Journal of Archaeological Science: Reports 59 (2024) 104806

Contents lists available at ScienceDirect

Journal of Archaeological Science: Reports

journal homepage: www.elsevier.com/locate/jasrep

Microscope agnosticism and the characterization of sedimentary abrasion of flint stone tools

Guillermo Bustos-Pérez [a,b,c,d,*], Andreu Ollé [c,d]

[a] Department of Human Origins, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
[b] Departamento de Prehistoria y Arqueología, Universidad Autónoma de Madrid, Campus de Cantoblanco, 28049 Madrid, Spain
[c] Institut Català de Paleoecologia Humana i Evolució Social (IPHES-CERCA), Zona Educacional 4, Campus Sescelades URV (Edifici W3), 43007 Tarragona, Spain
[d] Universitat Rovira i Virgili, Departament d'Història i Història de l'Art, Avinguda de Catalunya 35, 43002 Tarragona, Spain

ARTICLE INFO

Keywords:
Microscopic analysis
Lithic analysis
Lithic taphonomy
Postdepositional surface modifications

ABSTRACT

The surface of lithic stone tools from Paleolithic archaeological sites can undergo a range of different post-depositional alterations, including sedimentary erosion induced by water displacement or wind. The surface of flint artifacts can reflect these alterations as changes in texture. Microscopic analyses and grayscale images can be employed to obtain quantitative data to help determine the degree to which the surfaces of flint stone tools have been altered. However, surface quantitative values depend directly on the image capturing system of each microscope. This raises the question of whether the quantitative values are actually capturing the evolution of the surface, whether they are dependent on the type of microscope and its image capturing system, and whether the detection of the degree of abrasion might vary depending on the type of microscope. The present work sought to determine whether data extracted from images from two different microscopes point to the same trends in surface change due to postdepositional alterations. Surface photographs of a sample of 25 flakes were taken using a Dino-Lite Edge 3.0 AM73915MZT and a 3D Optical Profiler Sensofar S neox 090. These flakes represented three different stages of alteration (fresh, ten hours of experimentally-induced sedimentary erosion, and geological neocortex). Results from grayscale images indicate that, despite yielding different numeric ranges, the quantitative values of the images from both types of microscope reflect the same trends in surface change. The classification accuracy of the three stages of erosion did not vary between microscopes.

1. Introduction

Flint stone tools are among the most common remains recovered from Paleolithic sites. They provide information not only about chrono-cultural developments, but also about the behavioral and spatial organization of Paleolithic groups. However, stone tools from Paleolithic sites may be subjected to any number of postdepositional alterations, most commonly water flow or wind abrasion (Byers et al., 2015; Hosfield and Chambers, 2016; Michel et al., 2019; Petraglia and Potts, 1994; Schick, 1986). These postdepositional processes can disrupt archaeological remains, resulting in horizontal and vertical mixing of artifacts and, consequently, unreliable chrono-cultural interpretations. Therefore, a solid analysis of the integrity of a lithic assemblage is needed prior to its interpretation.

Postdepositional alterations and their intensity are recorded on the surface of stone tools, most commonly observed in the form of increased ridge width and surface abrasion (Burroni et al., 2002; Bustos-Pérez et al., 2019; Bustos-Pérez and Ollé, 2023; Chambers, 2016; Shackley, 1974). Unaided visual assessments of sedimentary abrasion (rounding) can result in an error ratio of over 80 %, emphasizing the need for microscopic analyses with quantitative variables (Chambers, 2016).

The quantitative characterization of surfaces is an important part of many lithic microscopic analytical approaches. In recent traceological studies, surface texture quantification is often used to identify worked materials (Evans and Donahue, 2008; Ibáñez et al., 2019; Ibáñez and Mazzucco, 2021; Macdonald, 2014; Sferrazza, 2023; Stemp et al., 2016, 2008; Stemp and Chung, 2011; Stevens et al., 2010) or test different models of polish development (Ibáñez and Mazzucco, 2021). Another area of lithic microscopic analysis in which surface quantification plays a key role is in the identification, characterization and estimation of the intensity of postdepositional alterations on lithic artifacts (Burroni et al., 2002; Bustos-Pérez et al., 2019; Caux et al., 2018; Chambers, 2016; Chu

* Corresponding author at: Departament of Human Origins, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany.
E-mail address: guillermo.willbustos@gmail.com (G. Bustos-Pérez).

https://doi.org/10.1016/j.jasrep.2024.104806
Received 3 May 2024; Received in revised form 7 October 2024; Accepted 8 October 2024
Available online 16 October 2024

and Hosfield, 2020; Hiscock, 1985; Hosfield et al., 2000; Levi Sala, 1986). The development of these analyses has been incorporated in studies focusing on the integrity of the lithic artifacts in archaeological assemblages (Fraile-Márquez et al., 2022; Galland et al., 2019; Staurset et al., 2023). Fig. 1.Fig. 2.Fig. 3.

Recent approaches using sequential experimentation, grayscale images (Bustos-Pérez and Ollé, 2023; Sferrazza, 2023) and texture metrics (Haralick et al., 1973) have demonstrated the viability of quantifying surface changes caused by sedimentary abrasion on flint tools. However, the information contained in the pixels of a photograph may vary according to the capturing system of each microscope. This raises the issue of the importance of *microscope agnosticism* (meaning that metric values or trends do not vary in relation to the choice of microscope). In the present study, we explored four aspects to explore the issue of microscope agnosticism in relation to sedimentary abrasion.

1) Observed trends in quantitative metrics should be consistent independent of the choice of microscope. The combination of sequential experimentation and quantitative metrics (Bustos-Pérez and Ollé, 2023; Ibáñez and Mazzucco, 2021; Ollé and Vergès, 2014) has resulted in a certain degree of confidence regarding how the surface of flint changes progressively as result of a mechanical action. However, the current understanding of this change is directly related to the image acquisition procedure and the quantitative variables derived from those images. Ideally, quantitative trends would remain consistent regardless of the type of microscope used.

2) Collinearity or multicollinearity of metric features extracted from images should be considered. Several approaches (Bustos-Pérez and Ollé, 2023; Ibáñez and Mazzucco, 2021; Pedergnana et al., 2020; Sferrazza, 2023; Stevens et al., 2010) use machine learning (ML) classification algorithms on metric features extracted from images. Along with higher accuracy, one substantial advantage of ML algorithms is that they can provide insights into feature importance for classification. Collinearity is generally not considered problematic for classification metrics, provided that any collinearity present in the training set is also present in the predicted sample. However, collinearity is considered to have substantial effects on interpreting feature importance due to unstable coefficients or redundancy in feature selection. Thus, it is important to be aware of the presence of collinear or multicollinear variables when using quantitative features extracted from images. Images of the same area from different microscopes will vary depending on the image acquiring system, and as a consequence, the presence of collinearity among the extracted features is also expected to vary from microscope to microscope. A microscope generating images with fewer pairs of collinear variables can be considered more reliable for the quantitative characterization of a surface.

3) Classification accuracy should remain similar despite the use of variables obtained from the images of different microscopes. Accuracy can be affected when multicollinearity is present among the variables. A common approach to this problem is the use of dimensionality reduction methods (PCA, t-SNE; Naes and Mevik, 2001), which make it possible to combine multiple collinear variables while avoiding the loss of information. If multicollinearity is present, the accuracy of a model using raw variables should be compared with a model using dimensionally reduced variables.

4) Consistency of variable importance among classification algorithms and photographs obtained from different microscopes is a good indicator of microscope agnosticism. However, as previously indicated, variable importance can be affected by collinearity among predictors. This is an important consideration since the importance of a pair of variables might be a result of their collinearity. Additionally, it makes it possible to consider which groups of metric variables should be emphasized when analyzing surface change due to a given mechanical action.

2. Methods

2.1. Experimental sample and cleaning protocol

The sample consisted of 25 flakes experimentally knapped by one of the authors (GBP) using direct percussion with a hard hammer. The

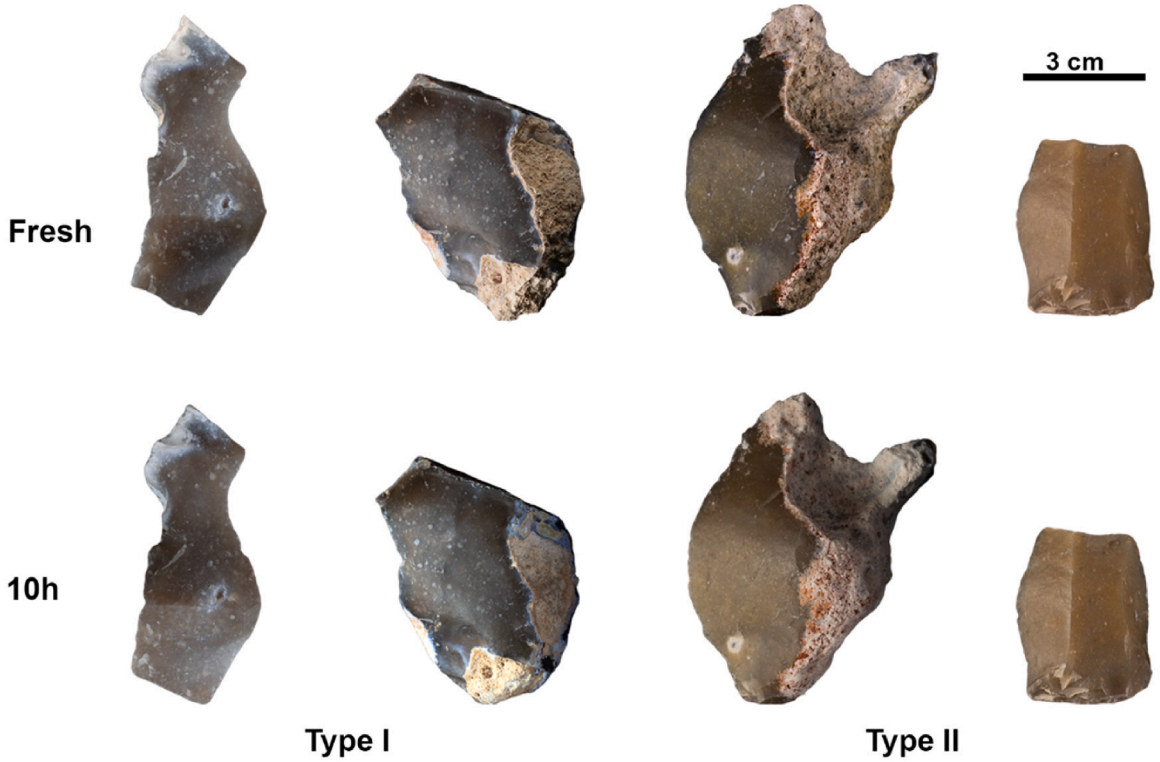

Fig. 1. Sample of experimental materials before and after 10 h of sedimentary abrasion (photographs by M. D. Guillén).

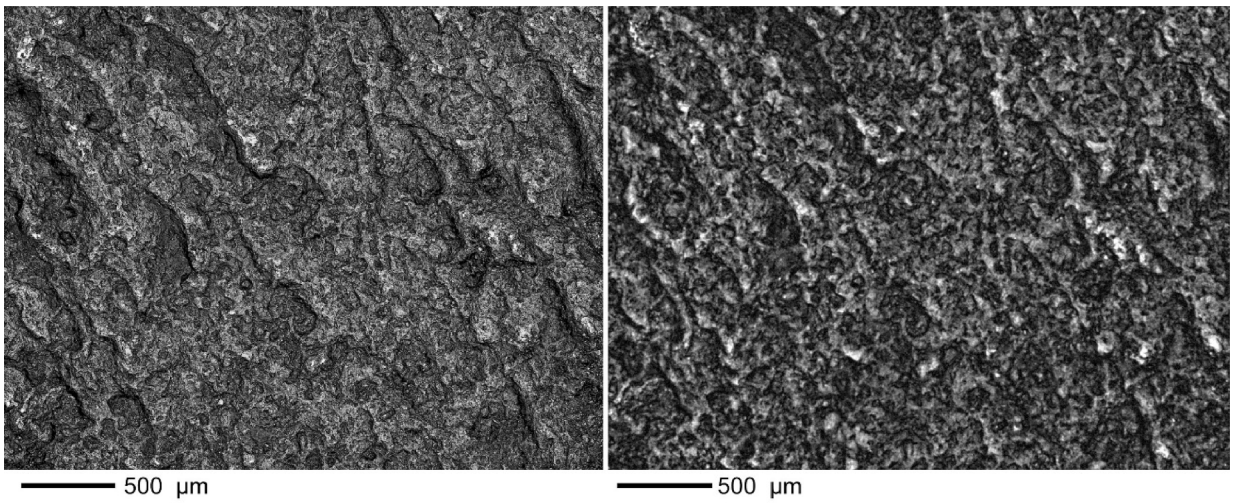

Fig. 2. Example of two images from the same neocortex surface. Left: Sensofar S neox 090; right: Dino-Lite Edge 3.0 AM73915MZT. Both images after processing using Fiji/ImageJ.

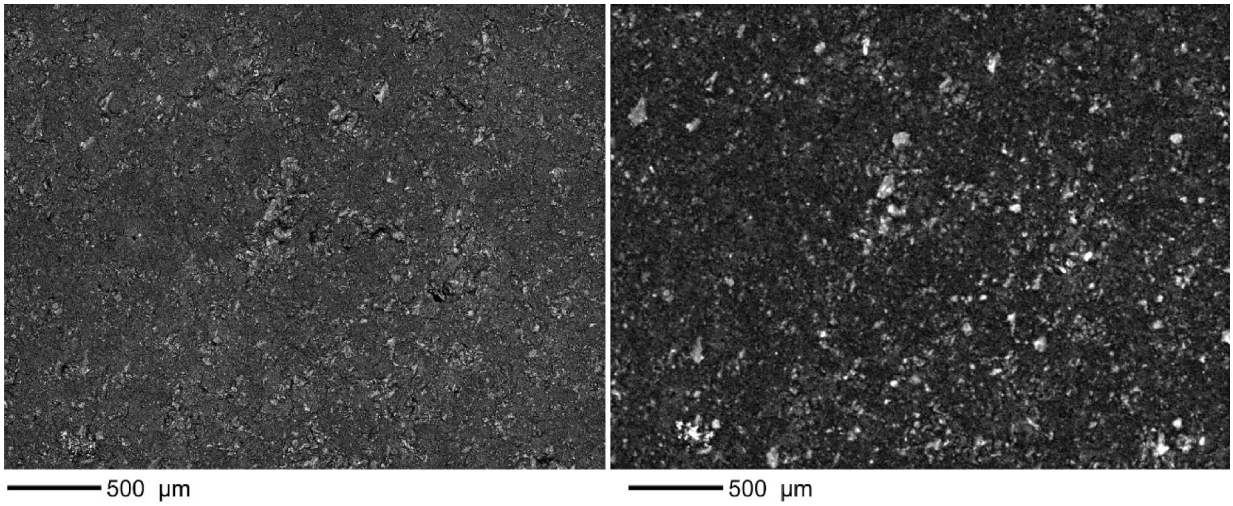

Fig. 3. Example of two images of the same flint fresh surface. Left: Sensofar S neox 090; right: Dino-Lite Edge 3.0 AM73915MZT. Both images after processing using Fiji/ImageJ.

flakes came from three different types of flint (Table 1), all of them south Madrid Miocene flint (Bustillo et al., 2012; Bustillo and Pérez-Jiménez, 2005) from different locations. South Madrid Miocene flints were formed by the replacement of sedimentary rocks which had filled the original basin, which is thought to have taken place under continental conditions such as alluvial plain deposits, shallow lacustrine waters, and marshes (Bustillo et al., 2012). Macroscopic analysis of the flints shows that they present a fine, opaque, homogeneous surface and blue/grey and reddish/ocher coloration. There is also a relative absence of opal in these flints, although geodes and pseudo-morphs are sometimes present.

Five type 1 and 2 flakes were analyzed to obtain images of the fresh surface, while four type 1 flakes and five type 2 flakes were submitted to ten hours of rounding in a tumbling machine (KT-3010 SUPER-TUMBLER). The sedimentary matrix employed to simulate rounding consisted of a mix of sand and water (a total weight of 5 kg of which 30–40 % was water). Sediment was obtained from the quaternary levels of the Madrid basin and was made up of fine quartz sands with silt and partial carbonation. The tumbler machine was set at continuous direction at 83 rpm. The average weight of the flakes introduced into the tumbler was 26.25 g.

Geological images of the neocortex were obtained from three type 3 flakes and three type 2 flakes (Table 2). These images of geological neocortex serve as reference samples for extreme levels of sedimentary abrasion.

Table 1
Number of flakes analyzed according to type of flint.

	Fresh	10 h of rounding	Neocortex
Type 1	5	4	
Type 2	5	5	3
Geological sample			3
N flakes	10	9	6

Table 2
Number of photographs taken by each microscope by time of abrasion.

Microscope	Exposure time	Number of photos
Sensofar S neox 090	Fresh	100
Sensofar S neox 090	10 h	105
Sensofar S neox 090	Geological Neocortex	71
Dino-Lite Edge	Fresh	100
Dino-Lite Edge	10 h	87
Dino-Lite Edge	Geological Neocortex	71

Possible contaminants were removed by means of a multi-step procedure based on a previous study (Pedergnana et al., 2016). The present study used a two-step procedure consisting of an ultrasonic bath (frequency of 40 kHz) in a 2 % neutral soap (Derquim) solution for 10 to 15 min, followed by a second ultrasonic bath in pure acetone for another 10 to 15 min. After each step, the lithic artifacts were placed in a tap water bath and finally dried using compressed air. During the cleaning protocol and microscopic analysis, all artifacts were handled using powder-free surgical gloves.

2.2. Image acquisition and processing

In order to compare images, the field of view (FOV) and pixel ratio of both microscopes must be as similar as possible. Parameters of the Dino-Lite Edge 3.0 AM73915MZT were kept the same as in previous experiments (Bustos-Pérez and Ollé, 2023), with a FOV of 3.28 x 2.46 mm and a pixel ratio of 2548 x 1918. As a result, each pixel measured 1.28 μm (width) by 1.28 μm (height). The Dino-Lite Edge 3.0 AM73915MZT microscope was mounted in a Dino-Lite RK-06-AE stand in order to ensure verticality, and a N3C-D2 diffuser cap was used to ensure the even distribution of light. In the process of taking each photograph, the region of interest on the flint was manually positioned as horizontally as possible (Calandra et al., 2022). To avoid problems due to focus variation, each surface was photographed several times at different heights, and the sequences obtained were mounted using a Helicon Focus 7.7.2. Normally, satisfactory stacking required between two and four images, although additional images were employed when needed.

The Sensofar S neox 090 zoom was manually adjusted to the most similar FOV (3.18 x 2.65 mm) with 2x2 mosaics taken for each image. All images were obtained using a x10 objective lens (numerical aperture 0.30) in light scanning confocal mode (microdisplay scanning confocal microscopy) at 5 Mp resolution with at least 95 % of the information retrieved. The original Sensofar S neox 090 photographs had a pixel ratio of 4616 x 3848. This resulted in each pixel measuring 0.69 x 0.69 μm. In order to match the FOV and pixel ratio of the Sensofar S neox 090 zoom to those of MP previous studies (Bustos-Pérez and Ollé, 2023), the images were cropped and the pixels downsampled. Although the images from the Dino-Lite Edge 3.0 AM73915MZT microscope were slightly wider (0.10 mm), this resulted in almost identical FOV, pixel ratios and pixel widths/heights (Table 3).

Both sets of images underwent the same two-step image treatment procedure employed in a previous study (Bustos-Pérez and Ollé, 2023). First, the Fiji (Schindelin et al., 2015, 2012) "subtract background" plugin was used to minimize the effects of different lighting and differing flint coloration. Second, the "enhance contrast" function was used to desaturate the images by normalizing their histograms. This process provided a gray-level image for use as input for the statistical analysis. All analyzed images were in TIFF format.

Photographs of fresh and neocortex surfaces were obtained in the same areas with both microscopes. The surface of the flakes that had undergone 10 h of rounding was initially recognized with both types of microscopes and the most-developed surfaces were photographed. This is a common procedure when analyzing microscopic traces, as the most-developed area (Ibáñez and Mazzucco, 2021; Pedergnana et al., 2020) is targeted for photography. This ensured the maximum visibility of the abrasion for each microscope.

Table 3
Summary of the characteristics of the images from the two microscopes.

Microscope images	Image aspect ratio	FOV (mm)	Pixel ratio	Pixel width height (μm)
AM73915MZT	1.33	3.28 x 2.46	2548 x 1918	1.28 x 1.28
S neox 090 (original images)	1.19	3.18 x 2.65	4616 x 3848	0.69 x 0.69
S neox 090 (transformed)	1.33	3.18 x 2.46	2480 x 1918	1.28 x 1.28

2.3. Quantitative analysis

Three groups of metrics were extracted from the microscope images (Table 4). The first group corresponds to descriptive statistics of the gray-level values in each image, which can be divided into measures of central tendency (mean, mode and median), and measures of deviation and distribution (standard deviation, kurtosis and skewness).

The second group corresponds to measures of roughness. The Fiji/ImageJ (Collins, 2007; Schindelin et al., 2015, 2012) SurfCharJ plugin (Chinga et al., 2007; Chinga and Dougherty, 2002) was employed to obtain measures of Rq (root mean square deviation/roughness), Ra (arithmetical mean deviation), Rsk (skewness of the assessed profile) and Rku (kurtosis of the assessed profile). Profiles of the whole surface (those with the "R" prefix) were employed as input, and measures were calculated following the ISO 4287/2000 standard (Chinga et al., 2007; Chinga and Dougherty, 2002).

Texture measures take into consideration the spatial distribution and intensity values of the pixels from a grayscale image (Haralick et al., 1973). The spatial distribution and intensity are analyzed through a gray-level covariance matrix (GLCM). This process works in two steps. First, for a given distance and direction, a matrix is built that captures the relationship of intensity between pairs of pixels (reference and neighbor). Second, for every x and y it considers the co-occurrence of values, forming a new matrix. This matrix makes it possible to obtain a series of statistical descriptors (Haralick et al., 1973): the angular second moment (ASM), contrast (CONT), correlation (CORR), inverse different moment (IDM) and entropy (ENT). Based on previous studies (Bustos-Pérez and Ollé, 2023), four distances (5, 10, 15 and 20 pixels) and all four possible directions (north, east, south, west) were employed to create the GLCM and extract the features.

Two procedures were used to address the issue of collinearity among the predictors. The first of these was the removal of collinear variables using a pairwise cutoff discard procedure with a matrix representing the linear correlation (r^2) of each pair of variables. For pairs of variables presenting a correlation above a given threshold, the variable presenting the highest average correlation (among all variables) was removed. An arbitrary cutoff threshold of 0.9 was selected. A visual evaluation of variable relationships showed that this threshold prevented the exclusion of pairs of variables presenting polynomial or logarithmic

Table 4
Summary of variables and the group to which they belong. C: central tendency; D&D: deviation and distribution; R: roughness; T: texture.

Name	Acronym	Group	Description
Mean	$\bar{x}$	C	Central tendency of the sample
Modal	Mo	C	Most repeated value
Median	M	C	Value of at least half the sample
Standard Deviation	SD	D&D	Variation expected from the mean
Skewness	Sk	D&D	Asymmetry of the distribution
Kurtosis	Ku	D&D	Tailedness of the distribution
RMS deviation/ roughness	Rq	R	Indicator of surface roughness
Arithmetical mean deviation	Ra	R	Deviation of a surface from a mean height
Skewness assessed profile	Rsk	R	Indicator of the departure from surface symmetry
Kurtosis assessed profile	Rku	R	Sharpness of the peaks
Angular second moment	ASM	T	Measure of homogeneity in the image
Contrast	CONT	T	Indicative of local variations
Correlation	CORR	T	How a reference pixel is related to its neighbor
Inverse different moment	IDM	T	Closeness of the distribution of the GLCM elements to the GLCM diagonal.
Entropy	ENT	T	Amount of irremediable chaos or disorder in an image

relationships. In the second procedure, dimensionality was reduced through a principal component analysis (PCA), which identifies the linear combinations that best represent the variables in an unsupervised manner (James et al., 2013; Pearson, 1901). Principal components (PCs) capture as much variance as possible for the complete dataset. This is especially useful when several collinear variables will be combined as a single variable, as it ensures a minimum loss of information (James et al., 2013; Yang and Yang, 2003). Consistency among the variable trend of data from both microscopes was evaluated through visual exploratory analysis and statistical differences between each consecutive episode of abrasion using student's *t*-test (Student, 1908).

A previous study (Bustos-Pérez and Ollé, 2023) using the same type of metrics showed that the linear discriminant analysis (LDA) provided the best results for classification. LDA reduces the dimensionality of the data aiming to maximize the separation between classes while decision boundaries divide the predicted classes into regions (Fisher, 1936; James et al., 2013). In the present study, three sets of LDA models were trained for each group of images. The first model was trained using the complete set of variables, second model was trained using the set of variables remaining from the pairwise cutoff to avoid collinearity, and the third model was trained after conducting a PCA for dimensionality reduction.

Machine learning protocols divide the data into training (used to train the model) and test sets (used to evaluate the algorithm predictive power on unseen data). A k-fold cross validation divides the dataset in a *n* number of folds. During an iteration, one fold serves as test set and the rest of the folds serve as the training set. These iterations are repeated successively until each fold has served as test set. However, the composition of the folds depends on the random shuffling of the data. Once all folds have served as test sets, the data is randomly shuffled and a new cycle of cross validation starts. All models were evaluated using 10 x 50 k-fold cross-validation (10 folds and 50 cycles), which provided measures of accuracy. Using a 10-fold division, each fold from the Dino-Lite Edge 3.0 AM73915MZT images was composed of 26 images, while each fold from Sensofar S neox 090 was composed of 28 images. Each fold subsequently acted as a test set for a trained model. Although computationally more expensive, this guaranteed that all data points served as test sets. At the start of each of the 50 cycles, and prior to fold division, the images were randomly shuffled, ensuring that the composition of the folds varied in each cycle and that composition did not play a significant role in the evaluation of the models.

Two measures were selected for the evaluation of the machine learning models: accuracy and area under the curve (AUC). Accuracy indicates the success rate of a model, representing the proportion of times in which a class was correctly identified (Lantz, 2019). Accuracy is usually calculated using a 0.5 threshold for class assignation. However, classification thresholds can be modified to balance the ability of the model to detect true positives and avoid false positives (sensitivity and specificity). The receiver operating characteristic (ROC) curve makes it possible to systematically evaluate the ratio of detected true positives, while avoiding false positives, for a given threshold (Bradley, 1997; Spackman, 1989). The ROC curve makes it possible to calculate the area under the curve (AUC) with a value that can range from 1 (perfect classifier) to 0, and a value of 0.5 representing a random classifier. In the present study, only the general AUC value of model performance was considered. The general AUC was calculated from the average of each AUC class (Hand and Till, 2001; Robin et al., 2011).

Machine learning models make it possible to estimate the importance of variables for classification. In the present study, variable importance was calculated only for LDA models trained on non-collinear sets of variables.

The statistical study was conducted using R version 4.3.1 in the IDE RStudio version 2023.09.0 (R Core Team, 2019; RStudio Team, 2019). Data were managed and graphs created using the tidyverse v.2.0.0 package (Wickham et al., 2019). LDA models were trained using MASS (Modern Applied Statistics with S) v.7.3.60 (Venables and Ripley, 2002). The *k*-fold cross validation of all models, precision metrics, the pairwise discard of collinear variables, and variable importance were performed using the caret package v.6.0.94 (Kuhn, 2008). ROC curves and AUC values were obtained using the pROC v.1.18.5 package (Robin et al., 2011).

All data, code and the complete workflow needed to perform the analysis is made freely available as an open repository at Github Zenodo (https://doi.org/10.5281/zenodo.13918530) which provides a citable doi, and at Github (https://github.com/GuillermoBustosPerez/Two-microscopes-flint-abrasion) using an RMarkdown document (Xie, 2014; Xie et al., 2018). If the current draft is accepted for publication, the repository will be made available at Zenodo.

3. Results

3.1. Trends of metric variables

Table 5 summarizes the observed trends and consistency between microscopes for the variables employed in the analysis. Images from the Dino-Lite Edge 3.0 AM73915MZT show clear trends for all variables and for all three stages of mechanical action considered. The interpretation of the images from the Sensofar S neox 090 was not nearly as clear, with no statistical differences between stages of abrasion for several of the variables.

Clear consistency between microscopes was observed for seven variables (mean of pixel values, standard deviation, Sk, Ku, Rq, Ra and CONT) in the form of the same trends and marked statistical differences between values of variables according to stage of abrasion. Two additional variables (median and CORR) presented *p* values at the limit of significant statistical difference, with the presence of several outliers obfuscating the apparent trends among values.

Visual analysis of the values in box plots makes it possible compare the evolution of trends among the different groups of variables and compare microscope consistency (Fig. 5; Fig. 6; Fig. 7; Fig. 8). A tendency of increasing values was observed among the three measures of central tendency in the Dino-Lite Edge 3.0 AM73915MZT images (Fig. 5). The images of geological neocortex from the Sensofar S neox 090 tended to have higher mean and median values than images of fresh flakes and those exposed to 10 h of abrasion. However, no statistical difference was found between the median values of the images of fresh flint images and those of flakes after 10 h of rounding ($t = -0.73$; $p = 0.46$, $df = 191.32$), although there was a statistical difference for the mean values of the same two categories ($t = -9.80$; $p < 0.01$, $df = 99.42$). As expected, the modal of the distribution was the least reliable variable, with the mode consistently having a value of 0 for all Sensofar S neox 090 images.

Table 5
Schematic table of observed trends among variables and microscope consistency. Green: indicates a clear trend; yellow: pattern observed, but not statistically significant between one of the three stages; red: no pattern observed.

Feature	Sensofar	Dinolite	Consistency
Mean	Increase	Increase	Yes
Median	Increase	Increase	Yes
Modal	No trend	Increase	No
SD	Increase	Increase	Yes
Sk	Decrease	Decrease	Yes
Ku	Decrease	Decrease	Yes
Rq	Increase	Increase	Yes
Ra	Increase	Increase	Yes
Rsk	No trend	Decrease	No
Rku	No trend	Decrease	No
ASM	No trend	Decrease	No
CONT	Increase	Increase	Yes
CORR	Increase	Decrease	Yes
IDM	No trend	Decrease	No
ENT	No trend	Increase	No

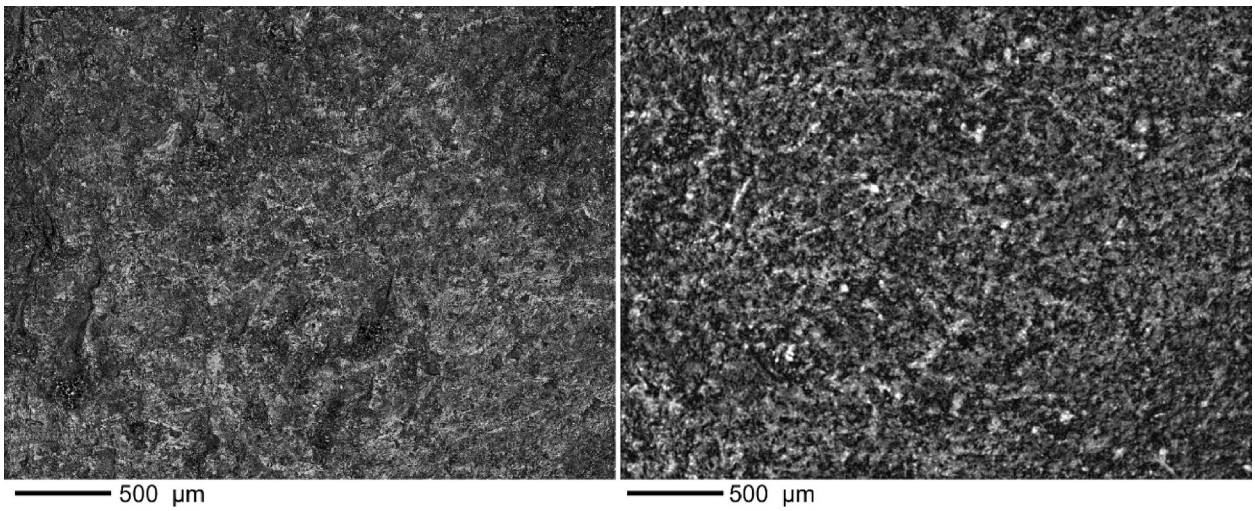

Fig. 4. Examples of heavily developed abrasion. Images are of the same flint. Both images after processing using Fiji/ImageJ.

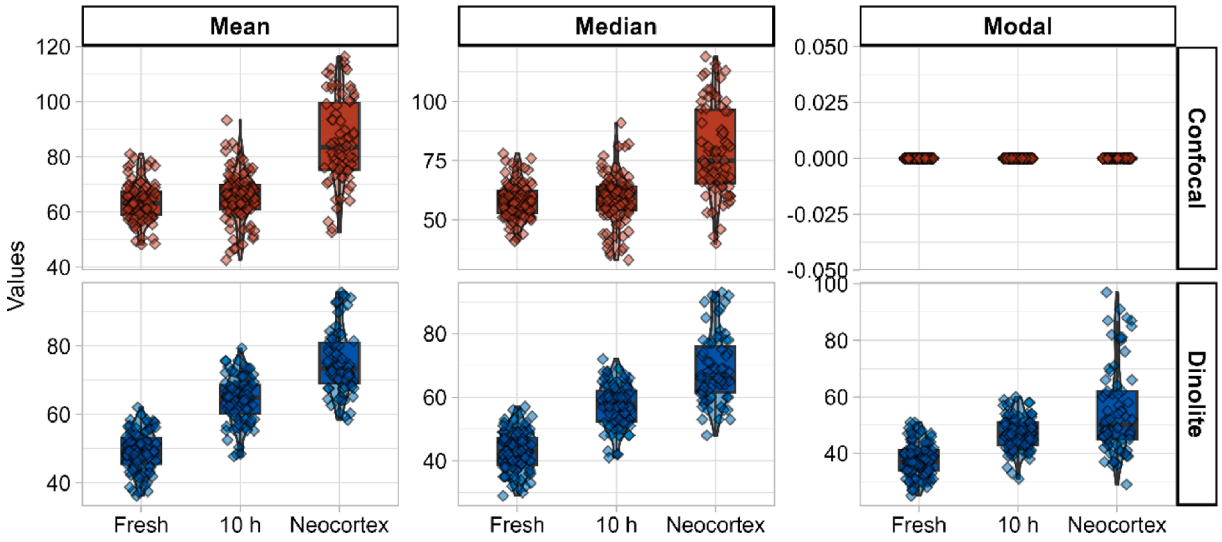

Fig. 5. Box plots presenting the distribution of values of the central tendency features.

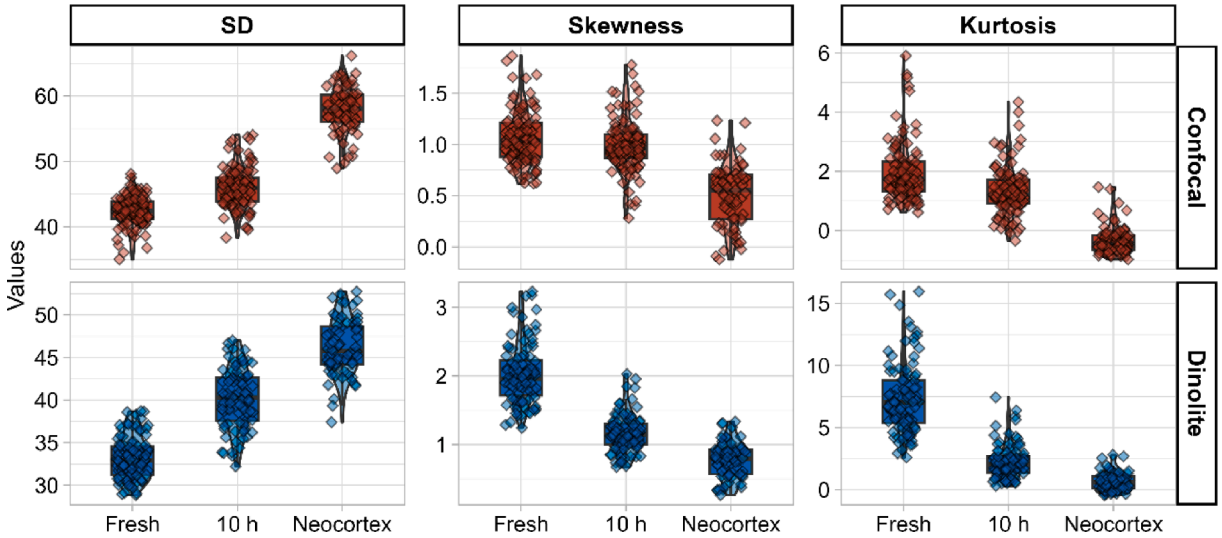

Fig. 6. Box plots presenting values of variables capturing deviation and distribution of pixel values.

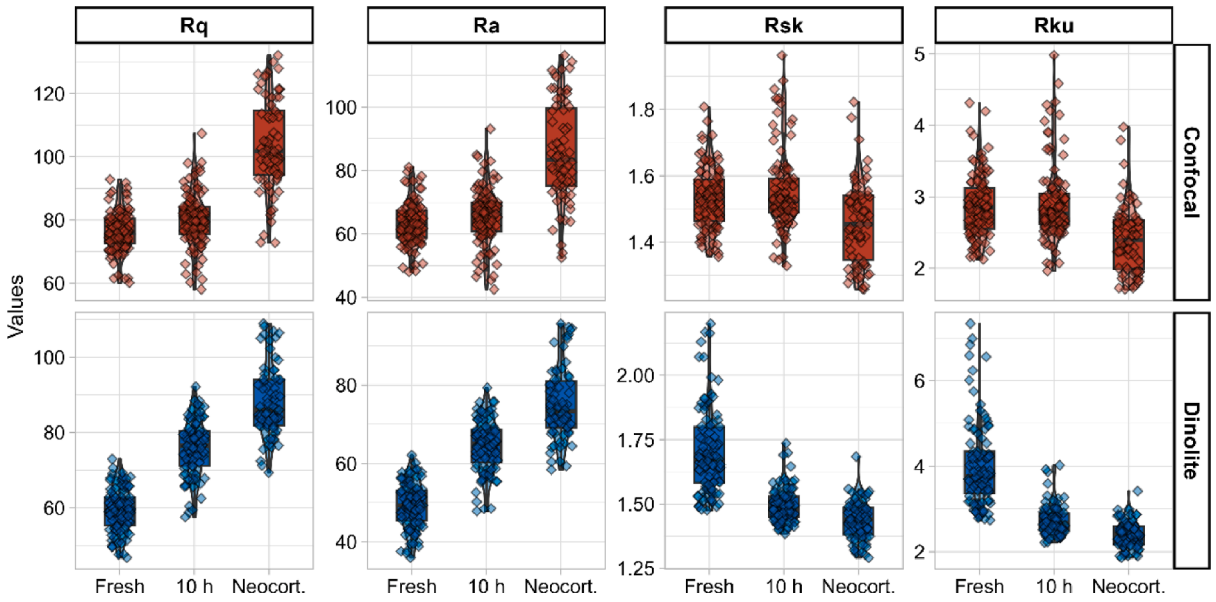

Fig. 7. Box plots presenting values of variables capturing roughness features.

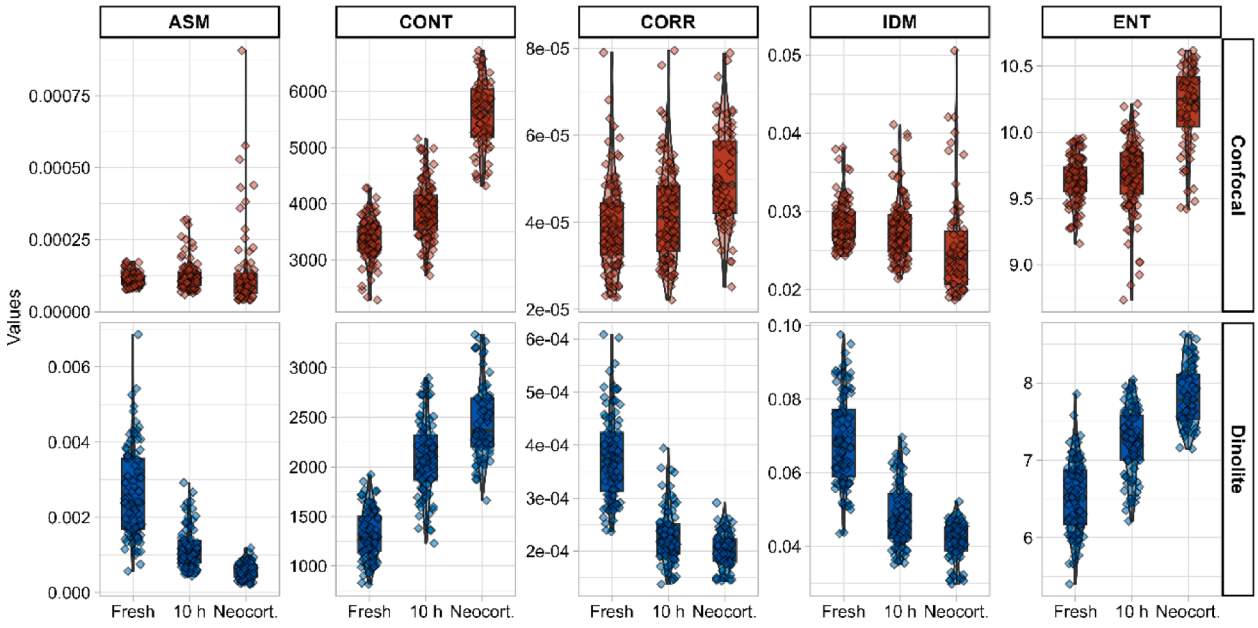

Fig. 8. Box plots presenting values of variables capturing texture features.

The images from both microscopes demonstrated a much better trend consistency for variables capturing the deviation and distribution of values (Fig. 6). Standard deviation and kurtosis showed marked statistically significant trends (increase and decrease respectively) for both types of microscopes. The skewness values of the Sensofar S neox 090 images for fresh flint and flint subjected to 10 h of rounding presented slight statistical differences ($t = 2.15$; $p = 0.03$, $df = 201.75$).

Only two features of roughness were consistent between the two microscopes (Rq and Ra). Images from the Dino-Lite Edge 3.0 AM73915MZT showed a clear trend of increasing values as sedimentary abrasion progresses. This trend was less marked in the Sensofar S neox 090 images, although there were statistically significant differences in Rq values ($t = -3.86$, $p < 0.01$, $df = 194.49$) between fresh flint and flint subjected to ten hours of rounding. The statistical significance between these stages of sedimentary abrasion was less marked for the Ra values ($t = -2.21$, $p = 0.02$, $df = 193.99$). Although Rsk, and Rku presented clear trends of diminishing values in the case of the Dino-Lite Edge 3.0 AM73915MZT images, no statistically significant difference was observed between fresh flint values and those of flint after 10 h of rounding for the Sensofar S neox 090 images ($t = -1.42$, $p = 0.16$, $df = 194.46$; $t = -0.47$, $p = 0.64$, $df = 197.98$).

One of the textural features (CONT) presented good consistency of trend evolution in images from both microscopes (Fig. 8), and a relatively good pattern was observed in a second such feature (CORR). Images from both microscopes presented a decrease in angular second moment (ASM) values, although this decrease was much more pronounced among the images obtained from the Dino-Lite Edge 3.0 AM73915MZT. For the Sensofar S neox 090, there was a statistically significant difference between the images of fresh flint and flint after 10 h of rounding ($t = -2.04$, $p = 0.04$, $df = 143.25$), although this significance was not present in images of flint after 10 h of rounding and neocortex ($t = -0.79$, $p = 0.43$, $df = 81.28$), which may be the result of

the presence of abundant outliers in the neocortex category. Contrast (CONT) presented the best example of consistency of all the variables, with the values of images from both microscopes presenting a clear increasing trend as sedimentary abrasion increased. Correlation (CORR) presented the clearest example of an inverse trend between the microscopes. Values of images from the Sensofar S neox 090 showed a clear increasing trend (within a wide range of distribution), while values from the Dino-Lite Edge 3.0 AM73915MZT images showed a decreasing trend and more concentrated values.

Inverse different moment (IDM) showed a decrease in values for both groups of images, although the neocortex images from the Sensofar S neox 090 presented a wider range of distribution values, with no significant difference between fresh flint and flint after 10 h of abrasion ($t = 1.37$; $p = 0.17$, $df = 190.54$). No statistically significant difference was found between fresh flint and flint after 10 h of abrasion in the Sensofar S neox 090 images ($t = -1.2$, $p = 0.23$, $df = 173.47$) for entropy (ENT) either.

3.2. Collinearity among features

Fig. 9 presents the results of correlation pairs between features for both types of microscopes. In general, a high level of correlation was observed among the features for both microscopes. For both microscopes, ASM presented very little correlation with other variables. The CORR feature presented little systematic correlation in the case of images from the Sensofar S neox 090 microscope, and ENT also presented little systematic correlation for the images from the Dino-Lite Edge 3.0 AM73915MZT microscope.

The Sensofar S neox 090 presented nine variables exceeding the $r^2 = 0.9$ cut-off threshold (standard deviation, median, Ku, Sk, Rsk, ASM, CORR, IDM and ENT), while the Dino-Lite Edge 3.0 AM73915MZT presented seven features below the cut-off threshold (median, modal, standard deviation, Rsk, ASM, CORR and ENT).

3.3. LDA accuracy

Dimensionality reduction through PCAs resulted in PC1 capturing 99.97 % of variance for the set of images from both microscopes. PC2 captured 0.028 % of variance for the Sensofar S neox 090 images and 0.022 % of variance for the Dino-Lite Edge 3.0 AM73915MZT images.

Table 6 presents the results of LDA models for the three different sets of variables selected (complete set of variables, non-collinear variables and first three PCs) by microscope. In general, all models presented outstanding AUC values, and differences between model performance were minimal.

In the Sensofar S neox 090 images, no differences were observed between the classification metrics of the LDA model using the complete set of variables (accuracy = 0.853; AUC = 0.957) and the LDA model using non-collinear variables (accuracy = 0.851; AUC = 0.956). The accuracy of the LDA model trained on PCs was slightly lower (0.810), although the AUC value was practically equal to that of the other models (0.944). The LDA model trained on the whole set of variables from the Dino-Lite Edge 3.0 AM73915MZT images presented slightly higher values than the other models (accuracy = 0.89; AUC = 0.977). The LDA model trained on non-collinear variables and PCs using the Dino-Lite Edge 3.0 AM73915MZT images presented very similar accuracy (0.859 and 0.844) and AUC values (0.963 and 0.949).

3.4. Feature importance

Mean variable importance is presented in Fig. 10. In general, there was good consistency in variable importance for classification, as standard deviation was considered the most important variable for classification in both sets of images. It is important to consider that the standard deviation was strongly correlated with the CONT textural feature in both groups of images ($r^2 = 0.98$ and 0.92). This indicates that the degree of deviation in general (standard deviation) and in local space (CONT) is highly important in the characterization of abraded flint surfaces. The

Table 6
Results of accuracy and AUC for the LDA models trained with different sets of variables.

Variables	Metric	Sensofar	Dinolite
Complete set	Accuracy	0.853	0.890
Complete set	AUC	0.957	0.977
Non-Collinear	Accuracy	0.851	0.859
Non-Collinear	AUC	0.956	0.963
PCs	Accuracy	0.810	0.746
PCs	AUC	0.944	0.911

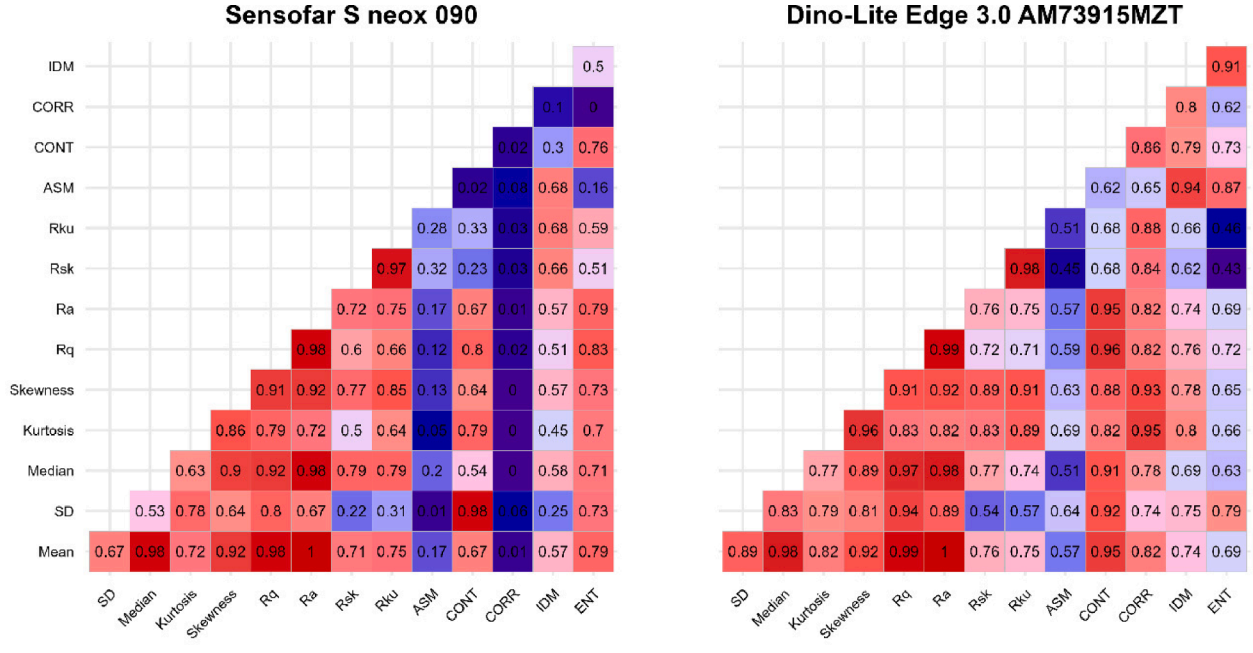

Fig. 9. Correlation plots between pairs of variables for both microscopes.

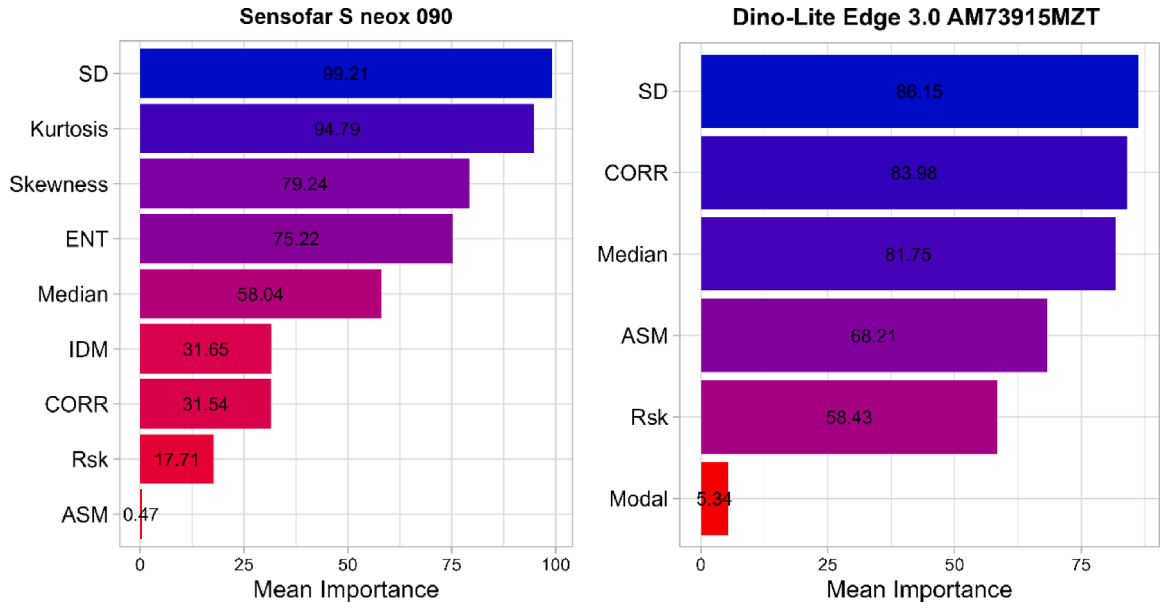

Fig. 10. Variable mean importance for the LDA models trained on non-collinear features.

following second and third most important variables for the Sensofar S neox 090 group of images were kurtosis and skewness. For the Dino-Lite Edge 3.0 AM73915MZT images, kurtosis and skewness were strongly correlated with the CORR textural feature ($r^2 = 0.95$ and 0.93), which is considered the second most important variable.

The LDA model trained on the Sensofar S neox 090 images considered ENT the fourth most important variable for discrimination. ENT on the Dino-Lite Edge 3.0 AM73915MZT LDA model was correlated with IDM ($r^2 = 0.91$) and with ASM ($r^2 = 0.87$), the latter of which was considered the fourth most important variable in that model. A measure of central tendency (median) was considered equally important by both LDA models. On the Sensofar S neox 090 images, the median was highly correlated with the mean, the Rq and the Ra ($r^2 = 0.98$, 0.98 and 1 respectively). This high level of correlation was shared in the Dino-Lite Edge 3.0 AM73915MZT LDA images along with the CONT textural feature (Fig. 9).

Exploratory visual analysis of the combination of variables and the degree of sedimentary abrasion in scatter plots makes it possible to observe a consistency in the evolution of quantitative features (Fig. 11; Fig. 12). For both sets of images there is a clear separation between images of fresh surfaces and images of neocortex. In addition to this consistency, the images of flint surfaces subjected to ten hours of rounding are located in between the fresh and neocortex images, indicating the directionality of the process.

4. Discussion

This study has two major and related outcomes: we demonstrated consistency in the quantitative characterization of a mechanical process (sedimentary abrasion of flints), and we compared the quality of the

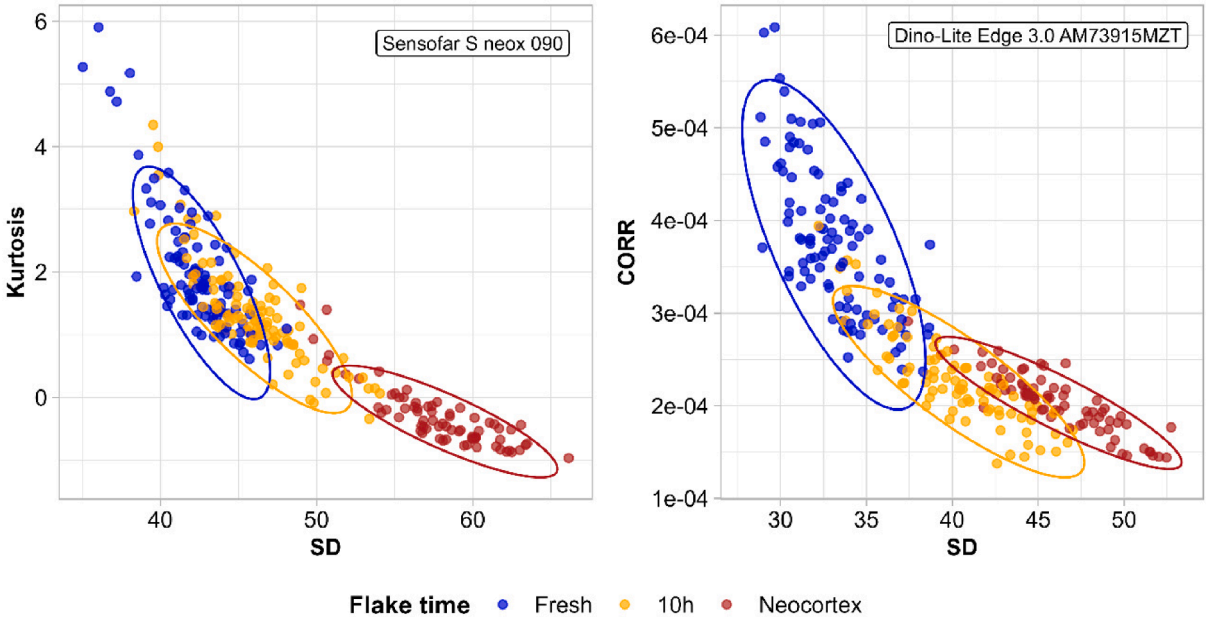

Fig. 11. Scatter plots showing the relationship of the first and second most important variables for each microscope set and according to degree of sedimentary abrasion.

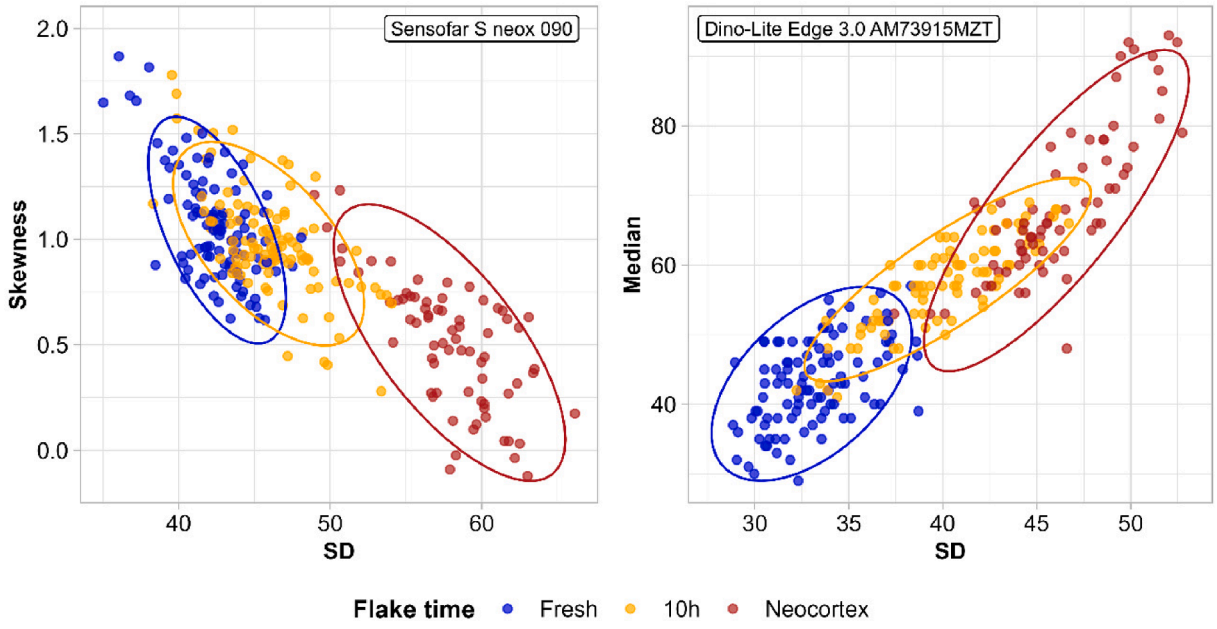

Fig. 12. Scatter plots showing the relationship of the first and third most important variables for each microscope set and according to degree of sedimentary abrasion.

quantitative features extracted from the images taken by two different microscopes.

The consistency of the quantitative characterization of a mechanical process was tested by comparing metric features extracted from the images taken by two different microscopes. Three aspects were found to contribute to this consistency: most of the extracted metric features presented the same trends in both microscopes; classification models (LDA) trained on images from the microscopes presented similar values of accuracy; and the same (or analogous/highly collinear) features were considered important for classification by the LDA models trained on each microscope. This indicates that the changes in metric features are correctly capturing changes in the surface due to a mechanical process.

The quality of the metric features extracted from the images from the two microscopes was compared through number of collinear features and classification accuracy of the LDA models. The Sensofar S neox 090 images can be considered to provide a better set of quantitative features, since there is a smaller number of collinear variables. Fewer correlated features are indicative of a more comprehensive characterization of a surface, since more distinct features are available (instead of collinear/redundant). However, the lower quality of the images (in the form of a higher presence of collinear variables) did not imply a drawback for the LDA models trained on the Dino-Lite Edge 3.0 AM73915MZT images, which presented classification performance values similar to those of LDA models trained on the Sensofar S neox 090 images. This indicates that the given task (classification of sedimentary abrasion stages) can be accurately performed with both microscopes despite differences in the quality of the quantitative features.

Pixel gray level values of the images might be the result of changes in the topography, presence of polish/abrasion, homogeneity of the raw material, or presence of reflective geological structures such as geodes, opals, etc. Ideally, for the present study, the image capture process of a microscope should not be affected by processes other than abrasion and changes in topography. Comparison of the images from both microscopes shows that Dino-Lite Edge 3.0 AM73915MZT is prone to be more affected by the presence of these reflective structures, thus affecting the pixel values of the grayscale level images. It can be considered that because of this, feature values extracted from Dino-Lite Edge 3.0 AM73915MZT images present a wider range and more marked trends, while still capturing the mechanical process of sedimentary abrasion. While the present workflow seems to adequately capture changes on flint surface due to sedimentary abrasion, it might not be suited (or require adaptation) for other more heterogeneous raw materials such as quartzites or quartz's.

Because chrono-cultural interpretations are derived from lithic archaeological assemblages, determining the integrity of such assemblages is a key aspect of Paleolithic archaeology (Dibble et al., 2006; Galland et al., 2019). Previous research has shown that unaided visual assessments of sedimentary abrasion (rounding) can result in an error ratio exceeding 80 % (Chambers, 2016). Therefore, the integrity of lithic assemblages must be determined through microscopic analysis, which, ideally, should study a range of variables and focus primarily on ridge width and surface abrasion (Burroni et al., 2002; Bustos-Pérez et al., 2019; Chambers, 2016; Chu and Hosfield, 2020; Shackley, 1974). Results from the present study reinforce the quantitative characterization of surface change due to sedimentary abrasion (Bustos-Pérez and Ollé, 2023). This ensures the reliability of the set of quantitative features extracted, and strengthens the methodological background of studies focusing on postdepositional processes using the quantification of surface alteration.

Although the three classes of surface studied here were *a priori* easily differentiable, there was a certain degree of confusion in their classification. However, this should not be considered a problem. Previous work (Bustos-Pérez and Ollé, 2023) has shown that abrasion develops unevenly among lithic artifacts. In some images, abrasion was minimally developed after ten hours (and classified as fresh), while in others abrasion was heavily developed after the same treatment (up to the point they were identified as neocortex).

Current research on use-wear combines quantitative features extracted from microscope images and machine learning models to generate information on feature importance (Ibáñez and Mazzucco, 2021; Pedergnana et al., 2020; Sferrazza, 2023). However, the present study has shown the importance of testing for multicollinearity among the quantitative features extracted from microscope images. Previous research (Bustos-Pérez and Ollé, 2023) considered CONT, Rq and SD as three of the four top variables according to mean importance. The current study shows that these three variables are highly correlated. While the present research shows that this does not affect results of classification, it does attenuate the interpretation of feature importance.

Therefore, although it does not affect classification performance, collinearity between selected features should always be considered prior to variable importance interpretation. This is a common aspect for machine learning workflows (James et al., 2013; Naes and Mevik, 2001; Yang and Yang, 2003).

When comparing the distribution values of the images taken by the two microscopes (Figs. 4 to 7), the Dino-Lite Edge 3.0 AM73915MZT images seem to be more sensitive to changes due to sedimentary abrasion with more visible trends in quantitative variable changes. For the Sensofar S neox images, no statistical differences were found in six of fifteen features on fresh flint and flint that had undergone 10 h of rounding. However, this did not result in a problem for the LDA models. One of the advantages of machine learning models resides in their ability to combine multiple features for classification. Features with no statistical differences between groups might become important when combined with an additional feature. Thus, it comes as no surprise that the LDA model trained on the Sensofar S neox selected a measure of central tendency (median) as an important feature for classification.

An additional possible explanation is that the variables employed for calculation were not adequate for the Sensofar S neox and are better suited for images from the Dinolite. Our observation of trends indicates that, for the given task, and under the given parameters, the set of quantitative variables seems to be better suited for images coming from the Dino-Lite Edge 3.0 AM73915MZT. Analyses using confocal mode commonly implement the ISO 25178 or ISO 25178–2 (Ibáñez and Mazzucco, 2021; Pedergnana et al., 2020) with surface parameters (Sq, Sv, Str, etc.) differing from the roughness parameters (Rq, Ra, Rsk, and Rku) employed in the present study (ISO 4287/2000). This could be indicative that the ISO 4287/2000 parameters are better suited for the Dino-Lite Edge 3.0 AM73915MZT, while those of the ISO 25178 are better suited for the Sensofar S neox.

5. Conclusions

Postdepositional alterations and assemblage integrity are fundamental analyses in the study of Paleolithic lithic assemblages. Microscopic analysis plays a crucial role in accurately assessing the extent of postdepositional alteration in lithic artifacts, and this degree of alteration can be determined by examining the abrasive erosion of flint surfaces. Therefore, it is critical to ensure that the quantitative features of this mechanical process are correctly characterized to obtain reliable results. In the present study, the metric features extracted from the images generated by two microscopes presented similar trends, similar classification accuracy, and similar variable importance. This made it possible to address the issue of *microscope agnosticism* regarding the development of surface abrasion. Additionally, using the workflow presented here, the quality of quantitative variables extracted from microscope images can be evaluated. In order to compare the quality of the quantitative features extracted from the images taken by two different microscopes, it is important to: 1) compare observed trends among the quantitative variables; 2) compare collinearity among variables (generally speaking, images from microscopes with fewer collinear variables are better); 3) compare classification accuracy among same models; 4) compare feature importance considering collinearity among the predictors.

CRediT authorship contribution statement

Guillermo Bustos-Pérez: Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization. **Andreu Ollé:** Writing – review & editing, Supervision, Resources, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The link to data and codee is shared at the file

Acknowledgments

The authors wish to thank both reviewers for their comments, time and very constructive feedback. Guillermo Bustos-Pérez is postdoctoral researcher at the Max Planck Institute for Evolutionary Anthropology at the Department of Human Origins lead by Tracy Kivell. The authors would also like to thank Juan Luis Fernández-Marchena for his comments and suggestions during the development of the present work. The following research has been possible thanks to the Program for the Requalification of the University System Margarita Salas (CA1/RSUE/2021-00743) financed through the Spanish "Recovery, Transformation and Resilience Plan" and managed from the Ministry of Universities (Ministerio de Universidades) and the Autonomous University of Madrid (Universidad Autónoma de Madrid). This work has been carried out with the financial support of the Generalitat de Catalunya, AGAUR agency (2021SGR01239 Research Group), Universitat Rovira i Virgili (2022PFR-URV-64), and the PID2021-122355NB-C32 project, funded by MCIN/AEI/10.13039/501100011033 and by "ERDF: A way of making Europe". The Institut Català de Paleoecologia Humana i Evolución Social (IPHES-CERCA) has received financial support from the Spanish Ministry of Science and Innovation through the "María de Maeztu" program for Units of Excellence (CEX2019-000945-M). The research technical support of Maria Dolors Guillén was supported by the Spanish Ministry of Science and Innovation through the "María de Maeztu" excellence accreditation (CEX2019-000945-M). This work is also related to project reference PID2022-138590NB-C42, financed by the Agencia Estatal de Investigación: "On the limits of diversity: Neanderthal behavior in the central and southern Iberian Peninsula (2)"

References

Bradley, A.P., 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recogn. 30, 1145–1159.

Burroni, D., Donahue, R.E., Pollard, A.M., 2002. The Surface Alteration Features of Flint Artefacts as a Record of Environmental Processes. J. Archaeol. Sci. 29, 1277–1287. https://doi.org/10.1006/jasc.2001.0771.

Bustillo, M.A., Pérez-Jiménez, J.L., 2005. Características diferenciales y génesis de los niveles silíceos explotados en el yacimiento arqueológico de Casa Montero (Vicálvaro, Madrid). Geogaceta 38, 243–246.

Bustillo, M.Á., Pérez-Jiménez, J.L., Bustillo, M., 2012. Caracterización geoquímica de rocas sedimentarias formadas por silicificación como fuentes de suministro de utensilios líticos (Mioceno, cuenca de Madrid). Revista Mexicana De Ciencias Geológicas 29, 233–247.

Bustos-Pérez, G., Díaz, S., Baena, J., 2019. An Experimental Approach to Degrees of Rounding Among Lithic Artifacts. J. Archaeol. Method Theory 26, 1243–1275. https://doi.org/10.1007/s10816-018-9409-8.

Bustos-Pérez, G., Ollé, A., 2023. The quantification of surface abrasion on flint stone tools. Archaeometry N/a. https://doi.org/10.1111/arcm.12913.

Byers, D.A., Hargiss, E., Finley, J.B., 2015. Flake Morphology, Fluvial Dynamics, and Debitage Transport Potential. Geoarchaeology 30, 379–392. https://doi.org/10.1002/gea.21524.

Calandra, I., Bob, K., Merceron, G., Blateyron, F., Hildebrandt, A., Schulz-Kornas, E., Souron, A., Winkler, D.E., 2022. Surface texture analysis in Toothfrax and MountainsMap® SSFA module: Different software packages, different results? Peer Community Journal 2, e77.

Caux, S., Galland, A., Queffelec, A., Bordes, J.-G., 2018. Aspects and characterization of chert alteration in an archaeological context: A qualitative to quantitative pilot study. J. Archaeol. Sci. Rep. 20, 210–219. https://doi.org/10.1016/j.jasrep.2018.04.027.

Chambers, J.C., 2016. Like a rolling stone? The identification of fluvial transportation damage signatures on secondary context bifaces. Lithics 24, 66–77.

Chinga, G., Dougherty, B., 2002. Roughness Calculation.

Chinga, G., Johnsen, P.O., Dougherty, R., Berli, E.L., Walter, J., 2007. Quantification of the 3D microstructure of SC surfaces. J Microsc 227, 254–265. https://doi.org/10.1111/j.1365-2818.2007.01809.x.

Chu, W., Hosfield, R., 2020. Lithic artifact assemblage transport and microwear modification in a fluvial setting: A radio frequency identification tag experiment. Geoarchaeology 35, 591–608. https://doi.org/10.1002/gea.21788.
Collins, T.J., 2007. ImageJ for microscopy. Biotechniques 43, S25–S30. https://doi.org/10.2144/000112517.
Dibble, H.L., McPherron, S.J.P., Chase, P., Farrand, W.R., Debénath, A., 2006. Taphonomy and the Concept of Paleolithic Cultures: The Case of the Tayacian from Fontéchevade. PaleoAnthropology 2006, 1–21.
Evans, A.A., Donahue, R.E., 2008. Laser scanning confocal microscopy: a potential technique for the study of lithic microwear. J. Archaeol. Sci. 35, 2223–2230. https://doi.org/10.1016/j.jas.2008.02.006.
Fisher, R.A., 1936. The use of multiple measurements in taxonomic problems. Ann. Eugen. 7, 179–188.
Fraile-Márquez, C., Diez-Martín, F., Duque-Martínez, J., Uribelarrea, D., Sánchez-Yustos, P., de Francisco, S., Baquedano, E., Mabulla, A., Domínguez-Rodrigo, M., 2022. Facing the palimpsest conundrum: an archaeo-stratigraphic approach to the intra-site analysis of SHK Extension (Bed II, Olduvai Gorge, Tanzania). Archaeol Anthropol Sci 14, 230. https://doi.org/10.1007/s12520-022-01691-3.
Galland, A., Queffelec, A., Caux, S., Bordes, J.-G., 2019. Quantifying lithic surface alterations using confocal microscopy and its relevance for exploring the Châtelperronian at La Roche-à-Pierrot (Saint-Césaire, France). J. Archaeol. Sci. 104, 45–55. https://doi.org/10.1016/j.jas.2019.01.009.
Hand, D.J., Till, R.J., 2001. A simple generalisation of the area under the ROC curve for multiple class classification problems. Mach. Learn. 45, 171–186. https://doi.org/10.1023/A:1010920819831.
Haralick, R.M., Shanmugam, K., Dinstein, I.H., 1973. Textural features for image classification. IEEE Trans. Syst. Man Cybern. 3, 610–621.
Hiscock, P., 1985. The Need for a Taphonomic Perspective in Stone Artefact Analysis. Queensland Archaeological Research 2, 82–95.
Hosfield, R.T., Chambers, J.C., 2016. Flake modifications during fluvial transportation: three cautionary tales. Lithics 24, 57–65.
Hosfield, R.T., Chambers, J.C., Macklin, M., Brewer, P., Sear, D., 2000. Interpreting secondary context'sites': a role for experimental archaeology. Lithics 21, 29–35.
Ibáñez, J.J., Lazuen, T., González-Urquijo, J., 2019. Identifying Experimental Tool Use Through Confocal Microscopy. J Archaeol Method Theory 26, 1176–1215. https://doi.org/10.1007/s10816-018-9408-9.
Ibáñez, J.J., Mazzucco, N., 2021. Quantitative use-wear analysis of stone tools: Measuring how the intensity of use affects the identification of the worked material. PLoS One 16, e0257266.
James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning with Applications in R, Second Edition. ed, Springer Texts in Statistics. New York: springer.
Kuhn, M., 2008. Building Predictive Models in R using the caret Package. J. Stat. Softw. 28, 1–26. https://doi.org/10.18637/jss.v028.i05.
Lantz, B., 2019. Machine learning with R: expert techniques for predictive modeling. Packt publishing ltd, Birmingham.
Levi Sala, I., 1986. Use Wear and Post-depositional Surface Modification: A Word of Caution. J. Archaeol. Sci. 13, 229–244.
Macdonald, D.A., 2014. The application of focus variation microscopy for lithic use-wear quantification. J. Archaeol. Sci. 48, 26–33. https://doi.org/10.1016/j.jas.2013.10.003.
Michel, M., Cnuts, D., Rots, V., 2019. Freezing in-sight: the effect of frost cycles on use-wear and residues on flint tools. Archaeol Anthropol Sci 11, 5423–5443. https://doi.org/10.1007/s12520-019-00881-w.
Naes, T., Mevik, B.-H., 2001. Understanding the collinearity problem in regression and discriminant analysis. J. Chemom. 15, 413–426. https://doi.org/10.1002/cem.676.
Ollé, A., Vergès, J.M., 2014. The use of sequential experiments and SEM in documenting stone tool microwear. J. Archaeol. Sci. 48, 60–72. https://doi.org/10.1016/j.jas.2013.10.028.
Pearson, K., 1901. On lines and planes of closest fit to systems of points in space. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 2, 559–572. https://doi.org/10.1080/14786440109462720.
Pedergnana, A., Asryan, L., Fernández-Marchena, J.L., Ollé, A., 2016. Modern contaminants affecting microscopic residue analysis on stone tools: A word of caution. Micron 86, 1–21. https://doi.org/10.1016/j.micron.2016.04.003.
Pedergnana, A., Calandra, I., Evans, A.A., Bob, K., Hildebrandt, A., Ollé, A., 2020. Polish is quantitatively different on quartzite flakes used on different worked materials. PLoS One 15, e0243295.
Petraglia, M.D., Potts, R., 1994. Water Flow and the Formation of Early Pleistocene Artifact Sites in Olduvai Gorge, Tanzania. J. Anthropol. Archaeol. 13, 228–254. https://doi.org/10.1006/jaar.1994.1014.
R Core Team, 2019. R: A language and environment for statistical computing.
Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., Müller, M., 2011. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinf. 12, 1–8.
RStudio Team, 2019. RStudio: Integrated Development for R.
Schick, K.D., 1986. Stone Age Sites in the Making. Experiments in the Formation and Transformation of Archaeological Occurrences. BAR International Series, Oxford.
Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., 2012. Fiji: an open-source platform for biological-image analysis. Nat. Methods 9, 676–682.
Schindelin, J., Rueden, C.T., Hiner, M.C., Eliceiri, K.W., 2015. The ImageJ ecosystem: An open platform for biomedical image analysis. Molecular Reproduction Devel 82, 518–529. https://doi.org/10.1002/mrd.22489.
Sferrazza, P., 2023. Grey level co-occurrence matrix and learning algorithms to quantify and classify use-wear on experimental flint tools. J. Archaeol. Sci. Rep. 48, 103869. https://doi.org/10.1016/j.jasrep.2023.103869.
Shackley, M.L., 1974. Stream abrasion of flint implements. Nature 248, 501–502. https://doi.org/10.1038/248501a0.
Spackman, K.A., 1989. Signal detection theory: Valuable tools for evaluating inductive learning, in: Proceedings of the Sixth International Workshop on Machine Learning. Elsevier, pp. 160–163.
Staurset, S., Coulson, S.D., Mothulatshipi, S., Burrough, S.L., Nash, D.J., Thomas, D.S.G., 2023. Post-depositional disturbance and spatial organization at exposed open-air sites: Examples from the Middle Stone Age of the Makgadikgadi Basin. Botswana. Quaternary Science Reviews 301, 107824. https://doi.org/10.1016/j.quascirev.2022.107824.
Stemp, W.J., Childs, B.E., Vionnet, S., Brown, C.A., 2008. The Quantification of Microwear on Chipped Stone Tools: Assessing the Effectiveness of Root Mean Square Roughness (Rq). Lithic Technol. 33, 173–189. https://doi.org/10.1080/01977261.2008.11721067.
Stemp, W.J., Chung, S., 2011. Discrimination of surface wear on obsidian tools using LSCM and RelA: pilot study results (area-scale analysis of obsidian tool surfaces). Scanning 33, 279–293. https://doi.org/10.1002/sca.20250.
Stemp, W.J., Watson, A.S., Evans, A.A., 2016. Surface analysis of stone and bone tools. Surf. Topogr. Metrol. Prop. 4, 013001. https://doi.org/10.1088/2051-672X/4/1/013001.
Stevens, N.E., Harro, D.R., Hicklin, A., 2010. Practical quantitative lithic use-wear analysis using multiple classifiers. J. Archaeol. Sci. 37, 2671–2678. https://doi.org/10.1016/j.jas.2010.06.004.
Student, 1908. The Probable Error of a Mean. Biometrika 6, 1–25. https://doi.org/10.2307/2331554.
Venables, W.N., Ripley, B.D., 2002. Modern applied statistics with S, Fourth, Edition. ed. Statistics and Computing, Springer, New York.
Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H., 2019. Welcome to the Tidyverse. JOSS 4, 1686. https://doi.org/10.21105/joss.01686.
Xie, Y., Allaire, J.J., Grolemund, G., 2018. R markdown: The definitive guide. CRC Press.
Xie, Y., 2014. knitr: A Comprehensive Tool for Reproducible Research in R, in: Implementing Reproducible Research. Chapman and Hall/CRC.
Yang, J., Yang, J.-y., 2003. Why can LDA be performed in PCA transformed space? Pattern Recognition, Biometrics 36, 563–566. https://doi.org/10.1016/S0031-3203(02)00048-1.