Received: 24 April 2023 | Accepted: 31 August 2023

DOI: 10.1111/arcm.12913

ORIGINAL ARTICLE

The quantification of surface abrasion on flint stone tools

Guillermo Bustos-Pérez[1,2,3] | **Andreu Ollé**[2,3]

[1]Departamento de Prehistoria y Arqueología, Universidad Autónoma de Madrid, Madrid, Spain

[2]Institut Català de Paleoecologia Humana i Evolució Social (IPHES-CERCA), Tarragona, Spain

[3]Departament d'Història i Història de l'Art, Universitat Rovira i Virgili, Tarragona, Spain

Correspondence
Guillermo Bustos-Pérez, Departamento de Prehistoria y Arqueología, Universidad Autónoma de Madrid, Campus de Cantoblanco, 28049 Madrid, Spain.
Email: guillermo.bustos@uam.es

Funding information
Program for the Requalification of the University System Margarita Salas, Grant/Award Number: CA1/RSUE/2021-00743; Ministry of Universities (Ministerio de Universidades); Autonomous University of Madrid (Universidad Autónoma de Madrid); Generalitat de Catalunya, AGAUR agency, Grant/Award Number: 2021SGR01239; Universitat Rovira i Virgili, Grant/Award Number: 2022PFR-URV-64; Spanish Ministry of Science and Innovation, Grant/Award Numbers: CEX2019-000945-M, PID2021-122355NB-C32; Agencia Estatal de Investigación, Grant/Award Number: HAR2016-76760-C3-2-P; Spanish National Plan for Scientific and Technical Research and Innovation, Grant/Award Number: ID2019-103987 GB-C33

Abstract

Lithic artifacts are some of the most common and numerous remains recovered from paleolithic archaeological sites. However, these materials can undergo multiple post-depositional alterations after their introduction into the archaeological record. Due to the high quantity of lithic remains recovered, a quick, flexible, and effective method for identifying degrees of alteration on the surface of lithic implements is highly desirable. The present study examines the use of gray level images to obtain quantitative data from the surface of flint artifacts and determine whether these images can detect the presence of post-depositional alterations. An experimental collection of flints was subjected to sequential episodes of rounding in a tumbling machine. After each episode, photographs were taken with a microscope, resulting in quantitative surface values using gray level values. The quantitative surface values were used as variables in machine learning models to determine time of exposure and the most salient variables for discrimination. Our results indicate that the extraction of metrics from gray level images successfully capture changes in the surface of flint artifacts caused by post-depositional processes. Additional results provide insight into which areas to sample when seeking post-depositional alterations and underscore the importance of particle size in the generation of alterations.

KEYWORDS

experimental archaeology, lithic taphonomy, machine learning

INTRODUCTION

Some of the most common remains recovered from paleolithic archaeological sites are lithic artifacts. Because eco-cultural inferences can be drawn from lithic artifacts, post-depositional alterations are a key factor to consider when analyzing an assemblage. Methods for determining assemblage integrity usually focus on the assemblage as a whole, and often include a spatial analysis of the artifacts as well as fabric analysis to determine whether water flow has reorganized the spatial distribution and orientations of the pieces (Lenoble & Bertran, 2004; McPherron, 2005, 2018; Petraglia & Potts, 1994; Schick, 1987). Also, the size distribution of the lithics is analyzed to determine whether post-depositional processes have resulted in the sorting of materials (Bertran et al., 2012; Byers et al., 2015; Hiscock, 2002; Maíllo Fernández, 1998). Another option is to focus specifically on individual lithic artifacts to determine the degree of alteration that they have undergone (Chu et al., 2015; Levi Sala, 1986). Although more time consuming, this type of analysis establishes an individual graduation in the degree of post-depositional alterations. It can also be combined with spatial analysis to potentially dissect distinct archaeological episodes or detect instances of recycling. The individual determination of the degree of alteration undergone by a lithic artifact makes use of a microscopic analysis to measure ridge width, a visual analysis of the surface to determine the existence and intensity of abrasion, and an examination of the edges to determine the presence of detachments caused by particle impact, or dulling in more extreme cases (Bustos-Pérez et al., 2019; Chambers, 2003; Shackley, 1974).

Broadly speaking, how post-depositional alterations can affect and interact with stone tools depends on two factors: first, the types of fluvial sedimentary processes an artifact might have encountered; and, secondly, the degree to which an artifact has been exposed to the source of alteration and when it enters into the archaeological record. Fluvial sedimentary processes are characterized by particle transport. Lithic artifacts can form part of these processes as another component, or remain static and be affected by the particles being transported across their surfaces. Commonly, three modes of particle transport are described in fluvial sedimentary contexts— rolling, sliding, and saltation (Alhusban & Valyrakis, 2021)—although alterations cause by saltation are rarely seen on stone tools (Petraglia & Potts, 1994). In addition to alterations resulting from fluvial contexts, stone artifacts might also be affected by aeolian particle transport, which results in wind abrasion (Stapert, 1976). The second factor of variability affecting the post-depositional alterations of stone tools is the degree of exposure they have had and the speed with which they enter the archaeological record (Petraglia & Potts, 1994; Schick, 1987; Schiffer, 1972). For example, partially buried artifacts with water and sediment flowing above them are expected to present modifications on the exposed surface, while the edges and the buried surface will remain semi-intact (Petraglia & Potts, 1994). Meanwhile, artifacts that are transported by rolling in coarse sediments will exhibit abrasion across their entire surface and have dulled edges. However, the impact from coarse particles might result in freshly detached surfaces and edges, which then undergo new modifications until the artifact enters the archaeological record (Harding et al., 1987; Petraglia & Potts, 1994). Finally, the same homogeneous lithic assemblage might be affected differentially by stream abrasion. This differential alteration is a consequence of the complex structure of water streams, in which the overall slope of the terrain, energy and charge of the flow, morphology of the channel (which also affects the lateral deposition of sediments), and several other factors result in complex and uneven structures (Jain et al., 2008; Montgomery & Buffington, 1997; Rust, 1972).

Use-wear studies have shown that quantifying surface alterations can offer a higher resolution of analysis (Boström & Lundin, 2022; Evans & Donahue, 2008; Ibáñez et al., 2016, 2019; Ibáñez & Mazzucco, 2021; Pedergnana et al., 2020; Stemp et al., 2009). Obtaining quantitative data on surface modifications caused by post-depositional processes can also increase the resolution of these analyses and complement other features of post-depositional studies. However, as previously mentioned, stone artifacts are some of the most common remains recovered from

paleolithic archaeological sites. Thus, a fast, versatile method is highly desirable since it would enable the analysis of large quantities of lithic materials. Previous approaches in the study of use-wear have used gray values to analyze the surfaces of different worked materials (Adán et al., 2003; Barceló et al., 2001; Bietti, 1996; Grace et al., 1985; Pijoan-López et al., 2002; Vila & Gallart, 1993). The present research seeks to determine whether changes in flint surface caused by sedimentary abrasion can be captured and quantified from grayscale images.

METHODS

Experimental sample and setting

Two blocks of flint representing two different types (Type I and Type II) were experimentally knapped. Both types are south Madrid Miocene flint (Bustillo et al., 2012; Bustillo & Pérez-Jiménez, 2005) although sourced from two different locations. South Madrid Miocene flints were formed by the replacement of sedimentary rocks which had filled the original basin. This replacement of the sedimentary rocks is considered to have taken place under continental conditions such as alluvial plain deposits, shallow lacustrine waters, and marshes (Bustillo et al., 2012). The macroscopic analysis of the flints evidences a fine opaque homogeneous surface that is blue/gray and reddish/ocher in color. There is also a relative absence of opal, although geodes and pseudo-morphs can be observed. From these two blocks, three flakes were selected from Type 1, and four flakes were selected from Type 2 (Figure S1).

In order to simulate sedimentary abrasion, the flakes were introduced in pairs into a tumbler machine (KT-3010 Super-Tumbler) along with a mix of sand and water (30/40% water with a total weight of 5 kg). The sediment was obtained from the quaternary levels of the Madrid basin and is made up of fine sands with silt and partial carbonation. The tumbler machine was set to turn in a continuous direction at 83 rpm.

All of the flakes were submitted to three cumulative cycles of tumbling with times set to 1, 5, and 10 h. Prior to placement in the tumbling machine, six photographs (three on each side) were taken of the surface of each flake to act as references for the texture metrics of the flint flakes. Previous research (Bustos-Pérez et al., 2019; Chu et al., 2015) suggests that for the development of post-depositional alterations decreased particle size results in increased heterogeneity in the speed, intensity, and location of their development. Thus, after the first cycle of rounding (1 h), each flint flake was screened using a Dino-Lite Edge 3.0 AM73915MZT USB microscope in order to determine and photograph areas that had developed sedimentary abrasion. These areas were photographed in the subsequent rounding cycles, thereby generating sequential images and data on surface changes for cumulative times of 1, 5, and 10 h. The dataset of images was supplemented by the inclusion of sample photographs of macroscopically recognizable geological neocortex of flints from the same formation, resulting in a total of 269 photographs: 42 of fresh surfaces, 68 from each cycle of rounding (i.e., 204 in total), and 23 of neocortex.

Cleaning protocol, image acquisition and processing

Prior to extracting the quantitative data from the images, the workflow employed in the present study included a series of steps: cleaning protocol, image capture, and image enhancement.

Multiple studies have emphasized the need for cleaning protocols to remove modern contaminants prior to analysis (Asryan & Ollé, 2020; Fernández-Marchena & Ollé, 2016; Ollé & Vergès, 2014; Pedergnana et al., 2016). A multi-step procedure based on previous studies was adopted (Pedergnana et al., 2016) in order to remove possible contaminants (Figure S2). This multi-step procedure included a sonic bath in 2% neutral soap (Derquim) solution for 10–15 min,

followed by a second sonic bath in pure acetone for a further 10–15 min. After each step, the lithic artifacts were placed in a water bath and finally dried with pressurized air. Surgical gloves were worn during the handling of all the artifacts during the cleaning protocol and microscopic analysis.

All surface photographs were taken using a Dino-Lite Edge 3.0 AM73915MZT USB microscope at 120 magnifications with a field of view of 3.28 × 2.46 mm and a pixel ratio of 2548 × 1918. The USB 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, 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.

A common problem of images obtained from USB microscopes is the lack of detail due to saturation in one of the gray level values and the effects of different lighting or surface color. This saturation is often observed as a general glaze in one of the gray level values and results in a low-quality image with poor detail (Figure 1, left). To increase detail and quality prior to the analysis, all images were subjected to a two-step process. First, the Fiji (Schindelin et al., 2012) "Subtract background" plugin was used to minimize the effects of different lighting and differing flint coloration. Second, we used the "Enhance contrast" function to desaturate the images by normalizing their histograms (Figure 1). This process provided a gray level image for use as input for the statistical analysis. All analyzed images were in TIFF format.

Statistical analysis

The present study used three sets of statistical metrics to analyze the images: descriptive statistics (mean, standard deviation, mode, median, skewness, and kurtosis); measures of roughness; and measures of intensity values.

The surface parameters were analyzed using the "R" prefix for the profiles as input. The present work used a Fiji/ImageJ plugin with which *R*-values were obtained over the whole surface

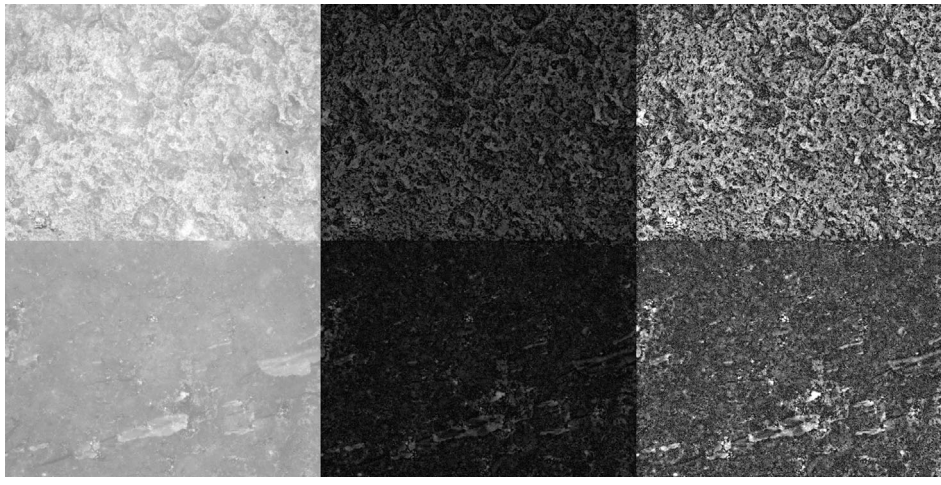

FIGURE 1 Two examples of image enhancement. Left: original images taken with the AM73915MZT USB microscope and using a N3C-D2 diffuser cap. Center: images after removing the background to avoid the effects of different lighting or color. Right: after normalizing the histogram to increase detail and prevent saturation. Top row: geological neocortex. Bottom row: fresh surface of an experimentally knapped flake.

following the ISO 4287/2000 standard (Chinga et al., 2007; Chinga & Dougherty, 2002). These measures were:

Root mean square deviation/roughness (Rq): indicator of surface roughness.
Arithmetical mean deviation (Ra): indicator of the deviation of a surface from a mean height.
Skewness of the assessed profile (Rsk): indicator of the departure from surface symmetry. Negative values indicate a surface made of deep valleys, and positive values indicate peaks and asperities.
Kurtosis of the assessed profile (Rku): indicator of the sharpness of the peaks. Low values indicate blunt peaks, while high values indicate sharp peaks.

The intensity values were analyzed with the gray level co-occurrence matrix (GLCM; Haralick et al., 1973), which takes into consideration the spatial distribution of the intensity values. The GLCM works in two steps (Haralick et al., 1973). First, using a given distance and direction, it builds a matrix 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. From this new matrix, a series of statistical descriptors are derived (Haralick et al., 1973):

Angular second moment (ASM) is a measure of homogeneity in the image. Homogeneous images (with low gray-tone transitions) will have fewer high-magnitude entries. Thus, homogeneous images will have high ASM values, while the opposite will be true for non-homogeneous images.
Contrast (CONT) is a value for local variations. High values indicate more local variation and low values indicate less local variation.
Correlation (CORR) measures gray-tone linear dependencies in the image. It indicates how a reference pixel is related to its neighbor. A value of 0 indicates that it is uncorrelated, while a 1 indicates a perfect correlation.
Inverse different moment (IDM), also referred to as **homogeneity**, measures the closeness of the distribution of the GLCM elements to the GLCM diagonal.
Entropy (ENT) is a measure of the amount of irremediable chaos or disorder in an image. High values of entropy indicate values of similar magnitude, while low values indicate unequal entries.

As previously mentioned, the use of the GLCM requires selecting the pixel distance between the reference and neighbor, and the direction at which to establish the distance (Haralick et al., 1973). To do this, it is common to test for different combinations of distances and directions, as images at different magnifications, different fields of view, and different resolutions might require variations in pixel distance (Bietti, 1996; Grace et al., 1985). For the present study, a preliminary test indicated that using four distances at 5, 10, 15 and 20 pixels in the four possible directions (north, east, south, and west) to set the GLCM presented the best results for discrimination.

All metrics were calculated using the free software Fiji (Schindelin et al., 2012). Roughness metrics were calculated with the "Roughness calculation" plugin (Chinga & Dougherty, 2002). GLCM and texture metrics were calculated using the "GLCM texture" plugin (Cabrera, 2006).

Machine learning models and evaluation

The data from the descriptive statistics and the roughness and texture analyses were used as variables for the training of machine learning models in order to predict the time of exposure to sedimentary abrasion. The 10 models tested are described below.

Linear discriminant analysis (LDA) reduces dimensionality for the purpose of maximizing the separation between classes, while decision boundaries divide the predictor range into regions (Fisher, 1936; James et al., 2013).

The K-nearest neighbor (KNN) classifies cases by assigning the class of similar known cases. The "K" in KNN refers to the number of cases (neighbors) to consider when assigning a class, which must be found by testing different values. Given that KNN uses distance metrics to compute nearest neighbors and that each variable is at different scales, the data must be scaled and centered prior to fitting the model (Cover & Hart, 1967; Lantz, 2019).

Logistic regression essentially adapts continuous regression predictions to categorical outcomes (Cramer, 2004; Walker & Duncan, 1967).

The decision tree with the C5.0 algorithm is an improvement on decision trees for classification (Quinlan, 1996, 2014).

A *random forest* is made up of decision trees with each tree grown from a random sample of the data and variables, allowing for each tree to grow differently and to better reflect the complexity of the data (Breiman, 2001).

The *generalized boosted model* (Greenwell et al., 2019; Ridgeway, 2007) implements the gradient boosted machine (Friedman, 2001, 2002), making it possible to detect learning deficiencies and increase model accuracy for a set of random forests.

Support vector machines (SVMs) fit hyperplanes into a multidimensional space with the objective of creating homogeneous partitions (Cortes & Vapnik, 1995; Frey & Slate, 1991). The present study tests SVMs with linear, radial, and polynomial kernels.

The *naïve Bayes classifier* computes class probabilities using Bayes' rule (Weihs et al., 2005).

All models were evaluated using 10×100 *k*-fold cross-validation (10 folds and 100 cycles), providing measures of accuracy. Using a 10-fold division, each fold has 43 data points. Each fold subsequently acts as a test set for a trained model. Although computationally more expensive, this guarantees that all data points will serve as test sets. The 100 cycles randomly shuffle the dataset prior to fold division, thus ensuring that the composition of the folds varies in each cycle and that composition does not play a significant role in the evaluation of the models.

Machine learning models commonly use a 0.5 classification threshold to assign categories. However, classification thresholds can be modified to balance the ability of the model to detect true positives and avoid false positives, which are respectively referred to as sensitivity and specificity. The receiver operating characteristic (ROC) curve is employed to systematically evaluate the ratio of detected true positives while avoiding false positives (Bradley, 1997; Spackman, 1989). The ROC curve allows model performance to be visually analyzed and the area under the curve (AUC) to be calculated from 1 (perfect classifier) to 0.5 (random classifier). The ROC and AUC are commonly applied in two-class problems, and are usually extended to multiclass problems through pairwise analysis (Figure S3). For multiclass problems, the AUC provides two groups of values: first, each class obtains an AUC value using a "one versus all" approach (Figure S3); second, a general AUC value of model performance is obtained from the average of each AUC class (Hand & Till, 2001; Robin et al., 2011). In the case of the ROC curve, individual curves of each class are plotted using the previously mentioned "one versus all" approach. The present study tested 10 different models with a three-class classification problem, which would involve a total of 30 different ROC curves (three curves per 10 models). In this paper, we have provided only the three ROC curves of the best model. When analyzing lithic materials, the use of thresholds to guarantee true positives and avoid false positives is of special interest. The use of thresholds better indicates the accuracy of a model considering these probability values. In the present study we have slightly varied the interpretation of the AUC values (Lantz, 2019) with intervals interpreted as follows:

0.9–1: outstanding
0.85–0.9: excellent
0.8–0.85: good
0.75–0.8: acceptable
0.7–0.75: fair
0.6–0.7: poor
0.5 to 0.6: no discrimination

High levels of correlation were found between the variables in this study (Figure S5). The issue of multicollinearity in classification remains a matter of debate. It is commonly suggested that for multiple linear regressions collinearity affects the interpretation of the coefficients (variables), but does not affect the quality of the predictions (Alin, 2010; Chan et al., 2022; James et al., 2013). Additional arguments indicate that if the collinearity between the variables of the training set is also present in the test set it should not be considered a problem. In this study, the variables presenting perfect levels of collinearity (mean and Ra) were excluded from the training of the machine learning models, and feature importance was explored among non-correlated features. After evaluating and determining the best model, an additional model on PCA values was trained in order to determine possible overfitting.

This study was conducted using an R version 4.2.2 in IDE RStudio version 2022.12 (R Core Team, 2019; RStudio Team, 2019). The data and graphs were managed using the tidyverse v.1.3.2 package (Wickham et al., 2019). LDA and KNN were trained using the MASS (Modern Applied Statistics with S) v.7.3.58.1 package (Venables & Ripley, 2002). The C5.0 tree was trained using the C50 v.0.1.7 package (Quinlan, 1996, 2014). The random forest was trained using the ranger v.0.14.1 package (Wright & Ziegler, 2017). The generalized boosted model was trained using the gbm v.2.1.8.1 package (Greenwell et al., 2019; Ridgeway, 2007). SVMs were trained using the e1071 v.1.7.12 package (Karatzoglou et al., 2004, 2006). The naïve Bayes model was trained using the klaR v.1.7.1 package (Weihs et al., 2005). The *k*-fold cross-validation of all models, precision metrics, and confusion matrix were obtained using the caret v.6.0.93 package (Kuhn, 2008). Machine learning models also provide insight into variable importance for classification. The caret package was used to extract variable importance after each *k*-fold cross-validation. ROC curves and AUC values were obtained using the pROC v.1.18.0 package (Robin et al., 2011). All data, code and the complete workflow to perform the analysis is made freely available as an open repository at Zenodo (https://doi.org/10.5281/zenodo.8359711).

RESULTS

Texture metrics

A general multivariate analysis of variance (MANOVA) considering all groups and variables showed marked statistically significant differences between groups (df = 4; $F = 8.947$; $p < 0.001$). A MANOVA analysis comparing a category with its subsequent time of exposure showed marked statistical differences between fresh materials and 1 h of rounding (df = 1; $F = 5.985$; $p < 0.001$); 1 h of rounding and 5 h (df = 1; $F = 2.761$; $p < 0.01$); and 10 h and neocortex (df = 1; $F = 19.799$; $p < 0.01$). When considering all variables, no statistical differences were found between the images of materials exposed to 5 and 10 h of rounding (df = 1; $F = 1.55$; $p = 0.09$). However, statistical differences between these two categories were documented for the mean (df = 1; $t = 6.4$, $p = 0.01$), median (df = 1; $t = 4.95$, $p = 0.03$), standard deviation (df = 1; $t = 7.34$, $p < 0.01$), kurtosis (df = 1; $t = 6.53$, $p = 0.01$), skewness (df = 1; $t = 6.29$, $p = 0.01$), Rq (df = 1; $t = 6.96$, $p < 0.001$), Ra (df = 1; $t = 6.65$, $p = 0.01$), Rku (df = 1; $t = 4.37$, $p = 0.04$), ASM ($df = 1$; $t = 6.67$, $p = 0.01$), contrast (df = 1; $t = 6.85$,

$p < 0.01$), correlation ($df = 1$; $t = 6.96$; $p = 0.02$), IDM ($df = 1$; $t = 6.87$, $p < 0.01$), and entropy ($df = 1$; $t = 6.8$, $p = 0.01$).

An exploratory visual analysis revealed a series of clear tendencies for the three groups of statistics employed (Figure 2). In general terms, as sedimentary abrasion increased, the images presented increasing mean, median, standard deviation, Ra, Rq, contrast (CONT), and entropy (ENT) values. In contrast, as sedimentary abrasion increased, the images presented decreasing kurtosis, skewness, Rku, Rsk, angular second moment (ASM), correlation between pixels (CORR), and inverse different moment (IDM) values. Central tendency measures (mean, median, and modal) were found to be the least reliable, since the three cumulative abrasion times show considerable overlapping values for these statistic variables. The effects of sedimentary abrasion were especially observable in the data dispersion variables (kurtosis, skewness, and standard deviation) or in the five textural features (angular second moment, ASM; contrast, CONT; correlation, CORR; entropy, ENT; and inverse different moment, IDM). As sedimentary abrasion increased, the images became less homogeneous with increasing amounts of local changes, which may be related to increased roughness.

Although the trends in surface changes caused by sedimentary abrasion are clear, the exploratory visual analysis also revealed considerable overlapping between exposure time categories.

A visual evaluation of the images showed that the development of sedimentary abrasion is heterogeneous on the lithic artifacts (Figures 3 and 4). Convex surfaces and areas close to the ridges and edges developed abrasion more quickly and more intensely than other areas.

Machine learning model results

Figure 5 presents the performance metrics (general precision and AUC) for each of the machine learning models after the 100 × 10-fold cross-validation. Although all of the models presented general precision values below 0.5, these values were in all cases substantially higher than the "no-information rate" (0.296). Of the 10 models tested, logistic regression presented the highest general precision value (0.485), though the value of the LDA was very similar (0.479). The random forest and decision tree with C5.0 presented the lowest general precision rates, with respective values of 0.381 and 0.402. The use of the ROC curves for model evaluation indicated that all of the models presented good or acceptable general AUC values, with the exception of the SVM with polynomial kernel, which presented a poor general AUC (0.68). LDA yielded the highest AUC value (0.83)—substantially higher than the logistic regression analysis (0.819). Thus, the LDA model performed the best in differentiating degree of sedimentary abrasion.

Figure 6 presents the ROC curves and AUCs obtained for each category of sedimentary abrasion using the LDA model. Individual AUC values for fresh surfaces and neocortex were outstanding (respective values of 0.9 and 0.98). Individual AUC values for the different exposure times varied from fair for 1 (0.71) and 10 (0.74) hours of exposure to poor for 5 h (0.62) of exposure. The confusion matrix (Figure 6, right) provides insight into the sources and directionality of the confusions. As exposure time increased, the number of photographs identified as fresh diminished, with a minimum portion (1.66) being identified as fresh after 10 h of rounding. Similarly, a small number of photographs of the materials after 10 h of rounding started to resemble the photos of neocortex (11.91). The confusion matrix reinforces previous observations through exploratory visual analysis and direct examination of the photographs that sedimentary abrasion does not develop evenly, as some surfaces from the same flake develop abrasion faster than others. Nevertheless, we found a clear directionality (as sedimentary abrasion increased, the number of photographs identified as fresh decreased). Additional training of an LDA model on values from the first five PCs (99% of the variance) showed little change in general precision (0.464) or general AUC (0.817) compared to the LDA model trained with all non-perfectly collinear variables.

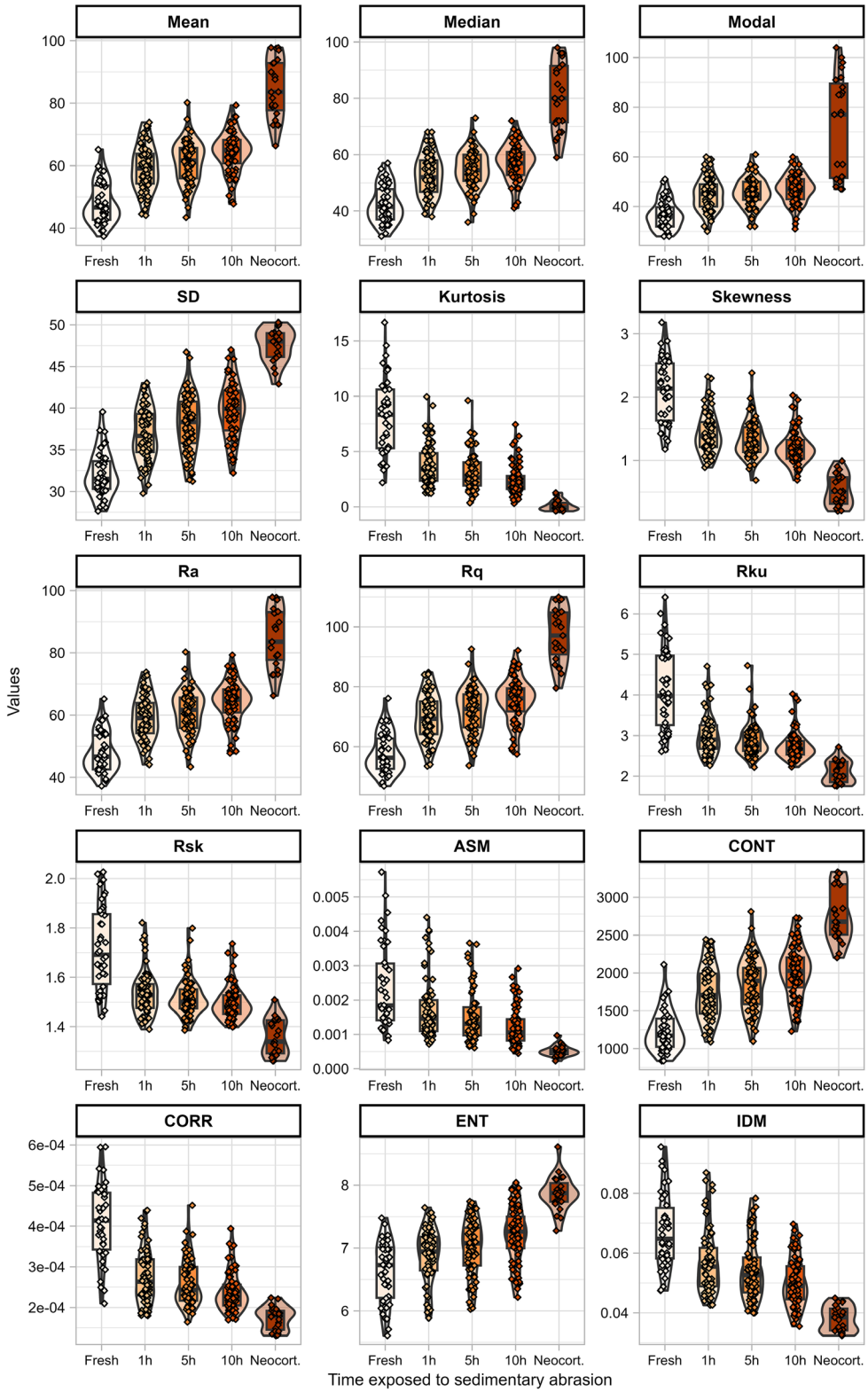

FIGURE 2 Box and violin plots of metrics employed in the study by exposure times.

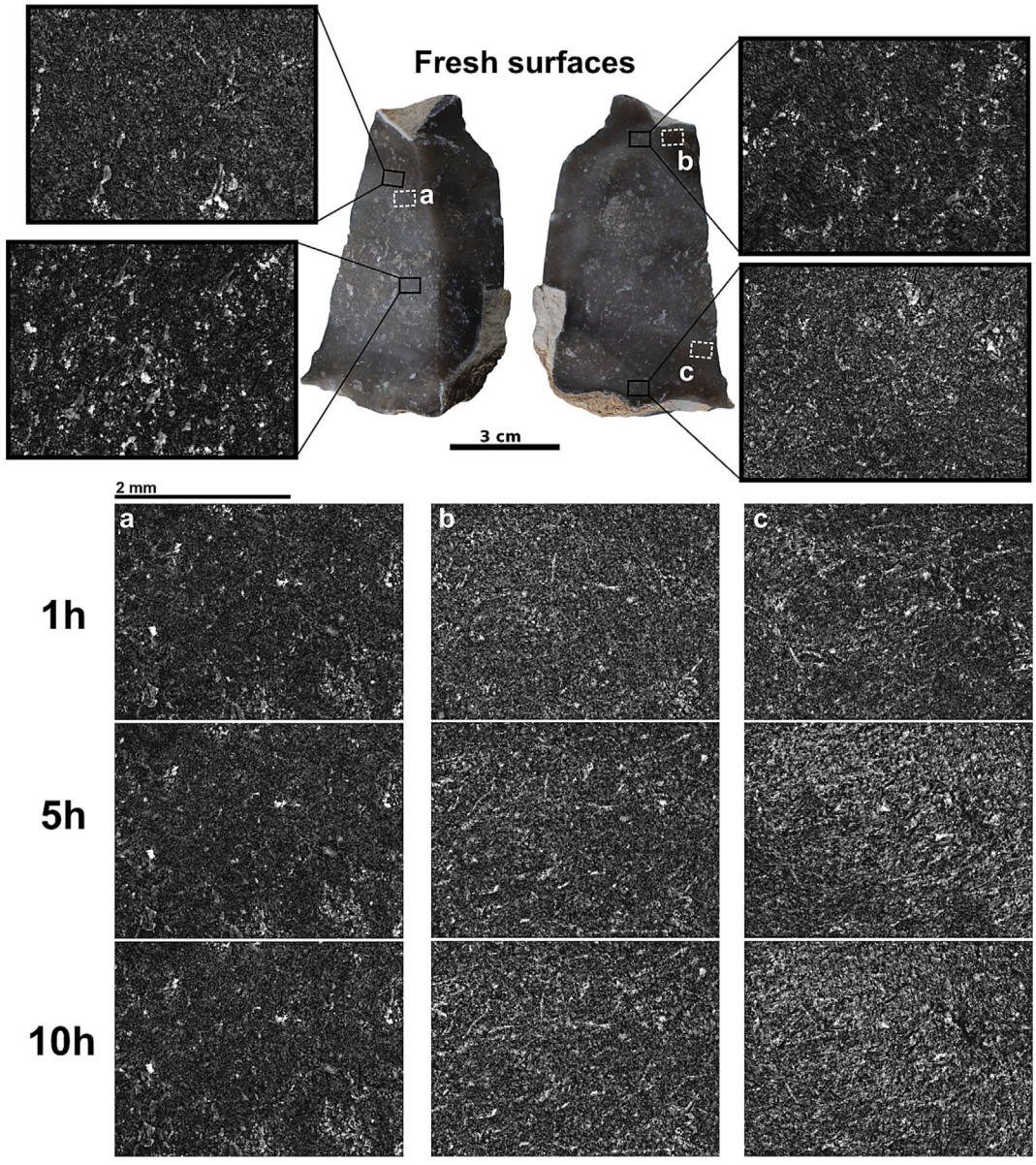

FIGURE 3 Examples of differential abrasion on the same artifact. Top: images of the fresh surface. Bottom: sequential images of sedimentary abrasion: (a) little or no abrasion is developed; (b) sedimentary abrasion is moderately developed; (c) sedimentary abrasion is heavily developed.

Feature importance

Figure 7 presents feature importance according to exposure time and average importance. In general, the LDA did not favor any group of statistics over another. The three most important features considered by the LDA model relate to the distribution of values (kurtosis), texture (contrast), and roughness (Rq). Although the remaining variables presented similar values of importance, it is important to consider the high levels of collinearity previously observed between features (Figure S4), which probably affect the interpretation of the other features.

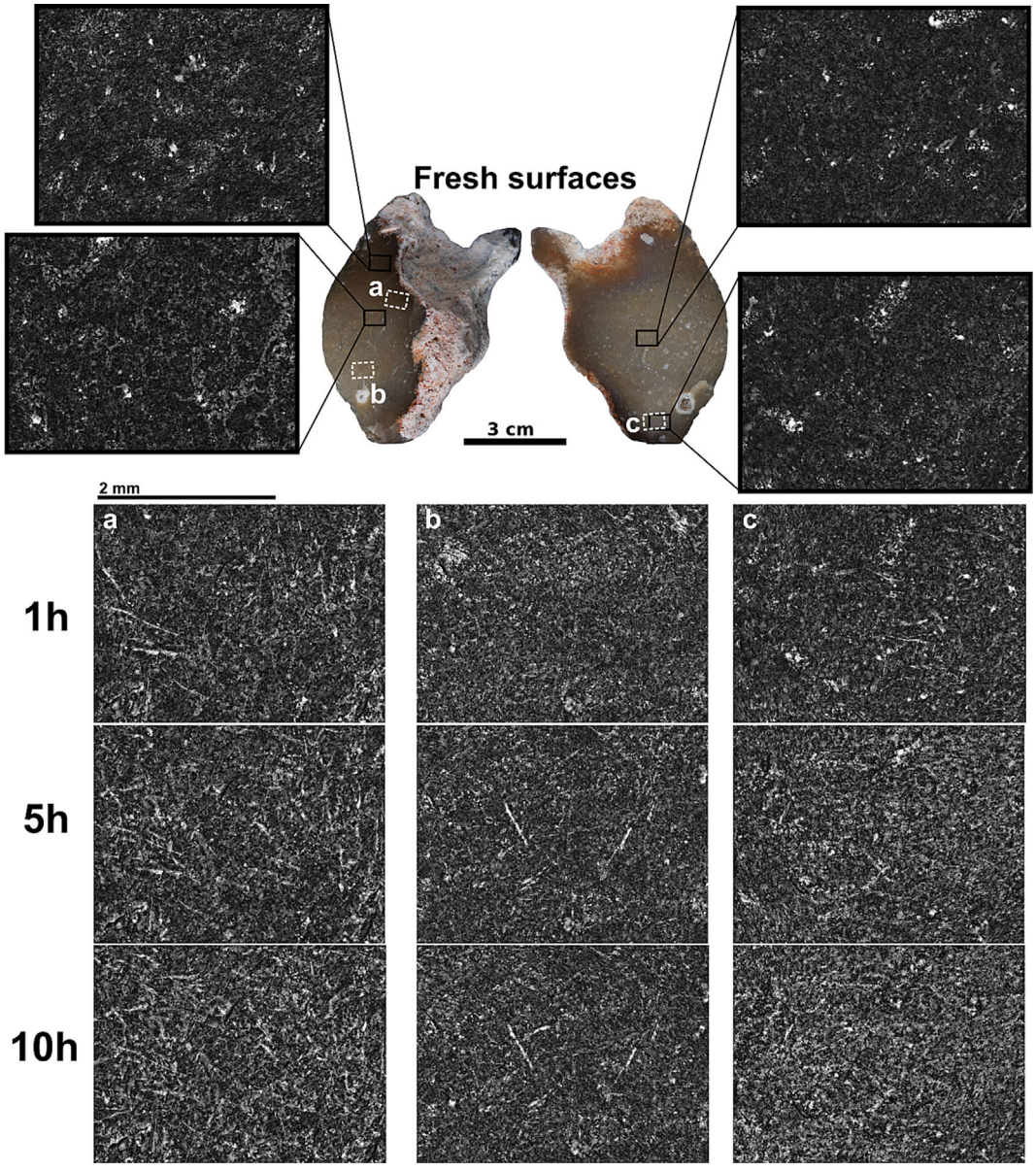

FIGURE 4 Examples of differential abrasion on the same artifact. Top: images of the fresh surface. Bottom: sequential images of sedimentary abrasion: (a) sedimentary abrasion is moderately/heavily developed; (b) sedimentary abrasion is lightly developed; (c) sedimentary abrasion is heavily developed.

Entropy (ENT) and angular second moment (ASM) presented clear trends in the visual exploratory analysis (Figure 2). However, the LDA model considered them less important, probably due to intense overlapping of values from different exposure times.

Figure 8 best illustrates the effects of sedimentary abrasion on non-collinear variables. Both plots present the relationship between the two most important variables considered by the LDA model (kurtosis and contrast), and their respective most uncorrelated features (Figure S4). In both cases, fresh flints presented the highest ranges of dispersion and, as exposure to sedimentary abrasion increased, the range of dispersion progressively decreased. In the case of kurtosis

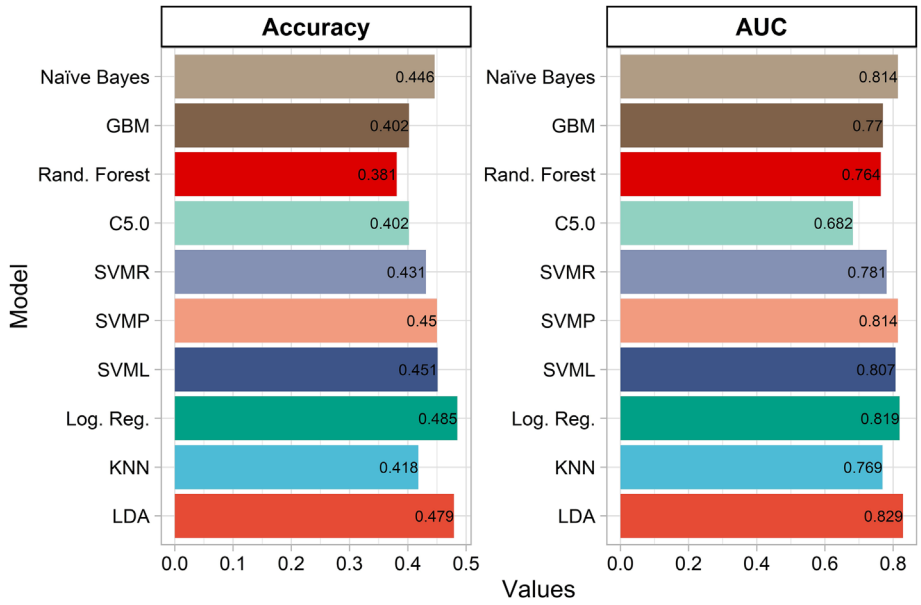

FIGURE 5 Performance metrics (general accuracy and AUC) of the machine learning models.

and entropy, increased exposure time resulted in decreasing values of kurtosis and increasing values of entropy. In the case of contrast (CONT) and angular second moment (ASM), increased sedimentary abrasion resulted in increasing values of the former and decreasing values of the latter.

DISCUSSION

The present work shows that gray level images can effectively be used to obtain quantitative values for changes on the surface of flint artifacts due to post-depositional alterations. The range of metrics employed to analyze the gray level images successfully captured surface changes and their directionality, effectively serving as discriminatory variables to differentiate between times of exposure. Additionally, the use of gray level images also provided insights into the differential development of sedimentary abrasion due to post-depositional processes. Our work shows that convex surfaces (bulb, parts of the dorsal surface, or hinge terminations) and surfaces close to the ridge or edges develop post-depositional alterations faster and more intensely. Overall, the results presented here indicate that increasing abrasion results in more heterogeneous surfaces. This can be observed in the evolution of the concentration of statistical values (such as kurtosis and standard deviation). The more heterogeneous nature of the surface is also observed in measures that take into consideration the spatial distribution of values (Haralick et al., 1973). This more heterogeneous nature of the surface is well observed in the increased values of contrast and entropy, while the values of angular second moment, correlation and inverse different moment (homogeneity) decreased.

The general low values of precision from the machine learning models were expected given the heterogeneous development of sedimentary abrasion. After 10 h of exposure, some surfaces had barely been abraded, resulting in minimal changes in the texture metrics, and were thus hardly distinguishable from fresh surfaces or surfaces exposed for a shorter time. The general confusion in the differentiation of sedimentary abrasion was not caused by the analytical

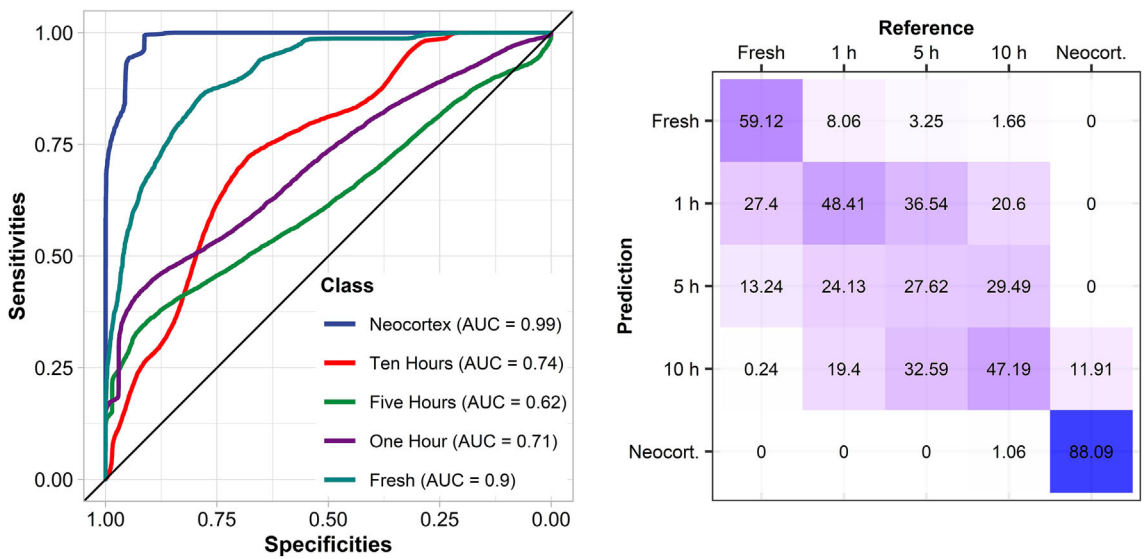

FIGURE 6 Left: ROC curves and AUCs from the LDA model for each category. Right: normalized confusion matrix from the LDA model.

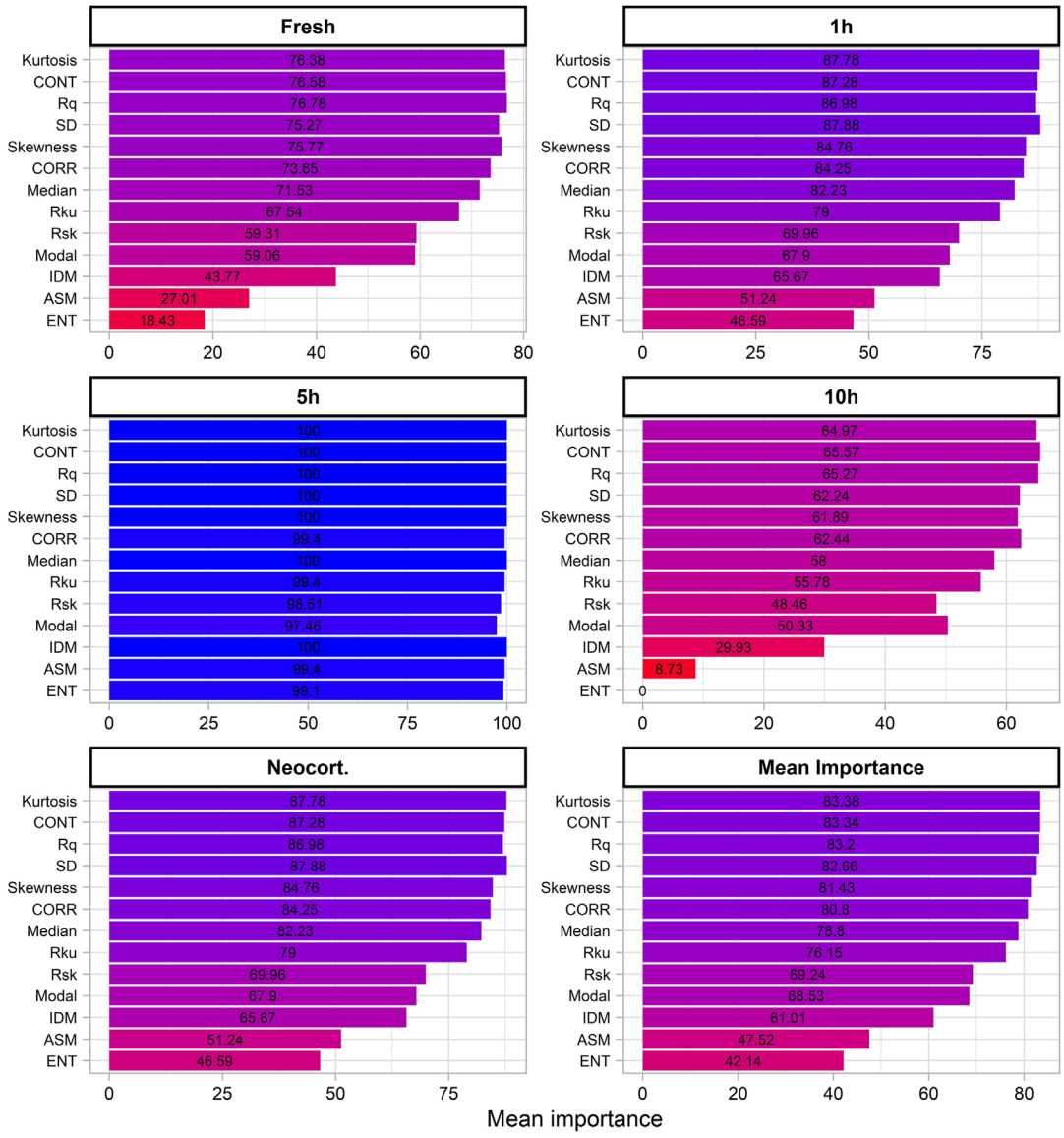

FIGURE 7 Feature importance according to each exposure time and average feature importance.

procedure (use of a Dino-Lite Edge 3.0 AM73915MZT USB and gray level images), but by the heterogeneous development of sedimentary abrasion.

Previous studies (Bustos-Pérez et al., 2019) have indicated that, as the particle size of the sediment decreases, heterogeneity in the development of the post-depositional process increases. However, this interaction remains unresolved. In the present study, the sediment particles used to induce sedimentary abrasion were small and the development of sedimentary abrasion was heterogeneous. This more heterogeneous development falls in line with previous studies (Bustos-Pérez et al., 2019) in which more fine-grained sediments resulted in increased ridge-width variability. Further research might contemplate subjecting lithic artifacts to similar experimental conditions, but using sediments with larger particle sizes. Use-wear analyses have shown that different worked materials exhibit quantifiable differences, with time not being a

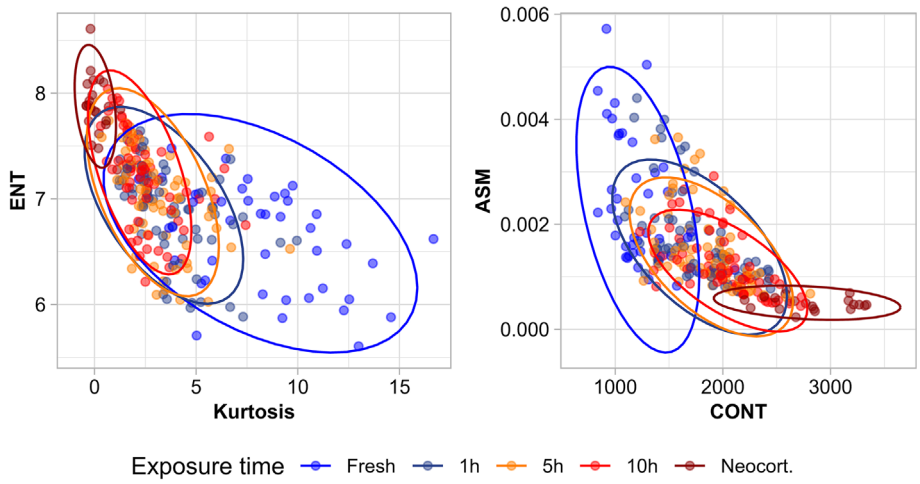

FIGURE 8 Scatter plot presenting the relationship between the two most important variables (kurtosis and contrast) and their respective less correlated variables.

factor that blurs or causes overlaps in differences (Bietti, 1996; Ibáñez & Mazzucco, 2021; Stemp et al., 2009). For post-depositional processes, it remains untested whether different particle sizes, types of transport of particles (fluvial or aeolian), or weathering induces quantifiable differentiable surfaces that can be identified with the use of grayscale images.

Results showing the heterogeneous development of abrasion on the same flint artifact are indicative of the need for caution when attributing post-depositional conditions. If the surface metrics of a lithic artifact are required, several sampling options can be adopted, such as only considering metrics from areas that would have potentially developed faster, or averaging the results from several areas of the same artifact. The present work employed complete photographs for the analysis of post-depositional abrasion in lithic artifacts. However, previous research has shown that sampling images on the most developed areas (Ibáñez & Mazzucco, 2021) can help improve the analysis by focusing on the areas of interest, or avoiding noise generated by irregularities such as fossils or geodes. Further research into the post-depositional study of lithic artifacts might benefit from applying these sampling procedures. However, the results presented here underscore the need to consider additional features of lithic artifacts, such as the width of the ridges or alterations on the edges (Burroni et al., 2002; Bustos-Pérez et al., 2019; Chu et al., 2015; Panera Gallego & Rubio Jara, 1996; Shackley, 1974).

In this study, we sampled only two types of flint that were similar in their surface appearance. Despite the heterogeneous development of abrasion on individual flakes, both flints reacted similarly to subsequent episodes of rounding. However, caution is advisable, since flints with pronounced imperfections in the form of geodes, carbonates, opals, or translucid areas might make obtaining reliable data challenging. In addition, digital microscopes using LED can reflect the imperfections on the surface of the material, which results in white pixels that are interpreted as having the maximum value (255) instead of a value corresponding to their surface height.

CONCLUSIONS

Determining the degree of post-depositional alteration in lithic materials is a key factor when evaluating the integrity of paleolithic archaeological assemblages. Using the workflow

presented here, gray level images can be used to obtain quantitative values reflecting surface changes caused by sedimentary abrasion. Further, sequential experimentation showed that these quantitative values also capture directionality in changes due to increased sedimentary abrasion. However, under the described conditions of this experiment, not all the surfaces of the artifacts developed abrasion at the same pace. Convex surfaces and areas close to the edges developed the alterations faster and more intensely (indicating that these areas should be inspected and sampled first when evaluating an artifact). Although studies focusing on post-depositional alterations of lithic artifacts are scarce, the present work highlights the potential of using gray level images and quantitative analyses of the surface of lithic artifacts in this area of inquiry.

ACKNOWLEDGEMENTS

The authors wish to thank the Editor and the two reviewers for their comments and suggestions. 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), the Universitat Rovira i Virgili (2022PFR-URV-64), and the Spanish Ministry of Science and Innovation (MICINN/FEDER project PID2021-122355NB-C32). The Institut Català de Paleoecologia Humana i Evolució 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 article is the result of the research projects "Como, Quien Y Donde?: Variabilidad De Comportamientos En La Captación Y Transformación De Los Recursos Liticos Dentro De Grupos Neandertales 2" (HAR2016-76760-C3-2-P), financed by Agencia Estatal de Investigación (AEI), Fondo Europeo de Desarrollo Regional (FEDER); and "En Los Limites De La Diversidad: Comportamiento Neandertal En El Centro Y Sur De La Penisula Iberica" (ID2019-103987 GB-C33) financed by the Spanish National Plan for Scientific and Technical Research and Innovation (2017–2020).

DATA AVAILABILITY STATEMENT

All data, code and the complete workflow has been made available throufgh a Zenodo repository (https://doi.org/10.5281/zenodo.8359711).

ORCID

Guillermo Bustos-Pérez https://orcid.org/0000-0002-1089-818X

REFERENCES

Adán, M., Barceló, J. A., Pijoan-López, J., Piqué, R., & Toselli, A. (2003). Spatial statistics in archaeological texture analysis. The Digital Heritage of Archaeology.

Alhusban, Z., & Valyrakis, M. (2021). Assessing and modelling the interactions of instrumented particles with bed surface at low transport conditions. *Applied Sciences*, *11*(16), 7306. https://doi.org/10.3390/app11167306

Alin, A. (2010). Multicollinearity. *Wiley Interdisciplinary Reviews: Computational Statistics*, *2*(3), 370–374. https://doi.org/10.1002/wics.84

Asryan, L., & Ollé, A. (2020). Results of a functional study on the middle to early upper Pleistocene lithic assemblages from the Azokh 1 cave site (South Caucasus). *Quaternary International*, *569–570*, 168–180. https://doi.org/10.1016/j.quaint.2020.05.028

Barceló, J. A., Pijoan, J., & Vicente, O. (2001). Image quantification as archaeological description. In Z. Stancic & T. Veljanovski (Eds.), *Computing Archaeology for Understanding the Past* (pp. 69–78). Archaeopress, Oxford.
Bertran, P., Lenoble, A., Todisco, D., Desrosiers, P. M., & Sørensen, M. (2012). Particle size distribution of lithic assemblages and taphonomy of Palaeolithic sites. *Journal of Archaeological Science*, *39*, 3148–3166. https://doi.org/10.1016/j.jas.2012.04.055
Bietti, A. (1996). Image processing in microwear studies of flint artifacts. *Archeologi e Calcolatori*, *7*, 387–396.
Boström, S., & Lundin, J. (2022). Quantifying use-Wear polish through 3D imaging software. *Lund Archaeological Review*, *26–27*, 7–22.
Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, *30*(7), 1145–1159. https://doi.org/10.1016/S0031-3203(96)00142-2
Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324
Burroni, D., Donahue, R. E., & Pollard, A. M. (2002). The surface alteration features of flint artefacts as a record of environmental processes. *Journal of Archaeological Science*, *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*(1), 233–247.
Bustos-Pérez, G., Díaz, S., & Baena, J. (2019). An experimental approach to degrees of rounding among lithic artifacts. *Journal of Archaeological Method and Theory*, *26*, 1243–1275. https://doi.org/10.1007/s10816-018-9409-8
Byers, D. A., Hargiss, E., & Finley, J. B. (2015). Flake morphology, fluvial dynamics, and debitage transport potential: Flakes, fluvial dynamics, and debitage transport. *Geoarchaeology*, *30*(5), 379–392. https://doi.org/10.1002/gea.21524
Cabrera, J. E. (2006). Texture analyzer. Java.
Calandra, I. (2022). A workflow for quality control in surface texture analysis applied to teeth and tools. *Journal of Archaeological Science: Reports*, *46*, 103692. https://doi.org/10.1016/j.jasrep.2022.103692
Chambers, J. C. (2003). Like a rolling stone? The identification of fluvial transportation damage signatures on secondary context bifaces. *Lithics*, *24*, 66–77.
Chan, J. Y.-L., Leow, S. M. H., Bea, K. T., Cheng, W. K., Phoong, S. W., Hong, Z.-W., & Chen, Y.-L. (2022). Mitigating the multicollinearity problem and its machine learning approach: A review. *Mathematics*, *10*(8), 1283. https://doi.org/10.3390/math10081283
Chinga, G., & Dougherty, B. (2002). *Roughness calculation*. Java.
Chinga, G., Johnsen, P. O., Dougherty, R., Berli, E. L., & Walter, J. (2007). Quantification of the 3D microstructure of SC surfaces. *Journal of Microscopy*, *227*(3), 254–265. https://doi.org/10.1111/j.1365-2818.2007.01809.x
Chu, W., Thompson, C., & Hosfield, R. (2015). Micro-abrasion of flint artifacts by mobile sediments: A taphonomic approach. *Archaeological and Anthropological Sciences*, *7*, 3–11. https://doi.org/10.1007/s12520-013-0157-0
Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297. https://doi.org/10.1007/BF00994018
Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, *13*(1), 21–27. https://doi.org/10.1109/TIT.1967.1053964
Cramer, J. S. (2004). The early origins of the logit model. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *35*(4), 613–626. https://doi.org/10.1016/j.shpsc.2004.09.003
Evans, A. A., & Donahue, R. E. (2008). Laser scanning confocal microscopy: A potential technique for the study of lithic microwear. *Journal of Archaeological Science*, *35*(8), 2223–2230. https://doi.org/10.1016/j.jas.2008.02.006
Fernández-Marchena, J. L., & Ollé, A. (2016). Microscopic analysis of technical and functional traces as a method for the use-wear analysis of rock crystal tools. *Quaternary International*, *424*, 171–190. https://doi.org/10.1016/j.quaint.2015.10.064
Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, *7*, 179–188. https://doi.org/10.1111/j.1469-1809.1936.tb02137.x
Frey, P. W., & Slate, D. J. (1991). Letter recognition using Holland-style adaptive classifiers. *Machine Learning*, *6*(2), 161–182. https://doi.org/10.1007/BF00114162
Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, *29*(5), 1189–1232. https://doi.org/10.1214/aos/1013203451
Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, *38*(4), 367–378. https://doi.org/10.1016/S0167-9473(01)00065-2
Grace, R., Graham, I. D. G., & Newcomer, M. H. (1985). The quantification of microwear polishes. *World Archaeology*, *17*(1), 112–120. https://doi.org/10.1080/00438243.1985.9979954
Greenwell, B., Boehmke, B., Cunningham, J., Developers, G. B. M., & Greenwell, M. B. (2019). Package 'gbm'. *R Package Version*, *2*(5), 1–39.

Hand, D. J., & Till, R. J. (2001). A simple generalisation of the area under the ROC curve for multiple class classification problems. *Machine Learning*, *45*(2), 171–186. https://doi.org/10.1023/A:1010920819831

Haralick, R. M., Shanmugam, K., & Dinstein, I. H. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, *3*(6), 610–621. https://doi.org/10.1109/TSMC.1973.4309314

Harding, P., Gibbard, P. L., Lewin, J., Macklin, M. G., & Moss, E. H. (1987). The transport and abrasion of flint handaxes in a gravel-bed river. In *The Human Uses of Flint and Chert: Proceedings of the Fourth International Flint Symposium Held at Brighton Polytechnic (eds. G. Sieveking, and M. Newcomer)* (pp. 115–126). Cambridge University Press.

Hiscock, P. (2002). Quantifying the size of artefact assemblages. *Journal of Archaeological Science*, *29*, 251–258. https://doi.org/10.1006/jasc.2001.0705

Ibáñez, J. J., Anderson, P. C., González-Urquijo, J., & Gibaja, J. (2016). Cereal cultivation and domestication as shown by microtexture analysis of sickle gloss through confocal microscopy. *Journal of Archaeological Science*, *73*, 62–81. https://doi.org/10.1016/j.jas.2016.07.011

Ibáñez, J. J., Lazuen, T., & González-Urquijo, J. (2019). Identifying experimental tool use through confocal microscopy. *Journal of Archaeological Method and Theory*, *26*(3), 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*(9), e0257266. https://doi.org/10.1371/journal.pone.0257266

Jain, V., Fryirs, K., & Brierley, G. (2008). Where do floodplains begin? The role of total stream power and longitudinal profile form on floodplain initiation processes. *Geological Society of America Bulletin*, *120*(1–2), 127–141. https://doi.org/10.1130/B26092.1

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning with applications in R* (2nd ed.). Springer. https://doi.org/10.1007/978-1-4614-7138-7

Karatzoglou, A., Meyer, D., & Hornik, K. (2006). Support vector machines in R. *Journal of Statistical Software*, *15*, 1–28. https://doi.org/10.18637/jss.v015.i09

Karatzoglou, A., Smola, A., Hornik, K., & Zeileis, A. (2004). Kernlab - An S4 package for kernel methods in R. *Journal of Statistical Software*, *11*, 1–20. https://doi.org/10.18637/jss.v011.i09

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, *28*(5), 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.

Lenoble, A., & Bertran, P. (2004). Fabric of Palaeolithic levels: Methods and implications for site formation processes. *Journal of Archaeological Science*, *31*, 457–469. https://doi.org/10.1016/j.jas.2003.09.013

Levi Sala, I. (1986). Use wear and post-depositional surface modification: A word of caution. *Journal of Archaeological Science*, *13*, 229–244. https://doi.org/10.1016/0305-4403(86)90061-0

Maíllo Fernández, J. M. (1998). Proporciones de debris en réplicas de talla experimental, Espacio Tiempo y Forma. *Serie I, Prehistoria Y Arqueología*, (11), 45–45. https://doi.org/10.5944/etfi.11.1998.4665

McPherron, S. J. P. (2005). Artifact orientations and site formation processes from total station proveniences. *Journal of Archaeological Science*, *32*, 1003–1014. https://doi.org/10.1016/j.jas.2005.01.015

McPherron, S. P. (2018). Additional statistical and graphical methods for analyzing site formation processes using artifact orientations. *PLoS ONE*, *13*(1), e0190195. https://doi.org/10.1371/journal.pone.0190195

Montgomery, D. R., & Buffington, J. M. (1997). Channel-reach morphology in mountain drainage basins. *Geological Society of America Bulletin*, *109*(5), 596–611. https://doi.org/10.1130/0016-7606(1997)109<0596:CRMIMD>2.3.CO;2

Ollé, A., & Vergès, J. M. (2014). The use of sequential experiments and SEM in documenting stone tool microwear. *Journal of Archaeological Science*, *48*, 60–71. https://doi.org/10.1016/j.jas.2013.10.028

Panera Gallego, J., & Rubio Jara, S. (1996). Propuesta de análisis tecnomorfológico para la industria lítica del Pleistoceno Medio, Espacio, tiempo y forma. *Serie I, Prehistoria Y Arqueología*, *9*, 33–76. https://doi.org/10.5944/etfi.9.1996.4628

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*(12), e0243295. https://doi.org/10.1371/journal.pone.0243295

Petraglia, M. D., & Potts, R. (1994). Water flow and the formation of early Pleistocene artifact sites in Olduvai Gorge. *Tanzania, Journal of Anthropological Archaeology*, *13*, 228–254. https://doi.org/10.1006/jaar.1994.1014

Pijoan-López, J., Barceló-Álvarez, J. A., Clemente-Conte, I., & Vila-Mitjà, A. (2002). Variabilidad estadística en imágenes digitalizadas de rastros de uso: resultados preliminares. In I. Clemente Conte, R. Risch, & J. F. G. Bao (Eds.), *Análisis funcional: su aplicación al estudio de sociedades prehistóricas* (pp. 55–64). Archaeopress.

Quinlan, J. R. (1996). Improved use of continuous attributes in C4.5. *Journal of Artificial Intelligence Research*, *4*, 77–90. https://doi.org/10.1613/jair.279

Quinlan, J. R. (2014). *C4. 5: Programs for machine learning*. Elsevier.

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
Ridgeway, G. (2007). Generalized boosted models: A guide to the gbm package, R package vignette.
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 Bioinformatics*, *12*(1), 1–8. https://doi.org/10.1186/1471-2105-12-77
RStudio Team. (2019). *RStudio: Integrated development for R*. RStudio, Inc.
Rust, B. R. (1972). Structure and process in a braided river. *Sedimentology*, *18*(3–4), 221–245. https://doi.org/10.1111/j.1365-3091.1972.tb00013.x
Schick, K. D. (1987). Experimentally-derived criteria for assessing hydrologic disturbance of archaeological sites. In D. T. Nash & M. D. Petraglia (Eds.), *Natural formation processes and the archaeological record* (pp. 86–107). BAR International Series. BAR Publishing.
Schiffer, M. B. (1972). Archaeological context and systemic context. *American Antiquity*, *37*(2), 156–165. https://doi.org/10.2307/278203
Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P., & Cardona, A. (2012). Fiji: An open-source platform for biological-image analysis. *Nature Methods*, *9*(7), 676–682. https://doi.org/10.1038/nmeth.2019
Shackley, M. S. (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* (pp. 160–163). Elsevier. https://doi.org/10.1016/B978-1-55860-036-2.50047-3
Stapert, D. (1976). Some natural surface modifications on flint in the Netherlands. *Paléo*, *18*, 9–41.
Stemp, W. J., Childs, B. E., Vionnet, S., & Brown, C. A. (2009). Quantification and discrimination of lithic use-wear: Surface profile measurements and length-scale fractal analysis. *Archaeometry*, *51*(3), 366–382. https://doi.org/10.1111/j.1475-4754.2008.00404.x
Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S fourth* (Edition ed.). Statistics and Computing. https://doi.org/10.1007/978-0-387-21706-2
Vila, A., & Gallart, F. (1993). Caracterización de los micropulidos de uso: Ejemplo de aplicación del análisis de imágenes digitalizadas, In Traces et fonction: Les gestes retrouvés (eds. P. C. Anderson, S. Beyries, M. Otte, and H. Plisson), 459–66: E.R.A.U.L. Etudes et recherches archéologiques de l'université de Liège. Série A 50, Service de Préhistoire, Université de Liège, Liège.
Walker, S. H., & Duncan, D. B. (1967). Estimation of the probability of an event as a function of several independent variables. *Biometrika*, *54*(1/2), 167–179. https://doi.org/10.1093/biomet/54.1-2.167
Weihs, C., Ligges, U., Luebke, K., & Raabe, N. (2005). klaR analyzing German business cycles. In *Data analysis and decision support* (pp. 335–343). Springer. https://doi.org/10.1007/3-540-28397-8_36
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., … Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686
Wright, M. N., & Ziegler, A. (2017). Ranger: A fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software*, *77*, 1–17. https://doi.org/10.18637/jss.v077.i01

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Bustos-Pérez, G., & Ollé, A. (2024). The quantification of surface abrasion on flint stone tools. *Archaeometry*, *66*(2), 247–265. https://doi.org/10.1111/arcm.12913