes classifier and (B) calculation of the main metrics taking juvenile as the class of interest.
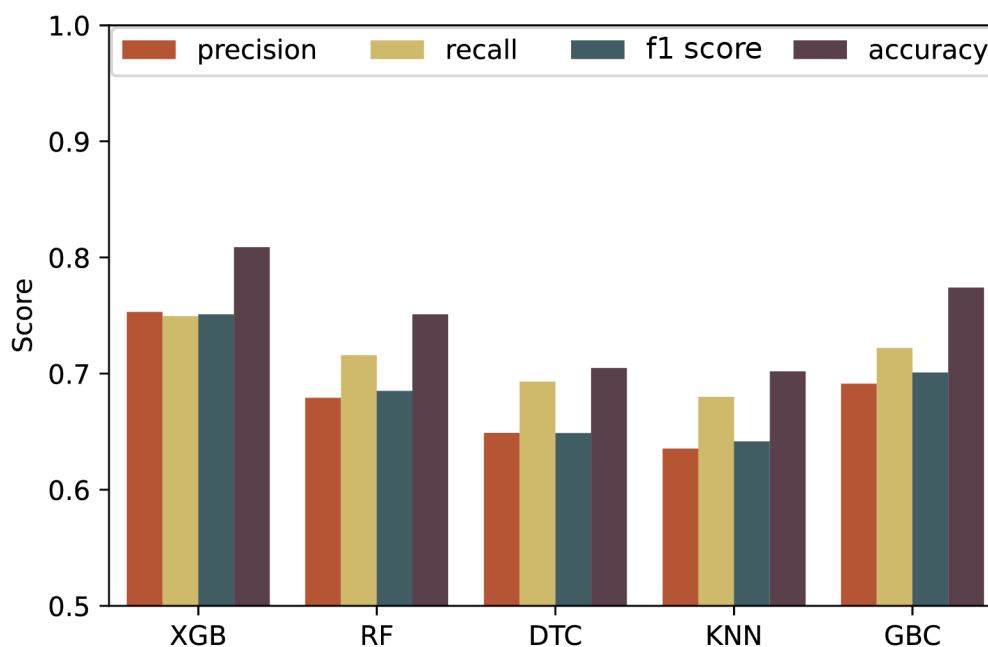
Figure S3. Evaluation of the models' performance with the test set after hyperparameter optimization based on the precision, recall, F1-score and accuracy.