The Contribution of 2D and 3D Geometric Morphometrics to Lithic Taxonomies: Testing Discrete Categories of Backed Flakes from Recurrent Centripetal Core Reduction

Guillermo Bustos-Pérez[1,2,3] · Brad Gravina[4,5] · Michel Brenet[5,6] · Francesca Romagnoli[1]

Abstract

Paleolithic lithic assemblages are usually dominated by flakes and display a high degree of morphological variability. When analyzing Paleolithic lithic assemblages, it is common to classify flakes into categories based on their morphological and technological features, which are linked to the position of the flake in the reduction sequence and how removals are organized in a given production method. For the analysis of Middle Paleolithic lithic assemblages, two categories of flakes are commonly identified: core–edge flakes and pseudo-Levallois points. A third type, core–edge flakes with a limited back, is also commonly found in the archaeological literature, providing an alternative category whose definition does not match the two previous types but shares many of their morphological and technological features. The present study addresses whether these three flakes constitute discrete categories based on their morphological and technological attributes. 2D and 3D geometric morphometrics are employed on an experimental set composed of the three categories of flakes to quantify morphological variation. Machine learning models and principal components biplots are used to test the discreteness of the categories. The results indicate that geometric morphometrics succeed in capturing the morphological and technological features that characterize each type of product. Pseudo-Levallois points have the highest discreteness of the three technological products, and while some degree of mixture exists between core edge flakes and core edge flakes with a limited back, they are also highly distinguishable. We conclude that the three categories are discrete and can be employed in technological lists of products for the analysis of lithic assemblages and that geometric morphometrics is useful for testing for the validity of categories. When testing these technological categories, we stress the need for well-defined and shared lithic analytical units to correctly identify and interpret the technical steps and decisions made by prehistoric knappers and to properly compare similarities and differences between stone tool assemblages. These are key aspects for current research in which open datasets are becoming more and more common and used to build interpretative techno-cultural models on large geographical scales. Now more than ever, lithic specialists are aware of the need to overcome differences in taxonomies between different school traditions.

✉ Guillermo Bustos-Pérez
   guillermo.bustos@uam.es

✉ Francesca Romagnoli
   francesca.romagnoli@uam.es

1 Departamento de Prehistoria y Arqueología, Universidad Autónoma de Madrid, Campus de Cantoblanco, 28049 Madrid, Spain

2 Institut Català de Paleoecologia Humana I Evolució Social (IPHES-CERCA), Zona Educacional 4, Campus Sescelades URV (Edifici W3), 43007 Tarragona, Spain

3 Departament d'Història I Història de L'Art, Universitat Rovira I Virgili, Avinguda de Catalunya 35, 43002 Tarragona, Spain

4 Musée National de Préhistoire, MC, 1 Rue du Musée, 24260 Les Eyzies de Tayac, France

5 Univ. Bordeaux, CNRS, MC, PACEA, UMR-5199, FF-33600 Pessac, France

6 Centre Mixte de Recherches Archéologiques, Domaine de Campagne, INRAP Grand Sud-Ouest, 242460 Campagne, France

🙋 Springer

Introduction

Lithic artifacts regularly constitute the most important and abundant remains found on Paleolithic sites. When analyzing lithic assemblages, in addition to taking metric measurements and noting attributes, it is common to classify unmodified flakes according to their morphologies and technological features. This is a crucial part of lithic analysis because it classifies flakes into technological categories in the sense that the retained features and morphology are indicative of the production method by which they were generated. These technological products usually reflect different knapping strategies, stages of reduction, and variations in the organization of removals and surface exploitation. Well-known examples of technological classifications of flakes include bipolar/on anvil flakes (Callahan, 1996; Hayden, 1980), overshot flakes (Cotterell & Kamminga, 1987), bifacial-thinning flakes (Raab et al., 1979), byproducts of blade production, such as core tablets or crested blades (Pelegrin, 1995; Shea, 2013a), Kombewa flakes (Tixier & Turq, 1999; Tixier et al., 1980), and lateral tranchet blows (Bourguignon, 1992). While the use of technological categories is common and helps increase the resolution of lithic analysis, it is important to bear in mind that lithic artifacts are characterized by a high degree of morphological variability, which, in many cases, results in overlapping features. Consequently, several categories remain underused because of this high morphological variability, their overlapping features, and their similar roles in the volumetric management of the core along the reduction process.

The Middle Paleolithic in western Europe is characterized by the diversification of an increase in knapping methods, resulting in what are generally flake-dominated assemblages (Delagnes & Meignen, 2006; Kuhn, 2013). For the analysis of Middle Paleolithic lithic assemblages, lists of technological products are common and generally reflect individual knapping methods, the organization of flake removals, and their morphology (Duran & Abelanet, 2004; Duran & Soler, 2006; Geneste, 1988; Shea, 2013b). These technological lists are usually dominated by categories of technological products related to Levallois and discoidal knapping methods (Boëda, 1993, 1995a; Boëda et al., 1990), which constitute an important part of Middle Paleolithic lithic variability. Various discoidal and Levallois products have been identified and first appear approximately at 400 ka, in a vast area from eastern Asia to the Atlantic Coast of western Europe through Siberia and Central Asia, the Levant, eastern and central Europe (see bibliography in Romagnoli et al., 2022), and Africa (Adler et al., 2014; Blinkhorn et al., 2021). The identification of discoidal and Levallois products therefore appears

widespread in lithic analysis across various research schools and is designed to create a comparable dataset, explore specific technological adaptations in different ecological contexts, and discuss long-term techno-cultural traditions and technological change. One special category of such products is backed flakes that exhibit remnants of the core on one of their lateral edges. Backed flakes are usually classified into two technological categories: "core–edge flakes" (*éclats débordants*) and "pseudo-Levallois points." A third category, "core–edge flakes with a limited back" (*éclats débordants à dos limité*), has also been defined (Liliane Meignen, 1993; Lilliane Meignen, 1996; Pasty et al., 2004), although its use is not widespread (Duran & Abelanet, 2004; Duran & Soler, 2006; Geneste, 1988; Shea, 2013b). One reason for this may be their overlapping features, including morphology, and a similar role in core reduction compared to classic core edge flakes. This usually results in their absorption into the group of core edge flakes when technological lists of products are employed.

The present study aims to evaluate whether "core–edge flakes with a limited back" represent a discrete technological category that can be easily separated from classic core–edge flakes and pseudo-Levallois points based on their morphological features. While this may seem a pleonastic technical exercise, refining stone tool taxonomy allows researchers to better describe lithic technology, which is the basis for documenting patterns of tool production, transport, maintenance, use, discard, and reuse. This is equally important for exploring differences in how past human groups adapted their technical knowledge and skills to resource constraints, economic strategies, and social dynamics. Furthermore, an improved and comprehensive use of technological types and sub-types within core trimming elements is fundamental to generating reliable comparisons between archaeological assemblages. Finally, core–edge flakes are present in multiple Palaeolithic techno-complexes, particularly those based around centripetal and recurrent reduction strategies. As such, better classifying types and sub-types within this technological category will be of use in lithic studies for multiple periods and regions. In addition to evaluating the discreteness of this specific artifact type, the present study also explores a workflow for testing lithic categories and compares the effectiveness of using data derived from 2 and 3D geometric morphometrics.

To test the discreteness of core–edge flakes categories, an experimental sample of backed flakes produced by discoidal and recurrent centripetal Levallois reduction is classified, following their technological definitions. Geometric morphometrics on 3D meshes are employed to quantify the morphological variability of the experimental assemblage. To test for the discreteness of these categories, machine learning algorithms are employed to classify the flakes according to

their technological category. Our hypothesis is that, although some degree of overlap is expected due to the high degree of morphological variability among lithic artifacts, machine learning models should easily differentiate the abovementioned categories. Testing this hypothesis would support the use of these backed flake categories in the classification of lithic assemblages. The creation of datasets based on the same analytical units and criteria will enable more accurate and reliable comparisons between stone tool assemblages, improving the definition of working hypotheses and interpretative models for technological, adaptive, and behavioral changes in prehistory (Romagnoli et al., 2022).

Methods

Experimental Assemblage

The present study uses an experimental assemblage comprising eight knapping sequences. Seven cores were knapped on Bergerac flint (Fernandes et al., 2012), and two cores were knapped on Miocene chert from south of Madrid (M. Á. Bustillo et al., 2012; M. A. Bustillo & Pérez-Jiménez, 2005). Three cores were knapped following the discoid *"sensu stricto"* concept, which strongly corresponds to Boëda's original technological definition of the knapping system (Boëda, 1993, 1995b), and five experimental cores were knapped following the Levallois recurrent centripetal system (Boëda, 1993, 1994, 1995b; Lenoir & Turq, 1995).

Six technological characteristics define the Levallois concept (Boëda, 1993, 1994): (1) the volume of the core is conceived as two convex asymmetric surfaces and (2) these two surfaces are hierarchical and not interchangeable. They maintain their roles as striking platforms and debitage (or exploitation) surfaces, respectively, throughout the entire reduction process. (3) The distal and lateral convexities of the debitage surface are maintained to obtain predetermined flakes. (4) The fracture plane of the predetermined products is parallel to the intersection between both surfaces. (5) The striking platform is perpendicular to the overhang (the core edge, at the intersection between the two core surfaces). (6) The technique employed during the knapping process is direct hard-hammer percussion. Depending on the organization of the debitage surface Levallois cores are usually classified into the preferential method (where a single predetermined Levallois flake is obtained from the debitage surface) or recurrent methods (where several predetermined flakes are produced from the debitage surface), with removals being either unidirectional, bidirectional, or centripetal (Boëda, 1995b; Boëda et al., 1990; Delagnes, 1995; Delagnes & Meignen, 2006).

According to (Boëda, 1993, 1995a, 1995b), there are six technological criteria that define the discoid "sensu stricto" method: (1) the volume of the core is conceived as two oblique asymmetric convex surfaces delimited by an intersecting plane; (2) these two surfaces are not hierarchical, being possible to alternate between the roles of striking platforms and exploitation surfaces; (3) the peripheral convexity of the debitage surface is managed to control lateral and distal extractions, thus allowing for a degree of predetermination; (4) the surfaces of the striking platforms are oriented in such a way that the core edge is perpendicular to the predetermined products; (5) the fracture planes of the products are secant; and (6) the technique employed is direct hard-hammer percussion.

A total of 139 unretouched backed flakes (independent of the type of termination) were obtained from the different experimental reduction sequences, 70 from discoidal reduction sequences and 69 from Levallois recurrent centripetal reduction sequences. The following criteria were monitored for the classification of backed flakes (Fig. 1).

Core–edge flakes/*éclat débordants* (Beyries & Boëda, 1983; Boëda, 1993; Boëda et al., 1990) have a cutting edge opposite and usually (although not always) parallel to an abrupt margin. This abrupt margin, or backed edge (*dos*), commonly results from the removal of a portion of the periphery of the core and can be plain, bear the scars from previous removals, be cortical, or present a mix of the three. Classic "core–edge flakes" (Boëda, 1993; Boëda et al., 1990), which are sometimes referred as "core–edge flakes with non-limited back"/ *"éclat débordant à dos non limité"* (Duran, 2005; Duran & Soler, 2006) have a morphological axis that follows the axis of percussion, although it may deviate slightly (Beyries & Boëda, 1983).

"Core–edge flakes with a limited back"/*"éclats débordants à dos limité"* share with core–edge flakes the morphological feature of having a cutting edge opposite a back. However, the main difference resides in a morphological axis clearly offset in respect to the axis of percussion (Meignen, 1993, 1996; Pasty et al., 2004). Because of this deviation from the axis of percussion, the backed edge is usually not completely parallel, nor does it span the entire length of the sharp edge.

Pseudo-Levallois points (Boëda, 1993; Boëda et al., 1990; Bordes, 1953, 1961; Slimak, 2003) are backed products in which the edge opposite to the back has a convergent morphology. This morphology is usually the result of the convergence of two or more previous removals. As with core–edge flakes, the back usually results from the removal of one of the lateral edges of the core and can be plain, retain the scars from previous removals, or more rarely be cortical. Pseudo-Levallois points share with core edge flakes with a limited back the deviation of symmetry from the axis of percussion, but they are clearly differentiable due to their triangular off-axis morphology.
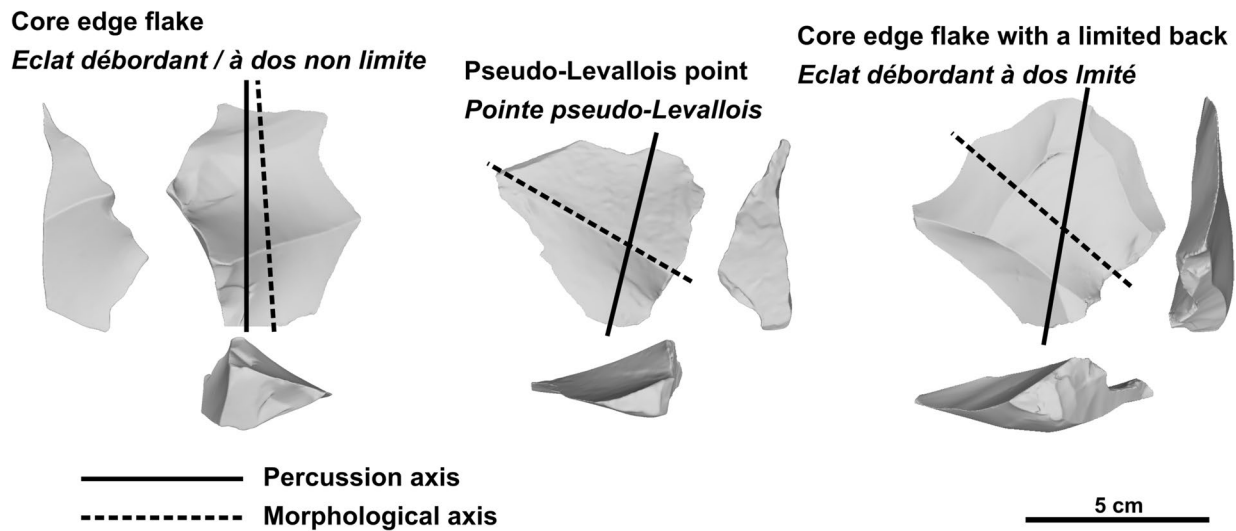
Fig. 1 Backed artifact classification types from the experimental assemblage and their classification

Table 1 Classification of backed flakes from the experimental assemblage

Strategy	Category	Count	Percent
Discoid	Core–edge flake	11	7.91
Discoid	Core–edge with limited back	47	33.81
Discoid	Pseudo-Levallois point	12	8.63
Levallois	Core–edge flake	19	13.67
Levallois	Core–edge with limited back	46	33.09
Levallois	Pseudo-Levallois point	4	2.88

Table 1 presents the distribution of backed flake types, following the previously established definitions. Due to the centripetal character of the knapping methods employed to generate the experimental assemblage, most of the backed flakes fall within the definition of core–edge flakes with a limited back (66.91%). Cortex distribution according to backed flake category (Fig. 2) shows that slightly (~ 25%) or non-cortical products dominate among the three categories, adding up to more than 65% in the three groups (90% core–edge flakes, 68.82% core–edge flakes with a limited back, and 87.5% pseudo-Levallois points).

Geometric Morphometrics

All flakes were scanned with an Academia 20 structured light surface scanner (Creaform 3D) at a 0.2-mm resolution. Flakes were scanned in two parts, automatically aligned (or manually aligned in the case automatic alignment failure) and exported in STL format. Cloudcompare 2.11.3 (https://www.danielgm.net/cc/) free software was employed to perform additional cleaning, mesh sampling, surface reconstruction and transformation into PLY files. Finally, all files were decimated to a quality of 50,000 faces using the Rvcg R package (Schlager, 2017a). The present work compares the use of 2D and 3D geometric morphometrics to test the limits of their application.

2D geometric morphometrics were done using screenshots (Cignoni et al., 2008) of the upper view of each flake orientated along the technological axis. One thin-plate spline (tps) was generated using tpsUtil v.1.82, and the tpsDig v.2.32 (Rohlf, 2015) outline tool was employed to automatically trace the perimeter of each flake. Each outline was resampled to 100 equidistant landmarks (Fig. 3) using Morpho v.2.11 (Schlager, 2017a).

The protocol for digitalizing 3D landmarks on flakes is based on previous studies (Archer et al., 2018, 2021). This included the positioning of a total of three fixed landmarks, 85 curve semi-landmarks, and 420 surface semi-landmarks (Bookstein, 1997a, 1997b; Gunz & Mitteroecker, 2013; Gunz et al., 2005; Mitteroecker & Gunz, 2009). This resulted in a total of 508 landmarks and semi-landmarks. The three fixed landmarks correspond to both laterals of the platform and the percussion point. The 85 curve semi-landmarks correspond to the internal and exterior curve outlines of the platform (15 semi-landmarks each) and the edge of the flake (55 semi-landmarks). Sixty surface semi-landmarks correspond to the platform surface. The dorsal and ventral surfaces of the flakes are defined by 180 semi-landmarks each. The workflow for digitalizing landmarks and semi-landmarks included the creation of a template/atlas on an arbitrary selected flake (Fig. 4: top). After this, landmarks and semi-landmarks were positioned in each specimen and relaxed to minimize bending energy (Fig. 4: bottom; (Bookstein, 1997a, 1997b). A complete workflow of landmark and
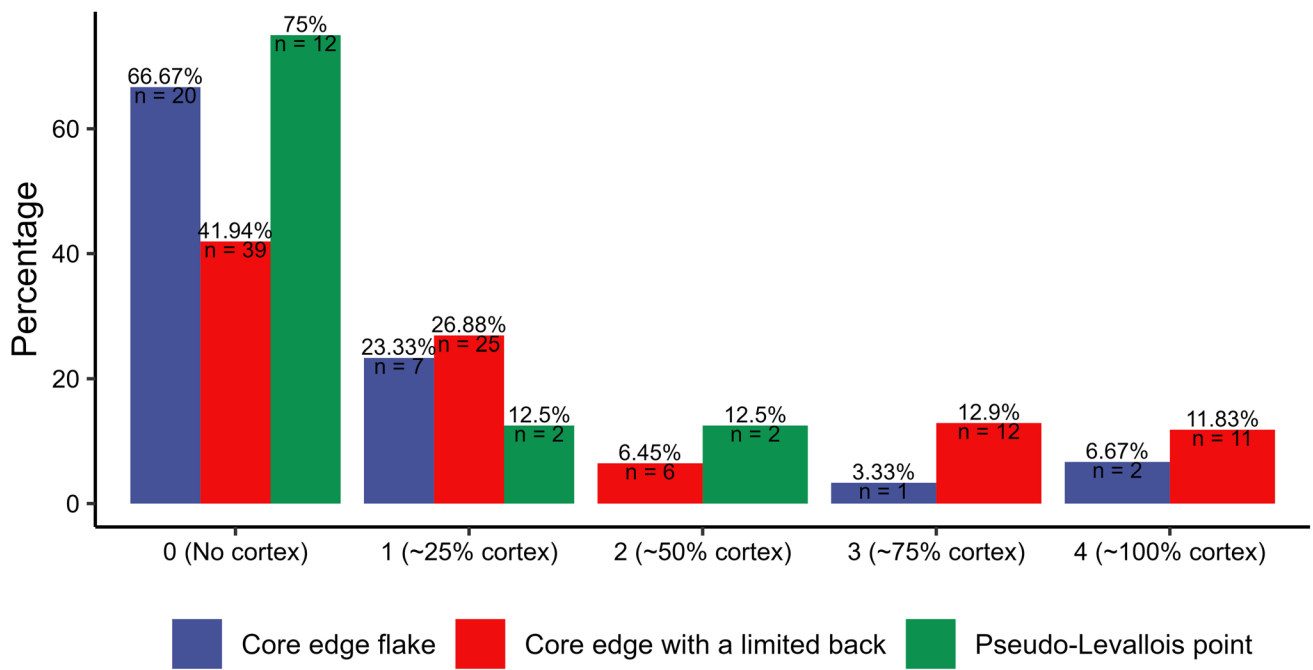
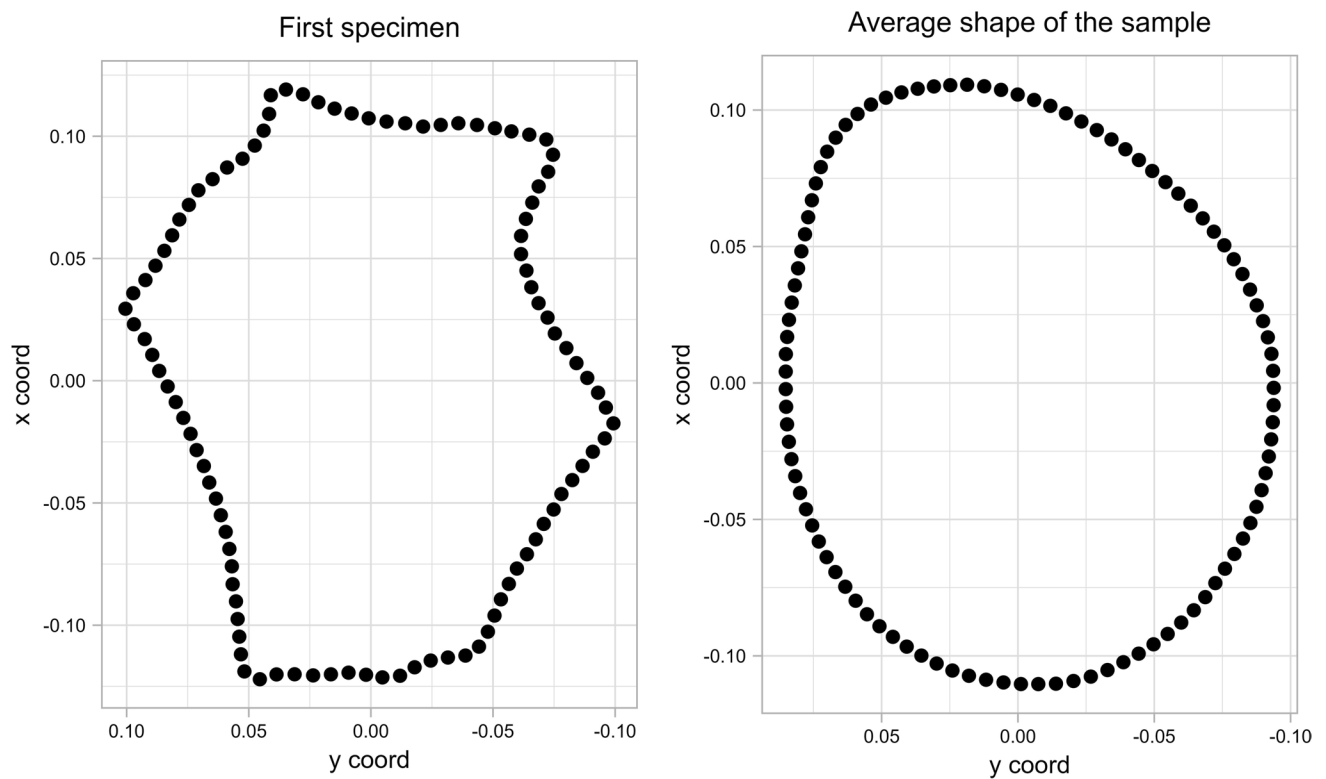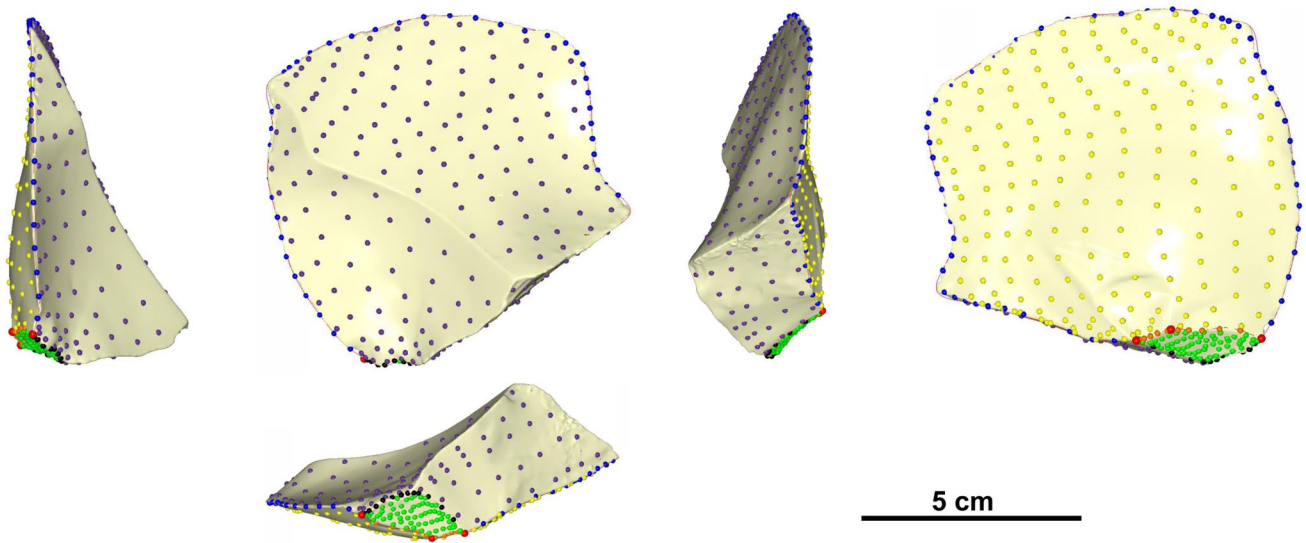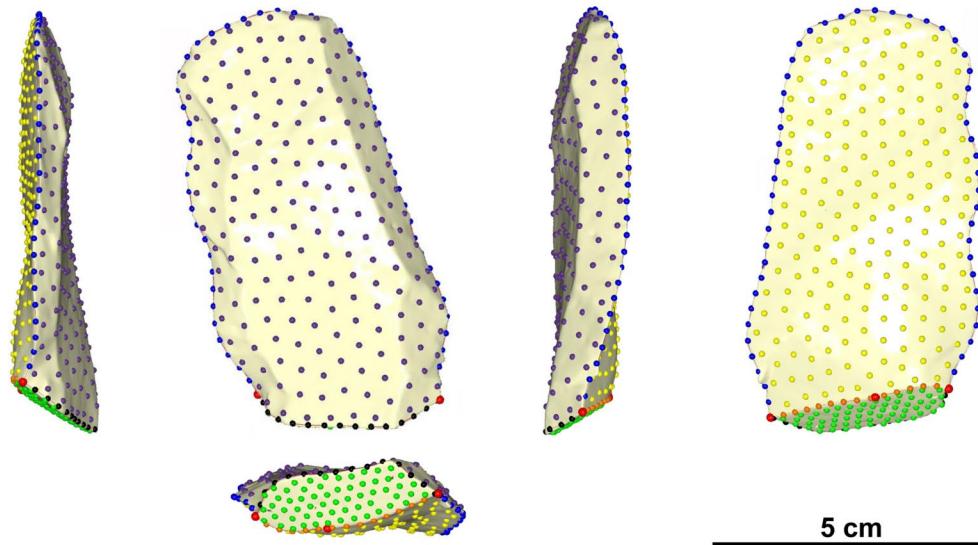**Fig. 2** Distribution of cortex according to backed flake category

Fig. 3 Example of the positioning of 100 equidistant landmarks along the perimeter of a core edge flake and the mean shape of the sample

Fig. 4 Top: template/atlas for a randomly selected flake with the defined landmarks, curves, and surfaces. Bottom: landmark positioning after sliding to minimize bending energy on a pseudo-Levallois point. Fixed landmarks are indicated in red

semi-landmark digitalization and relaxation to minimize bending energy was created in Viewbox Version 4.1.0.12 (http://www.dhal.com/viewbox.htm), and the resulting point coordinates were exported into.xlsx files.

Procrustes superimposition (Kendall, 1984; Mitteroecker & Gunz, 2009; O'Higgins, 2000) was performed using the Morpho v.2.11 package (Schlager, 2017a) on RStudio IDE (R Core Team, 2019; RStudio Team, 2019). Morpho package v.2.11 provides results from principal component analysis (PCA) allowing to reduce the dimensionality of the data (James et al., 2013; Pearson, 1901). There are multiple

reasons to use dimensionality reduction when dealing with high-dimension data on classification: to avoid having more predictors than observations (p > n), avoid collinearity of predictors, reduce the dimensions of the feature space, and avoid overfitting due to an excessive number of degrees of freedom (simple structure with lower number of variables). Principal component analysis achieves dimensionality reduction by identifying the linear combinations that best represent the predictors on an unsupervised manner. The principal components (PCs) of a PCA aim to capture as high a variance as possible for the complete data (James et al.,

2013), and PCs that capture the highest variance need not necessarily be the best for classification.

Debate exists on how many PCs from geometric morphometrics should be selected for classificatory analysis (Schlager, 2017b). Including all PCs up to an arbitrary percentage of variance can result in non-meaningful (noise) PCs pulling the classificatory analysis. This can be considered as a type of overfitting since the classification is not being driven by meaningful morphological trends. An alternative is to select PCs capturing a minimum percentage of variance. However, stone tools are notorious for their wide morphological variability, and increasing sample size results in diminishing variance captured by each PC. Usually the first two to three PCs will reflect ratios of elongation and width to thickness, while other meaningful PCs for classification (such as the angle between the internal or external surface of a flake) might be concealed in lower ranking PCs.

The problem of selecting a minimum variance is approached in two steps. A first round of models is trained using all PCs that represent up to 95% of variance. The threshold of 95% of variance is arbitrarily selected because it balances retaining most of the dataset variance with a reduced number of variables. This provides the most meaningful PCs for classification according to the best model. The effect on morphology of these meaningful PC was visually evaluated. PCs explaining little variance and with little effect on shape change are excluded. Based on this evaluation, the second and final round of models were trained using PCs which captured more than 3% of variance.

The identification of best PCs for classification is performed automatically by the machine learning models using the caret v.6.0.92 package (M. Kuhn, 2008). Morpho package v.2.11 additionally provides visualization of shape change according to PC. A previous work on the same dataset (Bustos-Pérez et al., 2022) performed PCA using the package stats v.4.2.2 (Venables & Ripley, 2002). The present work uses the PCA integrated in the Morpho v.2.11 (Schlager, 2017a). As a result of this, variance captured by PCs and their interpretation differs regarding previous analysis.

Machine Learning and Resampling Techniques

Different machine learning models treat the provided data differently. As a result, different models have different strengths and weaknesses. No universal model exists for all problems. Thus, testing several models is an important step in machine learning. It allows the performance of different models to be compared, the best model to be identified for a given task, and provides a general overview of the difficulty or ease of the problem. It can also serve as indication of possible underlying problems with the data (e.g., overfitting or unbalanced datasets). Machine learning is a quickly developing field where a high number of available models exist. The present work tests ten machine learning models for the classification of flake categories. These models cover some of the most commonly employed algorithms (Jamal et al., 2018) and provide an extensive analysis for the classification of backed flake categories.

- **Linear discriminant analysis (LDA):** reduces dimensionality in an attempt to maximize the separation between classes, while decision boundaries divide the predictor range into regions (Fisher, 1936; James et al., 2013).

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

- **Logistic regression:** essentially adapts the multiple linear regression by raising Euler's constant to its output in the numerator (plus one in the case of the denominator). This results in probability values ranging from 0 to 1which allows obtaining predictions for categorical outcomes (Cramer, 2004; Walker & Duncan, 1967).

- **Decision tree with C5.0 algorithm**: uses recursive partitioning to divide a dataset into homogeneous groups. The C5.0 algorithm uses entropy (measure of degree of mixture between classes) to determine feature values on which to perform the partitioning. The C5.0 algorithm is an improvement on decision trees for classification (J. R. Quinlan, 1996; J. Ross Quinlan, 2014).

- **Random forest:** uses an ensemble of decision trees. Each tree is grown from a random sample of the variables, allowing for each tree to grow differently and better reflect the complexity of the data and provide additional diversity. Finally, the ensemble of trees casts a vote to generate predictions (Breiman, 2001).

- **Gradient boosting machine (GBM):** uses a sequential ensemble of decision trees. After training an initial decision tree, GBM trains subsequent trees on a resampled dataset were the weight of observations difficult to classify is increased based in a gradient (Greenwell et al., 2019; Ridgeway, 2007). The subsequent trained trees complement decisions and allow for the detection of learning deficiencies and increase model accuracy (Friedman, 2001, 2002).

- **Supported vector machines (SVM):** fits hyperplanes into a multidimensional space with the objective of creating homogeneous partitions. The fitting of the hyperplanes is done in order to obtain the maximum margin of separation between classes. The maximum margin

of separation is reached by minimizing the cost (value applied for each incorrect classification). SVM's can use different kernels to transform data into linearly separable cases. The kernel selected for the transformations needs to be specified and plays a key role in the training of a SVM model (Cortes & Vapnik, 1995; Frey & Slate, 1991). The present study tests SVM's with linear, radial, and polynomial kernels.

- **Naïve Bayes**: computes class probabilities using Bayes' rule (Weihs et al., 2005).

As mentioned above, 66.91% of flakes fall into the definition of core–edge flakes with a limited back, resulting in an unbalanced dataset. To counter the unbalanced nature of the experimental dataset, up-sampling was undertaken for the two minority classes, and down-sampling was undertaken for the majority class. Up-sampling categories in a dataset can be considered inappropriate for training machine learning models (Calder et al., 2022; McPherron et al., 2022) because it increases the overfit (samples used in the test set to evaluate the model are likely to have already been used in the training set). However, here, the up-sampling of the two minority groups increases the discreteness of these groups but does not affect the potential overlap with the majority class (core–edge flakes with limited backs). On the other hand, down-sampling

results in missing information because some of the data are removed.

Random up- and down-sampling is conducted to obtain a balanced dataset and train the models. After each random sampling, each model is evaluated using a k-fold cross validation using 10 folds and 50 cycles. Each fold consisted of 7 flakes, with the exception of the last fold, which had 6 flakes. Because model performance metrics depend on random up- and down-sampling, this process is repeated 30 times, extracting model performance metrics each time, and averaging the values. The model with the best performance metrics is then trained again with thirty cycles of up- and down-sampling. The reported variable importance and confusion matrix from which model metric performance are extracted are obtained from these additional cycles of down- and up-sampling.

Results

PCA and Machine Learning Model Performance

Results from the PCA on the 2D data show that the nine first principal components account for 95% of the variance in the dataset, with PC1 accounting for 39.33% of variance and PC9 for 0.98% of variance (Fig. 5). For the 3D data the 22
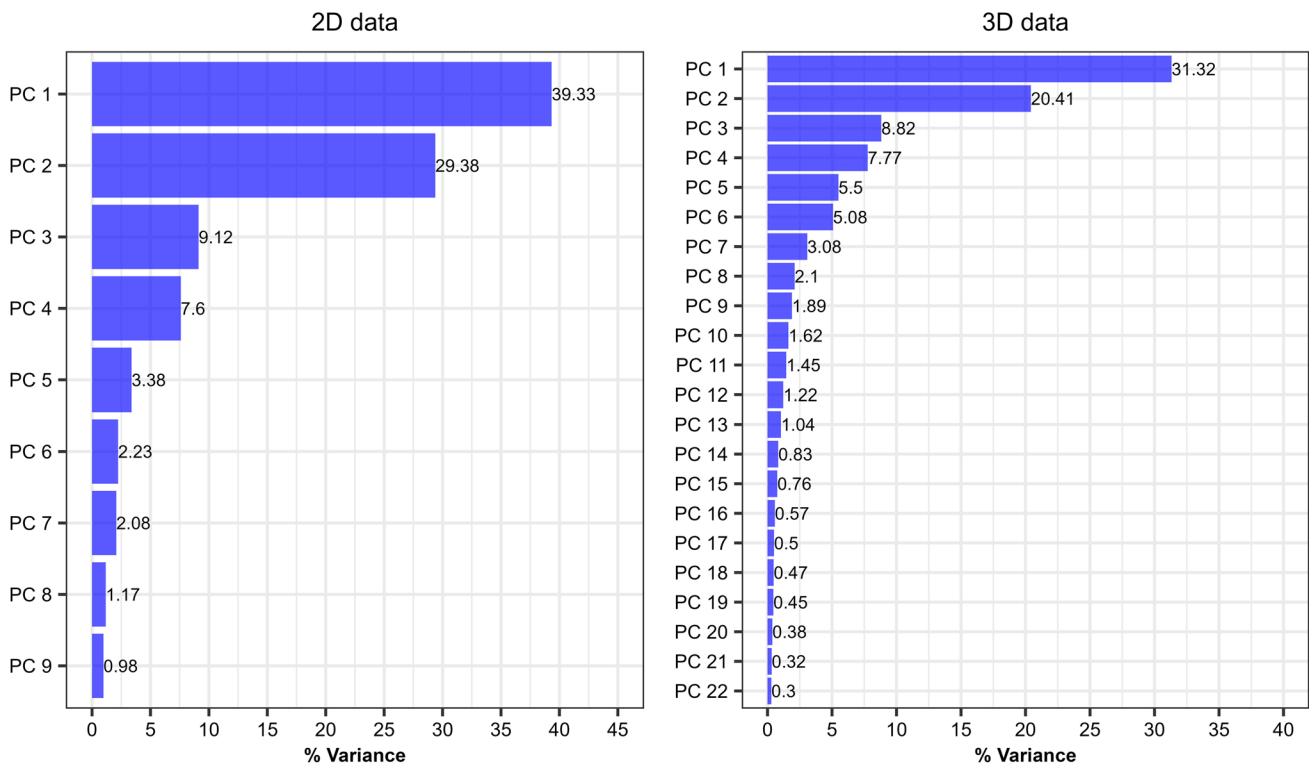

Fig. 5 Proportion of variance for the first PC which add up to 95% of variance in the 2D and 3D data

first principal components account for 95% of variance, with PC1 accounting for 31.32% and PC22 accounting for 0.3% of variance (Fig. 5). This represents a substantial reduction in dimensionality from the original number of variables of the 2D data (200 original variables) and the 3D data (1524 original variables) and is lower than the sample size (139).

Figure 6 presents the accuracy values for each model after their respective 30 cycles of random up- and down-sampling.
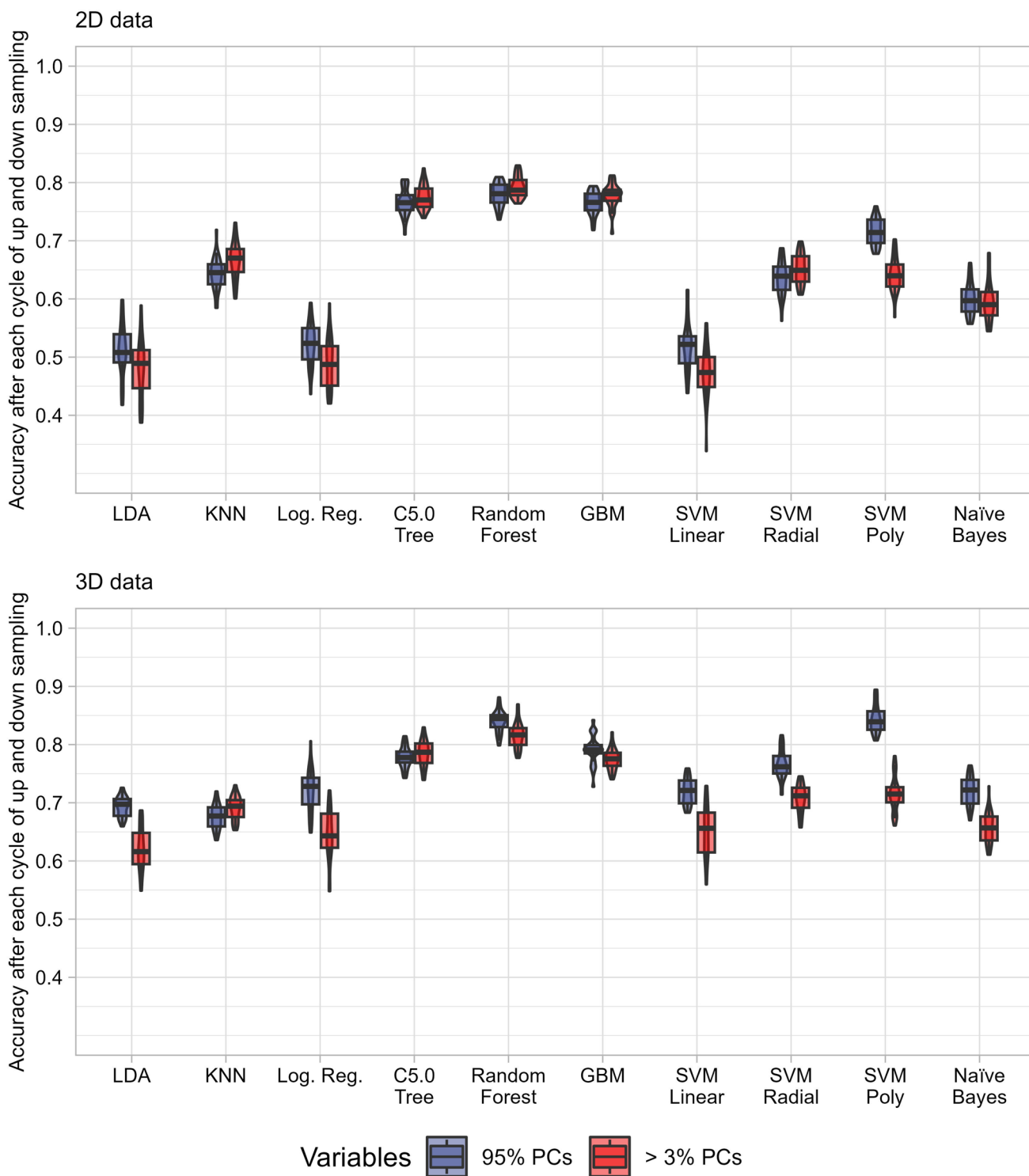
Fig. 6 Box and violin plots of model accuracy after each 30 cycles of random up- and down-sampling according to type of data (2D and 3D data) and number of variables employed (up to 95% of variance, or capturing more than 3% of variance)

These results are sorted according to the use of all variables up to 95% of variance, or only variables capturing more than 3% of variance.

Regarding the number of variables employed, important differences can be observed (Fig. 6). Models trained on 3D data presented a more general decrease on accuracy than models trained on 2D data. For both types of data (2D and 3D), tree-based machine learning models (C5.0 Tree, random forest and GBM) were the least affected when a limited number of variables was used. Contrary to this, the precision of the SVM models is strongly affected when a reduced number of variables is employed. SVM's using linear and polynomial kernels presented lower accuracy values when the number of variables is reduced. This is especially noteworthy in the case of the 3D data, since when variables summing 95% of variance are employed the SVM with polynomial kernel presents the highest accuracy values. However, when the number of variables is limited, the random forest presents the highest precision values. In general, models trained on 3D data presented higher overall precision metrics (independently of the number of variables employed for training the machine learning models).

In the case of the 2D data with a limited number of variables, the random forest model had the highest average value for general accuracy (0.785), closely followed by the decision tree (0.771) and the GBM (0.771). The LDA had the lowest average value for accuracy (0.496), followed by the SVM with linear kernel (0.497).

For the 3D data with limited number of variables, again the random forest presented the highest average value for general accuracy (0.828), followed by the GBM model (0.783). The LDA model had the lowest average value for accuracy (0.658), followed by logistic regression model (0.681).

Tables 2 and 3 present performance metrics of the random forest model on 2D data and on 3D data for the classification of the three products. The prevalence/no information ratio was kept constant at 0.33 for all three categories as a result of random up- and down-sampling to obtain balanced datasets. General performance metrics values (F1 and balance accuracy) of pseudo-Levallois points were similar for models trained on the 2D and 3D data. General performance metrics for the identification of core edge flakes and core edge flakes with a limited back did increase when 3D data was employed instead of 2D data.

Feature Importance

Figure 7 presents average variable importance for the three products after 30 cycles of up- and down-sampling and *k*-fold cross-validation when PC summing up to 95% of variance are employed. The random forest trained on 2D data consider four sets of principal components important in terms of classification. The SVM with polynomial kernel trained on 3D data considers five sets of principal components in terms of classification.

PC2 (29.38% of variance) is considered the most important variable for discrimination when using the 2D data, followed by PC1 (39.33% of variance), PC5 (3.38% of variance) and PC3 (9.12% of variance). Figure 8 presents shape change according to these PCs. PC1 and PC2 capture the elongation of flakes along an asymmetric axis, but with different orientation. Positive values of PC1 or PC2 result in wider flakes. PC3 captures variance of flakes where the proximal part is much wider than the distal portion of the flake. PC5 appears to capture pointed extremes resulting from concave delineations of the middle portion of the laterals. This might be representing the presence of convergent extremes at either end of the flake.

In the case of 3D data (Fig. 9), PC5 (5.5% of variance) is considered the most important variable for discrimination, followed by PC1 (31.32% of variance), PC6

Table 2 Performance metrics of random forest for backed artifacts using 2D data and limited number of variables

Class	Sensitivity	Specificity	Precision	F1	Prevalence	Balanced accuracy
Core–edge flake	0.760	0.847	0.760	0.736	0.333	0.804
Limited back	0.607	0.898	0.607	0.670	0.333	0.752
ps-Lev. point	0.950	0.914	0.950	0.895	0.333	0.932

Table 3 Performance metrics of random forest for backed artifacts using 3D data and limited number of variables

| Class	Sensitivity	Specificity	Precision	F1	Prevalence	Balanced accuracy
Core–edge flake	0.849	0.888	0.792	0.819	0.333	0.869
Limited back	0.662	0.909	0.785	0.719	0.333	0.786
ps-Lev. point	0.936	0.926	0.864	0.898	0.333	0.931
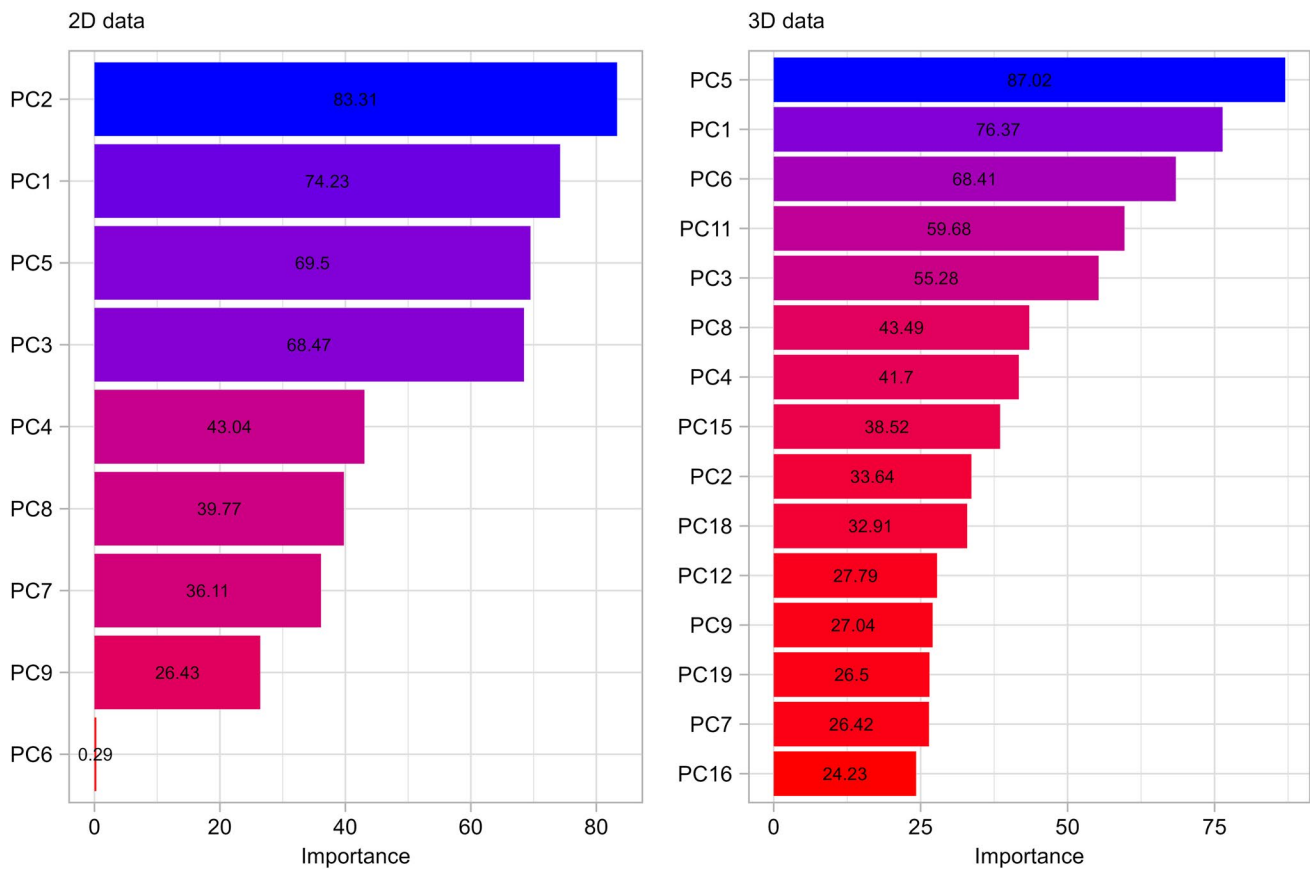
Fig. 7 Average feature importance after 30 cycles of up- and down-sampling using PC summing 95% of variance

(5.06% of variance), PC11 (1.45% of variance), and PC3 (8.82% of variance). PC11 presented an average importance value of 59.68. However, the low variance captured by this PC (1.45), and visual evaluation of shape change, indicate that its effect is minimum and should be excluded from analysis.

PC5 is driven by the interaction of platform depth and flake thickness. Increasing values of PC5 result in flakes with platforms much wider than deep, with the width of the platform finding their continuation in one of the abrupt laterals. Increasing PC5 values also result in thinner flakes. The negative space of PC5 captures flakes with platforms much more deep than wide and thicker flakes. PC1 largely captures elongation along with platform size and thickness. Positive PC1 values result in very wide flakes with a reduced length and bigger platforms. This increase in width is slightly accompanied with an increase in thickness. Negative space values result in thin elongated flakes with a distal convergent edge and small platforms. Positive values of PC6 represent the convergence of one of the distal laterals into a pointed end. The negative space of PC6 results in flakes with a wide proximal portion which

becomes narrower in the towards the distal part of the flake. PC3 represents transversal flake morphology and the relationship between thickness, width and asymmetry. Increasing values of PC3 result in thicker and narrower flakes with a marked asymmetry which results from a thick back located at the left lateral.

The combination of variance captured by each PC (Fig. 5), importance of a PC for classification (Fig. 7), and visualization of shape change according to PC (Figs. 8 and 9) indicate that a threshold of 3% variance captured is adequate for the present dataset. PC 5 of the 2D data presents an example of feature to be included in the analysis: despite capturing little more than 3% of variance (Fig. 5), it is considered important by the random forest as a classification feature (Fig. 7), and it does have an effect on sample shape (Fig. 8). PC11 of the 3D data presents an opposite example, were a variable should be excluded from the training of the final models. While it is considered an important variable for classification by the SVM with polynomial kernel, the low variance captured (1.45%) and the little effect on shape change prevent its inclusion into the final set of machine learning models.
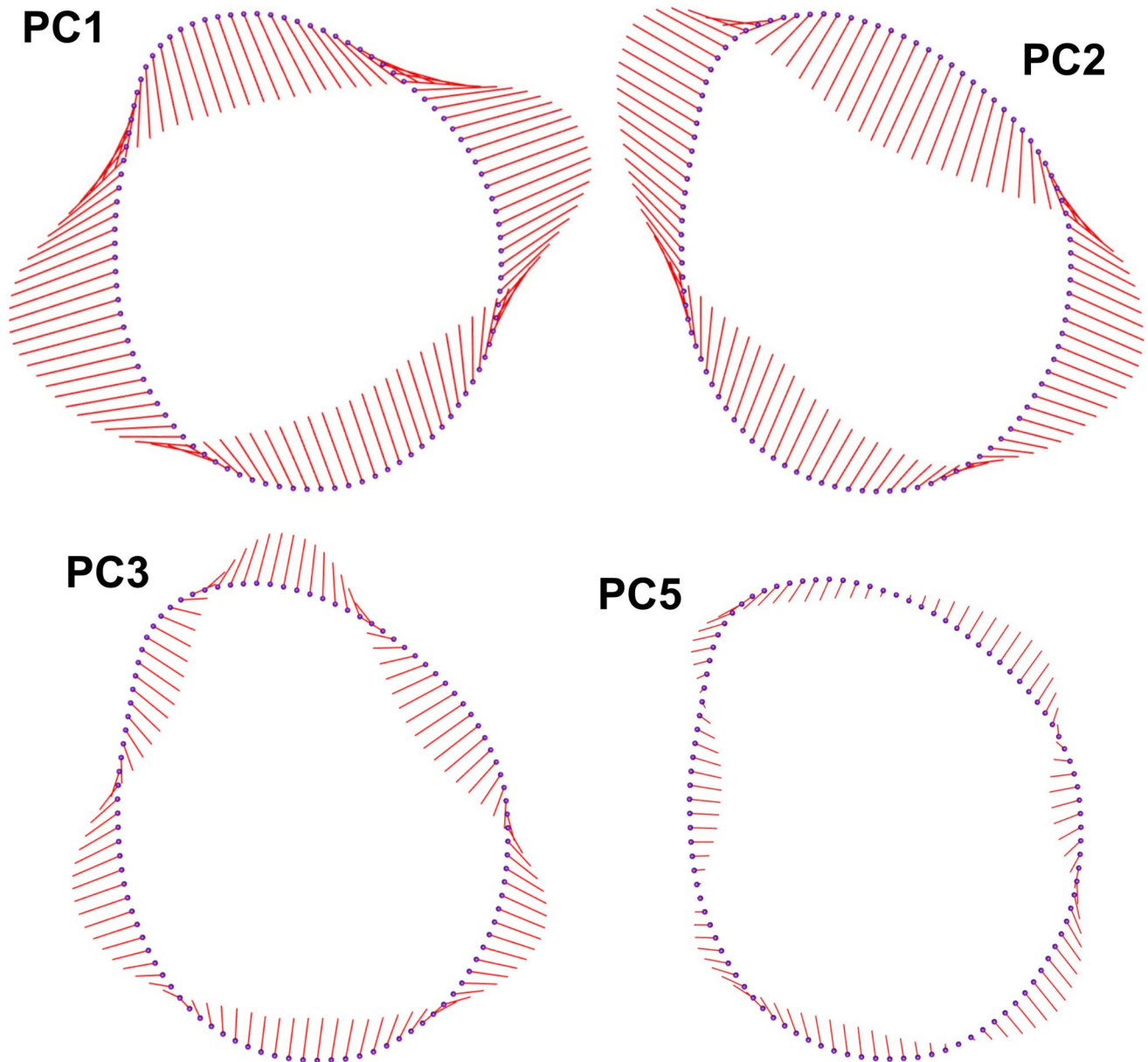
Fig. 8 Visualization of shape change according to PC for the 2D data. Differences towards mean shape have been magnified by a factor of three

Group Discreteness Through Confusion Matrix and PCA Biplots

The confusion matrixes of final models trained with limited number of variables (Fig. 10) illustrate the directionality of confusion between the predicted and true values of classified technological products for the 2D and 3D data. In both types of data pseudo-Levallois points have the best identification, in accordance with the reported sensitivity and specificity. In general, it is very difficult to mistake pseudo-Levallois points for any of the two considered technological products. Wrongly considering a pseudo-Levallois point as a core–edge flake is very unlikely for the 2D data (2.4) and

minimal for the model on 3D data (1.59). Although mistaking a pseudo-Levallois point for a core–edge flake with a limited back is slightly more common for both types of data, there is still a very low confusion value.

Core–edge flakes and core–edge flakes with a limited back offer slightly higher frequencies of misidentifications, although they maintain high values for sensitivity and specificity for both types of data. For the random forest model based on the 2D data, it is more common to mislabel core–edge flakes with a limited back as core edge flakes (28.18) than the reverse (17.81).

In the case of the random forest based on 3D data, the confusion between both categories diminishes slightly
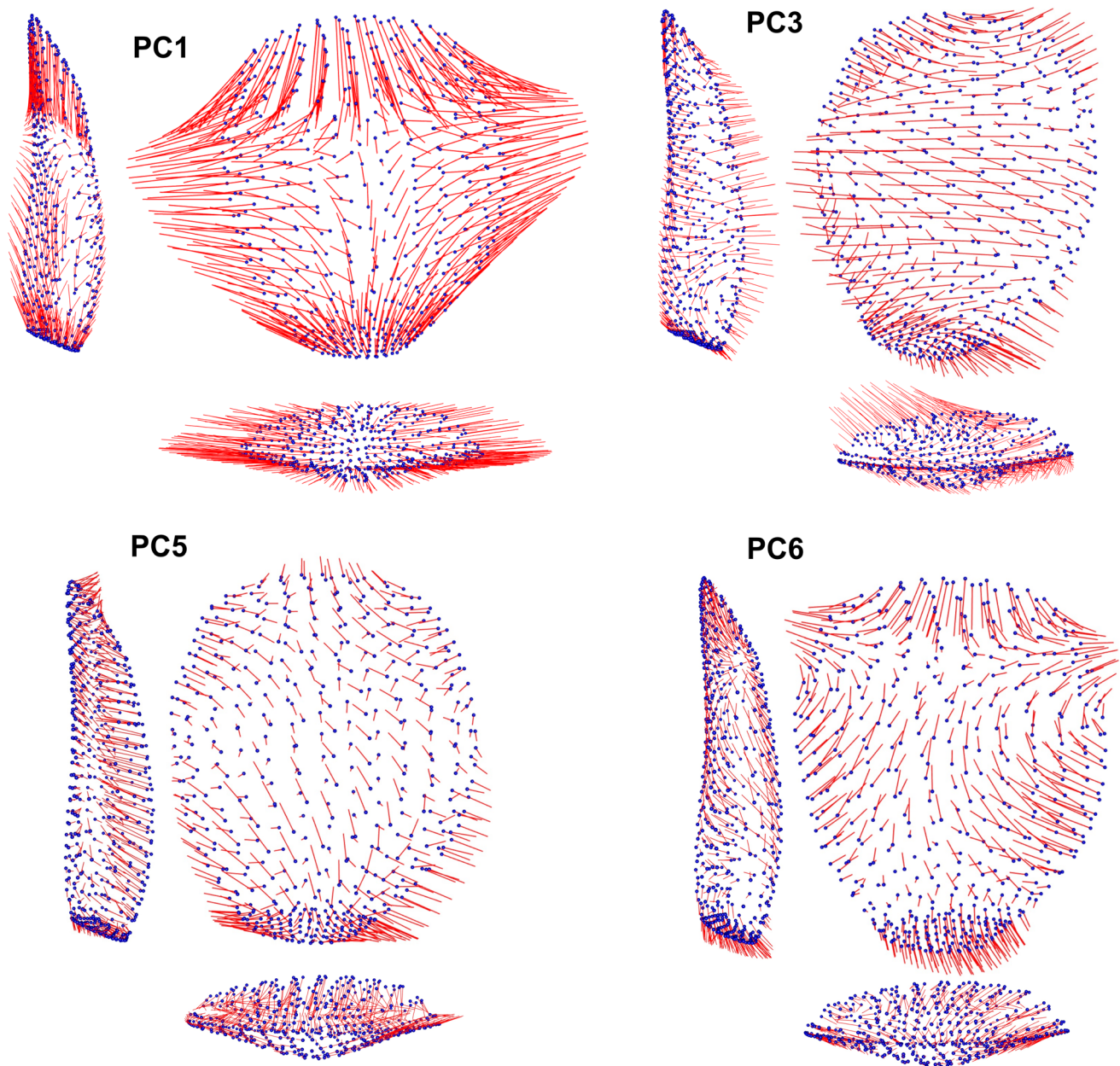
Fig. 9 Visualization of shape change according to PC of the 3D data. Differences towards mean shape have been magnified by a factor of three

(20.78 and 13.3). In general, the 3D data better identifies core–edge flakes with a limited back and core–edge flakes than the model based on the 2D data. This greater precision is the result of a lower frequency in the incorrect identification of core edge flakes with a limited back as core–edge flakes. The incorrect identification of backed flakes as pseudo-Levallois points is slightly more common when 2D data is employed. Their incorrect identification as pseudo-Levallois flakes is minimal in the case of the 3D data, although this incorrect identification has a higher frequency in core–edge flakes with a limited back (4.82) than in core–edge flakes (1.59) (Figs. 11, 12, and 13).

The above interpretation of the PCs, biplot visualization (Figs. 8 and 9), and descriptive statistics of PC values (Fig. 14) allow us to evaluate the morphological features captured by the 2D and 3D geometric morphometrics and the characterization of each type of technological backed flake. In general, biplots from 2D data show much more overlap than biplots from 3D data (Figs. 11 and 12). The higher overlap observed on the PC biplot from the 2D data is also observed when a group PCA is performed using the most important variables for classification (Fig. 13). Visual analysis of group PCA biplot from variables of the 3D data shows much less overlap between the three categories.
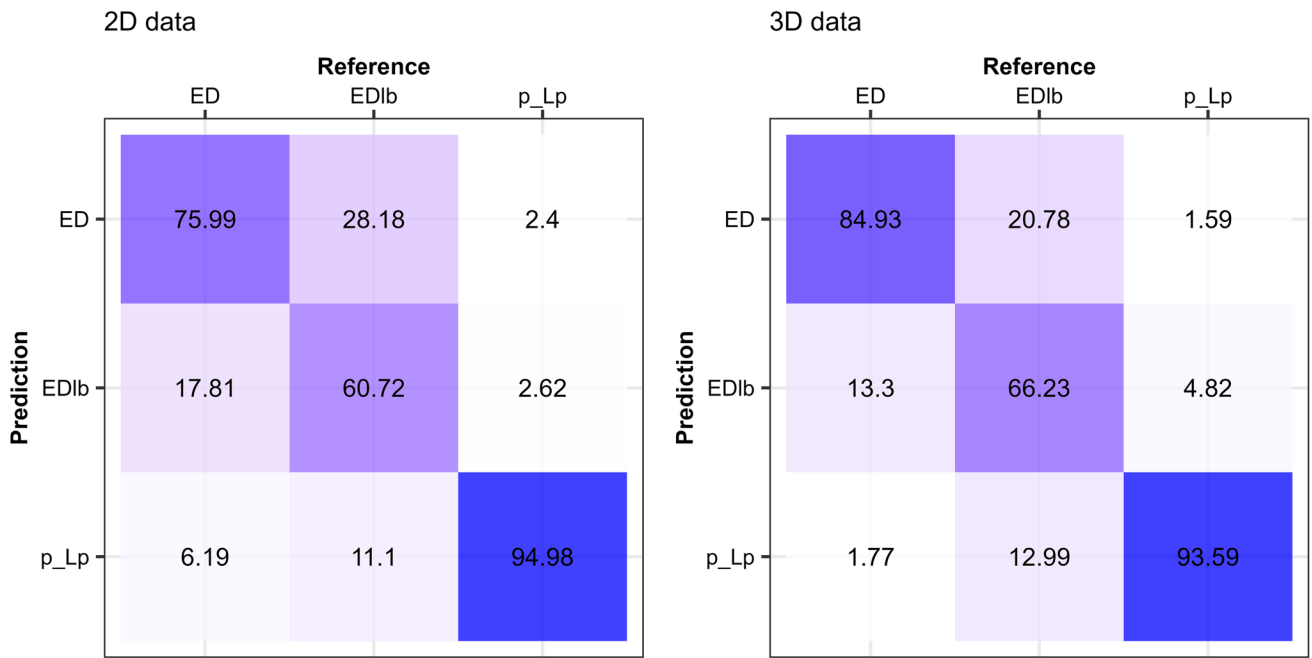
Fig. 10 Normalized confusion matrix of random forest for the 2D and 3D data when only PC capturing more than 3% of variance are employed (ED = Eclat débordant; EDlb = éclat débordant with limited back; p_Lp = pseudo-Levallois point)
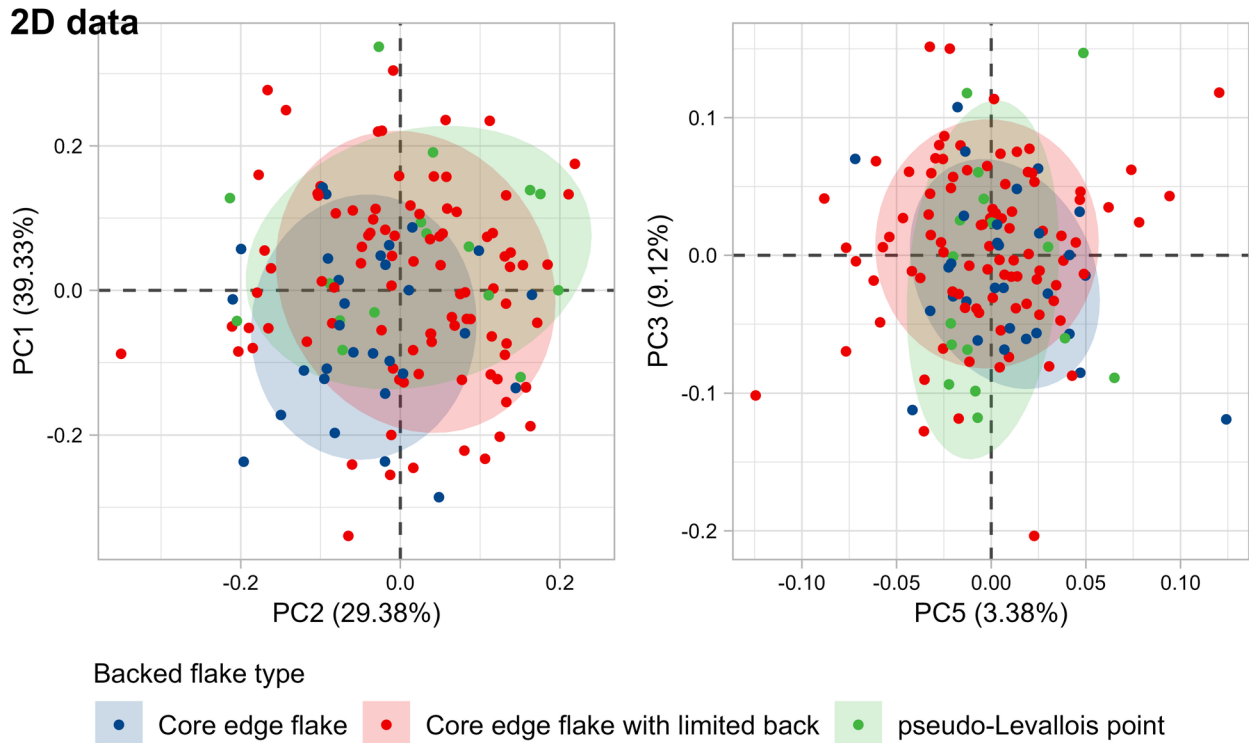
Fig. 11 Biplots of PC2, PC1, PC3, and PC5 according to flake category on the 2D data. Ellipses indicate the confidence level at 80%
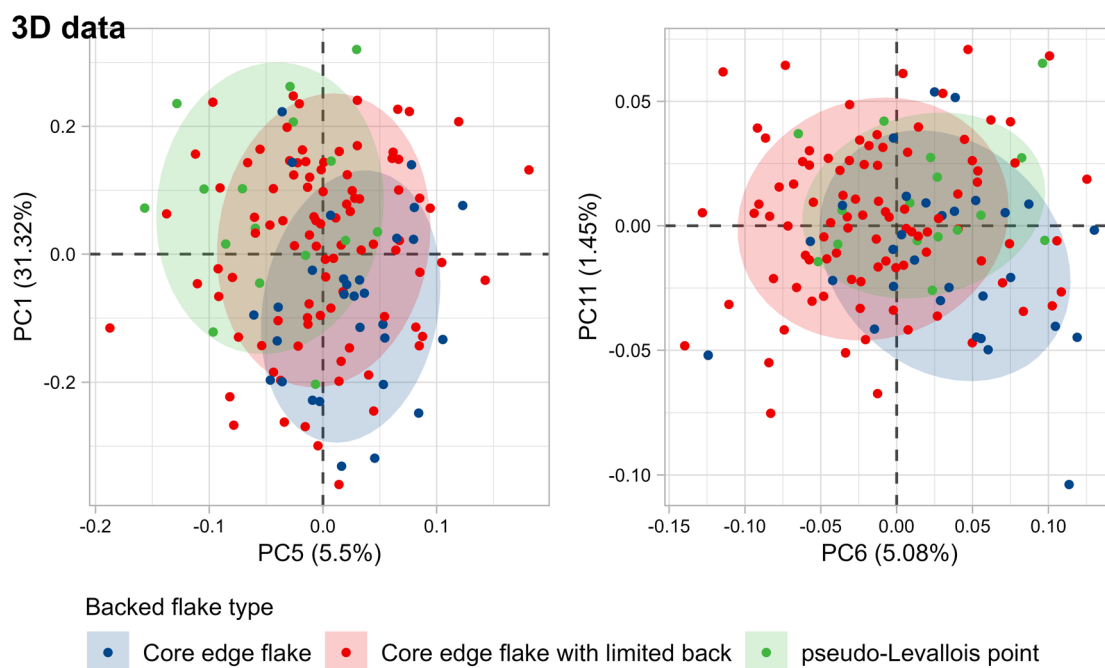
Fig. 12 Biplots of PC7, PC1, PC3, and PC4 according to flake category on the 2D data. Ellipses indicate the confidence level at 80%
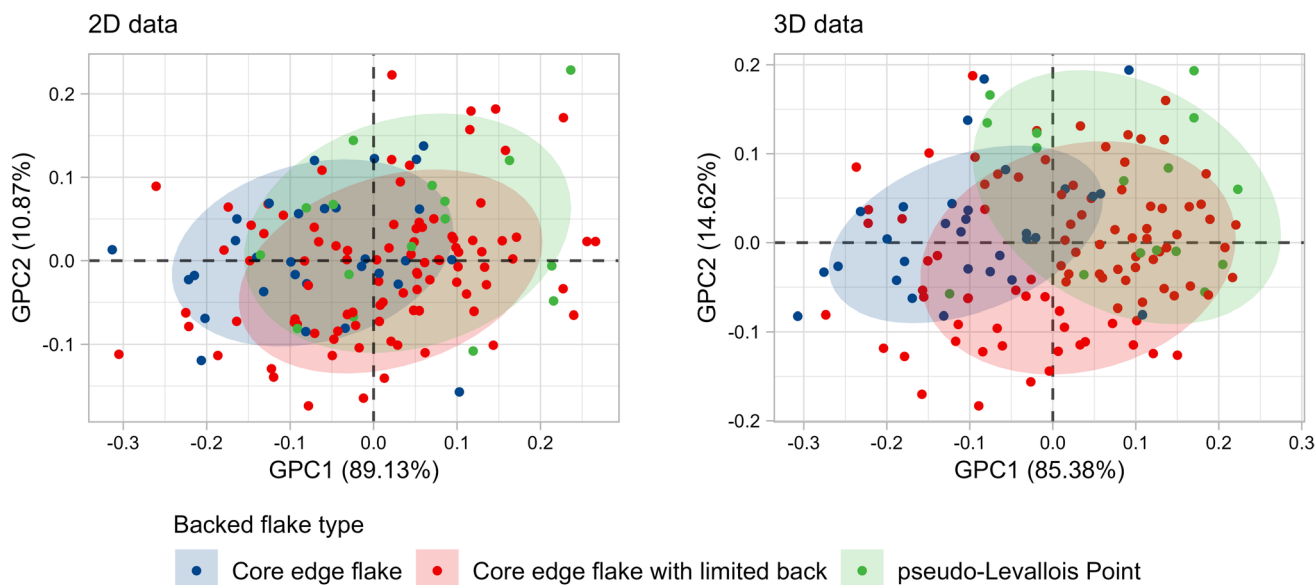

Fig. 13 Biplots of group PCA using the most important features for classification

While core–edge flakes and pseudo-Levallois points show little overlap between each other, core–edge flakes are situated as an intermediate product, pointing out to their wide morphological variability.

For the 2D data, pseudo-Levallois points are characterized by having positive values of PC1 (mean = 0.53; SD = 0.116) and PC2 (mean = 0.168; SD = 0.128). These

generally positive values are indicative of the low elongation of pseudo-Levallois points. The triangular off-axis morphology characteristic of pseudo-Levallois points is weakly expressed for the 2D data and is only slightly captured by PC5 (mean = 0.002; SD = 0.028). On the contrary, 2D geometric morphometrics is useful to detect features of core–edge flakes, although the overlap with core edge
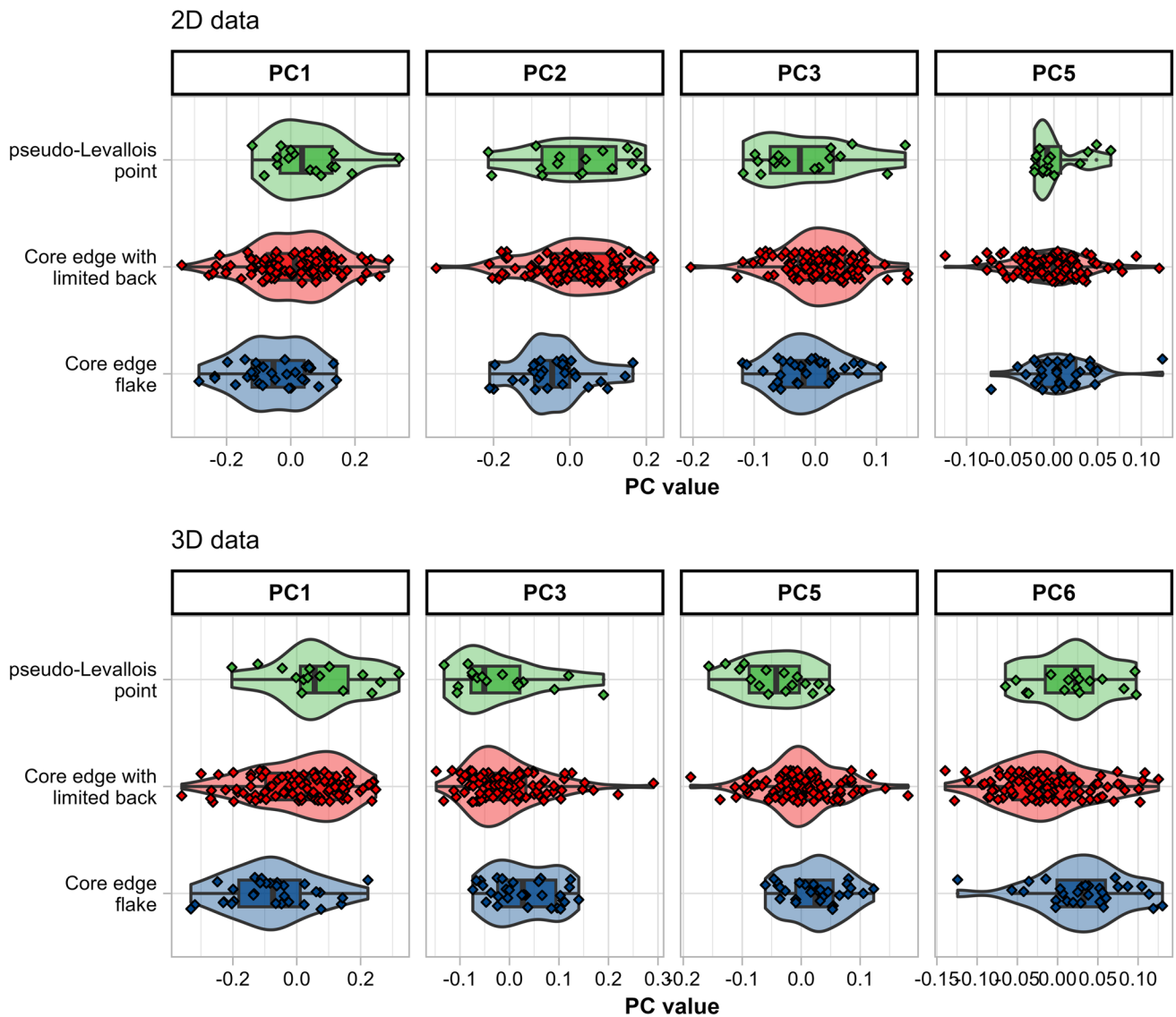
Fig. 14 Box and violin plots of PC values according to each backed flake category

flakes with a limited back is very high. In the 2D data core edge flakes are characterized by having low values of PC1 (mean = −0.052; SD = 0.111) and PC2 (mean = −0.043; SD = 0.092) which are indicative of being elongated products. Core edge flakes (mean = −0.014, SD = 0.055) and pseudo-Levallois points (mean = −0.014, SD = 0.079) tend to have negative values of PC3. This indicates that, in general, neither of the two products had a distal portion narrower than the proximal one. Core–edge flakes with a limited back presented slightly positive values of PC1 (mean = 0.008; SD = 0.131) and PC2 (mean = 0.011; SD = 0.111), indicating that their elongation and asymmetry falls between core–edge flakes and pseudo-Levallois points. Additionally, core–edge flakes with a limited back presented slightly positive values of PC3 (mean = 0.007;

SD = 0.06), indicating that it is more common for these types of products to present a narrower distal portion.

For the 3D data, pseudo-Levallois points were characterized by high values for PC1 (mean = 0.031; SD = 0.055), low values for PC3 (mean = 0.024; SD = 0.09) and negative values for PC5 (mean = −0.046; SD = 0.059). Finally, pseudo-Levallois points also exhibit intermediate positive values for PC6 (mean = 0.018; SD = 0.05). The combination of these PCs and their values indicates that the geometric morphometrics is capturing the low elongation (PC1), asymmetry (PC3), and to some extent, the triangular morphology resulting from the convergence of two edges (PC5 and PC6). Core–edge flakes exhibited negative values of PC1 (mean = −0.078; SD = 0.136), positive values of PC5 (mean = 0.024; SD = 0.048) and PC6 (mean = 0.031;

SD = 0.055). These PCs capture the elongated nature (PC1) of core edge flakes (in comparison to the other two categories), their lower ratio of length to thickness (PC5) and the presence of a distal transverse edge which can result in distal pointed portions. Core–edge flakes with a limited back are characterized by intermediate values of PC1 (mean = 0.012, SD = 0.142), PC3 (mean = − 0.006, SD = 0.078) and PC5 (mean = 0.00; SD = 0.061) and slightly negative values of PC6 (mean = − 0.013; SD = 0.058). The combination of values from these PC reflects the wide morphological variability of core–edge flakes with a limited back were the main features are a low elongation (PC1), varying ratios of thickness and platform morphology (PC5) and strong variability of the upper view (PC6).

Discussion

Our results show that geometric morphometrics, along with machine learning models, can easily differentiate between core edge flakes, core edge flakes with a limited back, and pseudo-Levallois points from discoidal and recurrent centripetal Levallois reduction sequences. The best model trained on the 2D data (random forest) obtained a good general average precision value (0.785), while the best model trained on the 3D data (also random forest) obtained a high general average precision value (0.828). As expected, this indicates that when dealing with morphological objects whose classification requires several views, more precise results will be obtained using 3D data. Despite the differences in average precision values and performance measures for each of the categories, 2D geometric morphometrics were able to capture some of the morphological features which characterize the three categories. Thus, looking to methods that can facilitate the proper comparison between stone tool assemblages, based on corresponding classifications, and reducing the margin of error in analyzing the similarities between data samples, 2D geometric morphometrics are not suitable for generating lithic taxonomies, at least in relation with the three technological categories analyzed in this study. Core–edge flakes, core–edge flakes with a limited back, and pseudo-Levallois points are relevant categories when looking at patterns of core reductions and variability within centripetal core knapping methods, whether due to raw material constraints or the search for specific tool types.

Several studies have employed machine learning models based on attribute analysis (Bustos-Pérez et al., 2023; González-Molina et al., 2020; Presnyakova et al., 2015) or geometric morphometrics (Archer et al., 2021; Bustos-Pérez et al., 2022). SVM with polynomial kernel and random forest stand out as the most common or accurate models when performing classificatory tasks among lithic artifacts. Previous studies have outlined the advantages of SVM's when dealing

with the classification of lithic products (Bustos-Pérez et al., 2022, 2023). Because of their features (hyperplane fitting, use of margins to find best separation, and use of a cost value for each misclassification) SVM's seem well suited for the classification of lithic artifacts. In the present study, for the 3D data SVM with polynomial kernel outperformed the rest of the machine learning models. However, when the number of variables employed as predictors was restricted, its general accuracy was reduced to a value of 0.78, making the random forest the model with the best average accuracy. Random forest was also selected as the best option when dealing with 2D data independent of the number of variables employed. Random forest are common in lithic analysis for classification tasks (Archer et al., 2021; González-Molina et al., 2020). The results from model performance on the 2D and 3D data indicate that random forests along with SVM with polynomial kernel are good options for classification when dealing with the variability observed in lithic artifacts, although the number of variables introduced as predictors should be taken into account.

When considering each technological product individually, pseudo-Levallois points stood out as the most clearly differentiable of the three categories considered, with performance metrics above 0.9 in PCs derived from both the 2D or 3D data. Following pseudo-Levallois points, core–edge flakes were the most clearly identifiable technological products, with a notable sensitivity value. The sensitivity values for the detection of core–edge flakes were notably higher in the case of 3D data (84.9) than the ones obtained from the 2D data (0.76). The directionality of the confusion matrixes shows that the main drawback reducing the identification of core–edge flakes is their identification as core–edge flakes with a limited back in both types of data. Two underlying causes of confusion between core–edge flakes and core–edge flakes with a limited back can be considered: an increased deviation between the technological and morphological axes and an increased angle between the platform and backed edge, which results in changing the morphological axis. These two factors can occur at the same time or individually, blurring the division between products in cases of similarity. This overlap is inherent in the morphological variability and defining features of both technological products. However, despite these overlapping features, a high degree of separation between both products is achieved by the machine learning model.

This difference in the sensitivity values depending on the type of data is also observed in the case of core–edge flakes with a limited back. The sensitivity value of the 3D data (0.662) was higher than the one obtained from the 2D data (0.607). In the present study, core–edge flakes with a limited back were the category of interest and on which down-sampling was applied, thus preventing overfitting. This indicates that geometric morphometrics are capturing

the technological features defining each category and that they are discrete categories with little overlap. Thus, their use (and, more specifically, the use of the core–edge flakes with a limited back as a category) for analysis in lists of technological products is justifiable.

The present study included all backed flakes from a series of experimental recurrent centripetal cores. As a result, backed artifacts that fall within the definition of "core–edge flakes with a limited back" were the overwhelming majority (n = 93, 66.91%). Consequently, it was necessary to use up- and down-sampling techniques to obtain balanced datasets (Ganganwar, 2012; Kumar & Sheshadri, 2012). The most up-sampled product (pseudo-Levallois) is also the one with the highest values for identification metrics irrespective of the type of data (2D or 3D), probably because of overfitting (the model is classifying repeated samples from the training set). This overfitting is noteworthy in the case of the 2D data were the high level of sensitivity contrasts with the visual analysis of biplots obtained from PCA (Fig. 11) or group PCA (Fig. 13) where a high level of overlap can be observed. Unbalanced data is common in archaeological analysis and sampling approaches to overcome this drawback can result in overfitting when machine learning models are applied (Calder et al., 2022; Domínguez-Rodrigo & Baquedano, 2018; McPherron et al., 2022). Providing visual evaluation through biplots can help determine the extent of overfitting and the reliability of performance metrics from machine learning models. Further research applying sampling techniques to archaeological problems might consider alternative solutions to unbalanced datasets. An alternative is the use of down-sampling in combination with leave-one-out cross-validation (were each case of the downsized dataset serves as a single test set). However, this approach also comes with the drawback of losing information, which can result in machine learning models not capturing the features which characterizing each class.

Given the strict definition adopted to classify a backed artifact as either a core–edge flake or a pseudo-Levallois point, their morphological variability is limited, and the likelihood of them overlapping is small. This is logical given their definitions and can be observed in the biplots from 3D data of most important PCs for the classification of backed products (Figs. 12 and 13). In the two biplots and the group PCA biplot (Figs. 12 and 13), there is little overlap of confidence ellipses of core–edge flakes and pseudo-Levallois points. Moreover, the confidence ellipsis of core–edge flakes with a limited back does seem to be intermediate between the other two categories. Thus, although up-sampling imposes some limitations, it does not seem to affect the overall results regarding backed flake classification. Core–edge flakes with a limited back were not up-sampled, thus avoiding the risk of overfitting their classification (having an observation on the training set repeated

in the test set). The results show a very limited misidentification of core–edge flakes as pseudo-Levallois points and their moderate confusion with core–edge flakes. This indicates that core–edge flakes with a limited back are being correctly identified despite the probable overfitting in the identification of core edge flakes and pseudo-Levallois points.

The present study has compared the use of 2D and 3D geometric morphometrics. 2D geometric morphometrics have the advantage of being less time consuming, requiring less equipment, and being little affected by camera positioning (Cardini & Chiapelli, 2020; Macdonald et al., 2020). Our research showed that the 2D dataset was not sufficiently useful to discriminate between the lithic backed categories and, therefore, not reasonably informative, at least for the three technological categories analyzed. Some authors considered 2D geometrics a useful tool for lithic analysis when performing lithic taxonomic or comparative studies. Previous studies using 2D geometric morphometrics aimed at analyzing formal retouched tools (large tanged points; and Clovis, Folsom, and Plainview projectile points; Buchanan & Collard, 2010; Serwatka & Riede, 2016), testing the relationship between shape and function in informal flakes (Borel et al., 2017), and compare similarity of bone and stone Acheulean bifaces (Costa, 2010). The cited 2D studies were not comparing between 2 and 3D datasets when analyzing the same sample, while the present study aimed at testing methodological aspects to improve lithic taxonomic analysis and refine stone tool technological description and interpretation. Furthermore, these studies mainly analyzed morphological aspects quantitatively. It seems reasonable that, when dealing with bidimensional surfaces, 2D geometric morphometrics should be valuable. The present research has addressed a volumetric problem. Thus, the better performance of 3D geometric morphometrics falls within what was expected. In the same line, previous works have outlined how, when working with volumetric structures, the use of 2D geometric morphometrics can result in a lower resolution of the analysis (Buser et al., 2018; Cardini & Chiapelli, 2020). The present results indicate the preferable application of 3D techniques for the identification of technological backed categories related to core reduction strategies, suggesting that more studies are needed to test the best approach when dealing with different research questions and technological categories, including cores, retouched tool-types, unretouched items, and shaped elements.

Sullivan and Rozen (1985) previously called attention to the use of technological categories of flakes. Their critique focuses on the lack of consistency in defining and using technological categories of flakes, along with a lack of consistency regarding the attributes employed to define them. Although their critique concerns flakes from bifacial knapping, it represents an important word of caution. The present study has shown that the three analyzed categories

are well defined in terms of morphological and technological features (see also Faivre et al., 2017). These morphological and technological features can be captured by geometric morphometrics, especially 3D techniques, and employed for accurate classifications. Thus, little ambiguity exists in the categories employed to classify backed flakes, which are connected to two of the main Middle Paleolithic flaking strategies (discoidal and recurrent centripetal Levallois).

Meignen (1993, 1996) first proposed the category "core–edge flake with a limited back" because better individualized and characterized the predominant centripetal debitage present at the site of Les Canalettes (Aveyron, France). The term has, however, seen limited use since, mainly employed when characterizing lithic assemblages produced via recurrent centripetal methods (Bernard-Guelle, 2004; Bourguignon & Meignen, 2010; Duran, 2005; Duran & Abelanet, 2004; Duran & Soler, 2006). The use of the "core–edge flake with a limited back" category is usually overlooked (or merged into the core edge flake category) when lists of technological products are employed for the analysis of Middle Paleolithic lithic assemblages (Debénath & Dibble, 1994; Geneste, 1988; Shea, 2013b). Although their merger into the core–edge flake category is valid for analyzing lithic assemblages, differentiating between classic core–edge flakes and core–edge flakes with a limited back in lists of technological products can increase the resolution and enrich the analysis of lithic assemblages. This, of course, results in a better understanding of the technical choices and constraints faced by past groups and may allow for better understanding of the relationships between technology and raw material, as well as volumetric core management, and could contribute to improving comparisons between assemblages to avoid simple and overly generic technological definitions and to avoid merging what are clearly different reduction concepts. Such generic definitions are often used to discuss human behaviors and cultural traditions and even human migrations (Blinkhorn et al., 2021), but they remain difficult to interpret (Romagnoli et al., 2022). They do not always reflect the diversity and specificity of the technological choices and constraints past human groups faced. A more in-depth analysis of variability in the techno-morphological characteristics of stone artifacts, in association with other approaches, such as refitting and taphonomic analysis (see bibliography in Romagnoli & Vaquero, 2019; Romagnoli et al., 2018), could also allow archaeologists to better understand the processes of assemblage formation and interpret technological changes throughout human evolution.

Previous researchers have pointed out the morphological and technological differences between classic core–edge flakes and core–edge flakes with a limited back. Beyries and Boëda (1983) originally defined classic core–edge flakes as having very similar morphological and percussion axes. Meignen (1993, 1996) used the "core–edge flake with a limited back" category to classify core edge flakes in which the morphological and percussion axes were not aligned. Additionally, most of the examples presented (Meignen, 1996) had the flake back offset from the percussion axis. Slimak (2003) also points out that, as a morphological feature, the flake back of core–edge flakes with a limited back will be offset in regards the axis of percussion. These features of alignment between the flake back and percussion axis are captured by PC1 and PC3, with pseudo-Levallois and core–edge flakes with a limited back having similar values, indicative of an offset backed edge and lower elongation.

Slimak (2003) also indicates that classic core edge flakes will be elongated because of percussion running parallel to the core edge. Following Slimak (2003), core–edge flakes with a limited back will have length/width ratios close to 1 as a result of percussion encountering ridges perpendicular to its direction. Although in the 3D data PC1 is the second most important variable for discrimination between backed products in the present study, it is clearly capturing the feature of elongation for the differentiation of classic core–edge flakes. Classic core–edge flakes have, on average, negative PC1 values (both on the 2D and 3D data), which are indicative of elongation. On the other hand, core–edge flakes with a limited back and pseudo-Levallois points are characterized by, on average, positive PC1 values, which are indicative of similar values for width and length or even being wider than they are long.

Ambiguity and overlap between some technological categories of flakes are common. Combining quantitative methods and techniques has showed to be a useful approach to distinguish between backed products extracted during discoid and recurrent centripetal Levallois knapping strategies. The present research shows that geometric morphometrics along with machine learning models is an effective way also to test for the discreteness of categories, the possible directionality of confusions between categories and to quantify the features which best characterize and define each category.

Conclusions

The present work aimed to evaluate whether "core–edge flakes with a limited back" are a discrete category that can be separated from classic core–edge flakes and pseudo-Levallois points. These products are defined by a series of morphological and technological features (overall shape morphology, morphological symmetry, axis of percussion, and position and angle of the backed edge). These features and their variability can be quantitatively captured by geometric morphometrics and employed in machine learning

models to test for the discreteness of the categories. The results indicate that, while some overlap exists between classical core edge flakes and core edge flakes with a limited back (pseudo-Levallois points are clearly differentiated), in general, they are easily distinguishable. Additionally, geometric morphometrics and machine learning also succeed in capturing the PCs directly associated with morphological and technological features employed to define each technological category. However, the precision of machine learning models can be affected by the number PCs used as predictors. For the present sample and study, tree-based methods seem to be little affected by the reduction of variables. As expected, 3D geometric morphometrics better capture the features which characterize volumetric implements and better classify them. Core–edge flakes with a limited back are therefore clearly distinguishable from classic core–edge flakes and geometric morphometrics can be employed to test for the validity of defined technological categories, the directionality of confusions, and the features characterizing these categories.

Declarations

Competing interests The authors declare no competing interests.

Ethical Approval Not applicable.

References

Adler, D. S., Wilkinson, K. N., Blockley, S., Mark, D. F., Pinhasi, R., Schmidt-Magee, B. A., et al. (2014). Early Levallois technology and the lower to middle Paleolithic transition in the Southern Caucasus. *Science, 345*(6204), 1609–1613. https://doi.org/10.1126/science.1256484

Archer, W., Pop, C. M., Rezek, Z., Schlager, S., Lin, S. C., Weiss, M., et al. (2018). A geometric morphometric relationship predicts stone flake shape and size variability. *Archaeological and Anthropological Sciences, 10*(8), 1991–2003. https://doi.org/10.1007/s12520-017-0517-2

Archer, W., Djakovic, I., Brenet, M., Bourguignon, L., Presnyakova, D., Schlager, S., et al. (2021). Quantifying differences in hominin flaking technologies with 3D shape analysis. *Journal of Human Evolution, 150*, 102912. https://doi.org/10.1016/j.jhevol.2020.102912

Bernard-Guelle, S. (2004). Un site moustérien dans le Jura suisse: la grotte de Cotencher (Rochefort, Neuchâtel) revisitée. *Bulletin De La Société Préhistorique Française, 101*(4), 741–769. https://doi.org/10.3406/bspf.2004.13066

Beyries, S., & Boëda, E. (1983). Étude technoloogique et traces d'utilisation des éclats débordants de Corbehem (Pas-de-Calais). *Bulletin De La Société Préhistorique Française, 80*(9), 275–279. https://doi.org/10.3406/bspf.1983.5455

Blinkhorn, J., Groucutt, H. S., Scerri, E. M. L., Petraglia, M. D., & Blockley, S. (2021). Directional changes in Levallois core technologies between Eastern Africa, Arabia, and the Levant during MIS 5. *Scientific Reports, 11*(1), 11465. https://doi.org/10.1038/s41598-021-90744-z

Boëda, E. (1993). Le débitage discoïde et le débitage Levallois récurrent centripède. *Bulletin De La Société Préhistorique Française, 90*(6), 392–404. https://doi.org/10.3406/bspf.1993.9669

Boëda, E. (1994). *Le concept Levallois: Variabilité des méthodes* (p. 280). CNRS.

Boëda, E. (1995a). Caractéristiques techniques des chaînes opératoires lithiques des niveaux micoquiens de Külna (Tchécoslovaquie). *Paléo. Supplément, 1*(1), 57–72. https://doi.org/10.3406/pal.1995.1380

Boëda, E. (1995b). Levallois: A volumetric construction, methods, a technique. In H. L. Dibble & O. Bar-Yosef (Eds.), *The definition and interpretation of Levallois technology* (Vol. 23, pp. 41–68). Prehistory Press.

Boëda, E., Geneste, J.-M., & Meignen, L. (1990). Identification de chaînes opératoires lithiques du Paléolithique ancien et moyen. *Paléo, 2*, 43–80.

Bookstein, F. L. (1997a). Landmark methods for forms without landmarks: morphometrics of group differences in outline shape. *Medical Image Analysis, 1*(3), 225–243. https://doi.org/10.1016/S1361-8415(97)85012-8

Bookstein, F. L. (1997b). *Morphometric tools for landmark data*. Cambridge University Press.

Bordes, F. (1953). Notules de typologie paléolithique II : Pointes Levalloisiennes et pointes pseudo-levalloisiennes. *Bulletin De La Société Préhistorique Française, 50*(5–6), 311–313. https://doi.org/10.3406/bspf.1953.3057

Bordes, F. (1961). *Typologie du Paléolithique Ancien et Moyen*. CNRS.

Borel, A., Cornette, R., & Baylac, M. (2017). Stone tool forms and functions: A morphometric analysis of modern humans' stone tools from Song Terus Cave (Java, Indonesia): Stone tool forms and functions. *Archaeometry, 59*(3), 455–471. https://doi.org/10.1111/arcm.12264

Bourguignon, L. (1992). Analyse du processus opératoire des coups de tranchet latéraux dans l'industrie moustérienne de l'abri du Musée (Les Eyzies-de-Tayac, Dordogne). *Paléo, 4*(1), 69–89. https://doi.org/10.3406/pal.1992.1195

Bourguignon, L., & Meignen, L. (2010). Ioton (Gard) 30 ans après : nouvelles considérations technologiques et techno-économiques sur l'industrie moustérienne. *Bulletin De La Société Préhistorique Française, 107*(3), 433–451. https://doi.org/10.3406/bspf.2010.13944

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Buchanan, B., & Collard, M. (2010). A geometric morphometrics-based assessment of blade shape differences among Paleoindian projectile point types from western North America. *Journal of Archaeological Science, 37*(2), 350–359. https://doi.org/10.1016/j.jas.2009.09.047

Buser, T. J., Sidlauskas, B.L., & Summers, A.P. (2018). 2D or Not 2D? Testing the utility of 2D Vs. 3D landmark data in geometric morphometrics of the sculpin subfamily Oligocottinae (Pisces; Cottoidea). *The Anatomical Record, 301*(5), 806–818. https://doi.org/10.1002/ar.23752

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., Gravina, B., Brenet, M., & Romagnoli, F. (2022). Combining quantitative approaches to differentiate between backed products from discoidal and Levallois reduction sequences. *Journal of Archaeological Science: Reports, 46*, 103723. https://doi.org/10.1016/j.jasrep.2022.103723

Bustos-Pérez, G., Baena, J., & Vaquero, M. (2023). What lies in between: Levallois, discoid and intermediate methods. *Journal of Lithic Studies, 10*(2), 32. https://doi.org/10.2218/jls.7132

Calder, J., Coil, R., Melton, J. A., Olver, P. J., Tostevin, G., & Yezzi-Woodley, K. (2022). Use and misuse of machine learning in anthropology. *IEEE BITS the Information Theory Magazine, 2*(1), 102–115.

Callahan, E. (1996). The bipolar technique: The simplest way to make stone tools for survival. *Bulletin of Primitive Technology, 12*, 16–20.

Cardini, A., & Chiapelli, M. (2020). How flat can a horse be? Exploring 2D approximations of 3D crania in equids. *Zoology, 139*, 125746. https://doi.org/10.1016/j.zool.2020.125746

Cignoni, P., Callieri, M., Corsini, M., Dellepiane, M., Ganovelli, F., & Ranzuglia, G. (2008). MeshLab: An open-source mesh processing tool. In V. Scarano, R. De Chiara & U. Erra (Eds.), *Eurographics Italian Chapter Conference* (p. 129–136).

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297.

Costa, A. G. (2010). A geometric morphometric assessment of plan shape in bone and stone Acheulean bifaces from the Middle Pleistocene site of Castel di Guido, Latium, Italy. In S. J. Lycett, & P. R. Chauhan (Eds.), *New perspectives on old stones: Analytical approaches to paleolithic technologies* (pp. 23–41). New York: Springer.

Cotterell, B., & Kamminga, J. (1987). The formation of flakes. *American Antiquity, 52*(4), 675–708.

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

Debénath, A., & Dibble, H. L. (1994). *Handbook of paleolithic typology: Lower and middle paleolithic of Europe* (Vol. 1, p. 202). University of Pennsylvania Press.

Delagnes, A. (1995). Variability within uniformity: Three levels of variability within the Levallois system. In H. L. Dibble & O. Bar-Yosef (Eds.), *The definition and interpretation of Levallois technology* (Vol. 23, pp. 201–211). Prehistory Press.

Delagnes, A., & Meignen, L. (2006). Diversity of lithic production systems during the Middle Paleolithic in France. Are there any chronological trends? In E. Hovers & S. L. Kuhn (Eds.), *Transitions before the transition evolution and stability in the Middle Paleolithic and Middle Stone Age* (pp. 85–107). Springer.

Domínguez-Rodrigo, M., & Baquedano, E. (2018). Distinguishing butchery cut marks from crocodile bite marks through machine learning methods. *Scientific Reports, 8*(1), 5786. https://doi.org/10.1038/s41598-018-24071-1

Duran, J.-P. (2005). L'industrie moustérienne des Ànecs (Rodès, Pyrénées-orientales, France). *PYRENAE, 36*(2), 11–39.

Duran, J.-P., & Abelanet, J. (2004). Un Moustérien Méditerranéen à bifaces: Le gisement de Moutou-la-Joliette. *Préhistoire Anthropologie Méditerranéennes, 13*, 7–27.

Duran, J.-P., & Soler, N. (2006). Variabilité des modalités de débitage et des productions lithiques dans les industries moustériennes de la grotte de l'Arbreda, secteur alpha (Serinyà, Espagne). *Bulletin De La Société Préhistorique Française, 103*(2), 241–262.

Faivre, G.-P., Gravina, B., Bourguignon, L., Discamps, E., & Turq, A. (2017). Late Middle Palaeolithic lithic technocomplexes (MIS 5–3) in the northeastern aquitaine basin: advances and challenges. *Quaternary International, 433*, 116–131. https://doi.org/10.1016/j.quaint.2016.02.060

Fernandes, P., Morala, A., Schmidt, P., Séronie-Vivien, M.-R., & Turq, A. (2012). Le silex du Bergeracois: État de la question. *Quaternaire Continental D'aquitaine, Excursion AFEQ, ASF, 2012*(2012), 22–33.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics, 7*, 179–188.

Frey, P. W., & Slate, D. J. (1991). Letter recognition using Holland-style adaptive classifiers. *Machine Learning, 6*(2), 161–182.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics, 29*(5), 1189–1232.

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

Ganganwar, V. (2012). An overview of classification algorithms for imbalanced datasets. *International Journal of Emerging Technology and Advanced Engineering, 2*(4), 42–47.

Geneste, J.-M. (1988). Les Industries De La Grotte Vaufrey: Technologie du debitage, economie et circulation de la matiere premiere lithique. In J.-P. Rigaud (Ed.), *La grotte Vaufrey à Cenac et Saint-Julien (Dordogne): Paléoenvironnements, chronologie et activités humaines* (pp. 441–517). Société préhistorique française.

González-Molina, I., Jiménez-García, B., Maíllo-Fernández, J.-M., Baquedano, E., & Domínguez-Rodrigo, M. (2020). Distinguishing Discoid and Centripetal Levallois methods through machine learning. *PLoS One, 15*(12), e0244288. https://doi.org/10.1371/journal.pone.0244288

Greenwell, B., Boehmke, B., Cunningham, J., & Developers, G. B. M. (2019). Package 'gbm'. https://cran.r-project.org/web/packages/gbm/gbm.pdf

Gunz, P., & Mitteroecker, P. (2013). Semilandmarks: a method for quantifying curves and surfaces. *Hystrix, 24*(1), 103–109. https://doi.org/10.4404/hystrix-24.1-6292

Gunz, P., Mitteroecker, P., & Bookstein, F. L. (2005). Semilandmarks in three dimensions. *Modern morphometrics in physical anthropology* (pp. 73–98). Springer.

Hayden, B. (1980). Confusion in the bipolar world: Bashed pebbles and splintered pieces. *Lithic Technology, 9*(1), 2–7. https://doi.org/10.1080/01977261.1980.11754456

Jamal, S., Goyal, S., Grover, A., & Shanker, A. (2018). Machine learning: What, why, and how? In A. Shanker (Ed.), *Bioinformatics: Sequences, structures, phylogeny* (pp. 359–374). Singapore: Springer. https://doi.org/10.1007/978-981-13-1562-6_16

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning with applications in R* (2nd edn., p. 607). New York: Springer.

Kendall, D. G. (1984). Shape manifolds, procrustean metrics, and complex projective spaces. *Bulletin of the London Mathematical Society, 16*(2), 81–121. https://doi.org/10.1112/blms/16.2.81

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, *28*(5). https://doi.org/10.18637/jss.v028.i05

Kuhn, S. L. (2013). Roots of the Middle Paleolithic in Eurasia. *Current Anthropology, 54*(S8), S255–S268. https://doi.org/10.1086/673529

Kumar, M., & Sheshadri, H. S. (2012). On the classification of imbalanced datasets. *International Journal of Computer Applications, 44*(8), 1–7.

Lantz, B. (2019). *Machine learning with R: Expert techniques for predictive modeling* (p. 375). Birmingham, Packt publishing ltd.

Lenoir, M., & Turq, A. (1995). Recurrent centripetal debitage (Levallois and Discoidal): Continuity or discontinuity? In H. L. Dibble & O. Bar-Yosef (Eds.), *The definition and interpretation of Levallois Technology* (Vol. 23, pp. 249–256). Prehistory Press.

Macdonald, D. A., Royal, K., & Buchanan, B. (2020). Evaluating the effects of parallax in archaeological geometric morphometric analyses. *Archaeological and Anthropological Sciences, 12*(7), 149. https://doi.org/10.1007/s12520-020-01111-4

McPherron, S. P., Archer, W., Otárola-Castillo, E. R., Torquato, M. G., & Keevil, T. L. (2022). Machine learning, bootstrapping, null models, and why we are still not 100% sure which bone surface modifications were made by crocodiles. *Journal of Human Evolution, 164*, 103071. https://doi.org/10.1016/j.jhevol.2021.103071

Meignen, L. (1993). Les industries lithiques de l'abri des Canalettes: cuche 2. In L. Meignen (Ed.), *L'abri des Canalettes. Un habitat moustérien sur les grands Causses (Nant-Aveyron)* (p. 238–328). Paris: Monographie du CRA., CNRS Ed.

Meignen, L. (1996). Persistance des traditions techniques dans l'abri des Canalettes (Nant-Aveyron). *Quaternaria Nova, 6*, 449–464.

Mitteroecker, P., & Gunz, P. (2009). Advances in geometric morphometrics. *Evolutionary Biology, 36*(2), 235–247. https://doi.org/10.1007/s11692-009-9055-x

O'Higgins, P. (2000). The study of morphological variation in the hominid fossil record: biology, landmarks and geometry. *Journal of Anatomy, 197*(1), 103–120. https://doi.org/10.1046/j.1469-7580.2000.19710103.x

Pasty, J.-F., Liegard, S., & Alix, P. (2004). Étude de l'industrie lithique du site paléolithique moyen des Fendeux (Coulanges, Allier). *Bulletin De La Société Préhistorique Française, 101*(1), 5–25. https://doi.org/10.3406/bspf.2004.12945

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*(11), 559–572. https://doi.org/10.1080/14786440109462720

Pelegrin, J. (1995). *Technologie lithique: Le Châtelperronien de Roc-de-Combe (Lot) et de la Côte (Dordogne)*. CNRS.

Presnyakova, D., Archer, W., Braun, D. R., & Flear, W. (2015). Documenting differences between Early Stone Age flake production systems: an experimental model and archaeological verification. *PLoS One, 10*(6), e0130732. https://doi.org/10.1371/journal.pone.0130732

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* (p. 312). Elsevier.

R Core Team. (2019). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/. Accessed 2019-09-03.

Raab, L. M., Cande, R. F., & Stahle, D. W. (1979). Debitage graphs and archaic settlement patterns in the Arkansas Ozarks. *Midcontinental Journal of Archaeology, 4*(2), 167–182.

Ridgeway, G. (2007). Generalized boosted models: A guide to the gbm package. R package vignette.

Rohlf, F. (2015). The tps series of software. *Hystrix, the Italian Journal of Mammalogy*, *26*(1). https://doi.org/10.4404/hystrix-26.1-11264

Romagnoli, F., & Vaquero, M. (2019). The challenges of applying refitting analysis in the Palaeolithic archaeology of the twenty-first century: an actualised overview and future perspectives. *Archaeological and Anthropological Sciences, 11*(9), 4387–4396. https://doi.org/10.1007/s12520-019-00888-3

Romagnoli, F., Nishiaki, Y., Rivals, F., & Vaquero, M. (2018). Time uncertainty, site formation processes, and human behaviours: New insights on old issues in high-resolution archaeology. *Quaternary International, 474*, 99–102. https://doi.org/10.1016/j.quaint.2018.04.033

Romagnoli, F., Rivals, F., & Benazzi, S. (2022). *Updating neanderthals: understanding behavioral complexity in the Late Middle Paleolithic* (p. 361). London: Academic Press.

RStudio Team. (2019). RStudio: Integrated development for R. Boston, MA: RStudio, Inc. http://www.rstudio.com/. Accessed 2019-09-12.

Schlager, S. (2017a). Morpho and Rvcg – Shape analysis in R. In G Zheng, S. Li, & G. Székely (Eds.), *Statistical shape and deformation analysis* (pp. 217–256). Elsevier. https://doi.org/10.1016/B978-0-12-810493-4.00011-0

Schlager, S. (2017b). Morpho and Rvcg – shape analysis in R. In *Statistical shape and deformation analysis* (pp. 217–256). Elsevier. https://doi.org/10.1016/B978-0-12-810493-4.00011-0

Serwatka, K., & Riede, F. (2016). 2D geometric morphometric analysis casts doubt on the validity of large tanged points as cultural markers in the European Final Palaeolithic. *Journal of Archaeological Science: Reports, 9*, 150–159. https://doi.org/10.1016/j.jasrep.2016.07.018

Shea, J. J. (2013a). The Upper Paleolithic. *Stone tools in the Paleolithic and Neolithic near East : A guide* (pp. 117–160). Cambridge University Press.

Shea, J. J. (2013b). The Middle Paleolithic. *Stone tools in the Paleolithic and Neolithic near East : A guide* (pp. 82–116). Cambridge University Press.

Slimak, L. (2003). Les Debitages discoïdes mousteriens: Evaluation d'un concept technologique. In M. Peresani (Ed.), *Discoid lithic technology. Advances and implications* (Vol. 1120, pp. 33–65). Oxford: Archaeopress.

Sullivan, A. P., & Rozen, K. C. (1985). Debitage analysis and archaeological interpretation. *American Antiquity, 50*(4), 755–779. https://doi.org/10.2307/280165

Tixier, J., & Turq, A. (1999). Kombewa et alii. *Paléo*, *11*, 135–143. https://doi.org/10.3406/pal.1999.1174

Tixier, J., Inizan, M.-L., & Roche, H. (1980). *Préhistoire de la pierre taillée. 1.Terminologie et Technologie* (2nd ed., p. 120). Cercle de Recherches et d'études Préhistoriques.

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (Fourth Edition.). New York: Springer. https://www.stats.ox.ac.uk/pub/MASS4/. Accessed 2022-01-18.

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/10.2307/2333860

Weihs, C., Ligges, U., Luebke, K., & Raabe, N. (2005). klaR analyzing German business cycles. In D. Baier, R. Decker, & L. Schmidt-Thieme (Eds.), *Data analysis and decision support. Studies in classification, data analysis, and knowledge organization* (pp. 335–343). Heidelberg: Springer Berlin. https://doi.org/10.1007/3-540-28397-8.