VolcAshDB: a Volcanic Ash DataBase of classified particle images and features

Damià Benet, Fidel Costa, Christina Widiwijayanti, John Pallister, Gabriela Pedreros, Patrick Allard, Hanik Humaida, Yosuke Aoki, Fukashi Maeno

Bulletin of Volcanology (2024) 86:9
https://doi.org/10.1007/s00445-023-01695-4

RESEARCH ARTICLE

VolcAshDB: a Volcanic Ash DataBase of classified particle images and features

Damià Benet[1,2,3] · Fidel Costa[1,2,3] · Christina Widiwijayanti[1,2] · John Pallister[4] · Gabriela Pedreros[5] · Patrick Allard[3] · Hanik Humaida[6] · Yosuke Aoki[7] · Fukashi Maeno[7]

Received: 30 April 2023 / Accepted: 12 December 2023

Abstract

Volcanic ash provides unique pieces of information that can help to understand the progress of volcanic activity at the early stages of unrest, and possible transitions towards different eruptive styles. Ash contains different types of particles that are indicative of eruptive styles and magma ascent processes. However, classifying ash particles into its main components is not straightforward. Diagnostic observations vary depending on the magma composition and the style of eruption, which leads to ambiguities in assigning a given particle to a given class. Moreover, there is no standardized methodology for particle classification, and thus different observers may infer different interpretations. To improve this situation, we created the web-based platform Volcanic Ash DataBase (VolcAshDB). The database contains > 6,300 multi-focused high-resolution images of ash particles as seen under the binocular microscope from a wide range of magma compositions and types of volcanic activity. For each particle image, we quantitatively extracted 33 features of shape, texture, and color, and petrographically classified each particle into one of the four main categories: free crystal, altered material, lithic, and juvenile. VolcAshDB (https://volcash.wovodat.org) is publicly available and enables users to browse, obtain visual summaries, and download the images with their corresponding labels. The classified images could be used for comparative studies and to train Machine Learning models to automatically classify particles and minimize observer biases.

Keywords Volcanic ash · Database · Image analysis · PCA

Introduction

With more than one billion people around the globe threatened by volcanic eruptions (Freire et al. 2019), volcanologists have tried for long to answer the basic questions of when, where, and how big is the next eruption going to be. The main approach to anticipating and tracking the evolution of eruptions has been the monitoring of geophysical and geochemical signals, e.g., seismicity (Chouet 2003), ground deformation (Dzurisin 2006), as well as the composition and flux of gas emissions (Aiuppa 2015). However, many volcanoes worldwide remain poorly monitored instrumentally, which hampers accurate interpretation of the processes occurring at depth and makes forecasting uncertain (Newhall and Punongbayan 1996; Doyle et al. 2014).

An additional piece of information that can be used to address these challenges is studying the characteristics of volcanic ash particles. The occurrence of ash emissions already implies eruptive activity, but its origin and style can vary widely over time, ranging from minor discrete

Editorial responsibility: J. Eychenne

✉ Damià Benet
dbenet@ipgp.fr

1 Asian School of the Environment, Nanyang Technological University, Singapore, Singapore

2 EOS, Earth Observatory of Singapore, Nanyang Technological University, Singapore, Singapore

3 Institut de Physique du Globe de Paris, Université Paris Cité, Paris, France

4 Volcano Disaster Assistance Program, U.S. Geological Survey, Vancouver, WA, USA

5 Observatorio Vulcanológico de los Andes del Sur, Servicio Nacional de Geología y Minería, Temuco, Chile

6 BPPTKG (Balai Penyelidikan Dan Pengembangan Teknologi Kebencanaan Geologi), PVMBG, Geology Agency, Yogyakarta, Indonesia

7 Earthquake Research Institute, The University of Tokyo, Tokyo, Japan

Published online: 05 January 2024

phreatic explosions to powerful magmatic or phreatomagmatic eruptions (Gaunt et al. 2016; Gunawan et al. 2019). Many eruptions go through various phases of activity, and these can change significantly during a single eruption which has been used for forecasting intra-eruption activity (Bebbington and Jenkins 2019). Because the characteristics of ash particles depend on both the nature of the rock source(s) (e.g., wall-rock fragments from previous eruptions or old lava dome material that has been hydrothermally altered) and the mechanisms of magma ascent and fragmentation, ash monitoring can give clues to anticipate future changes in eruptive activity, even before magma arrives at the surface (e.g., Watanabe et al. 1999; Cashman and Hoblitt 2004; Suzuki et al. 2013; Benet et al. 2021; Re et al. 2021). In particular, a robust identification of the so-called juvenile ash particles (those that derive from fresh magma) provide crucial indication of magma close to the surface, which helps making more informed hazard assessment and emergency planning during a volcanic crisis (e.g., Taddeucci et al. 2002; Hincks et al. 2014; Gaunt et al. 2016).

The traditional approach to classify ash particles is with visual observations of their color, texture (e.g., vesicularity or crystallinity), and shape under the binocular microscope (Dellino and Volpe 1995; White and Houghton 2006; Miwa et al. 2013; Pardo et al. 2014; Gurioli et al. 2015; Gaunt et al. 2016). These observations are often complemented by the particles' external surface and internal microstructures using the scanning electron microscope (SEM) (e.g., Andronico et al. 2013; D'Oriano et al. 2014; Pardo et al. 2014). In some studies, further chemical analyses of the particles are done with electron microprobe (e.g., Nakagawa and Ohba 2002; Ohba and Nakagawa 2002; Németh 2010; Hornby et al. 2018), X-ray diffraction (Yaguchi et al. 2022), mass spectrometry (Rowe et al. 2008), or spectroscopic analysis (Bardelli et al. 2020). Finally, image analysis techniques have been employed to capture particle attributes in a systematic and relatively fast manner. These include analyses of particles' shape (Liu et al. 2015; Dürig et al. 2018), grain-size (Verolino et al. 2018), textural complexity of their surface (Ersoy et al. 2006), and/or color (Yamanoi et al. 2008).

However, classifying ash particles into different types is not straightforward, especially when distinguishing between those originating directly from the fresh magma (juvenile) and the relatively older particles associated with magma(s) emplaced before the ongoing eruptive activity (lithic). A given particle type can include particles with a wide range in shapes and colors, and the classification criteria are often valid on a case-by-case sample basis. Moreover, there is no standardized set of observations to discriminate between particle main types, making classification subject to various interpretations depending on the observers. This can lead to inconclusive evidence for robust identification, e.g., at Mt. Tongariro, 2012 (Pardo et al. 2014) and to different classification of particles depending on the observer, which has had critical implications for hazard assessment, e.g., at Soufrière de Guadeloupe, 1975–1977 (Feuillard et al. 1983).

To address these problems we have created a new Volcanic Ash DataBase (VolcAshDB), which hosts a curated dataset of ash particles from a wide range of eruptive activities and volcanoes. In this study we aim to (i) obtain a standardized dataset of classified particle images and features, (ii) describe the contents of VolcAshDB, and (iii) explore the potential for classification of the extracted features through Principal Component Analysis. VolcAshDB offers accessibility through a web-based platform which could be used for comparative studies between eruptions. It could also serve as a basis for automatic, objective classification of ash particles by applying machine learning, as has been already done in geological sciences for sand particles (Li and Iskander 2022), mineral grains (Maitre et al. 2019; Latif et al. 2022), and even for classification of shapes of volcanic ash (Shoji et al. 2018).

Methodology

Building VolcAshDB

To obtain the images and characteristics of ash particles that constitute VolcAshDB we used the following steps: i) sample preparation, ii) particles image acquisition and processing, iii) feature extraction, iv) classification by the petrologist, and v) data archiving (Fig. 1). Information on the analyzed samples is provided in the section "VolcAshDB contents" and additional details in Table S1.

Laboratory procedures and image acquisition

The samples were cleaned ultrasonically in cycles of 15 s to avoid glass shard damage, dried overnight at 60 °C, and sieved using four meshes of varying pore-sizes on the Phi (ϕ) scale. The Phi scale is defined as Phi (ϕ) = -log2(particle diameter in mm). The four pore-sizes we used are 0ϕ (1 mm), 1ϕ (0.5 mm), 2ϕ (0.25 mm), and 3ϕ (0.125 mm). We prepared multiple glass slides for a given sample, each consisting of 100 to 300 individual particles, from the coarser available grainsize fraction (mostly 0ϕ–1ϕ). Particles were deposited on top of a transparent, 3 M 9415PC Removable Repositionable Tape that was glued on a glass slide of standard dimensions (75 mm by 25 mm). To improve the separation of individual particles on the slide, we used a mesh of a finer pore size than the particle's size of interest. We also manually separated any touching particles with a needle.

The glass slide was then positioned on top of an opaque, white plate, and automatically scanned using a binocular

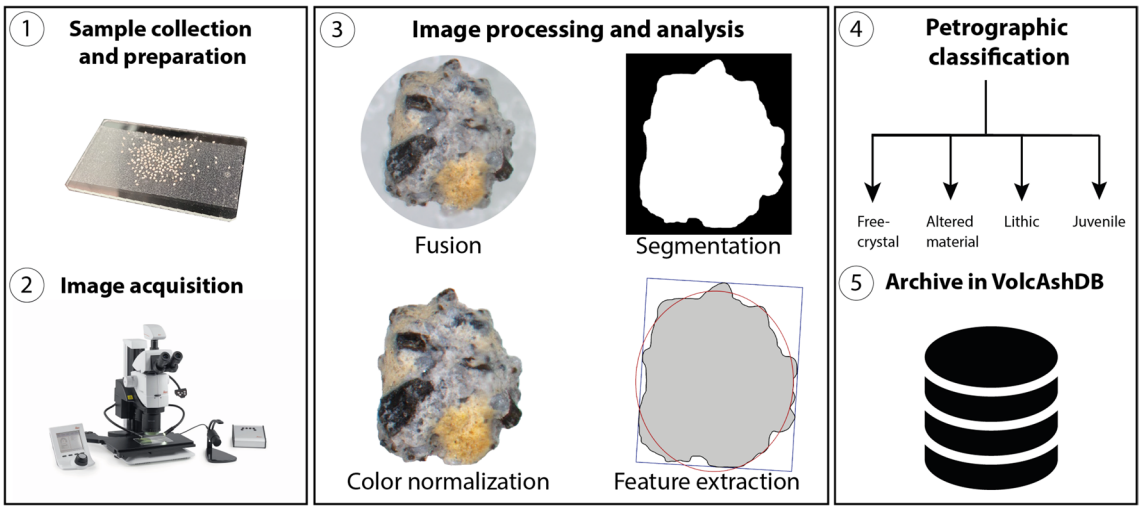

Fig. 1 Workflow and data acquisition method for creation of VolcAshDB. (1) ash particles are spread on a glass slide, (2) a scan of each slide with many images of individual particles is obtained using a binocular scanning stage (3) the scans are processed by image fusion, segmentation and color normalization, and analyzed by extracting 33 features related to the shape, texture and color. (4) Each particle image is classified by the petrologist, and (5) the particle image, its main characteristics, and its classification are stored in the database which are shared in a public web-based platform

microscope, with an episcopic (from above) ring light to guarantee uniform illumination, and a stage system by Leica (LMT260 XY Scanning Stage) equipped with the Leica LAS X imaging software available at Nanyang Technological University (NTU), Singapore. We used a Leica AX carrier to obtain 25 aligned scans at different focal depths to visualize the morphology of the particles top-to-bottom. The imaging software conditions to scan the 0ϕ-1ϕ, 1ϕ-2ϕ and 2ϕ-3ϕ fractions were at 5x, 6 × and 8 × magnifications, exposure values of 95, 105 and 120, without gain. This procedure is relatively fast, with acquisition times between 25 to 45 min to scan each glass slide. The resulting full scans have a high resolution of approximately 25,000 × 35,000 pixels, each occupying roughly 3 GB of storage space. Additionally, the associated temporary files can accumulate to a size of up to 140 GB per full scan. The scanned glass slides were then stored for reproducibility purposes. We also observed one slide per sample using a JEOL JSM-7600F Scanning Electron Microscope (SEM at NTU) to aid particle identification (see section "Labeling of the particles" for more detail). Operating conditions for the SEM analysis were at low vacuum (50 MPa), 15 kV of accelerating voltage, 8 nA of probe current, at a working distance of 20 mm. We used a pixel resolution of 1024 × 2048, obtaining about 5×10^6 pixels per particle of the exported images, and dwell time of 60 s.

Image processing

We processed the acquired images in three steps: image fusion, image segmentation, and color normalization (Fig. 1). With this procedure, we obtained multi-focused, segmented, and normalized particle images which are the primary type of images in VolcAshDB and will be referred to as "multi-focused images" henceforth. The steps were automated with a Python program that was run using the Gekko cluster at the Nanyang Technological University (NTU) High Performance Computing Center.

Image fusion consists of combining focused regions from multiple images of the same 3D object into one 2D array to obtain a multi-focused image. We fused the scans using the open-source model SESF-Fuse (Ma et al. 2021) which has been pretrained using tens of thousands of images through Deep Learning (DL). The training set consists of pairs of images that either have blurry foreground and focused background, or vice versa. The already trained model (Ma et al. 2021) is available in the GitHub repository: https://github.com/Keep-Passion/SESF-Fuse. To decrease the run time, we split each scan into ten smaller arrays and ran them separately, obtaining an overall run time per scan of < 3 h with ~ 90% of the images fused, while the remaining images were discarded.

The multi-focused scans were then segmented using the DL model named U^2-NET (Qin et al. 2020) with a Python code implementation accessible from the GitHub repository: https://github.com/OPHoperHPO/image-background-remove-tool. This model is based on about 20,000 images of single or multiple objects which are positioned in front of a background with variable textures and colors, and it automatically produces a binary mask where background pixels take a value of zero while the object of interest gets a value of one. To run U^2-NET on our dataset, we split the multi-focused scans (10–40 kilopixels square) into smaller arrays

(e.g., 5,000 × 1,000 pixels), obtaining run times between 5–7 days for each complete scan with ~ 80% of the particles properly segmented. The remaining images were discarded. Upon completion of this process, we obtained multi-focused images of individual particles from 12 different eruptions (see Table S1 for more details) with resolutions of ~ 2.5×10^6 pixels per particle image (pxls/p) and ~ 1,800 pixels per millimeter (pxls/mm) for the grain-size fraction 0ϕ–1ϕ, and resolutions of ~ 1.9×10^6 pxls/p and ~ 2,000 pxls/mm for the finer grain-size fraction 1ϕ–2ϕ. The segmentation algorithm by Qin et al. (2020) may not capture microscale irregularities (e.g., < 10 μm vesicles) of the particle outline at the image resolution we used.

Variations in the background brightness can be measured by pixel intensity and is subject to changes in experimental conditions, such as scan magnification and environmental light. We used the same white opaque plate as a background to obtain all the images, and as calibration to normalize the color of the particle images. We rescaled all image pixels to a background of pixel intensity of 200 (the pixel scale color varies from 0 to 255) to accommodate pixel values that are brighter than the background (e.g., crystal reflections). About 6% of images contained artefacts and were manually discarded. After this step, we obtained an average of 525 multi-focused images per eruption, and a total of 6,304.

Quantitative feature extraction from the multi-focused images

We measured 33 physical particle properties, hereby referred to as *features* (e.g., *Elongation*, with the first letter capitalised and in italics), that are related to the particles' shape, texture, and color. Please note that we refer to "texture" as the spatial arrangement of pixel intensity values in particle images. The shape features were obtained from the particle silhouette as projected in each multi-focused image, and are responsive to perimeter-based irregularities, particle-scale cavities, and/or overall particle form. The textural features were calculated from local pixel intensity distributions on the particle surface, providing insights into the spatial distribution of the grayscale pixel intensity values. These features aim to characterize, for instance, the heterogeneous surface of hydrothermal aggregates (high textural complexity) or the uniformly smooth surfaces of glass shards (high textural smoothness). The color features were extracted from both the Red–Green–Blue (RGB) and Hue-Saturation-Value (HSV) channel distributions of the color-normalized particle images. These features are sensitive to chromaticity, indicative of dominant color hues, intensity, and brightness. These properties are primarily influenced by factors such as the particle density (presence of vesicles), the nature of the minerals present (type and size), and the chemistry/color of the glass, and thus we expect properties to vary depending on the particle type. The steps for feature extraction detailed below were automated with a Python program that uses various functions from the open-source packages Scikit-image and OpenCV (code available at https://github.com/dbenet-ntu/Volcash-Project). The program was executed on the Gekko cluster of the High-Performance Computing Center at NTU.

Shape features Shape features were extracted from the particle outline, or silhouette, of the multi-focused particle images. To compute them, it is first necessary to measure some basic morphological properties (Figure S1 in supplementary and Table 1). The particle outline was obtained from the binary segmented image in VolcAshDB images (i.e., the alpha channel which gives images the transparency). Using the Scikit-image's function *regionprops*, we measured the particle area and perimeter, the area and perimeter of the convex hull (the minimum area that bounds the particle outline), the width and height of the bounding rectangle, the Feret maximum diameter which is the maximum distance between two parallel lines tangential to the particle outline, and the major ellipse axis (E_{maj}) which is the longest perpendicular axis of the enclosing ellipse (Dürig et al. 2018). These properties were then used to calculate 9 shape features which have been well-documented in the literature and categorized by their morphological sensitivity (Table 2).

Fine-scale roughness: *Convexity*, *Rectangularity, Circ_dellino* (Dellino and La Volpe 1996) and *Circ_cioni* (Cioni et al. 2014) are perimeter-based metrics that quantify the extent of fine-scale deviations from ideal geometric shapes in particle silhouettes. *Convexity* measures the ratio of the particle area to the area of its convex hull, *Rectangularity* quantifies the particle fit to its bounding rectangle, and *Circ_Dellino* and *Circ_Cioni* measure the proximity of the particle outline to a circle. *Circ_Dellino* and *Circ_Cioni* also reflect the particle form, with higher values indicating a greater resemblance to an equidimensional circle.

Particle-scale roughness: *Solidity* and *Compactness* are calculated from the particle area and an encompassing geometrical shape, and characterise the presence of cavities. *Solidity* quantifies how efficiently the particle occupies the convex hull, whereas *Compactness* how efficiently the particle fills the bounding rectangle.

Form: *Elongation, Roundness,* and *Aspect Ratio (Aspect_Rat)* provide insights into the overall form characteristics of particles. *Elongation* quantifies the degree of stretching, *Roundness* measures the proximity to a perfect circle, and *Aspect_Rat* captures length-to-width ratio.

Textural features Textural features were extracted using the Gray Level Cooccurrence Matrices (Haralick et al. 1973),

Table 1 List of symbols used as abbreviations of the measured properties that are used in the equations for feature extraction (Table 2)

Symbol	Definition	Reference
Shape		
A_p	Area of the particle	1
A_h	Area of the convex hull	1
P_p	Perimeter of the particle	1
P_h	Perimeter of the convex hull	1
W	Width of the bounding rectangle	1
H	Height of the bounding rectangle	1
$D_{MaxFeret}$	Feret maximum diameter: the maximum distance between two parallel lines tangential to the particle outline	1
E_{maj}	Major ellipse axis	1
Texture		
levels	Pixel intensity values from the ROI used for GLCM calculation	2
i,j	Origin pixel (i) and target pixel (j) for GLCM calculation	2
$P_d^{\theta}(i,j)$	It is the $(i,j)^{th}$ entry in the GLCM (see the sub-section "Textural features" for definition and Fig. 3 for an example), and represents the probability of finding a pixel pair with intensity values i and j at a given distance (d) and angle (θ) in the image	2
μ_i	GLCM mean	2
σ_i^2	GLCM variance	2
Color		
N	Total number of pixels	–
x_i	Pixel intensity value	–
$\bar{x}$	Mean of pixel intensity values	–

GLCM stands for 'Gray Level Co-occurrence Matrix'

1: as in Durig et al. (2018)

2: as in Hall-Beyer (2017)

a method that quantifies spatial patterns of pixel intensity distributions in the particle image (Tables 1 and 2). With these features, we intend to characterize the texture of the surface in terms of smoothness and complexity. Initially, the images were transformed from RGB to grayscale (one single channel with pixel intensity values ranging from 0 to 255) and rescaled to a maximum pixel intensity value of 15 to expedite computation. The images were then cropped into Regions Of Interest (ROI), radially distributed from the particle center, with sizes between 100–300 squared pixels and without the inclusion of background (Fig. 2). For each ROI, we computed the GLCM. In a GLCM, each element represents the frequency with which the value of a "starting" pixel is repeated with respect to the intensity of a "target" pixel. The spatial relation between the two pixels is defined by an angle (θ) and a distance (d) (Singh et al. 2017; see an example in Fig. 2). To construct the GLCM, the frequency is calculated using every possible pair of pixel values. We used several angles at steps of 11.25° and up to six different distances that gave a maximum of 90 GLCMs per ROI. For every individual GLCM, we used Scikit-image's functions *graycoprops* and *graycomatrix* to compute six well-established textural features in image analysis (Haralick et al. 1973; Hall-Beyer 2017) and which have been previously used for mineral classification from rock thin sections (Pereira Borges and Aguiar 2019). These six textural features were computed individually for all GLCMs and ROIs, which were then averaged to a single number that is used as representative for the particle image. The extracted textural features can be grouped into two categories of textural sensitivity (an example of how these features characterize four artificial textures can be found in Supplementary materials 2).

Textural smoothness: *Homogeneity*, *Energy,* Angular Second Moment, or *Asm*, and *Correlation* are calculated based on the diagonal elements of the GLCM, which represent the occurrence of pixel pairs with the same intensity value. These features emphasize the self-similarity or homogeneity of the texture within an image.

Textural complexity: *Contrast* and *Dissimilarity* are calculated based on the off-diagonal elements of the GLCM, which represent pixel pairs with different intensity values. These features capture variations and differences in texture as they reflect the occurrence of pixel pairs with varying intensities.

Table 2 List of the extracted features from each particle image, equations to calculate them, and references used

Feature	Equation	Sensitivity	Reference
Convexity	P_h/P_p	Fine-scale roughness	(Liu et al. 2015)
Rectangularity	$\frac{P_p}{2H+2W}$	Fine-scale roughness	(Dellino and La Volpe 1996)
Circ_Dellino	$\frac{P_p}{2\sqrt{\pi A_p}}$	Fine-scale roughness and form	(Dellino and La Volpe 1996)
Circ_Cioni	$\frac{4\pi A_p}{{P_p}^2}$	Fine-scale roughness and form	(Cioni et al. 2014)
Solidity	$\frac{A_p}{A_h}$	Particle-scale roughness	(Liu et al. 2015)
Compactness	$\frac{A_p}{HW}$	Particle-scale roughness	(Dellino and La Volpe 1996)
Elongation	$\frac{{D_{MaxFeret}}^2}{E_{maj}}$	Form	(Liu et al. 2015)
Roundness	$\frac{4A_p}{\pi {D_{MaxFeret}}^2}$	Form	(Liu et al. 2015)
Aspect_Rat	W/H	Form	(Leibrandt and Le Pennec 2015)
Contrast	$\sum_{i,j=0}^{levels-1} P_d^{\theta}(i-j)^2$	Textural complexity	(Hall-Beyer 2017)
Dissimilarity	$\sum_{i,j=0}^{levels-1} P_d^{\theta}\lvert i-j\rvert$	Textural complexity	(Hall-Beyer 2017)
Homogeneity	$\sum_{i,j=0}^{levels-1} \frac{P_d^{\theta}(i,j)}{1+(i-j)^2}$	Textural smoothness	(Hall-Beyer 2017)
ASM	$\sum_{i,j=0}^{levels-1} P_d^{\theta}(i,j)^2$	Textural smoothness	(Hall-Beyer 2017)
Energy	$\sqrt{ASM}$	Textural smoothness	(Hall-Beyer 2017)
Correlation	$\sum_{i,j=0}^{levels-1} P_d^{\theta}\left[\frac{(i-\mu_i)(j-\mu_j)}{\sqrt{(\sigma_i^2)(\sigma_j^2)}}\right]$	Textural smoothness	(Hall-Beyer 2017)
Channel[1] *mean*	$\frac{1}{N}\sum_{i=i}^{n} x_i$	Average pixel intensity	(Maitre et al. 2019)
Channel standard dev	$\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(x_i-\bar{x}\right)^2}$	Pixel intensity variation	(Maitre et al. 2019)
Channel mode	Computationally found as the most common value in the array	Most frequent pixel intensity value	(Mehbodniya et al. 2022)

[1]Channel in any of the Red–Green–Blue and Hue-Saturation-Value color spaces. As explained in the main text, if the channel is Red, Green, Blue or Hue, the measured feature is associated with the chromaticity; if the channel is the Saturation, the feature is associated with the color intensity; and if the channel is the Value, the feature is associated with the brightness

Color features Color features were extracted from each image using six channels from two color spaces: (1) Red, Green, and Blue (RGB), and (2) Hue, Saturation, and Value (HSV). We chose these channels because they are sensitive to color properties such as the chromaticity, intensity, and brightness (Sural et al. 2002; Ibraheem et al. 2012). These properties are mainly controlled by the particle density (presence of vesicles), the nature of the minerals present (type and size), and the chemistry/colour of the glass, and thus we expect properties to vary depending on the particle type. In digital images, the channels represent discrete pixel intensity values, typically ranging from 0 to 255 (except for the Hue channel, which ranges from 0 to 179). We computed the RGB pixel values of the normalized, multi-focused images using the Python library OpenCV, and then transformed these values into the HSV color space (Fig. 3). For each channel, we binned the pixel intensity distribution into as many bins as possible (e.g., 255 bins for the Red channel) to obtain pixel intensity frequency histograms. From these histograms, we calculated the *mean*, *mode*, and *standard deviation*, which were utilized as color features, as previously done for the recognition of mineral grains (Maitre et al. 2019) (Tables 1 and 2). We extracted color features that are predominantly sensitive to the properties described below.

Chromaticity: The Red, Green, Blue, and Hue channels capture the color composition of the particle image.

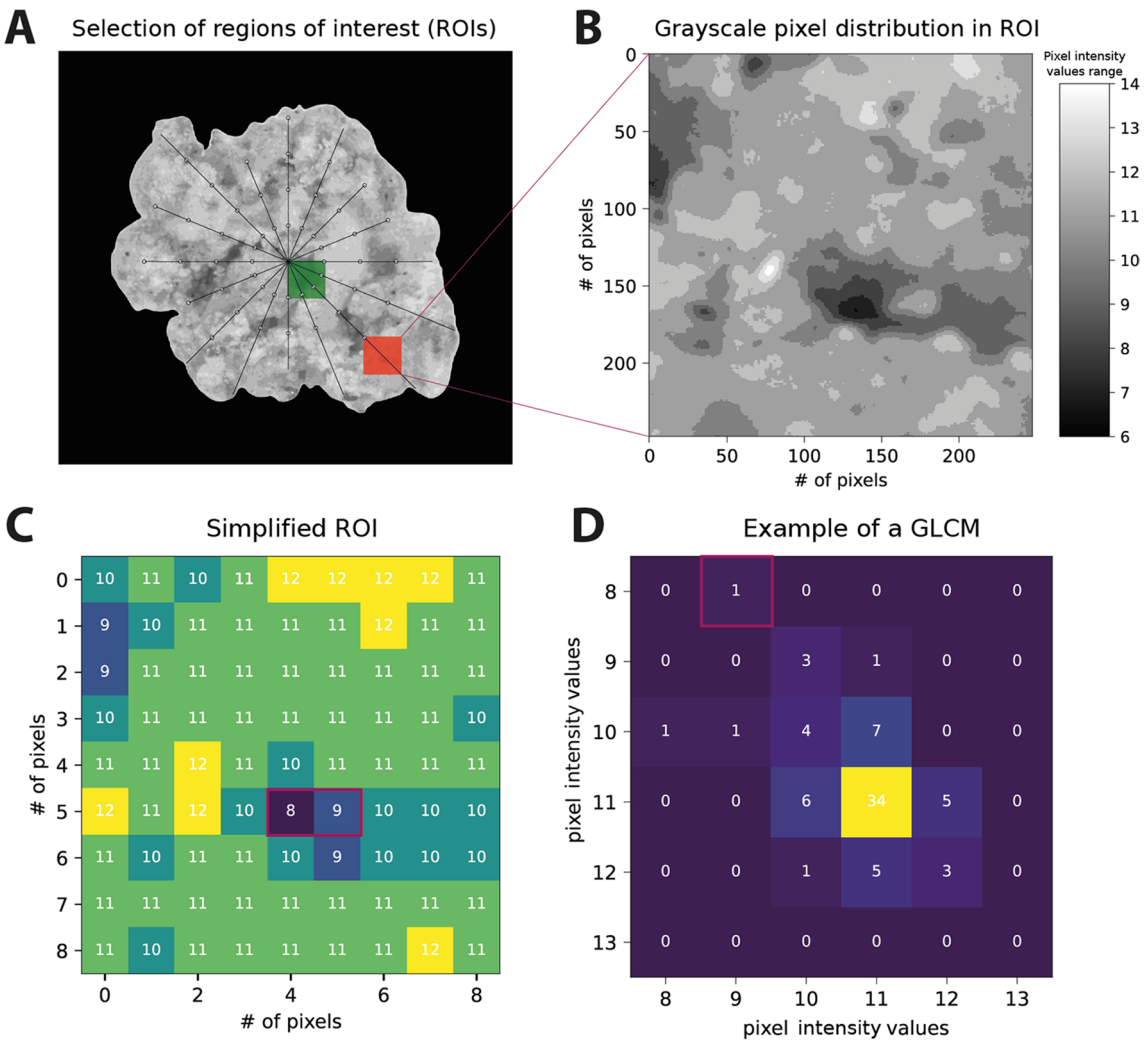

Fig. 2 Calculation of the Gray Level Co-occurrence Matrix (GLCM) for textural characterization includes four steps. **A** Grayscale image of an ash particle. Starting from the particle center (green square), an array of Regions of Interest (ROIs) is defined concentrically (black lines) and within the particle outline (each white dot corresponds to the top-left corner of a ROI). **B** Each ROI (red square) is rescaled to pixel intensity values between 0 and 15 to improve the computational efficiency. The ROI has a range of pixel intensity values from 6 to 14. **C** Simplified representation of image shown in panel (B) and expressed as a heat color map according to the pixel intensity values (numbers inside the pixel). **D** Calculation of the GLCM based on the pixel distribution shown in (C). For this example we used a distance (d) of 1, and an angle of 0°. To calculate the number of pairs between pixel intensities of 8 and 9, i.e., the element (8, 9) in the GLCM outlined in red in panel (D), the algorithm checks whether a 9 is found right next (d = 1) to an 8 at 0° (i.e., at the right-hand side). As there is only one occurrence (C; red rectangle), the element (8,9) of the GLCM takes a value of 1. Following the same process, the algorithm finds 34 occurrences (squared in yellow in diagram D) of 11 being on the right side of 11 in (C). This process is repeated for every possible pixel combination at various depths and angles, obtaining an array of GLCMs, from which texture features listed in Table 2 were calculated (see Supplementary material 2 for more information about the meaning and calculation of the texture features)

It is important to clarify that the RGB channels also incorporate information about brightness and saturation, which, in this procedure, affect the particle images in the same order of magnitude due to color normalization. In contrast, the Hue channel measures chromaticity while excluding the influence of saturation and brightness. The *mean* (e.g., *Hue mean*) serves as a global chromatic indicator of particle, the *mode* (e.g., *Hue mode*) is the most frequently occurring pixel value, and the *standard deviation* (e.g., *Hue standard dev*) reflects the spread of the data around the mean.

Intensity: The Saturation (S) channel in the HSV space quantifies the intensity or vividness of colors of the particle image. The *Saturation mean* represents the overall color intensity of the particle, whereas the *Saturation mode* represents the most frequently occurring pixel value in terms of color intensity. The *Saturation standard dev* provides insights into the variation in color intensity across the image.

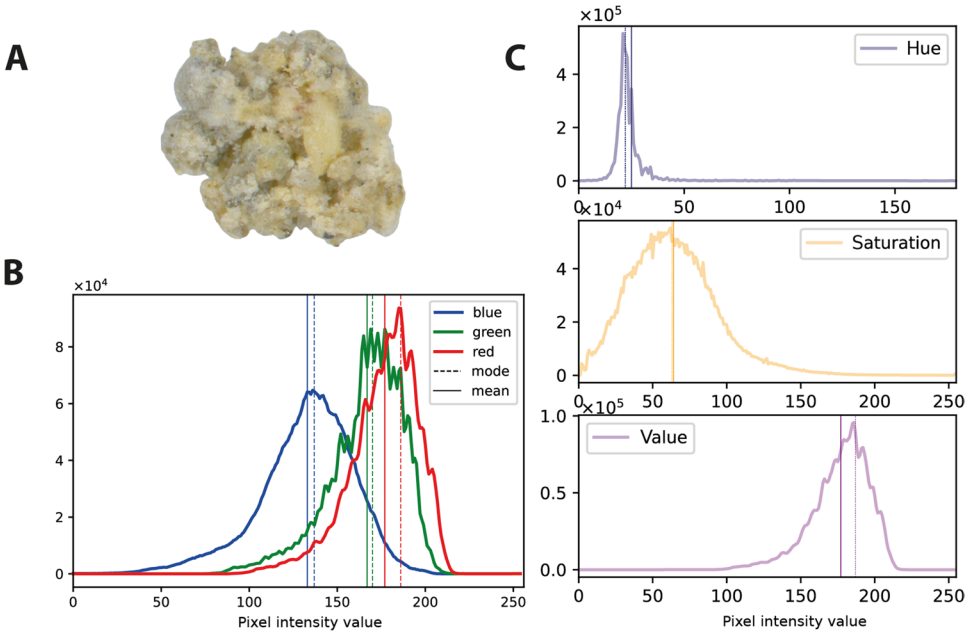

Fig. 3 Example of particle image and extracted histograms of the color features. **A** multi-focused color image of the ash particle, **B** histograms from decomposition in Red–Green–Blue channels. **C** RGB image transformed into the Hue-Saturation-Value (HSV) space from which we also computed the histograms. Vertical solid lines are the mean and dashed lines are the modes in each channel (calculated following the equations in Table 2). The mode, mean and standard deviation values were recorded in the database and used as color features for subsequent analysis

Brightness: The Value (V) channel in the HSV space provides information about the brightness or luminance of the colors in the particle image. The *Value mean* represents the overall brightness, while the *Value mode* represents the most frequently occurring pixel value in terms of luminance, and *Value standard dev* reflects the variation in brightness across the image.

Main visual characteristics recorded from the particle images

When observing the particle images under the binocular and SEM, we paid special attention to several particle characteristics that have been used in the literature as classification indicators (Table 3). Properties related to the particle color, luster, shape, and texture were used to associate the particle with a particle main type. Some of these observables were used to assign a label (code of letters) of the particle as we explain in more detail below (abbreviations in italics in Table 3).

We identified a variety of particle colors qualitatively, and the most common include "transparent" (Fig. 4A), black or dark gray (Fig. 4I–J), white (Fig. 4H), and reddish (Fig. 4E) to yellowish (Fig. 4F); the latter two are typical of hydrothermally altered material (Minami et al. 2016). The reported colors may vary with the eyesight of the observer. In our case the classification was conducted by the lead author, who was found not to be color-blind according to a web-based test (https://eu.enchroma.com/pages/colour-blind-test). The luster has been shown to be critical for recognizing juvenile particles (Miwa et al. 2013; D'Oriano et al. 2014; Gaunt et al. 2016), which are typically glossy (Fig. 4M–P). In addition, we also identified particles with dull (Fig. 4I–J), vitreous (Fig. 4A), and waxy (Fig. 4G) lusters.

We qualitatively categorized the particles edge angularity into: (i) angular (Fig. 4N), (ii) subangular/subrounded (Fig. 4M), and (iii) rounded/well rounded (Fig. 4H), following the visual comparison chart of Russell, Taylor and Pettijohn (Muller 1967) (see Figure S2 in the supplementary materials). These categories are important for particle classification, as those with rounded edges could have been weathered, whereas those with angular, sharp edges might be fresh. Various terms have been proposed to describe the particle shapes since the first petrographic studies (Heiken and Wohletz 1985). Here, we used blocky (Fig. 4M) for relatively equant particles with perpendicular to sub-perpendicular edges, fluidal if smooth-surfaced with rounded walls (Fig. 4P), spongy for particles that contain abundant and relatively small vesicles (e.g., 20 μm diameter), highly-vesicular (Fig. 4N) where vesicles are less abundant but larger (e.g., 150 μm diameter), microtubular, where particles contain elongated hollows, and pumice-like (Fig. 4O) where the groundmass contains ubiquitous < 10 μm-sized vesicles, resulting in a characteristic appearance under the binocular. Furthermore, we observed whether the particle surface appears smooth-skinned or not, as it can be useful for discriminating some particle types, e.g., fluidal glassy particles (smooth), and granular hydrothermally altered particles (rough). We also recorded the relative abundance of glass and crystals in the groundmass as: low crystallinity for 0–20% (Fig. 4O), mid for 20–40% (Fig. 4M), and high for crystallinities above > 40% (Fig. 4J). We note that here we refer only to groundmass microcrystallinity, i.e., excluding phenocrysts (crystals larger than > 0.1 mm).

Table 3 List of observations made for each particle image according to the method used

Observations in binocular					Observations in SEM–EDS				
Color	Transparent (*tr*)	Black (*bl*)	Reddish/ Yellowish	White	Pitting	Absent	Low	Medium	High
Luster	Glossy	Vitreous	Metallic	Dull	Iron oxides lineation	Yes	No		
Edges	Angular	Subangular/ Subrounded	Rounded/ Well rounded		Glass greyscale	Homogeneous	Heterogeneous		
Shape	Blocky (*b*)	Fluidal (*f*)	Spongy (*s*)	Others[1]					
Surface	Smooth	Rough							
Crystallinity	Low (*lc*)	Medium (*mc*)	High (*hc*)						
Alteration[2]	Absent (*n*)	Low (*l*)	Medium (*m*)	High (*h*)					
Weathering signs[3]	Yes	No							

In brackets and lower-case shows the abbreviations of the observations used for labelling the particles

[1]Others include highly-vesicular (*hv*), microtubular (*mt*), and pumice-like (*p*) shapes

[2]Alteration refers specifically to the degree of hydrothermal alteration

[3]These refer to the presence of coatings of white minerals, dissolution textures, and evidence of recrystallization/devitrification

Fig. 4 Examples of images of various particle types with their main characteristics, and labels (see Tables 3 and 4 for the meaning of labels). Note that only some of the characteristics have been used to build the particle labels (*in italics*)

We categorized the amount of yellowish, reddish and white material adhered to the surface (Table 3), typical of hydrothermal origin (Minami et al. 2016) as: absent, if free of hydrothermal coatings (Fig. 4M–O); low, if the amount is very small (e.g., dust; Fig. 4I); medium, when the coatings are abundant and may form encrustations (Fig. 4L); and high, when the grain surface is entirely or almost entirely covered (Fig. 4E). We paid attention to features indicative of weathering, including coatings of white minerals (clays; Fig. 4H), dissolution textures (Fig. 4G), and evidence of recrystallization/devitrification. Moreover, for the particles observed under the SEM we also recorded the presence of pitting, a form of chemical alteration that generates micro-porosity, evidence of recrystallization, such as iron oxides lineations, and whether the pixel instensity of glassy groundmass is homogeneous or heterogenous (D'Oriano et al. 2014).

Labeling of the particles by the petrologist

Using the observational features noted above, each particle was classified into the four main types that are typically used in the literature (Suzuki et al. 2013; Gaunt et al. 2016; Ross et al. 2022): free crystals, altered material, lithic, and juvenile. In addition, we also classified the particles into a few sub-types (Table 4), and noted special characteristics such as crystallinity degree, degrees of hydrothermal material, and shapes.

We used a four-step process to classify the particles into the main types (Fig. 5):

(1) Features that are characteristic of free crystals (*F*; Fig. 4A–D) include planar structures (e.g., twinning) and well-faceted crystal habit. The free crystals in the database are mainly plagioclase and pyroxene, minor amphibole, and rarely native sulfur and olivine.
(2) Altered material (*A*) includes both hydrothermally altered as well as weathered particles. We looked for and noted evidence of major hydrothermal alteration. Particles that were partially or entirely covered by hydrothermal encrustations (medium or high degrees of hydrothermal alteration) were classified as hydrothermally altered (*AH*; Fig. 4E–F). When visible, we also noted their crystallinity. These hydrothermally altered particles typically have granular texture or form aggregates that are white, or yellowish to reddish. After discarding free crystals and hydrothermally altered particles, most of the particles that are left are generally glassy and variably altered. At this point, we identified features that are characteristic of weathered particles (*AW*; Fig. 4G–H). Under the binocular, these include a loss in shine (dull luster), round edges, and modifications of the original groundmass, such as recrystallization into secondary minerals (typically whitish clays) and dissolution textures.

Table 4 List of main particle types and sub-types we have used in the database

Main types	Sub-type	Particle label example	Meaning
Free-crystal (*F*[1])			
	Plagioclase (*PL*)		
	Pyroxene (*PX*)	–[2]	–
	Amphibole (*AMF*)		
Altered material (*A*)			
	Weathered material (*AW*)	–	–
	Hydrothermally altered material (*AH*)	*AH-bl-hc-m*	Hydrothermally altered material-black-high crystallinity-medium degree of hydrothermal alteration
Lithic (*L*)			
	Standard lithic (*LL*)	*LL-tr-mc-n*	Lithic-transparent-medium crystallinity-absent hydrothermal alteration
	Recycled juvenile particles (*LRJ*)	*LRJ-bl-lc-f-m*	Recycled juvenile-black-low crystallinity-fluidal-medium degree of hydrothermal alteration
Juvenile (*J*)			
	Standard juvenile (*JJ*)	*JJ-bl-hc-b*	Juvenile-black-high crystallinity-blocky
	Coated juvenile particles (*JC*)	*JC-tr-lc-l-s*	Coated juvenile-transparent-low crystallinity -low degree of hydrothermal alteration-spongy

The individual particle labels are made of a sequence of letters (in italics) that represent the abbreviations of the main particle type and some of their characteristics (Table 3). Note that the letters of the labels are separated by a dash to improve the readability, but are not present in the labels of the database

[1]In brackets and upper-case, the abbreviations used for image labelling

[2]The dash "– " indicates that any example takes the "Sub-type" label

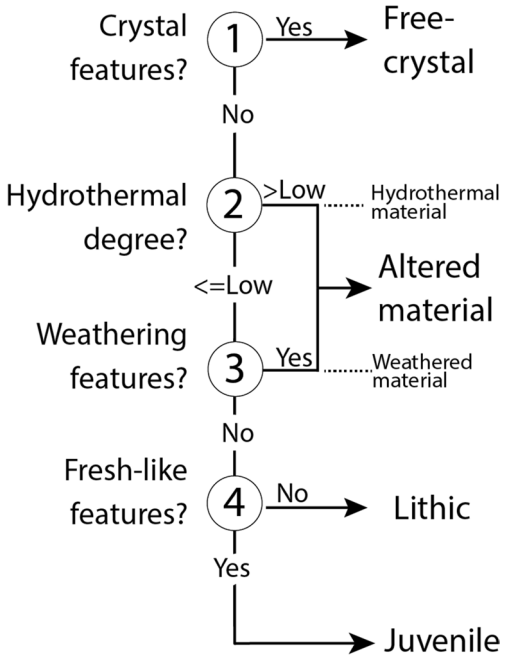

Fig. 5 To classify each ash image into the various types we used a dichotomous key that contains four steps. Note that particles classified as hydrothermal or weathered material belong to the main type 'Altered material'

Weathered particles are typically white, dull to waxy, and have rough surfaces. Particles containing weathering features at an early stage of development can be difficult to identify under the binocular microscope, and we recommend the observation of incipient palagonization, recrystallisation, and presence of secondary minerals by SEM.

(3) Most lithic particles (*L*) are typically dull, dark, with sub-angular to rounded edges, and contain limited signs of weathering or hydrothermal alteration (absent to low degrees). We further noted their crystallinity, and whether they are transparent or black. Lithic particles derive from already cooled magma that was emplaced before the arrival of the magma driving explosive activity. Non-magmatic fragments eroded from the subvolcanic basement are not comprised by the lithic group. If found, these would be grouped under "accidental fragments" (Fisher and Schmincke 1984). Recycled juvenile (*LRJ*), when observed under the binocular, often show a duller or metallic luster, sometimes with disseminated red patches (Fig. 4L), but the SEM is necessary to observe conclusive features such as recrystallization and the presence of iron oxides aligned around microphenocrysts, which may occur as particles fall back into the crater and are thermally altered in oxidizing conditions (D'Oriano et al. 2014). Because we don't know the time span between the LRJ fall and their ejection, we classified them as lithic component to prevent overestimating the juvenile component.

(4) Finally we paid special attention to features that are characteristic of fresh, juvenile particles (*J*; Fig. 4M–P). We mainly recorded five features, here referred as "fresh-like". These are based on a review of 35 articles from the literature (Fig. 6) and include: shiny gloss, sharp edges, smooth-skinned surface, and lack of weathering and alteration features (Fig. 4M–O). We avoided using specific names such as sideromelane and tachylite commonly referred for basaltic particles (Taddeucci et al. 2002, 2004) because these may have connotations related to the chemical composition. We also noted the particles' shape as these may indicate the mechanism of fragmentation, which is particularly valuable when analyzing temporal sequences of ash samples. Juvenile particles derive from fresh magma, and their presence is typically interpreted as evidence for shallowly emplaced magma, which has critical implications for hazards assessment. We thus also observed these particles using the SEM. We looked for homogeneous grayscale, smooth surface, sharp or stepped edges (Dürig et al. 2012; Pardo et al. 2020; Ross et al. 2022), and the lack of signs of weathering (e.g., etch pitting). Juvenile particles were further classified based on crystallinity, color, and according to the shape and presence of material on surfaces. Coated juvenile particles (JC) are classified as a subgroup, and are characterized by incipient and limited amount of coatings together with characteristics that strongly point towards a juvenile origin (e.g., the appearance of vesicular shapes). These are interpreted to form by syn-eruptive alteration of juvenile material by hot hydrothermal fluids with juvenile material (Alvarado et al. 2016) or by interaction with plume gases (Spadaro et al. 2002).

The particles are labelled with a sequence of letters that reflect the types and sub-types (Table 4). In some cases, this sequence also includes lower-case letter(s) that are the abbreviation(s) for special characteristics that are valuable for monitoring purposes (Table 3). The groundmass microcrystallinity may provide insights into the depth and cooling of the sampled magma; the particle shape offers information about the fragmentation mechanism; and the degree of hydrothermal alteration can provide hints on the prevalence of the hydrothermal system.

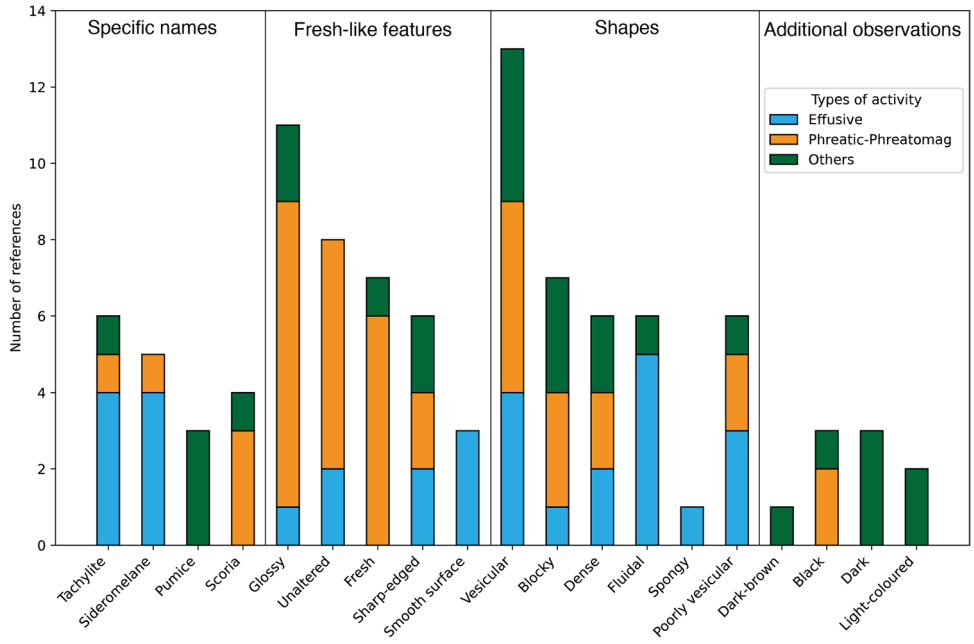

Fig. 6 Main characteristics of juvenile particles observed under the binocular microscope according to previous publications (Cioni et al. 1992; Taddeucci et al. 2002; Scasso and Carey 2005; D'Oriano et al. 2005, 2011, 2014, 2022; Ersoy et al. 2006; White and Houghton 2006; Savov et al. 2008; Ersoy 2010; Andronico et al. 2013, 2014; Miwa et al. 2013, 2021; Suzuki et al. 2013; Eychenne et al. 2015; Gaunt et al. 2016; Geshi et al. 2016; Lücke and Calderón 2016; Kurniawan et al. 2017; Troncoso et al. 2017; Gómez-Arango et al. 2018; Gorbach et al. 2018; Miyabuchi et al. 2018; Angkasa et al. 2019; Battaglia et al. 2019; Miyagi et al. 2020; Romero et al. 2020; Thivet et al. 2020; Matsumoto and Geshi 2021; Pistolesi et al. 2021; Benet et al. 2021; Minami et al. 2022) on a range of volcanic eruptions. Most observations are from basaltic effusive eruptions, and we have grouped the phreatic and phreatomagmatic explosions as they can be difficult to distinguish (Pardo et al. 2014). The category 'Others' includes subplinian, plinian and submarine eruptions

Uncertainties and sources of error

Uncertainties and errors in ash componentry and particle classification

The precision and accuracy of our observations in ash componentry and particle classification are affected by uncertainties and errors. Precision refers to the variability and spread of the data when the measurements are repeated, such as the proportion of juvenile particles in a given sample, and can be quantified with the standard deviation. Accuracy, on the other hand, refers to the difference between the experimental values (e.g., the classification of a particle as juvenile) and the true value (e.g., an actual juvenile particle), which is often unknown. It is generally assumed that errors affecting the precision are random, whereas those influencing the accuracy are systematic (Hughes and Hase 2010). In ash componentry studies, the proportion of the particle types are reported relative to the total number of particles counted within a presumed representative subsample or aliquot. The reported proportion of particles carries errors that depend on two key factors: (1) the number of particles counted, which relates to the precision of the measurements, and (2) the potential misclassification of particles by the observer, which corresponds to the accuracy of the measurements. These two topics are discussed in some detail below.

Precision in ash componentry determinations The error related to the precision in particle counting can be assumed to be random and varies according to the number and proportion of the particle types that are observed. The proportion (p) is the ratio between the number of a particle type and the total number of particles, and it can be reported in percentage or in decimal form. This error can be expressed as the margin of error (ME) (Tanur 2011), and is quantified with a confidence level that is associated with a z-score, z_i, which is obtained from the area under the gaussian curve (Mendenhall et al. 2012), a standard deviation (σ), and population size (n; the total number of particles measured for a given sample):

$$ME = z_i \sqrt{\frac{\sigma^2}{n}} \tag{1}$$

where we calculate the standard deviation as: $\sigma = \sqrt{p(1-p)}$ (Mendenhall et al. 2012). For example, for a 95% confidence level (i.e., z_i=1.96) and a measured proportion of 10% (p

=0.1) from a total of 400 particles, we obtain that $\sigma = 0.3$, hence, $ME = 1.96\sqrt{0.3^2/400} = 0.03$, which corresponds to 3%. This means that the proportion will be within the interval $10 \pm 3\%$, from which we can calculate the relative error to be 3/10, or 30%. There is a trade-off between the number of particles that we count and the precision that we need to make useful characterization of the sample componentry (Liu et al. 2017; Ross et al. 2022). We modelled the relationship between the number of particles, their proportions, and the precisions that we would obtain for a 95% confidence level (Figure S3). For example, if we wish a relative error < 30%, for a particle type with a proportion larger than 20%, we need to measure at least 200 particles (Figure S3). However, if we are dealing with a particle type that occurs in a low proportion such as about 1%, and we wish a relative error < 100% we need to measure at least 400 particles (Figure S3).

Particle classification errors and accuracy The errors related to the accuracy are much more difficult to quantify because we do not know a priori the true particle types. These errors can be random, as when the observer misclassifies a particle because the image is partly blurry or because of ambiguity of observations, or they can be systematic, when the observer systematically misclassifies particles from a given type into another. We have tried to quantify the random but non-systematic errors by classifying the particles from two aliquots of the same sample, which could in principle reflect the incorrect classification by the same observer due to random errors. The expectation is that, if the misclassification errors are small, the difference in the particle proportions between the two aliquots should be within the precision of the measurements as explained above. We did such exercises for ash samples of Kelud (2014) and Soufrière de Guadeloupe (1976; Figure S4) and found that the particles proportions from the two aliquots are within the margin of error. This suggests that the effect of random errors in particle misclassification is small, and thus not significant, but such inadvertent misclassification errors may vary from sample to sample. This exercise could be conducted as a standard step for the community to examine potential errors of classification.

Quantifying the accuracy for systematic errors of particle misclassification is difficult, as we don't know the true particle types, and although some particles have unequivocal traits for classification, others show inconclusive features. For example, classifying particles such as crystals, can be done with clear diagnostic observations such as cleavage, but classifying particles with limited signs of weathering as lithic or weathered material is not obvious, and will likely vary with the observer. We strived to limit the problems of misclassification by adopting the same observational characteristics of particles reported in the literature, especially for juvenile particles (Fig. 6). Proper quantification of the accuracy or misclassification could be done by expert elicitation procedures (Aspinall and Cooke 1998; Marzocchi and Bebbington 2012), where several experts classify the particles from the same sample, but this is currently beyond the scope of this contribution.

Uncertainties and errors in feature extraction from particle images

The binocular images of individual particles we used are of high-resolution (between 1,800 and 2,000 pxls/mm), multi-focused, and they capture certain physical properties of the actual particle related to its shape, texture and color. However, these have uncertainties and errors that affect the quality of the extracted features during image acquisition and segmentation.

Effects of the image type on shape analysis It is well-documented that the results of measurements of particle shape are significantly influenced by the image resolution (Liu et al. 2015; Saxby et al. 2020; Ross et al. 2022) and the method that is used to capture the particle contour. Notably, the measurement of the apparent 2D projected shape of the particles using the SEM or optical microscope, are different from those obtained from the 2D cross-sectional shape of the same particles (Liu et al. 2015; Buckland et al. 2018; Nurfiani and Bouvet de Maisonneuve 2018; Edwards et al. 2021; Comida et al. 2022). The difference is also found in the values of *Solidity* and *Convexity* we obtained, which are higher (meaning smoother particle contours) than those reported by Liu et al. (2015) who investigated the ash from the same eruption (although not the same exact samples) using 2D cross-sectional shape by SEM (Fig. 7). Moreover, the *Solidity* and *Convexity* values we obtained are also higher than those of the 2D projected particle shape reported by Nurfiani and Bouvet de Maisonneuve (2018) using a particle analyzer coupled with an optical microscope system on the same exact samples. The possible reasons for such differences are discussed below.

Effects of image segmentation methodology on particle shape Visual inspection of the images with the highest *Solidity* and *Convexity* values revealed that very small concavities (< 10 μm) that were present in the original particle images (with a grain-size fraction between 0ϕ–2ϕ) are absent in the images obtained after segmentation (Figure S5). This may be due to our original images not having enough resolution, or because the borders of the particle did not have a sharp enough contrast to allow for fine segmentation by the deep learning model we used (Qin et al. 2020) (see section "Image processing" for more details on the model).

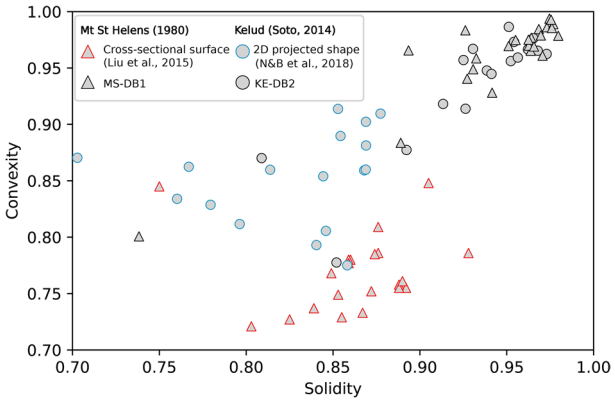

Fig. 7 Scatter plot of *Solidity* versus *Convexity* of samples from Mount St. Helens (*MS-DB1*) and Kelud (*KE-DB2*; see Table S1 for sample details) compared to two other studies (Liu et al. 2015; N&B: Nurfiani and Bouvet de Maisonneuve 2018). The values we obtained are much higher (i.e., contours are smoother) than those of Liu et al. (2015), as we use the apparent 2D projected shape instead of a 2D cross-sectional surface. Comparison with external shapes by Nurfiani and Bouvet de Maisonneuve (2018) shows that our values remain higher. We attribute this shift to smoothing of the contours by the segmentation algorithm combined with blurry particle borders. We also note that the compared studies used different image resolution and particle grain-sizes, and thus this may also play a role. Liu et al. (2015) used > 10^6 pixels per particle and size of 1ϕ–2ϕ (250–500 µm). Nurfiani and Bouvet de Maisonneuve (2018) used ~ 4×10^5 pixels per particle without grain-size selection. We used ~ 2.5×10^6 pixels per particle for a 0ϕ–1ϕ (1000–500 µm) for sample *MS-DB1*, and ~ $1.9 \times 10.^6$ pixels per particle for a 1ϕ–2ϕ (500–250 µm) for *KE-DB2*

We compared the *Solidity* and *Convexity* values for a pumice fragment and a glass shard measured through three different methods: (1) the 2D projected shape of the multi-focus binocular image followed by segmentation with the deep learning model, (2) the 2D projected shape of the SEM image followed by thresholding according to Liu et al. (2015), and (3) the 2D projected shape of the multi-focus binocular image followed by manual segmentation with Adobe Photoshop as recommended in Comida et al. (2022). The *Convexity* values obtained with these techniques are ~ 25% lower for the pumice fragment, and ~ 10% lower for the glass shard (Fig. 8) than those determined with our procedure. On the other hand, the *Solidity* values are very similar, which suggests that features sensitive to the particle-scale roughness are less affected. Variations in sensitivity between particle- and fine-scale features have also been documented in previous studies (Liu et al. 2015; Saxby et al. 2020). Therefore, it seems likely that the higher *Convexity* values are due to the smoothening of microscale irregularities upon segmentation, and thus care has to be taken when used for particle characterization and classification. Improved particle segmentation can be obtained by enhancing the resolution/focus of the original image, using an improved deep learning algorithm, or by manually segmenting with Adobe Photoshop or thresholding the SEM image of the projected particle, although these are very time-consuming, and unpractical for large number of particles.

Description of the features and Principal Component Analysis (PCA)

We conducted descriptive analysis and PCA of the 33 features extracted (Table 2) for a total of 6,304 particle images, and across eruptive activity types and samples. The analyses aimed to explore the characteristics of the features' distributions, identify those that contribute the most to the dataset variance (a metric that measures the dispersion of the data points from their central tendency), and also to gain insights

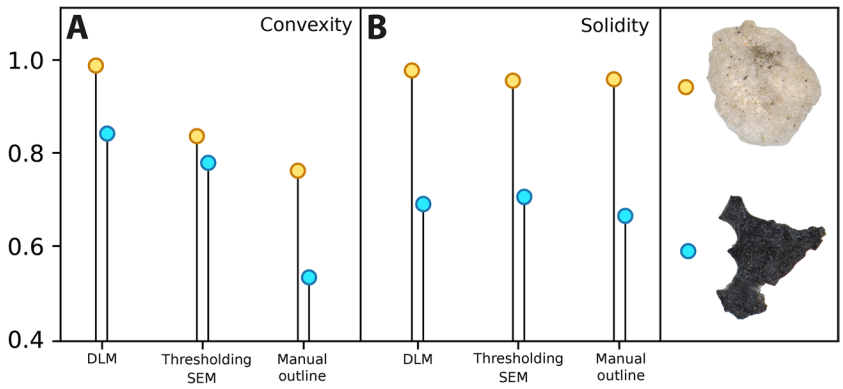

Fig. 8 Illustration of the effect of using different segmentation protocols on the retrieved values of (**A**) Convexity, and (**B**) Solidity of a pumice fragment (yellow circle) and a glass shard (light blue circle). We tested three methods. A Deep Learning Model ("DLM") to segment a binocular multi-focused image, "Thresholding SEM" results from thresholding the SEM image of the apparent 2D projected particle, and "Manual outline", results from using Adobe Photoshop to manually refine the outline of a binocular multi-focused image. The *Convexity* values of the pumice fragment obtained with DLM are much higher than those from the Manual Outline, as microvesicles were neglected during segmentation. On the other hand, Convexity values obtained from the glass shard are more similar. The *Solidity* values obtained from different methods are similar, suggesting that values sensitive to roughness at particle-scale are robust

on the dataset structure which may be relevant for particle classification.

We performed PCA to identify the features with a high contribution to the dataset variance. High-variance features capture a broader range of variations in the data, and are often selected for classification tasks (e.g., Khan 2018; Phillips and Abdulla 2021; Shehzad et al. 2022). In PCA, a new set of variables called principal components (PCs) are constructed as linear combination of the features to retain the maximum variance in the dataset (Smith 2002). The variance captured by each PC is termed "explained variance" and is expressed as a percentage of the total feature variance. Furthermore, PCA assigns a "loading" to each feature, representing the coefficient of its contribution to each principal component. PCA has been extensively used in volcanic ash studies for dimension reduction of shape features and has allowed for instance to identify different morphological types of particles (Scasso and Carey 2005; Liu et al. 2015; Nurfiani and Bouvet de Maisonneuve 2018).

We first standardized the features to have a mean at 0 and a variance at 1 by applying Scikit-learn's *StandardScaler*. This pre-processing step is commonly used to prevent features with larger range of values from dominating the PCA results (Hastie et al. 2009). We chose to extract 3 PCs by eigen decomposition (Dürig et al. 2021) using the Python package *pca* (https://erdogant.github.io/pca/) based on observed explained variance. A first round of PCA with 10 PCs showed that beyond the 3rd principal component, the explained variance dropped below 10% (Supplementary 3), and thus we prioritized the first 3 PCs for an overview of dataset variance and simplicity in interpretation. We conducted the PCA for the entire dataset, and within activity types and samples. For each PCA, we computed the explained variance of the 3 PCs and the feature with the highest loading. Moreover, we visualized the 3 PCs in a 3D plot which allowed us to better discuss the potential for classification of different features. It should be noted that PCA may not consider low-variance features that hold valuable information for classification, and does not account for non-linear relationships. To address these aspects, machine learning models designed to capture non-linearities might be employed, but this goes beyond the scope of our current study.

Results

VolcAshDB contents

We analyzed 12 samples from 8 volcanoes and 11 eruptions (see Table S1 for further sampling details), from which we obtained 6,304 images of particles that were classified in the different types (Table 5). Our collection of ash samples derive from a wide spectrum of volcanic activities:

(1) Phreatic eruptions: those of la Soufrière de Guadeloupe (Lesser Antilles) in 1976 and 1977 (Le Guern et al. 1980; Feuillard et al. 1983), during the early unrest of Mt. Pinatubo (Philippines) in April 1991 (Paladio-Melasantos et al. 1996), and from Ontake (Japan) in 2014 (Miyagi et al. 2020);
(2) Lava dome explosions: those from Nevados de Chillán volcanic complex (Chile), from the beginning of the eruptive period in December 2016 and after a dome extrusion in April 2018 (Benet et al. 2021), explosions from Merapi volcano (Indonesia) in July and November 2013 (Nurfiani and Bouvet de Maisonneuve 2018);
(3) Basaltic lava fountaining: at Cumbre Vieja (Canary Islands) in October 2021 (Romero et al. 2022); and
(4) Plinian to sub-plinian eruptions: two samples from different locations (KE-DB2 and KE-DB3) of Kelud (Indonesia) in 2014 (Maeno et al. 2019; Utami et al. 2021), and one sample from the main explosive stage of Mount St. Helens (USA) on 8 May 1980 (Scheidegger et al. 1982).

The total number of particles per sample vary between 142 and 1142, and the relative precisions for each particle type varies between 196% (1 ± 2) for the free-crystal component of Cumbre Vieja down to 2% (255 ± 5) for the juvenile component of the Mount St. Helens sample (Table 5). The largest number of particles we classified (22% of the total) are from dome eruptions of Nevados de Chillán volcanic complex, whereas samples from each of the other volcanoes represent 10–20% of the total database. The most abundant particle is altered material (47%), followed by juvenile (27%), lithic (19%) particles, and free crystals (7%; Fig. 9B). On a volcano-by-volcano level, the proportions range between two endmembers (Fig. 9C): one entirely made of altered material (Ontake, 2014), and the other dominated by juvenile particles (Mount St. Helens, 1980). The lithic particle content varies from low (Pinatubo, 1991) to very rich (Nevados de Chillán, 2016–2018).

Componentry variations in the dataset

The diversity in proportions of particle types reflects the wide range of activity types in our collection. A closer look reveals that ash from samples of the same activity type share certain characteristics (Fig. 10):

(1) Ash from the phreatic events (Ontake, 2014; Soufrière de Guadeloupe, 1976–1977; and Pinatubo in the early

Table 5 Main sample characteristics; number of particles per component and associated errors in brackets

Volcano/Samples	Eruption date	Magma composition[1] (reference)	Volcano type	Eruptive Activity type	Measured grain-size in phi (ϕ) scale and microns (μm)	Number of particles per component and associated error				Total
						Altered material	Free-crystal	Juvenile	Lithic	
Cumbre Vieja										
CV-DB1	19/10/21	Tephrite (Romero et al. 2022)	Cinder cone	Lava fountaining (Romero et al. 2022)	0ϕ–1ϕ (1000–500 μm)	3 (±3)	1(±2)	719(±30)	352(±30)	1075
Kelud										
KE-DB2	14/2/14	Basaltic Andesite (Maeno et al. 2019)	Stratovolcano	Subplinian (Utami et al. 2021)	1ϕ–2ϕ (500–250 μm)	50(±13)	4(±4)	268(±13)	3(±3)	325
KE-DB3	14/2/14	Basaltic Andesite (Maeno et al. 2019)	Stratovolcano	Subplinian (Utami et al. 2021)	1ϕ–2ϕ (500–250 μm)	162(±18)	59(±14)	54(±13)	65(±14)	340
Merapi		Basaltic Andesite (Gertisser et al. 2023)								
ME-DB1	22/7/13	–	Stratovolcano	Dome explosion (N&B[2] 2018)	0ϕ–1ϕ (1000–500 μm)	232(±16)	13(±7)	0	78(±15)	323
ME-DB2	22/11/13	–	Stratovolcano	Dome explosion (N&B 2018)	0ϕ–1ϕ (1000–500 μm)	595(±23)	76(±16)	4(±4)	100(±18)	775
Sourfière de Guadeloupe		Basaltic andesite-Andesite (Metcalfe et al. 2021)								
SG-DB1	8/7/76	–	Stratovolcano	Phreatic (Le Guern et al. 1980)	0ϕ–1ϕ (1000–500 μm)	222(±17)	54(±13)	0	66(±14)	342
SG-DB2	1/3/77	–	Stratovolcano	Phreatic (Le Guern et al. 1980)	0ϕ–1ϕ (1000–500 μm)	134(±5)	8(±5)	0	0	142
Nevados de Chillán		Basaltic-Andesite (Dixon et al. 1999)								
NC- DB15	3/4/18	–	Dome complex	Dome explosion (Benet et al. 2021)	0ϕ–1ϕ (1000–500 μm)	224(±26)	77(±17)	92(±18)	749(±31)	1142
NC-DB2	29/12/16	–	Dome complex	Dome explosion (Benet et al. 2021)	0ϕ–1ϕ (1000–500 μm)	99(±16)	12(±7)	14(±7)	171(±17)	296
Ontake		Andesite (Miyagi et al. 2020)								
ON-DB1	27/9/14	–	Stratovolcano	Phreatic (Miyagi et al. 2020)	0ϕ–1ϕ (1000–500 μm)	777(±0*)	0	0	0	777
Pinatubo		Dacite (Bernard et al. 1996)								
PI-DB1	2/4/91	–	Caldera	Phreatic (Paladio-Melasantos et al. 1996)	0ϕ–1ϕ (1000–500 μm)	386(±19)	104(±18)	0	16(±8)	506
Mount St Helens										
MS-DB1	18/5/80	Dacite (Scheidegger et al. 1982)	Stratovolcano	Plinian (Scheidegger et al. 1982)	0ϕ–1ϕ (1000–500 μm)	4(±4)	0	255(±58)	2(±3)	261
				Total		2888(±78)	408(±38)	1406(±65)	1602(±68)	6304

The error in absolute value was calculated according to the margin of error (*ME*) as defined in Eq. (1) with a confidence interval of 95%

[1]Where available, we refer to the composition of the juvenile magma from the eruption for which we have the sample(s). If not available (e.g., in phreatic deposits), we use a dash "– " and refer to the predominant magma composition erupted from the volcano

[2]N&B:Nurfiani & Bouvet de Maisonneuve 2018

* We obtained an absolute error of 0 because *ON-DB1* consists of one single proportion (100% altered material) and the *ME* is calculated for a population with multiple proportions

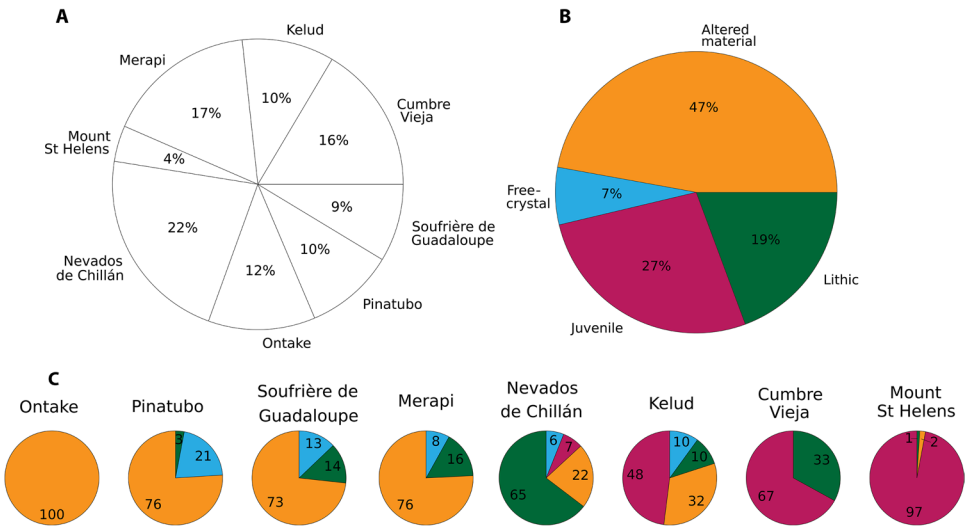

Fig. 9 Charts to illustrate the proportions of particles per volcano, per particle type and combined volcano and particle type that are currently stored in VolcAshDB. Pie charts showing (**A**) the percentage of the total number of particles in the database per volcano, (**B**) the overall proportion of each particle types in the database expressed in percentages, and (**C**) the same as in (**B**) but per volcano, with each particle type shown with different color as in panel B. Values in (**C**) are also percentages but the sign "%" has been deleted to optimize space

activity of April 1991) is dominated by altered material (86%), with minor free-crystals and lithics, and is free of juvenile grains. The particles are typically white or red to yellowish, with irregular surfaces, rounded edges, and may form aggregates (Fig. 10A). The particle proportions we report are consistent with previous studies of phreatic events (e.g., Ontake, 2014, Miyagi et al. 2020; Soufrière de Guadeloupe, Heiken et al. 1980), and which have been interpreted to be driven by gas or steam accumulation and interaction with a shallow active hydrothermal system. However, our samples do not include those from low-silica mafic magmas, nor from interaction with crater lakes.

(2) The samples from dome explosions (Nevados de Chillán, 2016–2018; Merapi, July and November 2013) are characterized by abundant altered material (45%) and lithic (43%) grains, with small amounts of free-crystals and juvenile types. Often the particles are coated by hydrothermal material, or they are dark, with a massive appearance and high crystallinity (Fig. 10B). Samples from lava dome explosions can vary in componentry from predominantly lithic with abundant fragments from an old lava dome and a low juvenile content (e.g., 15%), to samples with high juvenile content when dome extrusion starts, e.g., > 80% (Benet et al. 2021; Primulyana et al. 2018). It should be remarked that our componentry results are subject to a specific grain-size fraction at a particular sampling site. Improved componentry per eruption would include the addition of multiple grain-size fractions and samples at varying distances from the vent, e.g., at Tungurahua eruptive deposits of 2006 (Eychenne et al. 2013).

(3) The ash particles from lava fountaining (Cumbre Vieja, 2021) contain abundant juvenile grains from fresh magma but also juvenile grains that have been recycled, typically by falling back into the crater (LRJ, Fig. 10C), with a lesser amount of lithic and free-crystal types. The juvenile particles are dark, with fluidal to highly vesicular shapes, whereas the recycled juvenile particles (LRJ) appear duller and featured by modifications on the surface, such as metallic luster and disseminated red patches. Based on previous experiments and other case studies, the fluidal shape is indicative of magma breakup hydrodynamically (Gonnermann 2015; Comida et al. 2022, 2023), whereas the highly-vesicular particles indicate effective degassing during fragmentation, as for example at Etna (Taddeucci et al. 2002). Particles from this activity type have been well documented at Etna (Polacci et al. 2019) and Stromboli (Cannata et al. 2014).

(4) The ash samples we analyzed from a subplinian eruption (Kelud, 2014) contain abundant juvenile particles, and can be recognized by their low-crystallinity and pumice-like vesicularity (Fig. 10D). Moreover, these samples contain variable amounts of lithic and altered material, and minor free-crystals. The presence of pumice-like shape indicates efficient fragmentation due to syn-eruptive volatile expansion and exsolution (Taddeucci and Wohletz 2001).

(5) Our ash sample from a plinian eruption (Mount St Helens, 1980) is clearly dominated by the same type of pumice-like juvenile particles (> 95%; Fig. 10E) as in subplinian samples. Further grouping of the pumice-like type based on the vesicularity shape and density can provide important details of the fragmentation mechanism (Taddeucci and Wohletz 2001), but such a detailed analysis is beyond the scope of this study. Previous studies of plinian eruptive deposits, including Mount St Helens, 1980 (Carey and Sigurdsson 1982) and the Minoan eruption of Santorini (Druitt 2014),

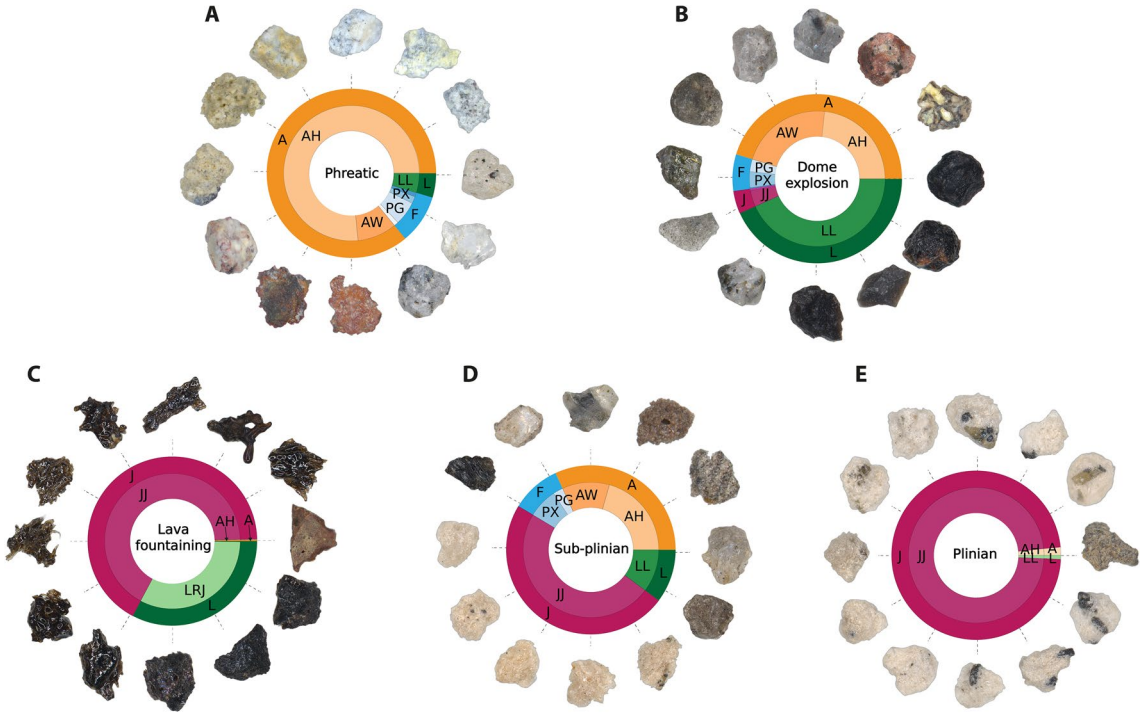

Fig. 10 Examples of particle images for different eruptive activity and according to the particle main types and subtypes shown as pie charts (Table 4). Examples are shown for the most predominant subtype particle in each pie chart (panel A to E). **A** "Phreatic" includes three samples from Ontake, 2014 (*ON*-DB1), Pinatubo (*PI-DB1*) and Soufrière de Guadeloupe (*SG-DB1*), **B** "Dome explosion" includes four samples from Nevados de Chillán, 2016–2018 (*NC-DB2* and *NC-DB15*) and Merapi (*ME-DB1* and *ME-DB2*). **C** "Lava fountaining" includes one sample from Cumbre Vieja, 2021 (*CV-DB1*). **D** "Sub-plinian" includes two samples from Kelud, 2014, one collected at Solo (*KE-DB2*) and the other at Dieng (*KE-DB3*). **E** consists of one sample from Mount St. Helens, 1980 (*MS-DB1*). Note the differences in particles' aspect across eruptive activity, e.g., particles from lava fountaining are darker and more elongated, whereas those from plinian events are yellow and microvesicular. All our samples belong to the 0ϕ–1ϕ (1000–500 µm) grain-size fraction, except for those from Kelud (Sub-plinian) that belong to 1ϕ–2ϕ (500–250 µm). Abbreviations as in Table 4: A = Altered material, F = Free-crystal, J = Juvenile, L = Lithic, AH = Hydrothermally altered material, AW = Weathered material, JJ = Standard juvenile, LL = Standard lithic, LRJ = Recycled juvenile, PG = Plagioclase, PX = Pyroxene. The values used for this figure can be consulted in Table S2

report abundant lithics, and thus we expect to add ash samples from multiple stratigraphic levels, distance from the vent and grain sizes in the future.

Description of the particles' features

We extracted a total of 33 features for each particle image which were incorporated in VolcAshDB. In this section, first we describe histograms for the different features of the entire database, and then we examine histograms categorized by activity types and particle types to identify the existence of distinctive subpopulations that may only appear within specific subgroups, reflecting variations in feature sensitivity.

Overall feature distributions of the particles The aggregated values of the shape, texture and color features for all particles are generally unimodal but with a wide range of variability. The variability was measured as the standard deviation of the normalized feature values (ranging between 0 to 1). The feature *Elongation* has the lowest variability (Fig. 11A), suggesting that the length–width ratio of the particles is rather similar within the dataset, whereas *Homogeneity* has the largest (Fig. 11B), indicating that particle surfaces exhibit a large variety of textural smoothness. On the other hand, the features related to color show a large range of distributions, with multimodality and a wider variability. For example, the *Hue mean* shows two modes (Fig. 11C), the *Red mean* has three (Fig. 11D), the *Blue mode* has four or more, (Fig. 11E) and the *Value mode* has the highest variability (Fig. 11F). The well-defined local maxima in these multimodal distributions highlight the presence of subpopulations of particles that have characteristic feature values. These subpopulations may correspond to specific particle types, as our features were designed to capture the distinct shape, texture and color properties that particle types exhibit under the binocular microscope. We consider these multimodal distributions to have more diagnostic power for classification than the low-variability features.

Feature distributions by particle type and volcanic activity type The feature histograms were split into main particle types and into activity types to qualitatively assess whether

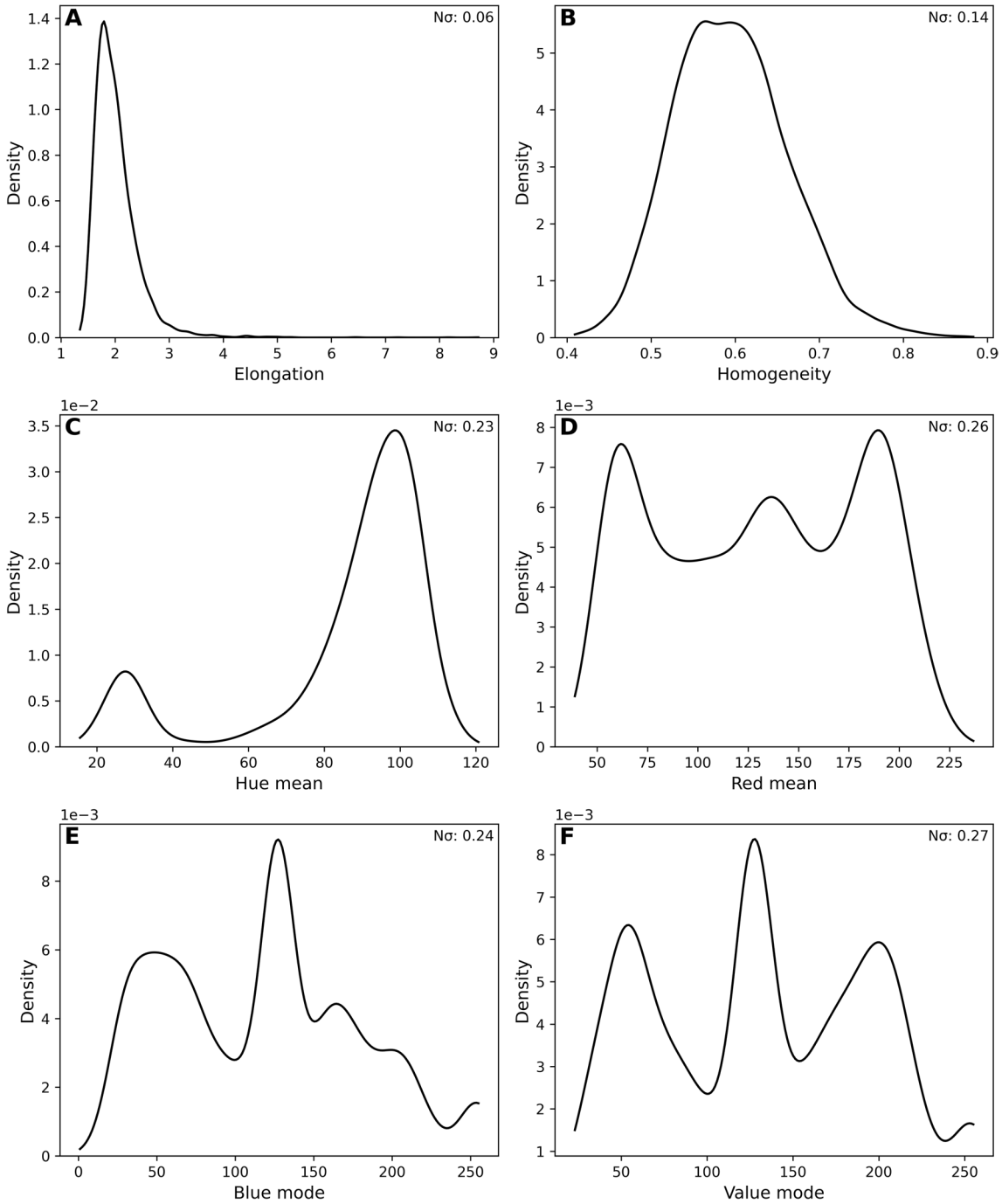

Fig. 11 Examples of density plots of six features for all ash particles in the database (6,304 in total). At the top right corner of each panel, the normalized standard deviation (Nσ) is shown which has been calculated after rescaling the feature values from 0 to 1. This allows for comparison of variability between different features (e.g., a value of 0.27 corresponds to 27% of relative standard deviation; see Methods section). Shape and texture features are generally unimodal with but a range in standard deviation as shown by difference between the narrower curve of *Elongation* (**A**), and the wider of *Homogeneity* (**B**). In contrast, color features show multiple modes and a much wider variability. **C** *Hue mean* shows a bimodal distribution, **D** *Red mean* is trimodal, **E** the distribution of the feature *Blue mode* has multiple modes, and (**F**) the *Value mode*, which relates to the intensity or luminance of a particle, has the largest variability of the dataset. Note that "Density", in the y-axis, is the "Probability Density" of the Kernel Density Estimate plot, which is a non-parametric way to estimate the probability density function of a continuous random variable and can take values above 1. Note that some density curves have been truncated at their extreme range of data points, as the function used for plotting (*seaborn.kdeplot*) passes a gaussian kernel that expands the curves beyond the actual range

certain features are characteristic of one or more of these subgroups. We found that some subgroups have distinctive distributions depending on the feature, as illustrated by the *Convexity* (shape), *Homogeneity* (texture), and *Value mean* (color).

(1) The *Convexity* values are similar (mode ~ 0.98) for the main particle types, except for the juvenile one, which has lower values (mode ~ 0.95 Fig. 12A). Filtering by activity type reveals that the lower *Convexity* values are particles produced by lava fountaining (Fig. 12B). Comparing between main particle types within the lava fountaining style, we found abundant juvenile particles with *Convexity* values < 0.9 (Fig. 12C), which are the most vesicular ones of the overall juvenile type (Fig. 12A).

(2) The *Homogeneity* values are similar across particle types, although free crystals have a greater variance (Fig. 12D). The *Homogeneity* values vary between activity types (Fig. 12E), from low values of the lava fountaining and phreatic eruptions (mode ~ 0.55) to higher values of the Plinian (mode ~ 0.72; Fig. 12E). The higher values can be explained by the abundance of pumice in the plinian ash, which has a similar and uniform appearance under the binocular microscope, whereas the low values of particles from lava fountaining can be explained by the scattered light reflections of glass shards. The phreatic samples show a low *Homogeneity* group corresponding to altered material (mode ~ 0.55; Fig. 12F), which possibly is the highly heterogeneous material we refer as hydrothermal aggregates, and is abundant in the samples of Ontake (2014) and Soufrière de Guadeloupe (1976–1977) eruptions. The higher values of *Homogeneity* correspond to free crystals, typically with well-defined crystallographically controlled surfaces, although their variance in the dataset is large (Fig. 12D), as they are often adhered to another component.

(3) The values of the feature *Value mean* (from the HSV space), which relates to the intensity of the color, shows three bimodalities and one trimodality depending on the particle type (Fig. 12G). The bimodality in free crystals reflects the dark (mode at ~ 100) and light (mode at ~ 220) appearance of pyroxene and plagioclase. Similarly, the bimodality of lithic particles may correspond to the presence of black (mode at ~ 60), unaltered lava fragments, typically from dome eruptions, versus lighter (mode at ~ 150) modified surfaces. The *Value mean*, when categorized by activity types, separates the juvenile component bimodality into lava fountaining (mode at ~ 70) and plinian (mode at ~ 205; Fig. 12H). This suggests that the feature effectively captured the significant contrast in brightness between the dark glass shards and the bright pumice-like particles. A value threshold around 125 of the *Value mean* could be used to differentiate between the two activity types. A closer look into the subplinian samples reveals that the juvenile component, with a high *Value mean* (mode ~ 200) due to its high brightness, could be almost fully discriminated from the other components by setting a threshold above 175 (Fig. 12I). However, it is worth noting that samples containing more plagioclase crystals, which often exhibit bright reflections, may overlap with the juvenile mode, and thus, a more complete dataset is required for drawing general conclusions.

Exploring feature contributions through PCA

We performed PCA of the particles' features to gain insights on the underlying dataset structure and identify the high-variance features that drive PCA. For each PCA, we obtained three principal components (PCs), and computed their explained variance and the feature with highest loading (for more detail see the methods section "Features' description and PCA"). The aggregated explained variance of the 3 PCs for the whole database is 71%, which means that a significant amount of the variance from the original features is captured by the 3 PCs. The most contributing feature (see Supplementary material 4 for all loadings and contributing features) to the PC1 is *Saturation standard dev* (Table 2 for the meaning and details of each feature; Table 6 for PCA results), followed by *Blue standard dev* for the PC2, and *Circ_Dellino* for the PC3. These findings suggest that the particle properties that vary more in the database are those sensitive to the variation of the pixel intensity of the color and the fine-scale roughness. The highest aggregated explained variance across activity types is 72% and corresponds to the plinian activity. This indicates that features capture broader variations of particle data (Table 6), which may reflect the presence of distinctive subpopulations (e.g., pumice-like particles; Fig. 12I). Aggregated explained variance differs across the rest of activity types including subplinian (70%), phreatic (69%), lava fountaining (66%) and dome explosion (65%), suggesting a more uniform distribution of the extracted features. To investigate the relationship between the explained variance, most contributing features, and their potential for classification we looked at the sample level and found the following:

(1) Two of our samples from plinian (*MS-DB1*, from Mount St. Helens) and subplinian activity (*KE-DB2* from Kelud) have the highest aggregated explained variance (72%; Table 6) and share the same three most

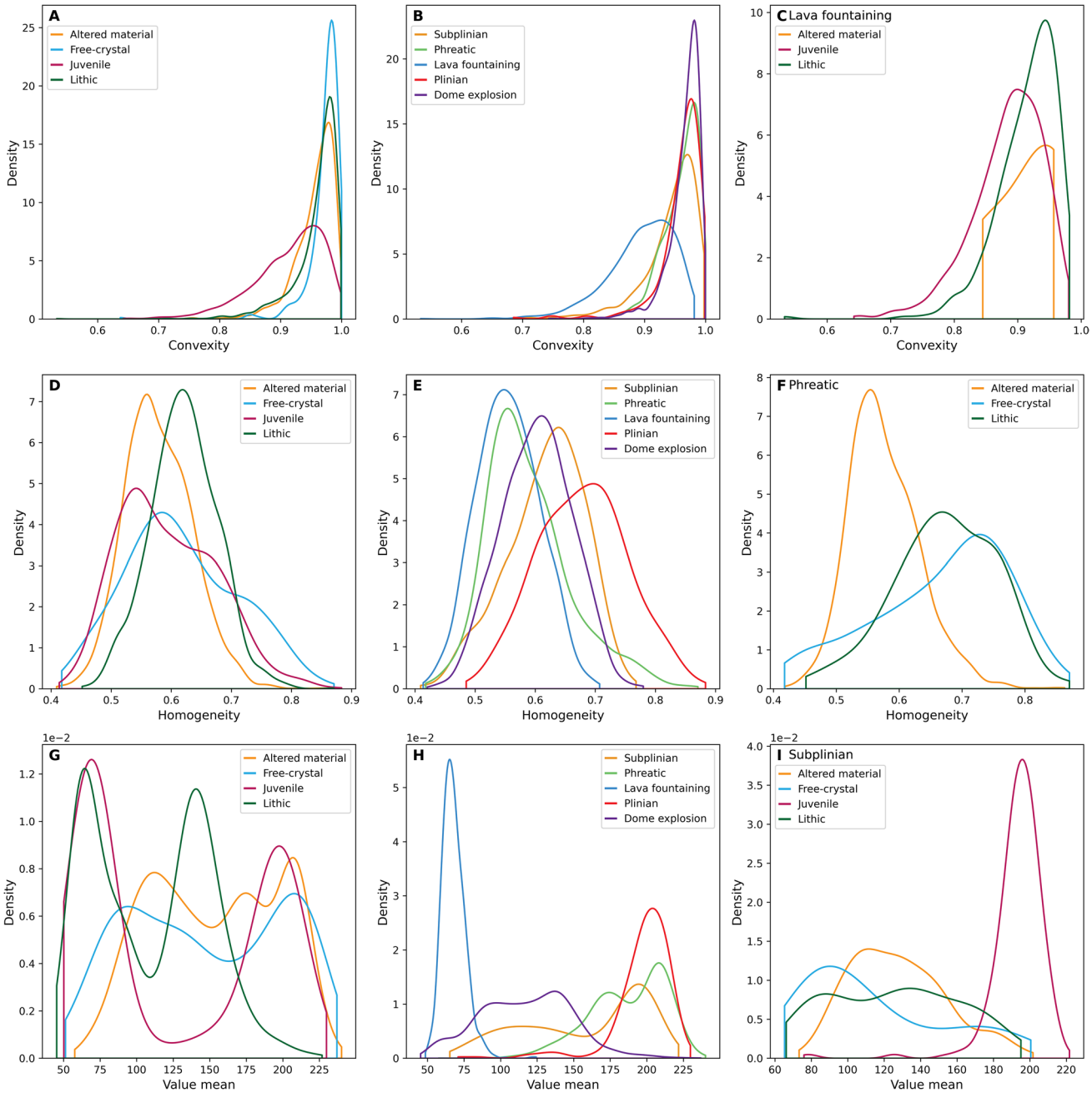

Fig. 12 Density plots of *Convexity* (**A**, **B** and **C**), *Homogeneity* (**D**, **E** and **F**) and *Value mean* (**G**, **H** and **I**) across particle types (**A**, **D**, **G**), eruptive styles (**B**, **E**, **H**), and both a given particle type and eruptive activity type (**C**, **F**, **I**). Note the increase in dispersion of the modes from top to bottom. Whereas the *Convexity* discriminates slightly one subgroup of particles (**A** and **B**), the *Value mean* (i.e., mean of the Value channel of the HSV space) can successfully separate between lava fountaining and plinian (**H**), and almost isolates juvenile particles in the subplinian samples (**I**). Note that some density curves have been truncated at their extreme range of data points, as the function used for plotting (*seaborn.kdeplot*) passes a gaussian kernel that expands the curves beyond the actual range

contributing features to the PCs: For the PC1, the *Green mean*, which is sensitive to particle color, for the PC2, the *Circ_Dellino*, and for the PC3, the *Saturation mean*, which is sensitive to the color intensity. The PC1 of *KE-DB2* retained almost half (48%) of the total variance of the sample. Representation of *KE-DB2* in the 3D PCs' space (Fig. 13A) reveals two distinct clusters along the PC1 cluster, one of which consisting almost entirely of juvenile particles.

(2) The four samples from dome explosions (from Merapi and Nevados de Chillán) have the lowest aggregated explained variance in the dataset: 62% for *ME-DB1*,

Table 6 Results of the principal component analysis for the entire VolcAshDB, and for the different eruptive activity types and samples

	PC1		PC2		PC3		
	Exp var (%)[1]	Top feature (type)	Exp var (%)	Top feature (type)	Exp var (%)	Top feature (type)	Aggr var (%)[2]
Database	40	*Saturation std* (C)	18	*Blue std* (C)	13	*Circ_Dellino* (S)	71
Activity types							
Plinian	40	*Green mean* (C)	20	*Circ_Dellino (S)*	13	*Saturation mean (C)*	72
Subplinian	41	*Green mean* (C)	20	*Circ_Cioni* (S)	8	*Red std* (C)	70
Phreatic	30	*Red mean* (C)	23	*Blue std* (C)	16	*Circ_Dellino* (S)	69
Lava fountaining	39	*Red std* (C)	15	*Blue mean* (C)	13	*Green mode* (C)	66
Dome explosion	27	*Green mean* (C)	23	*Blue std* (C)	15	*Circ_Dellino* (S)	65
Samples							
MS-DB1	40	*Green mean* (C)	20	*Circ_Dellino* (S)	13	*Saturation mean* (C)	72
KE-DB2	48	*Green mean* (C)	15	*Circ_Dellino* (S)	8	*Saturation mean* (C)	72
SG-DB1	39	*Green mean* (C)	15	*Circ_Dellino* (S)	14	*Saturation mean* (C)	68
PI-DB1	44	*Green mean* (C)	16	*Circ_Cioni* (S)	9	*Hue mode* (C)	68
KE-DB3	35	*Red mean* (C)	23	*Circ_Cioni* (S)	11	*Green std* (C)	68
SG-DB2	36	*Dissimilarity* (T)	20	*Circ_Cioni* (S)	11	*Saturation mean* (C)	68
CV-DB1	39	*Red std* (C)	15	*Blue mean* (C)	13	*Green mode* (C)	66
ON-DB1	29	*Homogeneity* (T)	20	*Blue mode* (C)	17	*Circ_Dellino* (S)	65
ME-DB2	25	*Blue mean* (C)	21	*Dissimilarity* (T)	18	*Hue std* (C)	65
NC-DB15	30	*Blue mode* (C)	21	*Value mean* (C)	13	*Circ_Dellino* (S)	65
NC-DB2	27	*Red mean* (C)	24	*Dissimilarity* (T)	13	*Circ_Cioni* (S)	64
ME-DB1	27	*Homogeneity* (T)	23	*Value mean* (C)	12	*Circ_Cioni* (S)	62

[1] "Exp var" is the % of the explained variance of each principal component

[2] "Aggr var" is the % of aggregated explained variance and is calculated as the sum of the explained variance of PC1, PC2 and PC3

64% for *NC-DB2*, and 65% for *NC-DB15* and *ME-DB2*. The PC1 of *ME-DB1* has relatively low explained variance (27%) and its most contributing feature is *Homogeneity*, followed by the PC2 (23%) dominated by *Circ_cioni*, and the PC3 (11%) dominated by the *Value mean*, a feature sensitive to the particle brightness (Table 6). Visualization of *ME-DB1* in the 3D PCs' space (Fig. 13B) reveals a decrease in lithics and increase in altered material along the PC1 but without these forming distinct clusters, and thus with limited potential for classification.

(3) PCA on the lava fountaining sample (*CV-DB1*) achieves 66% of aggregated explained variance. The PC1 retained 37% of explained variance and is driven by *Red standard dev*, a feature that measures the dispersion of the different pixel intensity values of the Red channel (from the RGB space). Graphical representation of *CV-DB1* shows that there is a transition from juvenile to lithic particle types as PC1 values decrease, although there is some overlap (Fig. 13C). The juvenile component consists of glass shards that often exhibit a wide range of white reflections (Fig. 10C), whereas the lithic component, with predominant recycled juvenile (LRJ) particles, has a duller or metallic luster under the binocular. Thus, it seems likely that the *Red standard dev* has effectively captured two distinctive subpopulations.

(4) Our four samples from phreatic activity from Ontake (*ON-DB1*), Soufrière de Guadeloupe (*SG-DB1* and *SG-DB2*) and Pinatubo (*PI-DB1)* range in aggregated explained variability between 65 and 68%. The most contributing features vary across samples. For example, PCA on *SG-DB2* uses the *Dissimilarity*, *Circ_cioni*, and *Saturation mean*, whereas PCA on *PI-DB1* leverages on *Green mean*, *Circ_cioni* and *Hue mode*. Visualization of *PI-DB1* in the 3D PCs' space reveals a transition in altered material and free crystals, where the latter take high and low values along the PC1, whereas the former take intermediate values (Fig. 13D). This suggests that *Green mean* captured the distinctive dark and bright colors from pyroxene and plagioclase crystals.

Our PCA highlights that high-variance features vary across samples, some of which show promise for classification. Some high-variance features capture shape, textural, and color properties of specific subpopulations depending on the sample and activity type, indicating that (1) a single

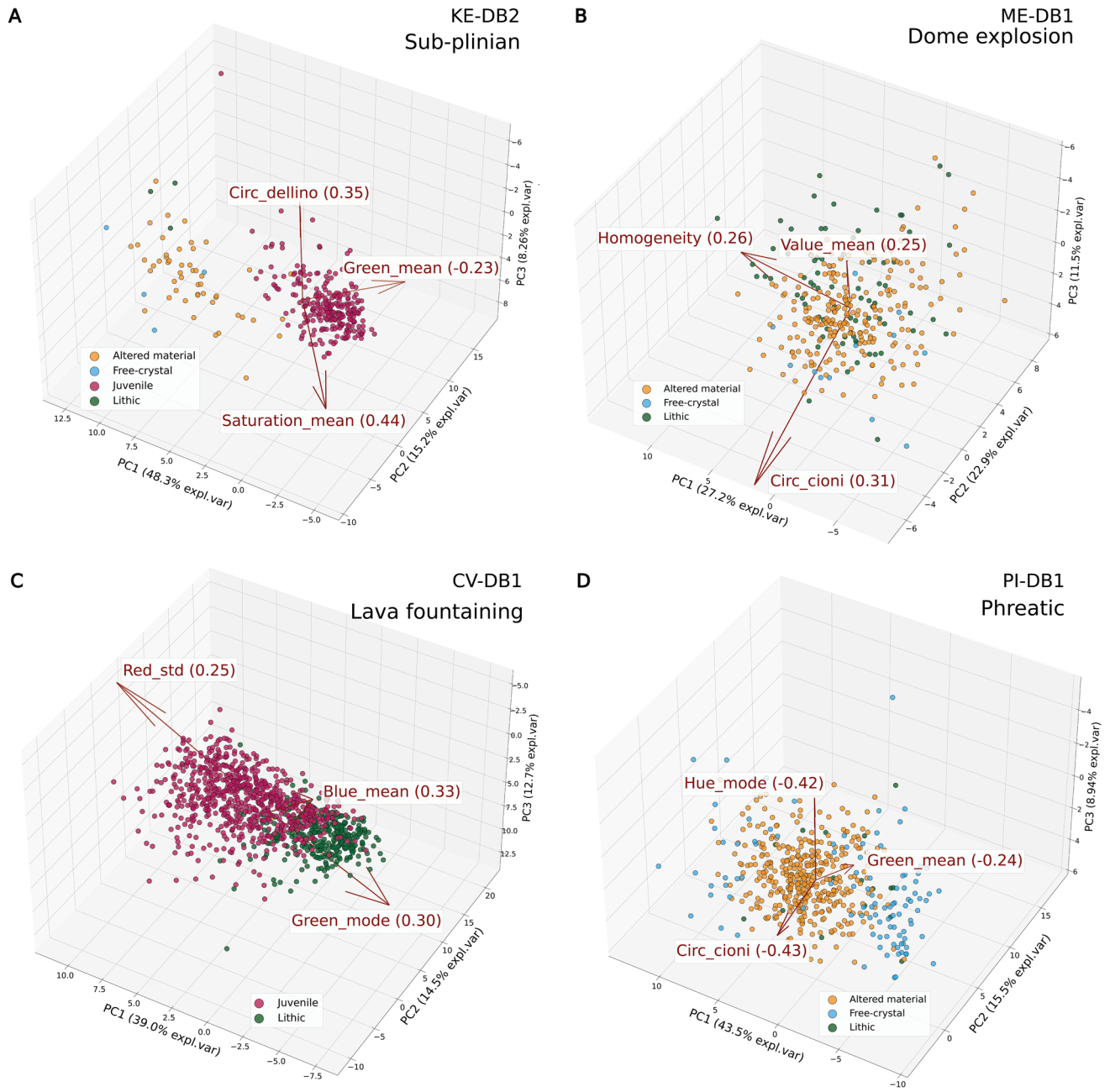

Fig. 13 Illustrations of the distribution of the transformed feature values from PCA in a 3-dimensional Principal Components' space (PC1, PC2, and PC3) for four samples. In the panels (**A**, **B**, **C** and **D**), red arrows are used to represent the contribution of the feature dominating each principal component. The angle of the arrow indicates the direction of the feature contribution (e.g., parallel alignment to PC1 axis would indicate a strong impact on PC1), whereas the arrow's length indicates the feature impact, as measured by the loading value (displayed in brackets). **A** The sample from Kelud (*KE-DB2*) has a distribution with two distinct clusters. The cluster with lower values of PC1 and higher values of PC2 consists of juvenile particles. **B** The sample from Merapi (*ME-DB1*) has the lowest aggregated explained variance and does not display a clear distinction between particle types in its distribution. **C** The sample from Cumbre Vieja (*CV-DB1*) shows a transition from lithic to juvenile particles as PC1 values increase. **D** The distribution of the preclimactic sample from Pinatubo (PI-DB1) is characterized by a spread along the PC1 axis. Altered material particles take intermediate PC1 values, whereas the free-crystals' values concentrate at the extremes. Note that the loading values and other features contributing to the PCs can be found in Supplementary material 4

classification criteria may not be possible, and (2) that information extracted from binocular images, with their inclusion of color and texture in addition to the shape, can be a valuable tool to separate between the different types (Yamanoi et al. 2008; Miwa et al. 2015). Note that we used PCA as an exploratory tool. To assess the full potential of the features to classify particles, machine learning models have been proved to perform best, as they account for non-linearities and are not variance-driven (Verdhan 2020).

VolcAshDB Web platform

VolcAshDB is an open-access, web-based platform that hosts the curated dataset of the high-resolution, multi-focused images that we have discussed hereto. Each image is linked to: (1) a summary label of the main type, subtype and some of the special characteristics, (2) the measured physical features of shape, color and textures, and (3) the metadata such as the image magnification, the grain-size or the sample collector. Users can browse through the whole image dataset, or use filters to only visualize particles according to their type, activity type, or volcanoes. The images, their classification and the 33 measured features can be downloaded from the web site in various file formats at https://volcash.wovodat.org/database/catalogue. We have also created an app for data visualization at https://volcash.wovodat.org/analytic. Users can visualize basic relationships of the features, as well as their distributions depending on the activity type, particle type, amongst others. The graphs are interactive to allow users choosing a specific volcano, activity type, sample and feature for display.

The database content of the platform is stored in a server, using the database manager MongoDB, as it is cost-effective, flexible and can handle many data types. The server infrastructure to receive and process the browser's requests is located under WOVOdat (Newhall et al. 2017), which is a comprehensive global database on volcanic unrest (https://www.wovodat.org/). The backend uses several technologies, including JavaScript Object Notation (JSON), which holds the database, and the open-source libraries Node.js and Flask to execute tasks, such as opening a file on the computer's file system. The frontend, where the user interacts with the app, uses the open-source JavaScript library React.

Discussion

Limitations of the current database

VolcAshDB contains data for about > 6,300 particles from 12 samples and five activity types, for which we obtained the main types and some sub-types proportions, special characteristics, and a list of quantitatively measured features. However, the petrologic classification of each particle has been conducted by only one observer, and hence the classification could be biased, although we used diagnostic observations from the literature as a basis for classification (Fig. 6). To improve this in the future, classification should result from the aggregated knowledge and experience of various experts in the field. This could be accomplished via workshops and publications where several researchers classify the same particles. In addition, expert elicitation (Aspinall and Blong 2015) would allow to treat the problem in probabilistic terms. This approach has been successfully done in other volcanological studies dealing with highly uncertain situations. Another limitation is that each particle has been classified as belonging 100% to a given type, which implies 100% certainty. A more robust classification could include a percentage of a given particle to belong to a given class without these being mutually exclusive. For instance, if a particle exhibits four out of five fresh-like features, and the weights are equally distributed, the particle could be assigned 80% of probability of being juvenile. A third limitation of the database is that the componentry per eruption is obtained from one grain-size fraction collected at one sampling site. More representative componentry should account for various grain-size fractions and samples collected at strategic distances from the vent (Eychenne et al. 2013). Other limitations of the database are the range of activity types and magma compositions. We currently have not yet incorporated ash particles from vulcanian or strombolian eruptions, or from phreatomagmatic events driven by water-magma interactions. In terms of magma compositions, we are also lacking andesites. A future goal of the database is to make it more complete by incorporating data from our own samples, but also to make the platform open for any user to upload ash image samples that would be classified into the different types so that the database could grow by the community as it is the case for WOVOdat (Costa et al. 2019). If the uploaded images are multi-focused, high-resolution (e.g., above 1,500 pixels per millimeter (pxls/mm), and taken under the same white plate, these could be normalized (as explained in section "Image processing") and added in our curated multi-focused dataset. Binocular images consisting in the "standard" single-focus binocular images, cross-sectional and external particle SEM images will be maintained in a separate repository, and their use for statistical analysis will require additional curation efforts.

Applications to comparative studies

Nothwithstanding the limitations highlighted above, our results show that the ash samples from different volcanoes and activity types have distinctive particle main type proportions. This means that we can make meaningful comparisons

between samples and quantify their similarity at specific variables (e.g., % of juvenile content). It is possible to compare between eruptions by their proportion of juvenile, altered material, and lithic components and visualize their distributions using a ternary diagram (Fig. 14). Such ternary diagram shows that samples from activity that does not directly involve magma (phreatic) or only to a limited extent (dome explosions) plot closer (along the isoline of 0% of juvenile content), than those from lava fountaining, subplinian, and plinian samples (between the isolines 20–100% of juvenile content). By incorporating different grain-sizes, more samples per eruption at different distances from the vent, and including Strombolian and Vulcanian eruptive styles, particles features, such as shape, texture and color, could be used to apply statistical clustering and quantify their similarity in the context of analogue volcanoes. Users could plot the proportions of the ash samples that they are studying and compare with the eruptions we have produced in the database. It should be also possible to plot a time series of ash samples that visually shows how the volcano may be going from a mainly phreatic phase towards a magmatic eruption as for instance reported in ashes from Shinmoe-dake eruption in 2011 (Suzuki et al. 2013).

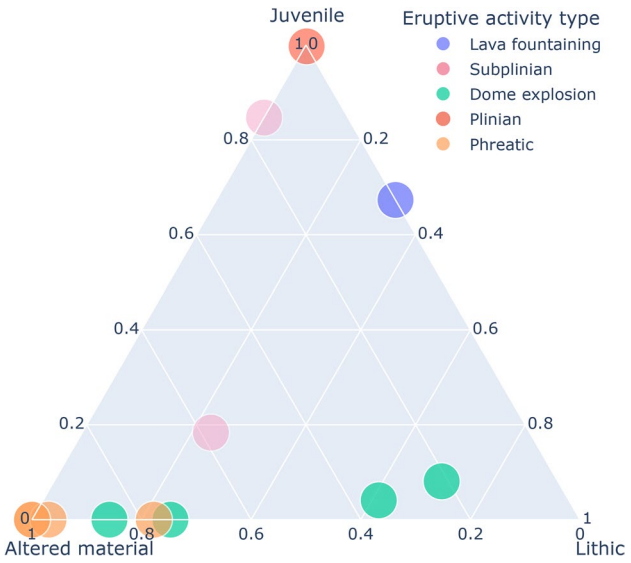

Fig. 14 Ternary diagram of the particles' proportions normalized to 1 excluding free crystals, and separated by eruptive activity types. Samples from activity types with limited or absent involvement of magma are near the isoline Juvenile 0%, whereas magmatic eruptions transition towards the juvenile apex. This diagram could be used in comparative studies to identify whether certain areas are characteristic to a given eruptive activity type, which could be used for petrologic monitoring (Suzuki et al. 2013). Reported componentry data belongs to the 0ϕ–1ϕ (1000–500 μm) grain-size fraction, except for Kelud (Subplinian) samples that belong to 1ϕ–2ϕ (500–250 μm). As noted in the main text, more representative componentry per eruption should include various grain-size fractions and collection of samples at various distances from the vent

Our expectation is that external datasets will be imported to VolcAshDB to examine their relationship with the ash particle features we have compiled qualitatively (e.g., plots) or quantitatively by using statistical tools, e.g., dendrograms (Dürig et al. 2020). Alternatively, users can download locally the datasets currently in VolcAshDB and compare on their own environment. If such datasets are acquired with a similar methodological protocol and standardized statistical tests (Dürig et al. 2021), the data from VolcAshDB can help researchers with classifying their own particles from any given volcano and eruption they are studying.

Applications for automatic classification and machine learning

The particle features we extracted capture specific values of subpopulations that may be associated with the particle main types, as shown by multimodality in the histograms, and by the formation of distinctive clusters in the 3D principal components' space. These features could be used to train machine learning (ML) models, such as decision trees, random forest and boosted trees (Benet et al. 2024) commonly used for multi-class classification via supervised learning. The images of particles could also be used for training models such as deep neural networks or Vision Transformers (Benet et al. 2024). Developing accurate ML models would allow automatic particle classification that would be standardized and in principle independent of the observer, and thus the ash componentry between different eruptions can be compared in an objective and reproducible manner. Object classification via image observations is a problem that has been addressed by applying machine learning in several fields, e.g., plants based on their leaves (Aakif and Khan 2015), variable stars (Hosenie et al. 2019), and also in volcanic ash for classifying into their shapes (Shoji et al. 2018). These often require a large amount of standardized and widely representative data about particle features, images, or even chemical composition and the database that we have started could be the basis for such an endeavor.

Conclusions

Volcanic ash provides critical insights to the state of a volcano, but requires classifying the particles into different types and there is no standardized methodology to do so. In this contribution we developed a standardized methodology to obtain high-quality images of a large number of particles and classified the particles following the diagnostic observations from previous studies. We additionally measured a range of physical parameters from the particle images related to their shape, texture, and color. The whole dataset includes 6,304 ash particles from a range of activity

types which constitutes VolcAshDB, a public web-based platform that allows users to browse and download various data types. We believe such platform and dataset are useful in comparative studies and the basis for Machine Learning algorithms towards a systematic ash particle classification.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00445-023-01695-4.

Acknowledgements D.B. is very grateful to Charles Tran, Khai Truong and Bui Quang for their invaluable help on the development of the VolcAshDB web-app and the MongoDB collection, to Edwin Tan for support on the Gekko cluster, to Caroline Bouvet de Maisonneuve for insightful discussions, and to Vanesa Burgos for help on the PCA. The manuscript has greatly benefited from thorough revisions by Hannah Buckland, an anonymous reviewer, and the editor Julia Eychenne. This research was supported by the Earth Observatory of Singapore via its funding from the National Research Foundation Singapore and the Singapore Ministry of Education under the Research Centres of Excellence initiative. F.C. acknowledges also support by a Chaire d'Excellence grant by the Université Paris Cité.

Authors contributions Benet led the development, acquired and processed the data, and wrote the original manuscript of the project, which was supervised and revised by Costa. Widiwijayanti and Benet led the development of front-end and the back-end of the web-platform. Pallister, Pedreros, Allard, Humaida, Aoki and Maeno provided some samples. The final version of the manuscript has been read, reviewed and approved by all co-authors.

Data availability The curated dataset of particle features described in this paper are available at https://volcash.wovodat.org/catalogue.

Code availability The Python program for feature extraction and for PCA, as well as the Java Script code for the front-end of the web are publicly accessible at the github repository https://github.com/dbenet-ntu/Volcash-Project. The Python code was developed using its version 3.9. Data wrangling and analysis was done using Numpy v.1.21.6, Pandas v. 1.3.5, image processing using Scikit-image v.0.19.2 and opencv v.3.4.2, and plotting by matplotlib v.3.5.1 and seaborn v.0.11.2.

Declarations

Competing interest The authors declare no competing interests.

References

Aakif A, Khan MF (2015) Automatic classification of plants based on their leaves. Biosys Eng 139:66–75. https://doi.org/10.1016/j.biosystemseng.2015.08.003

Aiuppa A (2015) Volcanic-gas monitoring. Cambridge University Press, pp 81–96

Alvarado GE, Mele D, Dellino P, de Moor JM, Avard G (2016) Are the ashes from the latest eruptions (2010–2016) at Turrialba volcano (Costa Rica) related to phreatic or phreatomagmatic events? J Volcanol Geoth Res 327:407–415. https://doi.org/10.1016/j.jvolgeores.2016.09.003

Andronico D, Lo Castro MD, Sciotto M, Spina L (2013) The 2010 ash emissions at the summit craters of Mt Etna: Relationship with seismo-acoustic signals. J Geophys Res: Solid Earth 118(1):51–70. https://doi.org/10.1029/2012JB009895

Andronico D, Scollo S, Lo Castro MD, Cristaldi A, Lodato L, Taddeucci J (2014) Eruption dynamics and tephra dispersal from the 24 November 2006 paroxysm at South-East Crater, Mt Etna, Italy. J Volcanol Geotherm Res 274(November 2006):78–91. https://doi.org/10.1016/j.jvolgeores.2014.01.009

Angkasa SS, Ohba T, Imura T, Setiawan I, Rosana MF (2019) Tephrastratigraphy and ash componentry studies of proximal volcanic products at mount Tangkuban Parahu, Indonesia: An insight to holocene volcanic activity. Indones J Geosci 6(3):235–253. https://doi.org/10.17014/ijog.6.3.235-253

Aspinall W, Blong R (2015) Volcanic Risk Assessment. In The Encyclopedia of Volcanoes (Second Edi). Elsevier Inc.https://doi.org/10.1016/B978-0-12-385938-9.00070-5

Aspinall W, Cooke RM (1998) Expert Judgement and the Montserrat Volcano Eruption. Proc 4th Int Conf Probabilistic Safety Assess Manag PSAM4 3:2113–2118

Bardelli F, Giuli G, Di BF et al (2020) Spectroscopic study of volcanic ashes. J Hazard Mater 400:123213. https://doi.org/10.1016/j.jhazmat.2020.123213

Battaglia J, Hidalgo S, Bernard B, Steele A, Arellano S, Acuña K (2019) Autopsy of an eruptive phase of Tungurahua volcano (Ecuador) through coupling of seismo-acoustic and SO 2 recordings with ash characteristics. Earth Planet Sci Lett 511:223–232. https://doi.org/10.1016/j.epsl.2019.01.042

Bebbington MS, Jenkins SF (2019) Intra-eruption forecasting. Bull Volcanol 81:1–7

Benet D, Costa F, Pedreros G, Cardona C (2021) The volcanic ash record of shallow magma intrusion and dome emplacement at Nevados de Chillán Volcanic complex, Chile. J Volcanol Geotherm Res 417. https://doi.org/10.1016/j.jvolgeores.2021.107308

Benet D, Costa F, Widiwijayanti C (2024) Volcanic ash classification through Machine Learning. Geochemistry, Geophysics, Geosystems. https://doi.org/10.1029/2023GC011224

Bernard A, Knittel U, Weber B, Weis D, Albrecht A, Hattori K, Klein J, Oles D (1996) Petrology and geochemistry of the 1991 eruption products of Mount Pinatubo. In: Fire and mud: eruptions and lahars of Mount Pinatubo, Philippines, pp 767–797

Buckland HM, Eychenne J, Rust AC, Cashman KV (2018) Relating the physical properties of volcanic rocks to the characteristics of ash generated by experimental abrasion. J Volcanol Geoth Res 349:335–350. https://doi.org/10.1016/j.jvolgeores.2017.11.017

Cannata CB, Rosa RD, Donato P, Taddeucci J (2014) Ash Features from Ordinary Activity at Stromboli Volcano. Int J Geosci 05(11):1361–1382. https://doi.org/10.4236/ijg.2014.511111

Carey SN, Sigurdsson H (1982) Influence of particle aggregation on deposition of distal tephra from the May 18, 1980, eruption of Mount St. Helens volcano. J Geophys Res 87:7061–7072

Cashman KV, Hoblitt RP (2004) Magmatic precursors to the 18 May 1980 eruption of Mount St. Helens, USA. Geology 32(2):141–144. https://doi.org/10.1130/G20078.1

Chouet B (2003) Volcano seismology. Pure Appl Geophys 160(3–4):739–788. https://doi.org/10.1007/PL00012556

Cioni R, Pistolesi M, Bertagnini A, Bonadonna C, Hoskuldsson A, Scateni B (2014) Insights into the dynamics and evolution of the 2010 Eyjafjallajökull summit eruption (Iceland) provided by volcanic ash textures. Earth Planetary Sci Lett 394(May 2010):111–123. https://doi.org/10.1016/j.epsl.2014.02.051

Cioni R, Sbrana A, Vecci R (1992) Morphologic features of juvenile pyroclasts from magmatic and phreatomagmatic deposits of Vesuvius. J Volcanol Geotherm Res 51(1–2):61–78

Comida PP, Ross P-S, Zimanowski B, Büttner R, Dürig T (2023) Controls on juvenile ash morphologies in lava fountains: insights from laboratory experiments. Bull Volcanol 85(4):1–24. https://doi.org/10.1007/s00445-023-01637-0

Comida PP, Ross PS, Dürig T, White JDL, Lefebvre N (2022) Standardized analysis of juvenile pyroclasts in comparative studies of primary magma fragmentation: 2. Choice of size fraction and method optimization for particle cross-sections. Bull Volcanol 84(1):1–24. https://doi.org/10.1007/s00445-021-01517-5

Costa F, Widiwijayanti C, Nang TZW, Fajiculay E, Espinosa-Ortega T, Newhall C (2019) WOVOdat – the global volcano unrest database aimed at improving eruption forecasts. Disaster Prev Manag: Int J 28(6):738–751. https://doi.org/10.1108/DPM-09-2019-0301

D'Oriano C, Bertagnini A, Cioni R, Pompilio M (2014) Identifying recycled ash in basaltic eruptions. Sci Rep 4. https://doi.org/10.1038/srep05851

D'Oriano C, Bertagnini A, Pompilio M (2011) Ash erupted during normal activity at Stromboli (Aeolian Islands, Italy) raises questions on how the feeding system works. Bull Volcanol 73(5):471–477. https://doi.org/10.1007/s00445-010-0425-0

D'Oriano C, Del Carlo P, Andronico D, Cioni R, Gabellini P, Cristaldi A, Pompilio M (2022). Syn-eruptive processes during the January–February 2019 ash-rich emissions cycle at Mt. Etna (Italy): implications for petrological monitoring of volcanic ash. Front Earth Sci 10(February 2019). https://doi.org/10.3389/feart.2022.824872

D'Oriano C, Poggianti E, Bertagnini A, Cioni R, Landi P, Polacci M, Rosi M (2005) Changes in eruptive style during the A.D. 1538 Monte Nuovo eruption (Phlegrean Fields, Italy): The role of syn-eruptive crystallization. Bull Volcanol 67(7):601–621. https://doi.org/10.1007/s00445-004-0397-z

Dellino P, La Volpe L (1996) Image processing analysis in reconstructing fragmentation and transportation mechanisms of pyroclastic deposits. The case of Monte Pilato-Rocche Rosse eruptions, Lipari (Aeolian islands, Italy). J Volcanol Geotherm Res 71(1):13–29. https://doi.org/10.1016/0377-0273(95)00062-3

Dellino P, Volpe LL (1995) Fragmentation versus transportation mechanisms in the pyroclastic sequence of Monte Pilato-Rocche Rosse (Lipari, Italy). J Volcanol Geoth Res 64(3–4):211–231. https://doi.org/10.1016/0377-0273(94)00084-T

Dixon HJ, Murphy MD, Sparks SJ, Chávez R, Naranjo JA, Dunkley PN, Young SR, Gilbert JS, Pringle MR (1999) The geology of Nevados de Chillán volcano, Chile. Rev Geol Chile 26(2):227–253

Doyle EEH, McClure J, Paton D, Johnston DM (2014) Uncertainty and decision making: Volcanic crisis scenarios. Int J Disaster Risk Reduction 10(PA):75–101. https://doi.org/10.1016/j.ijdrr.2014.07.006

Druitt TH (2014) New insights into the initiation and venting of the Bronze-Age eruption of Santorini (Greece), from component analysis. Bull Volcanol 76(2):1–21. https://doi.org/10.1007/s00445-014-0794-x

Dürig T, Bowman MH, White JDL, Murch A, Mele D, Verolino A, Dellino P (2018) Particle shape analyzer Partisan - An open source tool for multi-standard two-dimensional particle morphometry analysis. Ann Geophys 61(6). https://doi.org/10.4401/ag-7865

Dürig T, Mele D, Dellino P, Zimanowski B (2012) Comparative analyses of glass fragments from brittle fracture experiments and volcanic ash particles. Bull Volcanol 74(3):691–704. https://doi.org/10.1007/s00445-011-0562-0

Dürig T, Ross PS, Dellino P, White JDL, Mele D, Comida PP (2021) A review of statistical tools for morphometric analysis of juvenile pyroclasts. Bull Volcanol 83(11). https://doi.org/10.1007/s00445-021-01500-0

Dürig T, Schmidt LS, White JDL, Bowman MH (2020) DendroScan: an open source tool to conduct comparative statistical tests and dendrogrammatic analyses on particle morphometry. Sci Rep 10(1):1–14. https://doi.org/10.1038/s41598-020-78698-0

Dzurisin D (2006) Volcano deformation: new geodetic monitoring techniques. Springer Science & Business Media

Edwards MJ, Eychenne J, Pioli L (2021) Formation and dispersal of ash at open conduit basaltic volcanoes: lessons from Etna. Front Earth Sci 9:709657

Ersoy O (2010) Surface area and volume measurements of volcanic ash particles by SEM stereoscopic imaging. J Volcanol Geoth Res 190(3–4):290–296. https://doi.org/10.1016/j.jvolgeores.2009.12.006

Ersoy O, Chinga G, Aydar E, Gourgaud A, Evren Cubukcu H, Ulusoy I (2006) Texture discrimination of volcanic ashes from different fragmentation mechanisms: A case study, Mount Nemrut stratovolcano, eastern Turkey. Comput Geosci 32(7):936–946. https://doi.org/10.1016/j.cageo.2005.10.013

Eychenne J, Le Pennec JL, Ramón P, Yepes H (2013) Dynamics of explosive paroxysms at open-vent andesitic systems: High-resolution mass distribution analyses of the 2006 Tungurahua fall deposit (Ecuador). Earth Planet Sci Lett 361:342–355. https://doi.org/10.1016/j.epsl.2012.11.002

Eychenne J, Houghton BF, Swanson DA, Carey RJ, Swavely L (2015) Dynamics of an open basaltic magma system: The 2008 activity of the Halema'uma'u Overlook vent, Kīlauea Caldera. Earth Planet Sci Lett 409:49–60. https://doi.org/10.1016/j.epsl.2014.10.045

Feuillard M, Allegre CJ, Brandeis G, Gaulon R, Le Mouel J, Mercier JC, Pozzi JP, Semet M (1983) The 1975–1977 crisis of la Soufrière de Guadeloupe (F.W.I): A still-born magmatic eruption. J Volcanol Geoth Res 16:317–334

Fisher RV, Schmincke HU (1984) Pyroclastic Rocks. Springer, Berlin, Heidelberg

Freire S, Florczyk AJ, Pesaresi M, Sliuzas R (2019) An improved global analysis of population distribution in proximity to active volcanoes, 1975–2015. ISPRS Int J Geo-Inf 8(8):341

Gaunt HE, Bernard B, Hidalgo S, Proaño A, Wright H, Mothes P, Criollo E, Kueppers U (2016) Juvenile magma recognition and eruptive dynamics inferred from the analysis of ash time series: The 2015 reawakening of Cotopaxi volcano. J Volcanol Geoth Res 328:134–146. https://doi.org/10.1016/j.jvolgeores.2016.10.013

Gertisser R, Marmol M, Newhall C, Preece K, Charbonnier S, Andreastuti S, Handley H, Keller J (2023) Geological History, Chronology and Magmatic Evolution of Merapi. In: Gertisser R, Troll VR, Walter TR, Nandaka IGMA, Ratdomopurbo A (eds) Merapi Volcano. Springer, pp 137–193

Geshi N, Iguchi M, Shinohara H (2016) Phreatomagmatic eruptions of 2014 and 2015 in Kuchinoerabujima Volcano triggered by a shallow intrusion of magma. J Nat Dis Sci 37(2):67–78

Gómez-Arango JA, Murcia H, Borrero C (2018) Finding eruptive mechanisms through pyroclasts from the current eruption (1984 – present) at Nevado del Ruiz Volcanic Complex, Colombia. J Volcanol Geotherm Res. https://doi.org/10.1016/j.jvolgeores.2018.09.003

Gonnermann HM (2015) Magma fragmentation. Annu Rev Earth Planet Sci 43(1):431–458. https://doi.org/10.1146/annurev-earth-060614-105206

Gorbach NV, Plechova AA, Manevich TM, Portnyagin MV, Philosofova TM, Samoilenko SB (2018) The composition of volcanic ash and the dynamics of the 2013–2016 Zhupanovsky volcano eruption. J Volcanol Seismolog 12(3):155–171. https://doi.org/10.1134/S0742046318030028

Gunawan H, Surono Budianto A, Kristianto Prambada O, McCausland W, Pallister J, Iguchi M (2019) Overview of the eruptions of Sinabung Volcano, 2010 and 2013–present and details of the 2013 phreatomagmatic phase. J Volcanol Geotherm Res 382:103–119. https://doi.org/10.1016/j.jvolgeores.2017.08.005

Gurioli L, Andronico D, Bachelery P, Balcone-Boissard H, Battaglia J, Boudon G, Burgisser A, Burton MR, Cashman K, Cichy S, Cioni R, Di Muro A, Dominguez L, D'Oriano C, Druitt T, Harris AJL, Hort M, Kelfoun K, Komorowski JC, Thordarson T (2015) MeMoVolc consensual document: a review of cross-disciplinary approaches to characterizing small explosive magmatic eruptions. Bull Volcanol 77(6). https://doi.org/10.1007/s00445-015-0935-x

Hall-Beyer M (2017) GLCM texture: a tutorial. 17th Int Symp Ballistics 2(March):18–19

Haralick RM, Dinstein I, Shanmugam K (1973) Textural features for image classification. IEEE Trans Syst, Man Cybern SMC-3(6):610–621. https://doi.org/10.1109/TSMC.1973.4309314

Hastie T, Tibshirani R, Friedman JH (2009) The elements of statistical learning: data mining, inference, and prediction., Second edi. Springer

Heiken G, Crowe B, McGetchin T, West F, Eichelberger J, Bartram D, Peterson R, Wohletz K (1980) Phreatic eruption clouds: the activity of La Soufrière de Guadeloupe, FWI, August—October, 1976. Bull Volcanol 43:383–395

Heiken G, Wohletz K (1985) Volcanic ash. University Presses of California, Chicago, Harvard & MIT

Hincks TK, Komorowski JC, Sparks SR, Aspinall WP (2014) Retrospective analysis of uncertain eruption precursors at La Soufrière volcano, Guadeloupe, 1975–77: Volcanic hazard assessment using a Bayesian Belief Network approach. J Appl Volcanol 3(1). https://doi.org/10.1186/2191-5040-3-3

Hornby AJ, Lavallée Y, Kendrick JE, et al (2018) Volcanic ash generation mechanisms : Fingerprints in phase distribution highlighted by QEMSCAN particle mineralogical analysis. Earth Arxiv 1–22. https://doi.org/10.17605/OSF.IO/XMV9B

Hosenie Z, Lyon RJ, Stappers BW, Mootoovaloo A (2019) Comparing multiclass, binary, and hierarchical machine learning classification schemes for variable stars. Mon Not R Astron Soc 488(4):4858–4872. https://doi.org/10.1093/mnras/stz1999

Hughes I, Hase T (2010) Measurements and their uncertainties: a practical guide to modern error analysis. OUP Oxford

Ibraheem NA, Hasan MM, Khan RZ, Mishra PK (2012) Understanding color models : a review. ARPN J Sci Technol 2(3):265–275

Khan S (2018) An effective framework for driver fatigue recognition based on intelligent facial expressions analysis. IEEE Access 6:67459–67468. https://doi.org/10.1109/ACCESS.2018.2878601

Kurniawan IA, Sakakibara M, Suparka E (2017) Petrological studies of volcanic ash from Sakurajima volcano in 2013, Southern Kyushu, Japan. IOP Conf Ser: Earth Environ Sci 71(1). https://doi.org/10.1088/1755-1315/71/1/012008

Latif G, Bouchard K, Maitre J, Back A, Bédard LP (2022) Deep-learning-based automatic mineral grain segmentation and recognition. Minerals 12(4). https://doi.org/10.3390/min12040455

Le Guern F, Bernard A, Chevrier RM (1980) Soufrière of guadeloupe 1976–1977 eruption — mass and energy transfer and volcanic health hazards. Bull Volcanol 43(3):577–593. https://doi.org/10.1007/BF02597694

Leibrandt S, Le Pennec JL (2015) Towards fast and routine analyses of volcanic ash morphometry for eruption surveillance applications. J Volcanol Geoth Res 297:11–27. https://doi.org/10.1016/j.jvolgeores.2015.03.014

Li L, Iskander M (2022) Use of machine learning for classification of sand particles. Acta Geotech 0123456789. https://doi.org/10.1007/s11440-021-01443-y

Liu EJ, Cashman KV, Rust AC (2015) Optimising shape analysis to quantify volcanic ash morphology. GeoResJ 8:14–30. https://doi.org/10.1016/j.grj.2015.09.001

Liu EJ, Cashman KV, Rust AC, Höskuldsson A (2017) Contrasting mechanisms of magma fragmentation during coeval magmatic and hydromagmatic activity: the Hverfjall Fires fissure eruption, Iceland. Bull Volcanol 79(10). https://doi.org/10.1007/s00445-017-1150-8

Lücke OH, Calderón A (2016) Characterization of the ashes from the 2014-2015 Turrialba volcano eruptions by means of scanning electron microscopy and energy dispersive X-ray spectroscopy. Rev Geol Am Central 54:109–123. https://doi.org/10.15517/rgac.v54i0.23281

Ma B, Zhu Y, Yin X, Ban X, Huang H, Mukeshimana M (2021) SESF-Fuse: an unsupervised deep model for multi-focus image fusion. Neural Comput Appl 33(11):5793–5804. https://doi.org/10.1007/s00521-020-05358-9

Maeno F, Nakada S, Yoshimoto M, Shimano T, Hokanishi N, Zaennudin A, Iguchi M (2019) A sequence of a plinian eruption preceded by dome destruction at Kelud volcano, Indonesia, on February 13, 2014, revealed from tephra fallout and pyroclastic density current deposits. J Volcanol Geoth Res 382:24–41. https://doi.org/10.1016/j.jvolgeores.2017.03.002

Maitre J, Bouchard K, Bédard LP (2019) Mineral grains recognition using computer vision and machine learning. Comput Geosci 130:84–93. https://doi.org/10.1016/j.cageo.2019.05.009

Marzocchi W, Bebbington MS (2012) Probabilistic eruption forecasting at short and long time scales. Bull Volcanol 74(8):1777–1805. https://doi.org/10.1007/s00445-012-0633-x

Matsumoto K, Geshi N (2021) Shallow crystallization of eruptive magma inferred from volcanic ash microtextures: a case study of the 2018 eruption of Shinmoedake volcano. Jpn Bull Volcanol 83(5):1–14. https://doi.org/10.1007/s00445-021-01451-6

Mehbodniya A, Lazar AJP, Webber J, Sharma DK, Jayagopalan S, Kousalya K, Singh P, Rajan R, Pandya S, Sengan S (2022) Fetal health classification from cardiotocographic data using machine learning. Expert Syst 39(6):1–13. https://doi.org/10.1111/exsy.12899

Mendenhall W, Beaver RJ, Beaver BM (2012) Introduction to probability and statistics. Cengage Learning

Metcalfe A, Moune S, Komorowski J, Kilgour G (2021) Magmatic processes at La Soufrière de Guadeloupe: insights from crystal studies and diffusion timescales for eruption onset. 9(April):1–28https://doi.org/10.3389/feart.2021.617294

Minami Y, Imura T, Hayashi S, Ohba T (2016) Mineralogical study on volcanic ash of the eruption on September 27, 2014 at Ontake volcano, central Japan: Correlation with porphyry copper systems the Phreatic Eruption of Mt. Ontake Volcano in 2014. Volcanology. Earth, Planets Space 68(1). https://doi.org/10.1186/s40623-016-0440-2

Minami Y, Matsumoto K, Geshi N, Shinohara H (2022) Influence of hydrothermal recharge on the evolution of eruption styles and hazards during the 2018–2019 activity at Kuchinoerabujima Volcano, Japan. Earth, Planets Space 74(1). https://doi.org/10.1186/s40623-022-01580-y

Miwa T, Geshi N, Shinohara H (2013) Temporal variation in volcanic ash texture during a vulcanian eruption at the sakurajima volcano, Japan. J Volcanol Geoth Res 260:80–89. https://doi.org/10.1016/j.jvolgeores.2013.05.010

Miwa T, Geshi N, Tanada T, Iguchi M (2021) Automatic onsite imaging of volcanic ash particles with VOLCAT : Towards quasi-real-time eruption style monitoring. J Volcanol Geoth Res 416:107267. https://doi.org/10.1016/j.jvolgeores.2021.107267

Miwa T, Shimano T, Nishimura T (2015) Characterization of the luminance and shape of ash particles at Sakurajima volcano, Japan, using CCD camera images. Bull Volcanol 77(1). https://doi.org/10.1007/s00445-014-0886-7

Miyabuchi Y, Iizuka Y, Hara C, Yokoo A, Ohkura T (2018) The September 14, 2015 phreatomagmatic eruption of Nakadake fi rst crater, Aso Volcano, Japan : Eruption sequence inferred from ballistic, pyroclastic density current and fallout deposits (A). J Volcanol Geoth Res 351:41–56. https://doi.org/10.1016/j.jvolgeores.2017.12.009

Miyagi I, Geshi N, Hamasaki S, Oikawa T, Tomiya A (2020) Heat source of the 2014 phreatic eruption of Mount Ontake, Japan. Bull Volcanol 82(4). https://doi.org/10.1007/s00445-020-1358-x

Muller G (1967) Methods in sedimentary petrology. In Section in: Degree of roundness according to Russel-Taylor-Pettijohn (after Schneiderhohn) New York, Hafner Publishing Company. 100–101. (p. 1967)

Nakagawa MI, Ohba TS (2002) Minerals in volcanic ash 1: primary minerals and volcanic glass. Glob Environ Res 6(2):41–52

Németh K (2010) Volcanic glass textures, shape characteristics and compositions of phreatomagmatic rock units from the Western Hungarian monogenetic volcanic fields and their implications for magma fragmentation. Cent Eur J Geosci 2:399–419. https://doi.org/10.2478/v10085-010-0015-6

Newhall CG, Costa F, Ratdomopurbo A, Venezky DY, Widiwijayanti C, Win NTZ, Tan K, Fajiculay E (2017) WOVOdat – An online, growing library of worldwide volcanic unrest. J Volcanol Geoth Res 345:184–199. https://doi.org/10.1016/j.jvolgeores.2017.08.003

Newhall CG, Punongbayan RS (1996) The narrow margin of successful volcanic-risk mitigation. Monitoring and mitigation of volcano hazards. Springer Science & Business Media, pp 807–838

Nurfiani D, Bouvet de Maisonneuve C (2018) Furthering the investigation of eruption styles through quantitative shape analyses of volcanic ash particles. J Volcanol Geoth Res 354:102–114. https://doi.org/10.1016/j.jvolgeores.2017.12.001

Ohba T, Nakagawa M (2002) Minerals in volcanic ash 2: non-magmatic minerals. Glob Environ Res 6(2):53–60

Paladio-Melasantos ML, Solidum RU, Scott WE, Quiambao RB, Umbal JV, Rodolfo KS, Tubianosa BS, Delos Reyes PJ, Alonso RA, Ruerlo HB (1996) Tephra falls of the 1991 eruptions of Mount Pinatubo. In: Newhall CG (Editor) & Others, Fire and Mud; Eruptions and Lahars of Mount Pinatubo, Philippines, Philippine Institute of Volcanology and Seismology, Quezon City, layer D, 413–535. https://doi.org/10.1159/000153100

Pardo N, Avellaneda JD, Rausch J, Jaramillo-Vogel D, Gutiérrez M, Foubert A (2020) Decrypting silicic magma/plug fragmentation at Azufral crater lake, Northern Andes: insights from fine to extremely fine ash morpho-chemistry. Bull Volcanol 82(12). https://doi.org/10.1007/s00445-020-01418-z

Pardo N, Cronin SJ, Németh K, Brenna M, Schipper CI, Breard E, White JDL, Procter J, Stewart B, Agustín-Flores J, Moebis A, Zernack A, Kereszturi G, Lube G, Auer A, Neall V, Wallace C (2014) Perils in distinguishing phreatic from phreatomagmatic ash; insights into the eruption mechanisms of the 6 August 2012 Mt. Tongariro eruption, New Zealand. J Volcanol Geoth Res 286:397–414. https://doi.org/10.1016/j.jvolgeores.2014.05.001

Pereira Borges H, de Aguiar MS (2019) Mineral classification using machine learning and images of microscopic rock thin section. In Mexican international conference on artificial intelligence. Springer, Cham, 2019. https://doi.org/10.1007/978-3-030-33749-0_6

Phillips T, Abdulla W (2021) Developing a new ensemble approach with multi-class SVMs for Manuka honey quality classification. Appl Soft Comput 111:107710. https://doi.org/10.1016/j.asoc.2021.107710

Pistolesi M, Rosi M, Malaguti AB, Lucchi F, Tranne CA, Speranza F, Albert PG, Smith VC, Di Roberto A, Billotta E (2021) Chronostratigraphy of the youngest (last 1500 years) rhyolitic eruptions of Lipari (Aeolian Islands, Southern Italy) and implications for distal tephra correlations. J Volcanol Geoth Res 420:107397. https://doi.org/10.1016/j.jvolgeores.2021.107397

Polacci M, Andronico D, de' Michieli Vitturi M, Taddeucci J, Cristaldi A (2019) Mechanisms of ash generation at basaltic volcanoes: the case of Mount Etna, Italy. Front Earth Sci 7(August). https://doi.org/10.3389/feart.2019.00193

Primulyana S, Kern C, Lerner AH, Saing UB, Kunrat SL, Alfianti H, Marlia M (2018) Gas and ash emissions associated with the 2010-present activity of Sinabung Volcano, Indonesia. J Volcanol Geotherm Res. https://doi.org/10.1016/j.jvolgeores.2017.11.018

Qin X, Zhang Z, Huang C, Dehghan M, Zaiane OR, Jagersand M (2020) U2-Net: Going deeper with nested U-structure for salient object detection. Pattern Recognit 106. https://doi.org/10.1016/j.patcog.2020.107404

Re G, Corsaro RA, D'Oriano C, Pompilio M (2021) Petrological monitoring of active volcanoes: A review of existing procedures to achieve best practices and operative protocols during eruptions. J Volcanol Geotherm Res 419:107365. https://doi.org/10.1016/j.jvolgeores.2021.107365

Romero JE, Aguilera F, Delgado F, Guzmán D, Van Eaton AR, Luengo N, Caro J, Bustillos J, Guevara A, Holbik S, Tormey D, Zegarra I (2020) Combining ash analyses with remote sensing to identify juvenile magma involvement and fragmentation mechanisms during the 2018/19 small eruption of Peteroa volcano (Southern Andes). J Volcanol Geoth Res 402:106984. https://doi.org/10.1016/j.jvolgeores.2020.106984

Romero JE, Burton M, Cáceres F, Taddeucci J, Civico R, Ricci T, Pankhurst MJ, Hernández PA, Bonadonna C, Llewellin EW, Pistolesi M (2022) The initial phase of the 2021 Cumbre Vieja ridge eruption (Canary Islands): Products and dynamics controlling edifice growth and collapse. 431(July). https://doi.org/10.1016/j.jvolgeores.2022.107642

Ross PS, Dürig T, Comida PP, Lefebvre N, White JDL, Andronico D, Thivet S, Eychenne J, Gurioli L (2022) Standardized analysis of juvenile pyroclasts in comparative studies of primary magma fragmentation; 1. Overview and workflow. Bull Volcanol 84(1):1–29. https://doi.org/10.1007/s00445-021-01516-6

Rowe MC, Thornber CR, Kent AJR (2008) Identification and evolution of the juvenile component in. a volcano rekindled: the renewed eruption of Mount St. Helens 2004–2006, 2004–2006

Savov IP, Luhr JF, Navarro-Ochoa C (2008) Petrology and geochemistry of lava and ash erupted from Volcán Colima, Mexico, during 1998–2005. J Volcanol Geoth Res 174(4):241–256. https://doi.org/10.1016/j.jvolgeores.2008.02.007

Saxby J, Rust A, Beckett F, Cashman K, Rodger H (2020) Estimating the 3D shape of volcanic ash to better understand sedimentation processes and improve atmospheric dispersion modelling. Earth Planet Sci Lett 534:116075. https://doi.org/10.1016/j.epsl.2020.116075

Scasso RA, Carey S (2005) Morphology and formation of glassy volcanic ash from the August 12–15, 1991 eruption of Hudson volcano, Chile Latin American. J Sedimentology Basin Anal 12(1):3–21

Scheidegger KF, Federman AN, Tallman AM (1982) Compositional heterogeneity of tephras from the 1980 eruptions of Mount St. Helens. J Geophys Res: Solid Earth 87(B13):10861–10881

Shoji D, Noguchi R, Otsuki S, Hino H (2018) Classification of volcanic ash particles using a convolutional neural network and probability. 1–12. https://doi.org/10.1038/s41598-018-26200-2

Singh S, Srivastava D, Agarwal S (2017) GLCM and its application in pattern recognition. 5th International Symposium on Computational and Business Intelligence, ISCBI 2017, November 2018, 20–25. https://doi.org/10.1109/ISCBI.2017.8053537

Shehzad F, Javaid N, Aslam S, Umar M (2022) Electricity theft detection using big data and genetic algorithm in electric power systems. Electr Power Syst Res 209:107975. https://doi.org/10.1016/j.epsr.2022.107975

Smith LI (2002) A tutorial on principal components analysis. Department of Computer Science, University of Otago

Spadaro FR, Lefèvre RA, Ausset P (2002) Experimental rapid alteration of basaltic glass: Implications for the origins of atmospheric particulates. Geology 30(8):671–674. https://doi.org/10.1130/0091-7613(2002)030%3c0671:ERAOBG%3e2.0.CO;2

Sural S, Qian G, Pramanik S (2002) Segmentation and histogram generation using the HSV color space for image retrieval. IEEE Int Conf Image Process 2:589–592. https://doi.org/10.1109/icip.2002.1040019

Suzuki Y, Nagai M, Maeno F, Yasuda A, Hokanishi N, Shimano T, Ichihara M, Kaneko T, Nakada S (2013) Precursory activity and evolution of the 2011 eruption of Shinmoe-dake in Kirishima volcano-insights from ash samples. Earth, Planets Space 65(6):591–607. https://doi.org/10.5047/eps.2013.02.004

Taddeucci J, Pompilio M, Scarlato P (2002) Monitoring the explosive activity of the July-August 2001 eruption of Mt. Etna (Italy) by ash characterization. Geophys Res Lett 29(8):71–1–71–74. https://doi.org/10.1029/2001GL014372

Taddeucci J, Pompilio M, Scarlato P (2004) Conduit processes during the July-August 2001 explosive activity of Mt. Etna (Italy): Inferences from glass chemistry and crystal size distribution of ash particles. J Volcanol Geotherm Res 137(1–3 SPEC. ISS.):33–54. https://doi.org/10.1016/j.jvolgeores.2004.05.011

Taddeucci J, Wohletz KH (2001) Temporal evolution of the Minoan eruption (Santorini, Greece), as recorded by its Plinian fall deposit and interlayered ash flow beds. J Volcanol Geoth Res 109(4):299–317. https://doi.org/10.1016/S0377-0273(01)00197-4

Tanur JM (2011) Margin of error. In international encyclopedia of statistical science (pp. 765–765). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-030-53743-2_3

Thivet S, Gurioli L, Di Muro A, Eychenne J, Besson P, Nedelec JM (2020) Variability of ash deposits at Piton de la Fournaise (La Reunion Island): insights into fragmentation processes at basaltic shield volcanoes. Bull Volcanol 82:1–20

Troncoso L, Bustillos J, Romero JE, Guevara A, Carrillo J, Montalvo E, Izquierdo T (2017) Hydrovolcanic ash emission between August 14 and 24, 2015 at Cotopaxi volcano (Ecuador): Characterization and eruption mechanisms. J Volcanol Geoth Res 341:228–241. https://doi.org/10.1016/j.jvolgeores.2017.05.032

Utami SB, Costa F, Lesage PH, Allard P, Humaida H (2021) Fluid fluxing and accumulation drive decadal and short-lived explosive basaltic andesite eruptions preceded by limited volcanic unrest. J Petrol 62(11):1–29. https://doi.org/10.1093/petrology/egab086

Verdhan V (2020) Supervised learning with python. Okänd. Irland: Apress

Verolino A, White JDL, Brenna M (2018) Eruption dynamics at Pahvant Butte volcano, Utah, western USA: insights from ash-sheet dispersal, grain size, and geochemical data. Bull Volcanol 80(11). https://doi.org/10.1007/s00445-018-1256-7

Watanabe K, Danhara T, Watanabe K, Terai K, Yamashita T (1999) Juvenile volcanic glass erupted before the appearance of the 1991 lava dome, Unzen volcano, Kyushu, Japan. J Volcanol Geotherm Res 89(1–4):113–121

White JDL, Houghton BF (2006) Primary volcaniclastic rocks. Geology 34(8):677–680. https://doi.org/10.1130/G22346.1

Yaguchi M, Ohba T, Hirayama Y, Numanami N (2022) Volcanic Ash from the June 17, 1962 Eruption of Yakedake volcano : Stereomicroscopic , XRD , and water-soluble components analyses. 17:257–262. https://doi.org/10.20965/jdr.2022.p0257

Yamanoi Y, Takeuchi S, Okumura S, Nakashima S, Yokoyama T (2008) Color measurements of volcanic ash deposits from three different styles of summit activity at Sakurajima volcano, Japan: Conduit processes recorded in color of volcanic ash. J Volcanol Geoth Res 178(1):81–93. https://doi.org/10.1016/j.jvolgeores.2007.11.013