frontiers | Frontiers in Ecology and Evolution

TYPE Original Research
PUBLISHED 11 February 2026
DOI 10.3389/fevo.2026.1681814



OPEN ACCESS

EDITED BY
Mírian Pacheco,
Federal University of São Carlos, Brazil

REVIEWED BY
Eric W. Dewar,
Suffolk University, United States
F.R. Da Silva,
Universidade Federal de São Carlos, Brazil

*CORRESPONDENCE
Trevor L. Keevil
trevor.keevil@colostate.edu

RECEIVED 07 August 2025
REVISED 12 January 2026
ACCEPTED 22 January 2026
PUBLISHED 11 February 2026

CITATION
Keevil TL, Pelissero AJ, Negash T, Orlikoff ER, Osborne I, Tolley AM, Pobiner B and Pante MC (2026) A comparative bone surface modification database for revealing the origins and evolution of human carnivory. *Front. Ecol. Evol.* 14:1681814. doi: 10.3389/fevo.2026.1681814



# A comparative bone surface modification database for revealing the origins and evolution of human carnivory

Trevor L. Keevil[1]*, Alex J. Pelissero[1], Tewabe Negash[1], Emily R. Orlikoff[2], Isabell Osborne[1], April M. Tolley[1], Briana Pobiner[3] and Michael C. Pante[1]

[1]Department of Anthropology and Geography, Colorado State University, Fort Collins, CO, United States, [2]Department of Anthropology, University of Michigan, Ann Arbor, MI, United States, [3]Human Origins Program, Department of Anthropology, Smithsonian Institution, Washington, DC, United States

Fossilized trace marks left by our ancestors as they processed animals for food are important clues to the emergence and intensification of human carnivory and tool use. When studied in tandem with fossilized tooth marks made by carnivorous predators, butchery marks also help reconstruct the larger ecological framework and trophic dynamics of paleoecosystems. However, some taphonomic processes produce bone surface modifications that mimic the morphology of butchery marks, introducing the potential for misclassification when relying on imprecise criteria. The implementation of digital modeling techniques that allow the collection of microscopic quantitative data has begun to improve the reliability of mark identification. Although many digital taphonomy methods appear promising, their broader applications are limited by a lack of replicable methods, unpublished and closed-source databases, and statistical methodologies that violate core assumptions for accurate model inference. In this paper, we present an open-source database of experimentally generated cut, percussion, tooth, and trample marks measured and analyzed using high-resolution confocal profilometry and a replicable quantification protocol. Statistical classificatory models using our taphonomic measurement database can distinguish between experimentally generated bone surface modifications with 74% to 83% accuracy, depending on the comparative groupings. Our aim for these classification models is to facilitate accurate identification of the processes that created fossilized bone surface modifications, which is the first step to resolving long-standing debates surrounding the origins and evolution of human carnivory. Additionally, we hope that publishing our open-source data and code underscores the need for more replicable, collaborative, and transparent methods in paleoanthropological research.

KEYWORDS

**butchery marks, confocal profilometry, digital taphonomy, equifinality, experimental archaeology, feeding trace, tooth marks, trampling marks**

# 1 Introduction

Fossilized feeding traces left by hominins and carnivores are evidence of the biotic ecological dynamics that made up a paleoecosystem. Assemblage-wide cut, percussion, and tooth mark frequencies can reveal the amount and type of animal foods our ancestors consumed and their ecological interactions with prey and predator species (Selvaggio, 1994; Blumenschine, 1995; Capaldo, 1997; Domínguez-Rodrigo and Barba, 2006; Pante et al., 2012, 2015; Domínguez-Rodrigo et al., 2014). These mark frequencies are empirical data that can be used to understand the conditions that led to key events in the evolutionary history of our lineage, especially as it relates to the morphological and technological evolution that led to the appearance of modern humans (Blumenschine, 1989; Bunn and Ezzo, 1993; Domínguez-Rodrigo and Pickering, 2003; Blumenschine and Pobiner, 2007; Bunn, 2007; Pante et al., 2018; Pobiner, 2020). However, the validity of these ecological models and broader evolutionary implications rest on the assumption that researchers can accurately identify who or what created fossilized bone surface modifications (BSMs) representative of different feeding traces. As James and Thompson (2015) emphasize, consensus on a standardized BSM identification methodology for relating fossilized BSMs to prehistoric and unobservable actions has yet to be achieved.

Accurately identifying whether a stone-tool-wielding butcher or non-hominin predator created a fossilized mark is complicated because many taphonomic processes create similarly shaped marks that researchers may mistake as a cut or tooth mark (e.g., rodent gnawing, ungulate trampling, and rockfall in caves) (Behrensmeyer et al., 1986; Olsen and Shipman, 1988; Domínguez-Rodrigo et al., 2009; Litynski and Pante, 2023; Marginedas et al., 2023). This morphological overlap in BSM shape, defined as taphonomic equifinality, has given rise to prominent, long-standing debates about the creators of many BSMs, including the earliest potential evidence of hominin carnivory on 3.4-million-year-old fossils from Dikika, Ethiopia (McPherron et al., 2010; Domínguez-Rodrigo et al., 2012; Thompson et al., 2015) and also on 2.5-million-year-old fossils from Bouri Hata, Ethiopia (de Heinzelin et al., 1999; Sahle et al., 2017). Similar debates about fossilized BSMs undermine efforts to establish consensus about the meat-eating behaviors of early hominins at the FLK 22 (*Zinjanthropus* level) site in Olduvai Gorge, Tanzania (Domínguez-Rodrigo and Barba, 2006; Blumenschine et al., 2007; Pante et al., 2012, 2015; Parkinson, 2018).

To overcome issues of equifinality, some researchers qualitatively compare macromorphological BSM characteristics to determine who or what created a mark (Bunn, 1981; Shipman and Rose, 1983; Andrews and Cook, 1985; Blumenschine and Selvaggio, 1988; Blumenschine, 1995; Blumenschine et al., 1996; Njau and Blumenschine, 2006; Domínguez-Rodrigo et al., 2009; Galán et al., 2009; de Juana et al., 2010). In experimental settings, Blumenschine et al. (1996) demonstrated that qualitative BSM identification methods can be effective for differentiating between cut, percussion, and mammalian carnivore tooth marks. However, the real-world applicability of these results is complicated by the many other processes impacting the surfaces of fossils over thousands to millions of years and recent realizations that trained analysts are not capable of replicating each other's results despite being trained in the same research tradition (Domínguez-Rodrigo et al., 2017). The latter is almost certainly the result of the subjective nature of qualitative criteria, which are often ambiguously defined (Domínguez-Rodrigo et al., 2017).

Recognizing that qualitative BSM assessments may be insufficiently equipped to discern morphologically ambiguous marks, some researchers now use computational methods to analyze and measure BSMs. Typically, quantitative BSM studies use a two-step approach. First, BSMs are digitally reconstructed in 3D using either digital or optical microscopy (Bello and Soligo, 2008; Bello et al., 2011; Boschin and Crezzini, 2012; Domínguez-Rodrigo et al., 2024, 2025b), confocal profilometry (Braun et al., 2016; Pante et al., 2017; Otárola-Castillo et al., 2018), structured light scanning (Maté-González et al., 2017; Yravedra et al., 2018; Courtenay et al., 2019c, 2021), or micro-photogrammetry (Maté González et al., 2015; Arriaza et al., 2017; Yravedra et al., 2017; Maté-González et al., 2018). Then, BSMs are measured to estimate the most likely action or agent that created the mark using either landmark-based geometric morphometrics (Arriaza et al., 2017, 2017; Otárola-Castillo et al., 2018; Yravedra et al., 2018; Courtenay et al., 2020a), Elliptic Fourier analysis (Arriaza et al., 2023), computer vision convolutional neural networks (Courtenay et al., 2020a; Jiménez-García et al., 2020), or traditional measurement and angle-based morphometrics (Bello and Soligo, 2008; Bello et al., 2011; Pante et al., 2017).

Despite the availability of different quantitative BSM modeling methods, these approaches lack widespread adoption for two reasons addressed by this paper. First, computational approaches inherently require a comparative database of experimentally generated BSMs replicative of potential prehistoric BSM-creating actions (e.g., carnivore feeding trials or simulated stone tool butchery) to compare to the shape of fossilized BSMs created by unknown and prehistoric actions. However, curating and digitizing experimental BSM databases is both time-consuming and can require expensive instruments. Second, with the exception of Pante et al. (2017), BSM modeling techniques have not tested for inter-observer replicability throughout the entire process of scanning, measuring, and analyzing a singular BSM. Scientific replicability is a key step in validating results, facilitating collaborations, and encouraging methodological acceptance. Without such tests, it is unclear if these BSM modeling methods are creating highly precise but, ultimately, inaccurate depictions of the actions that created fossilized marks.

In this paper, we present measurement data collected from 946 BSMs produced by simulated stone tool and percussive butchery, captive and wild carnivore feeding trials, and ungulate trampling experiments. We digitally scanned and measured all 946 of these marks following the experimental protocol developed in Pante et al. (2017), which was previously shown to be replicable in experimental settings. We will continue to add to this database as our experimental samples increase. Ultimately, we hope that by publishing the results of our experimental BSM measurement database, other researchers can use our data to analyze the shape

of their fossilized BSMs without needing to create their own experimental databases.

# 2 Materials and methods

## 2.1 Experimental BSM sample

Our BSM sample was generated through experiments replicating the actions of prehistoric agents that have created fossilized BSMs. Table 1 summarizes BSM data. Below, we outline the experimental protocols for generating these marks, and if relevant, include a reference to the original experimental publication.

### 2.1.1 Cut marks

Cut marks created by intentional bone cutting and actualistic butchery experiments are included in the measurement database.

Intentionally created cut marks come from two sources. First, 207 cut marks were produced using quartzite, basalt, phonolite, and chert simple flakes and Acheulean handaxes attached to a motorized cutting machine [see Keevil (2018) for experimental protocols]. Second, 50 cut marks were created using chert handaxe and flake tools and 22 using unmodified stone tools by a single butcher moving the tools at a 90° angle across the surface of a defleshed bone with constant pressure.

Cut marks created through actualistic butchery experiments come from two sources. First, 111 marks come from two butchers processing skinned, but fully fleshed, deer limbs using chert and obsidian flakes. An additional 21 cut marks were created by butchers using chert and quartzite flake and core tools to process skinned adult goat hindlimbs.

While many digital taphonomy studies analyze intentional or actualistic cut marks (e.g., Bello and Soligo, 2008; Maté González et al., 2015; Pante et al., 2017; Courtenay et al., 2020a), they rarely include both types within a single analysis. To evaluate the practical relevance of treating intentional and actualistic cut marks as a single population, we assessed group equivalence between the 132 actualistic and 279 intentional cut marks described above using permutational multivariate analysis of variance (PERMANOVA; Appendix A.1). Given the large sample sizes, we emphasized effect size over statistical significance to determine methodological relevance. As detailed in Appendix A.1, this analysis supports treating both cut mark types as a single analytical group. Nonetheless, intentional and actualistic group labels are retained as supplementary information to allow future analyses to explicitly compare differences if warranted.

### 2.1.2 Percussion marks

We include 90 percussion marks from marrow extraction experiments described in Benito-Calvo et al. (2018). Percussion marks were produced from a quartzite anvil and a combination of quartzite and basalt cobblestone hammerstones on cow limb bones. Following the methodology outlined in Blumenschine (1988), bones were placed on stone anvils before being struck with hammerstones to split the bones longitudinally along the axis and extract marrow.

### 2.1.3 Tooth marks

A total of 313 carnivore tooth marks created by mammalian and crocodilian predators were obtained from bones collected during naturalistic and captive feeding experiments.

Tooth marks created during actualistic feeding trials come from two studies, and detailed experimental protocols can be found within their respective publications. First, we include 29 tooth marks created by free-ranging lions described in Pobiner (2007). Second, we include 31 tooth marks created by free-ranging spotted hyenas described by Blumenschine (1988).

Captive feeding trials were conducted at three locations, as described in Muttart (2017) and Njau and Blumenschine (2006). First, we include a total of 143 tooth marks from African wild dogs, lions, spotted and striped hyenas, and North American brown bears at the Denver Zoo, Colorado, USA. We also include 32 grey wolf tooth marks from captive wolves at the Rist Canyon, W.O.L.F. Sanctuary, Colorado, USA. Finally, we also include 78 tooth marks produced during experimental feeding trials with Nile crocodiles (Njau and Blumenschine, 2006).

### 2.1.4 Trample marks

A total of 132 trample marks were measured on bones created during experimental cow trampling trials. To simulate actualistic trampling actions, we placed transversely sectioned cow femur midshafts that had their periosteum, meat, and marrow removed in empty animal corrals at a local farm in Colorado, USA. Sectioned femurs were used for this study because they were readily available from a local butcher and could easily be partially buried in the different corral sediments. Before each trampling experiment, we thoroughly inspected the cleaned bones for any butchery marks or protrusions using low magnification methods. Any marks found were subsequently highlighted with clear nail polish and marked to ensure they were not confused with marks created during the experiment.

After scattering the cleaned bones around the ground and partially burying some in the corral, we let 25–30 heifers weighing between 350 and 550 kgs into the corrals for 30 minutes of trampling with gentle prodding to encourage movement. To capture a larger array of potential trampling mark shape variation, we conducted experiments using three sediment types, including fine-grained sand, gravel, and coarse-grained soil. Bones were collected after each 30-minute trampling trial and inspected for evidence of trampling.

## 2.2 Profilometry methodology

3D models of BSMs were produced using either a Nanovea ST400 white-light non-contact profilometer or a Sensofar S-Neox 3D optical profiler. The Nanovea 3D models were produced with a 3 mm optical pen that has a z-axis resolution of 40 nm. The spatial resolution was set to 5 μm in the x-axis and 10 μm in the y-axis. 3D models produced with the Sensofar S-Neox were made using a 5x lens that has a z-axis resolution of 75 nm. The 5x lens has a numerical aperture of 0.15, a working distance of 23.5 mm, a field of

TABLE 1 Experimental BSM sources for the 411 cut marks, 90 percussion marks, 313 carnivore tooth marks, and 132 trample marks included in this study.

| Mark type | Experimental context | Actor | Effector characteristic | Count | Reference |
|---|---|---|---|---|---|
| Cut | Intentional Marking | Human butcher | Chert Flake | 50 | This paper |
| | | Human butcher | Unmodified Stone | 22 | This paper |
| | | Cutting Machine | Basalt Biface | 27 | (Keevil, 2018) |
| | | Cutting Machine | Basalt Flake | 27 | (Keevil, 2018) |
| | | Cutting Machine | Chert Biface | 25 | (Keevil, 2018) |
| | | Cutting Machine | Chert Flake | 26 | (Keevil, 2018) |
| | | Cutting Machine | Phonolite Biface | 25 | (Keevil, 2018) |
| | | Cutting Machine | Phonolite Flake | 25 | (Keevil, 2018) |
| | | Cutting Machine | Quartzite Biface | 26 | (Keevil, 2018) |
| | | Cutting Machine | Quartzite Flake | 26 | (Keevil, 2018) |
| | Actualistic Butchery | Human butcher | Obsidian & Chert Flake | 111 | This paper |
| | | Human butcher | Chert & Quartzite Flake | 21 | This paper |
| Percussion | Marrow Extraction | Human butcher | Quartzite Anvil | 31 | (Benito-Calvo et al., 2018) |
| | | Human butcher | Basalt Hammerstone | 29 | (Benito-Calvo et al., 2018) |
| | | Human butcher | Quartzite Hammerstone | 30 | (Benito-Calvo et al., 2018) |
| Tooth | Captive Feeding | African Wild Dog | Tooth | 28 | (Muttart, 2017) |
| | | Brown Bear | Tooth | 28 | (Muttart, 2017) |
| | | African Lion | Tooth | 37 | (Muttart, 2017) |
| | | Spotted Hyena | Tooth | 28 | (Muttart, 2017) |
| | | Striped Hyena | Tooth | 30 | (Muttart, 2017) |
| | | Grey Wolf | Tooth | 32 | (Muttart, 2017) |
| | | Nile Crocodile | Tooth | 78 | (Njau and Blumenschine, 2006) |
| | Wild Feeding | Spotted Hyena | Tooth | 31 | (Blumenschine, 1988) |
| | | African Lion | Tooth | 21 | (Pobiner, 2007) |
| Trample | Corralled Trampling | Cow | Soil sediment | 65 | This Paper |
| | | Cow | Gravel sediment | 46 | This Paper |
| | | Cow | Sand sediment | 21 | This Paper |

References are included for previously described BSM experiments.

view of 3400 μm x 2837 μm, a spatial sampling of 2.76 μm, and an optical resolution of 0.93 μm. The scale of these resolution differences does not significantly impact our analysis because they are orders of magnitude smaller than the scale of the measured differences between mark types. Further, these differences are smaller than the reported variability in measurements taken from a single mark using the same instrument (Pante et al., 2017; Appendix A.2).

Processing and analysis of 3D models were carried out using Digital Surf's Mountains® following Pante et al. (2017). Model processing included removing outliers, filling in missing data points, and removing the underlying form of the bone with the mark excluded from the form removal process (see Pante et al., 2017 for further detail). Data collected through the analysis from the entire 3D model of the experimental mark were volume, surface area, maximum depth, mean depth, maximum length, and maximum width. Additional data were collected from a profile taken from the deepest point of the mark, including area of the hole, depth of the profile, width, roughness ($R_a$), opening angle, and radius of the hole. Collection of profile data from percussion marks

that occur on the edge of a bone fragment (approximately half of the 90 marks in the present study) follow a different protocol than was described by Pante et al. (2017) for other mark types because it is not possible to take a complete profile through the deepest part of the mark when it has been split due to its location across the point of fracture. In these cases, the deepest profile was taken parallel to the crack across the width of the mark instead of the length.

## 2.3 Statistical methodology

Statistical analyses were carried out using R (Version 4.4.3; R Core Team, 2024) and the associated R packages described below. Code and data are included as supplementary information.

First, we calculated group-level summary statistics (mean, median, and standard deviation) for the 12 measurement variables. We also performed a Principal Components Analysis (PCA) using the *prcomp* function from the base stats package to assess measurement variance among these 12 variables. Second, we use two statistical classification methods – k-fold cross-validated discriminant analysis (DA) and random forest (RF) – to evaluate whether our experimental measurement dataset can recognize specific behaviors from a multivariate analysis of mark shape.

Before generating a DA model, we assessed underlying assumptions of correlation, normality, and homogeneity of covariance matrices. Anderson-Darling Tests of univariate normality and Royston's test of multivariate normality, implemented using the *MVN* package (Korkmaz et al., 2014), were used to determine necessary data transformations. For variables that violated normality assumptions, Box-Cox tests identified an appropriate power or logarithmic transformation. To prevent data leakage, Box-Cox transformations were independently performed on the training dataset within each DA fold and those values were used to transform the testing dataset. All measurements were log-transformed, except 3D maximum depth and mean depth, which were transformed by taking the inverse square root, and angle, which was cubed. Additionally, the mean depth was transformed using a logarithmic transformation in two folds.

Pooled within-group tests of correlation on the transformed dataset, using the "statsBy" function from the *Psych* package (Revelle, 2024), indicated that volume was highly correlated with 3D surface area (r = 0.95). Similarly, maximum profile depth was correlated with 3D maximum depth (r = -0.92) and profile area (r = 0.92). Profile width was also correlated with profile area (r = 0.91). Based on these exploratory results, we removed volume, profile width, and maximum depth of the cross-sectional profile from the DA model.

A Box's M test, conducted using the "boxM" function from the *heplots* package (Friendly, 2010), indicated that our data do not satisfy the assumption of equal covariance matrices. Based on these results, we performed a 10-fold cross-validated quadratic discriminant analysis (QDA) using the "qda" function from the *MASS* package (Venables and Ripley, 2002) to assess the accuracy with which our four experimental mark types could be discriminated based on a multivariate analysis of mark shape. Prior group probabilities were set to be uninformative and uniform across groups to mitigate the influence of experimental BSM class imbalances on classification precision.

Following the QDA, we generated one-vs-rest (OVR) receiver operating characteristic (ROC) curves for each group by averaging the ROC curves generated across the 10 cross-validation folds. ROC curves were calculated using the "roc" function in the *pROC* package (Robin et al., 2011). We also calculated the average area under the curve (AUC) across the ten OVR ROC curves for each mark type. These steps allowed us to evaluate the discriminatory performance of the QDA model for each mark type against the other three mark types.

We employed RF analysis using the "randomForest" function from the *randomForest* package (Liaw and Wiener, 2002). RF models are robust to non-normal data and multicollinearity, so each RF model includes all 12 untransformed measurements. Because the accuracy and interoperability of RF models are reduced when group sample sizes are imbalanced (Chen et al., 2004), we use a stratified per-tree downsampling approach with the minimum sample size set to 90, the smallest group size.

Due to the out-of-bag (OOB) error estimation in RF methods, we also do not use cross-validation. Furthermore, because our dataset is not high-dimensional and preliminary testing has not shown major increases in model performance, we choose not to tune hyperparameters using a train/validate split, which is more necessary for high-dimensional or noisy data (Probst et al., 2019). Therefore, we opted to use the default hyperparameter settings with the number of trees tested set to 1000, which are likely sufficient for our low-dimensional measurement dataset.

We generated three RF models for the present study. The first model included all four mark types. Based on previous qualitative observations that percussion and tooth marks are commonly misclassified as each other and that cut and trample marks exhibit a similar pattern, we generated two additional RF models that include only marks representative of these two commonly misclassified relationships.

# 3 Results

## 3.1 Summary statistics and PCA

The first three principal component (PC) axes explain approximately 67%, 13%, and 12% of the variance in our measurement data, respectively (Figure 1). In total, these three axes contribute over 92% of the variation in our dataset.

Measurements contributing to data variance along the first axis include volume, maximum and mean depth, width, and profile area. In general, these measurements are largest in percussion marks, followed by tooth marks. Trample marks and cut marks are both smaller in these dimensions and more similar to each other than tooth or percussion marks (Table 2). These trends are reflected along the first PC axes in Figure 1, where there is significant overlap in the confidence ellipse of cut and trample marks as well as tooth

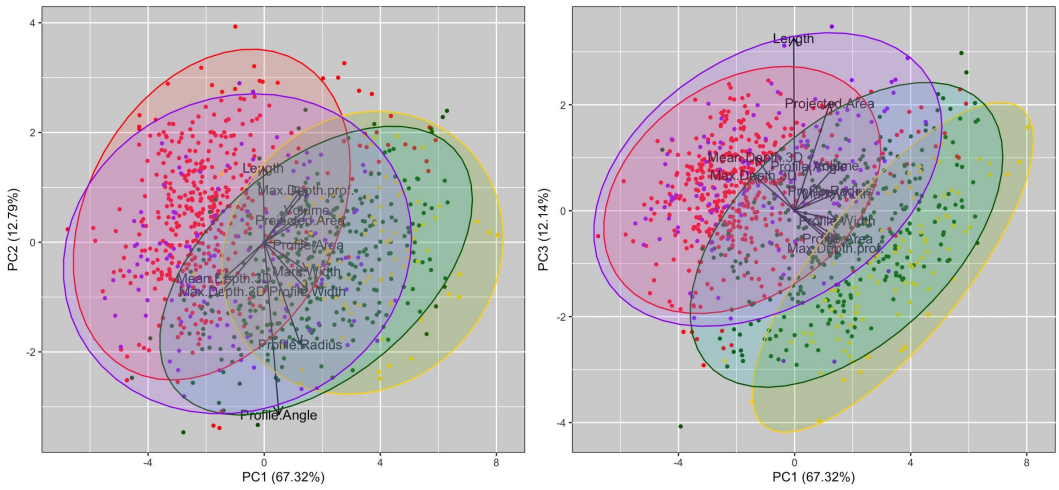


FIGURE 1
Principal component analysis of the BSM measurement dataset. The left scatterplot shows the distribution of measurement variance along the first two principal components (PC1 and PC2). The right scatterplot shows the distribution of measurement variance along the first and third principal components (PC1 and PC3). Normal data ellipses and data points are colored by mark type: red for cut marks, yellow for percussion marks, green for tooth marks, and purple for trample marks.

and percussion marks, marginal overlap between trample, tooth, and percussion marks, and the least amount of overlap between cut and percussion marks.

Variation in data on the second PC axis is primarily driven by angle, radius, and mark length measurements (Figure 1). In general, cut and trample marks are both relatively long; however, cut marks tend to be significantly more acute with a slimmer cross-sectional profile (Table 2). Tooth marks have the most obtuse opening angle and broad cross-sectional profile shape, with trample marks being only slightly more acute and slimmer on average (Table 2). Like tooth marks, percussion marks tend to be short with broad opening angles; however, these marks are also characterized by their very large profile radius values.

Variation in data on the third PC axis reflects a combination of mark surface area and length, which influenced variance on the second PC axis. In general, these features are highly similar between cut and trample marks, as well as between tooth and percussion marks, as shown by the overlapping cut and trample mark ellipses and tooth and percussion cut and trample mark ellipses in Figure 1.

## 3.2 Discriminant analysis

A 10-fold cross-validated QDA including the nine uncorrelated and transformed measurement variables could discriminate between our four taphonomic actions with approximately 76% classification accuracy (Table 3). In general, cut marks most frequently misclassify as trample marks (64 of the 89 misclassified cut marks), and trample marks misclassify as cut marks (40 of the 58 misclassified trample marks). Similarly, tooth marks most frequently misclassify as percussion marks (27 of the 60 misclassified tooth marks), and percussion marks most frequently misclassify as tooth marks (17 of the 21 misclassified percussion marks).

Cut, percussion, and tooth OVR ROC curves demonstrate excellent discriminatory power (AUCs of 0.94, 0.96, and 0.94, respectively), confirming the strong classificatory powers of our QDA model when classifying these mark types (Figure 2). Similarly, the OVR ROC curve for trample marks demonstrates very good but slightly lower discriminatory power (AUC of 0.85). ROC curves and AUC values illustrate how well our model separates different marks by testing model performance across all possible decision thresholds (Fawcett, 2006). In the context of QDA, these decision boundaries are varied by changing the necessary posterior probability threshold required to classify a data point into a particular group, and AUC values above 0.85 indicate excellent model performance. Overall, these high AUC values attest to the reliability and robustness of the classification precision obtained in the QDA model.

## 3.3 Random forests

A RF model including all 12 measurement variables and the four taphonomic actions studied in this paper produced an OOB error rate estimate of 25.8% or a predictive classification accuracy of approximately 74% (Table 4). Among the input measurement variables, importance analysis of mean GINI decrease shows that profile width, length, and profile area contributed most significantly to the predictive power of the RF model (Figure 3). In general, tooth and percussion marks tend to be shorter, wider, and have greater cross-sectional profile areas compared to cut and trample marks. All other variables showed slightly lower importance to the predictive powers of this RF mode (Figure 3). However, because the mean GINI decreases are greater than 10 for all measurements, it shows

TABLE 2 Summary statistics (mean and standard deviation) for BSM measurements, categorized by experimental action.

| BSM | Statistic | 3-D measurements | | | | | | Profile measurements | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Surface area ($\mu m^2$) | Volume ($\mu m^3$) | Maximum depth ($\mu m$) | Mean depth ($\mu m$) | Length ($\mu m$) | Width ($\mu m$) | Maximum profile depth ($\mu m$) | Profile area ($\mu m^2$) | Width ($\mu m$) | Radius ($\mu m$) | Angle (°) | Roughness ($R_a$) |
| Cut | Mean | 2159492.6 | 2010638542 | 83.3 | 28.5 | 9177.3 | 428.6 | 76.1 | 23936.7 | 333.2 | 576.5 | 125.7 | 3.5 |
| | Std. Dev | 4988971.5 | 10437408385 | 83.6 | 36.6 | 5618.9 | 399.9 | 79.8 | 83486.9 | 358.4 | 1937.8 | 25.9 | 3.7 |
| Percussion | Mean | 4339243.3 | 9864289617 | 456.5 | 150.5 | 3587.2 | 1603.6 | 371.4 | 791166.7 | 2901.5 | 7789.2 | 144.8 | 13.3 |
| | Std. Dev | 5940869.0 | 23854012032 | 375.3 | 105.7 | 2860.2 | 995.3 | 274.9 | 1702049.1 | 2454.5 | 12969.7 | 23.1 | 10.6 |
| Tooth | Mean | 5126881.0 | 6075151866 | 195.9 | 73.9 | 4042.5 | 1383.5 | 191.8 | 182560.0 | 1425.8 | 2678.9 | 148.5 | 6.9 |
| | Std. Dev | 8547317.0 | 13519575453 | 155.0 | 60.1 | 3326.7 | 951.2 | 163.1 | 292316.8 | 973.8 | 3980.1 | 18.6 | 5.5 |
| Trample | Mean | 3503978.8 | 1336779292 | 88.9 | 26.4 | 8011.5 | 759.2 | 78.8 | 24484.8 | 506.4 | 1955.9 | 142.4 | 4.6 |
| | Std. Dev | 5306077.7 | 2772410132 | 54.0 | 21.6 | 4937.5 | 749.5 | 58.1 | 40877.1 | 430.7 | 5928.1 | 23.1 | 3.7 |

that all variables are contributing to the predictive accuracy of this model.

The cut and trample subset model, including all 12 measurements, had an OOB error estimate of 21.0% or an approximate classification precision of 79% (Table 5). Variables that contributed most to the discriminatory power of this model were mark width at the widest point on the 3D reconstruction, as well as the depth, opening angle, and roughness of the cross-sectional profile (Figure 3). In general, trample marks are wider with rougher, deeper, and less acute cross-sectional profile shapes compared to cut marks.

The tooth and percussion subset model, including all 12 measurements, had an OOB error estimate of 16.6% or an approximate classification precision of 83% (Table 6). Variables that contributed most to the discriminatory power of this model were surface area and mean and maximum depth of the 3D reconstruction. Additionally, cross-sectional profile characteristics like area, radius, and roughness aided in discriminating between these two groups (Figure 3). In general, percussion marks have less surface area than tooth marks while tending to have larger maximum and mean depths.

## 4 Discussion

Digital bone surface modification modeling and analytical methodologies help reveal which ancient actions and taphonomic processes created marks on now fossilized bones. Among the various BSM categories, cut and tooth marks are a focus in many taphonomic studies due to their frequency in fossil deposits and relevance for understanding dynamic and unobservable interactions between prehistoric hominin butchers, predators, and prey animals during the origins and intensification of hominin carnivory (Blumenschine and Selvaggio, 1991; Capaldo, 1997; McPherron et al., 2010; Pante et al., 2012, 2015; Domínguez-Rodrigo et al., 2014; Parkinson, 2018). Although these two mark types are usually easily distinguished from each other, their morphological overlap with trample and percussion marks can lead to misclassifications. The relationship between cut, tooth, percussion, and trample marks is underexplored in digital BSM studies, raising the possibility that some fossilized marks characterized as resulting from a certain behavior could have been created by a different taphonomic action.

Issues created by taphonomic equifinality are the focal point of3several high-profile and long-standing debates in paleoanthropology. For example, researchers debate whether fossilized BSMs supporting depictions of hominin hunting and/or scavenging at the FLK *Zinjanthropus* site, Olduvai Gorge, Tanzania, were created by mammalian carnivores or another taphonomic process, such as bioerosion (Domínguez-Rodrigo and Barba, 2006; Blumenschine et al., 2007). Similarly, there is debate whether the 3.4-million-year-old marks from Dikika, Ethiopia are the earliest evidence of hominin carnivory or resulted from ungulate trampling or crocodilian feeding (McPherron et al., 2010; Domínguez-Rodrigo et al., 2012; Thompson et al., 2015). Potential BSM

TABLE 3 Confusion matrix for quadratic discriminant analysis (QDA).

| QDA | BSM | Predicted | | | | Total |
|---|---|---|---|---|---|---|
| | | Cut | Percussion | Tooth | Trample | |
| Actual | Cut | **322** | 1 | 24 | 64 | 411 |
| | Percussion | 3 | **69** | 17 | 1 | 90 |
| | Tooth | 11 | 27 | **253** | 22 | 313 |
| | Trample | 40 | 2 | 16 | **74** | 132 |
| | Total | 376 | 99 | 310 | 161 | 946 |

Bolded values indicate correctly identified BSMs.

misidentifications also generate doubt about the anthropogenic origin of marks on Late Pleistocene fossils that could rewrite the timing of humans in the Americas, including marks on the 130,000-year-old Cerutti Mastodon in California (Haynes, 2017; Holen et al., 2017a; Ferrell, 2019), 30,000-year-old fossils in Arroyo del Vizcaíno, Uruguay (Fariña et al., 2014; Holcomb et al., 2022), and 24,000-year-old fossils in Bluefish Caves, Canada (Bourgeon et al., 2017; Krasinski and Blong, 2020; Litynski and Pante, 2023).

Because BSMs are often the only direct evidence of hominin carnivory and interactions with carnivores and prey animals in Plio-Pleistocene sites (e.g., Pobiner et al., 2008; McPherron et al., 2010; Curran et al., 2025), misclassifications of these marks stymie palaeoecological models that are necessary to understand the evolution of hominin subsistence strategies (Gifford-Gonzalez, 1991; James and Thompson, 2015). Therefore, establishing methods to precisely identify the ancient actions that created fossilized BSMs is critical to understanding ecological and evolutionary trends in the hominin lineage. Despite the potential for quantitative and computational BSM modeling methods to clear up these long-standing debates, the widespread adoption of these methods is obstructed by a lack of open-source measurement databases investigating experimental BSM shape and methods that lack tests of reliability.

The present study addresses problems in quantitative BSM modeling by publishing measurement data for nearly 1, 000 experimental tooth, cut, percussion, and trample marks generated by replicating actions that may have resulted in fossilized BSMs. As described below, our goal in publishing these marks is primarily to gain further insight into the morphological complexities and shape variation present in BSMs created by different actions. However, we

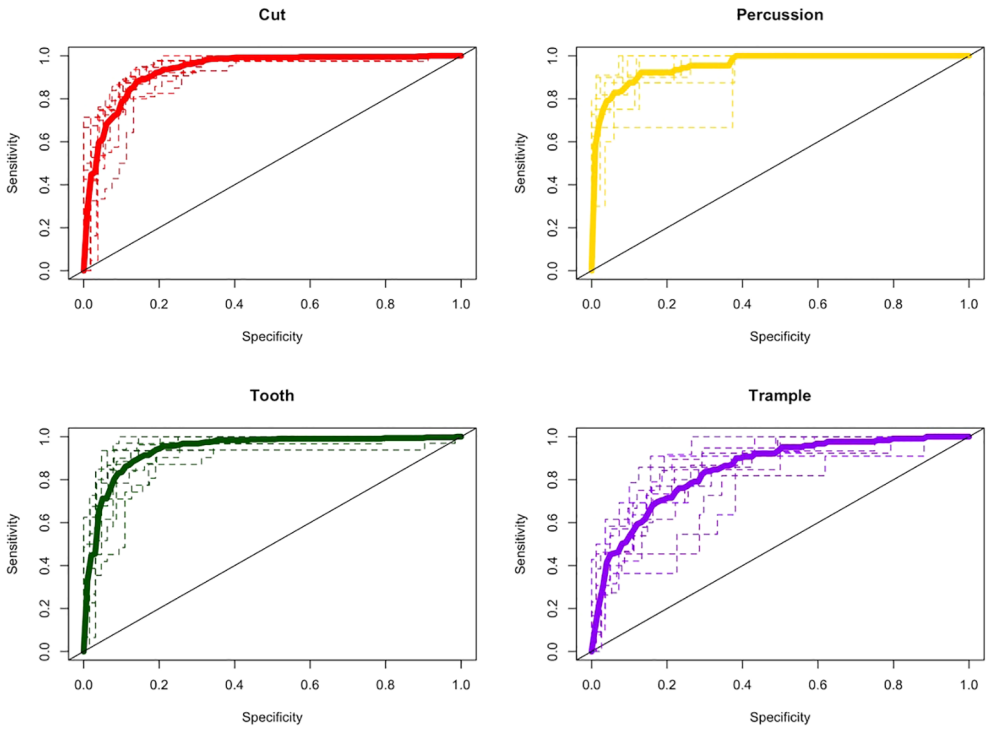


FIGURE 2
One-vs-rest receiver operating characteristic (ROC) curves for each mark type in the 10-fold cross-validated quadratic discriminant analysis model. Dashed lines show per-fold ROC curves; solid lines show the average ROC curve across the 10 folds.

TABLE 4 Confusion matrix for the random forest (RF) classifying the four BSM categories.

| RF | BSM | Predicted | | | | Total |
|---|---|---|---|---|---|---|
| | | Cut | Percussion | Tooth | Trample | |
| Actual | Cut | **318** | 2 | 23 | 68 | 411 |
| | Percussion | 1 | **68** | 21 | 0 | 90 |
| | Tooth | 13 | 41 | **241** | 18 | 313 |
| | Trample | 39 | 3 | 15 | **75** | 132 |
| | Total | 371 | 114 | 300 | 161 | 946 |

Bolded values indicate correctly identified BSMs.

also hope that other researchers find use in this now open-source database of 12 computationally measured features when analyzing their own fossilized BSMs.

## 4.1 Univariate descriptors of BSM morphology

Quantitative evaluations of mark shape provide context to appreciate the challenges researchers face when employing less precise qualitative techniques, that is, misidentifying the causal action behind a feeding trace. While this remains possible when using quantitative techniques, the known likelihood of error and posterior probabilities provided for individual classifications allow assessment of the reliability of mark identifications on fossils that are not possible with qualitative methods.

Results from our quantitative analysis of butchery, carnivore tooth, and trampling BSM morphology are consistent with qualitative descriptions of mark shape. Namely, we show that cut marks, on average, are the longest mark type, while also having the smallest cross-mark widths and most acute opening angles. These measurement data also emphasize the considerable degree of morphological overlap between trample and cut marks, as trample marks are often long with intermediary width measurements. However, our data also shows that trample marks can be morphologically diverse, having many features intermediate to and overlapping with cut, tooth, and percussion marks. This analysis underscores the complexity of qualitative trample mark identifications and the potential of misdiagnosing these marks as another BSM type.

Many measured characteristics of percussion and tooth marks morphologically overlap. The present quantitative assessment of mark shape corroborates qualitative assessments that tooth and percussion marks have more equal length:width ratios while also being deeper than other BSM types. Although quantitative descriptions highlight the many similarities between tooth and percussion marks, they also reveal that percussion marks are characterized by rougher, more complex cross-sectional profile shapes and larger mean depths. Still, all measurements overlap between these mark types, which could complicate attempts to determine whether stone tool percussion or carnivore feeding produced some fossilized BSMs.

It is well established that different actions and behaviors can create BSMs with overlapping morphologies, meaning that no individual morphological feature can serve as a definitive discriminator. Domínguez-Rodrigo et al. (2009) demonstrate this concept by showing that, among 14 qualitative criteria now used to distinguish cut and trample marks, no single characteristic is unique

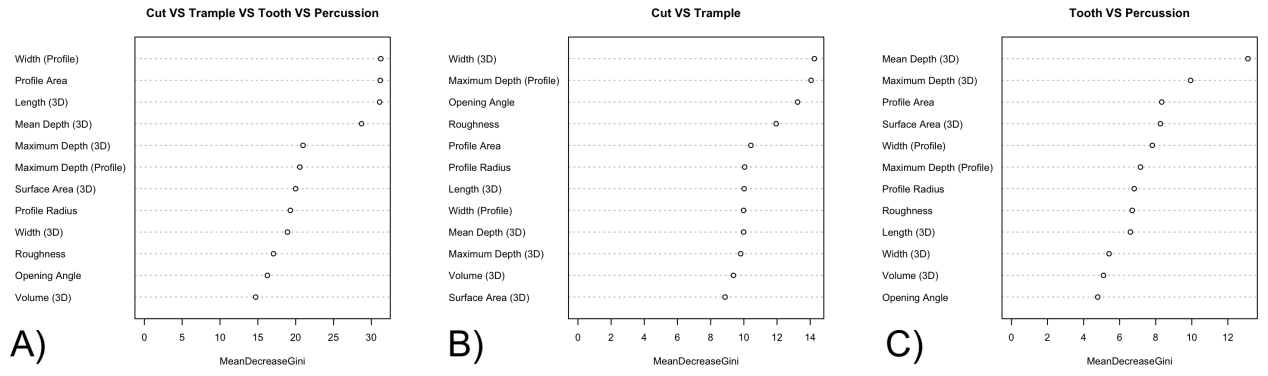


FIGURE 3
Variable importance chart based on mean decrease in Gini from the random forest analysis discriminating cut, percussion, trample, and tooth marks **(A)**, cut and trample marks **(B)**, and tooth and percussion marks **(C)**.

TABLE 5 Confusion matrix for the random forest (RF) classifying cut and trample marks.

| RF | BSM | Predicted | | Total |
|---|---|---|---|---|
| | | Cut | Trample | |
| Actual | Cut | **343** | 68 | 411 |
| | Trample | 46 | **86** | 132 |
| | Total | 389 | 154 | 543 |

Bolded values indicate correctly identified BSMs.

to either mark type. However, the use of binary or nominal scales to evaluate these qualitative criteria (e.g., U- vs V-shaped cross-sectional profile shape) obscures subtle differences between mark types that quantitative methods can more precisely measure. For example, our quantitative method described can use a combination of cross-sectional profile opening angle and radius measurements to describe BSM cross-sectional shape previously described as U- or V-shaped. The use of ratio-scale measurements can enhance the precision with which we capture morphological variation between mark types.

In addition to more thoroughly capturing microscopic details in BSM shape variability, our analytical methodology quantifies previously unconsidered, unobservable, and unmeasurable BSM shapes that might help determine what behaviors and actions created a mark. For example, we directly quantify the degree of cross-sectional profile "roughness" and the total mark surface area. While these new morphological criteria provide a more holistic representation of BSM shape, they also show that no single variable can reliably differentiate BSMs in experimental settings. This observation motivates the subsequent multivariate analysis of BSM shape.

## 4.2 Evaluating ML methods for BSM identification

Machine learning (ML) methods possess immense power to analyze vast amounts of multivariate data at scales inaccessible to individual analysts who rely on qualitative criteria. Consequently, these analytical tools are increasingly common in archaeological inquiry to classify artifact features using large geospatial, imaging, or microscopic datasets with complex and hidden variance patterns (Mantovan and Nanni, 2020; Bickler, 2021; Bellat et al., 2025). The expanded use of complex ML tools can likely, in part, be attributed to the introduction of "point-and-click" software that makes their application in analyses of large and complex datasets easy without the need to understand underlying statistical and methodological processes (Calder et al., 2022). Below, we demonstrate one application of ML methods for identifying prehistoric actions from fossilized BSM shape. However, we also show that, because ML tools exist within a statistical and analytical "black-box", there is a high risk that researchers will misapply these high-powered tools, which parallels current debates about the limitations of deep learning methods in resolving taphonomic equifinality associated with fossilized BSM identifications (Courtenay et al., 2024; Domínguez-Rodrigo et al., 2025a).

TABLE 6 Confusion matrix for the random forest (RF) classifying tooth and percussion marks.

| RF | BSM | Predicted | | Total |
|---|---|---|---|---|
| | | Tooth | Percussion | |
| Actual | Tooth | **268** | 45 | 313 |
| | Percussion | 22 | **68** | 90 |
| | Total | 290 | 113 | 403 |

Bolded values indicate correctly identified BSMs.

In the present study, we use two ML algorithms to investigate the potential for multivariate, quantitative descriptions of BSM shape to determine whether a hominin butcher, carnivore predator, or ungulate trampler created fossilized marks. In recent years, the "No Free Lunch" theorem, as described by Wolpert and Macready (1997), has been interpreted by some researchers as "one should use as many techniques as possible and determine which one(s) is (are) the best for the problem at hand" (Domínguez-Rodrigo, 2019, p. 2714), to justify applying upwards of ten different ML algorithms to a single dataset. However, this interpretation overlooks the spirit of this theorem. Model selection should not be done through brute-force experimentation but, instead, by familiarity with the variance in your dataset coupled with domain experience and statistical expertise. For example, bootstrapping a small sample of only a few hundred datapoints to create thousands of synthetic samples, solely to justify using ML algorithms like neural networks, does not resolve the fundamental limitations of small sample sizes that cause overfitting (Van Der Ploeg et al., 2014; Lones, 2021). Pursuing complex, data-hungry ML models is particularly questionable given the availability of many simpler algorithms that do not require data manipulation. Below, we explain our reasoning behind the two methods used in this study and explain model results in light of other BSM studies.

The first method we use to analyze BSM measurements is DA, as it handles multiple continuous predictor variables simultaneously to classify unknown datapoints while also being less prone to overfitting than many other ML methods (McLachlan, 2005; Khondoker et al., 2016; Nikita and Nikitas, 2020). A further motivator in adopting a DA approach is its long history in experimental studies of fossilized BSMs (e.g., de Juana et al., 2010; Bonney, 2014; Pante et al., 2017; Courtenay et al., 2018; Gümrükçü and Pante, 2018; Domínguez-Rodrigo, 2019) and archaeological inquiry (Kovarovic et al., 2011), facilitating discipline-wide accessibility of our results. The ubiquity of this method likely stems from its interpretability, as both input and output data are straightforward to understand. A further benefit of DA is its Bayesian framework that allows for the inclusion of prior qualitative information (Solberg, 1978; Srivastava et al., 2007). While we do not include qualitative information in our current model, we foresee Bayesian prior information being included in future DA models of BSMs using data derived from the past 50 years of qualitative BSM studies.

The second ML algorithm we use is RF. Given the structure of our dataset, RF algorithms present a useful ML method because they can produce precise classificatory models without needing data normalization or significant preprocessing (Breiman, 2001; Jiang et al., 2008). RF methods are also becoming increasingly frequent in BSM studies (e.g., Courtenay et al., 2019b; Domínguez-Rodrigo, 2019; Domínguez-Rodrigo et al., 2020), making broader dissemination and interpretation of results easier. However, downsides of RF models are their black-box programming, reducing data interpretability, problems with imbalanced datasets, and overfitting noisy datasets (Chen et al., 2004; Barreñada et al., 2024; Halabaku and Bytyçi, 2024). After accounting for these issues, we do not see them negatively impacting our analysis. However, we highlight them as a caution when interpreting the results of past and future BSM studies using RF methods without knowledge of the underlying algorithms.

ML results show that DA and RF algorithms can discriminate between cut, percussion, tooth, and trample marks using 3D and cross-sectional profile measurements with 76% and 74% accuracy, respectively. These classification accuracies are comparable to other ML BSM studies, which frequently report accuracies between 70 and 90% depending on the studied mark-creating actions (Yravedra et al., 2017, 2018; Courtenay et al., 2019b; Linares-Matás et al., 2019; Domínguez-Rodrigo et al., 2020). However, considering classification accuracy alone does not provide a meaningful assessment of a model's utility, as differences in the number of comparative groups or experimental procedures can influence classification accuracy, as described below.

Classification accuracy can decrease as the number of compared groups increases, which corresponds to a lower chance level of accuracy in the classificatory problem (Lones, 2021). Multi-class discrimination problems with four groups, as in the current study, have a baseline classification of 25% accuracy. This value represents the expected value when a dataset has zero underlying structure and the model, instead, relies on random sorting. Alternatively, two-class discrimination models with equal group sizes, which are common in BSM studies (e.g., Pante et al., 2017; Otárola-Castillo et al., 2018; Courtenay et al., 2020a, 2020b; Maté-González et al., 2023), have a higher uniform baseline accuracy of 50%. Therefore, achieving similarly high accuracies becomes more challenging when dealing with multi-class classification problems.

Experimental procedures also influence ML classification accuracy. Most algorithms can easily separate datapoints when there is either low intra-group or high inter-group variance. When considering BSM shape, low intra-group measurement variation can occur if experimental protocols generate identical or very similarly shaped marks that only capture a subset of total real-world variation. For example, Courtenay et al. (2020a) achieve high classification accuracies by comparing trample marks to cut marks intentionally created "by a single right-handed individual, perpendicular to the bone while the bone was fresh and the meat intact". From our experience, we anticipate this experimental procedure will produce uniformly shaped marks unreflective of the full spectrum of cut mark morphology.

In contrast to studies primarily focused on intentionally created BSMs, we study marks from actualistic experiments mimicking the real-world actions and behaviors that create bone surface markings (e.g., replicative stone tool butchery). This does not diminish the utility of intentionally made BSMs, as we include some of these marks in our database, but instead advocates for incorporating both intentional and actualistic marks. In general, this will increase intra-group variation, leading to lower expected classification accuracy. However, this protocol will, in turn, produce an ML model more reflective of real-world variability in mark morphology and have broader efficacy when identifying what created fossilized BSMs.

A further concern when comparing classification accuracies across BSM studies is a frequent misapplication of statistical models, combined with a lack of statistical literacy. This problem may be demonstrated by the multiple studies reporting 100% classification accuracies when identifying taphonomic actions using mark shape (Courtenay et al., 2019b; Domínguez-Rodrigo, 2019; Courtenay et al., 2020a; Domínguez-Rodrigo et al., 2022). While perfect discriminatory precision may not be entirely impossible, we view its repeated occurrence across multiple studies comparing microscopic and characteristically overlapping details with skepticism. Such consistency in precision across independent analyses may suggest that shared sampling biases, measurement constraints, or improper analytical frameworks influenced outcomes (McPherron et al., 2022; Courtenay et al., 2024). Therefore, rather than interpreting these reported BSM identification accuracies as methodologically superior to the method in the present study, we instead view it as indicative that further scrutiny is required to ensure reported precisions are not the artifacts of misapplied analyses.

Below, we discuss three major statistical and analytical issues in quantitative BSM studies that can artificially inflate experimental classification rates and, therefore, produce misinformed interpretations of what prehistoric actions created fossilized BSMs. We primarily comment on the recent publications by Domínguez-Rodrigo and Baquedano (2018; 2025) as examples of ongoing issues in BSM research. However, we emphasize these issues not to criticize any one research group, but to highlight why claims of 100% accuracy warrant further scrutiny, as such results may be more attributable to statistical bias than genuine discriminatory power.

McPherron et al. (2022) demonstrate that the 100% classification accuracy reported by Domínguez-Rodrigo and Baquedano (2018) when distinguishing cut, trample, and tooth marks likely stemmed from improper bootstrapping methods. Additionally, it seems likely that this high classification accuracy is due to data leakage caused by bootstrapping data before separating the testing and training datasets. Unfortunately, data leakage is common in ML studies of BSM shape, which may be creating over optimistic classification accuracies. For example, multiple geometric-morphometric studies follow nearly identical analytical protocols by reducing data dimensionality using PCA, then, often, bootstrapping the data, before, finally, separating it into testing and training datasets (e.g., Aramendi et al., 2019;

Courtenay et al., 2019b, 2020a, 2020b; Yravedra et al., 2022; Maté-González et al., 2023). As such, the testing and training data both inform the PC scores, leading to the training data in the model having direct knowledge of the test set and, potentially, biasing results (Moscovich and Rosset, 2022). It is difficult to fully assess the limitations of these studies because the statistical methods are rarely explained in detail. However, improper bootstrapping methods resulting in duplicated data in testing and training sets, alongside other issues of data leakage, almost certainly compromise model accuracy and create poorly informed interpretations of the fossil record.

While it is occasionally possible to trace methodological problems and identify statistical and analytical shortcomings in BSM studies, most studies do not report code, data, or explain their methodological design in sufficient detail to identify such problems. One recent exception is Domínguez-Rodrigo and Baquedano (2025), who publish their dataset and statistical code in response to McPherron et al.'s (2022) earlier criticisms. This renewed analysis shows that, when controlling for data leakage, their ML method can distinguish cut, tooth, and trample marks with up to 100% accuracy. While we applaud their open distribution of the dataset and code, we believe this dataset and code contain more fundamental issues in the data collection and statistical design that likely permeate most high-accuracy BSM identification studies without openly published data and code.

At a fundamental level, the dataset in Domínguez-Rodrigo and Baquedano (2025), published nearly a decade after their original 2018 paper and used in other studies (Domínguez-Rodrigo, 2019; Abellán et al., 2022), appears to contain typographical errors. Specifically, their binary present/absent "microstriation" variable has three levels.

A further concern with the dataset in Domínguez-Rodrigo and Baquedano (2025) is that it contains variable criteria that are not fully explained in either the original or this paper. For example, groove trajectories were originally coded as being straight, curvy, or sinuous by Domínguez-Rodrigo et al. (2009). However, a fourth category exists in the published dataset, which is potentially explained by Abellán et al. (2022, p.14) as being "variable". How a "variable" groove trajectory differs from sinuous grooves is unclear. The addition of new variates that are ambiguously defined further reduces replicability between analysts, which Domínguez-Rodrigo et al. (2017) showed was already an issue when coding for the 14 qualitative variables that Domínguez-Rodrigo et al. (2009) introduced for BSM identification.

Even if these structural dataset issues are overlooked, issues permeate how Domínguez-Rodrigo and Baquedano (2025) apply ML algorithms. After generating 1,000 RF models using three variables and randomized training/testing data splits, Domínguez-Rodrigo and Baquedano (2025) report a mean classification accuracy of 100% with a standard deviation of 0%. These values indicate that all 1,000 RF models could perfectly separate the cut, trample, and tooth marks in the randomized testing datasets every single time. However, inspection of their code reveals that they inadvertently included the label category (coded as "croc", "tramp", "rf", and "sf" for crocodile, trample, retouched cut marks, and simple cut marks, respectively) as a variable in the RF model, allowing their model to achieve perfect discrimination in the test dataset by using this single "variable". After correcting for this issue, the accuracy of this model decreases substantially to approximately 85%.

We do not presume that Domínguez-Rodrigo and Baquedano (2025) included the label variable in their openly published code intentionally, but we raise this issue as a cautionary tale that using ML methods without understanding the underlying algorithmic principles can easily lead to their misapplication. Researchers familiar with ML methods should recognize that 1,000 RF models reporting perfect classification accuracy is implausible, barring perfectly separable datasets. Here, domain knowledge also plays an important role in constructing ML models, as taphonomic researchers should also recognize that there will always be morphological overlap between these types of BSMs that would reduce model accuracy.

Ultimately, issues plaguing studies of fossilized butchery marks stem from a lack of transparent data and methods, as well as a lack of understanding of how to properly apply statistical methods. The present study overcomes these issues by publishing our raw BSM measurement dataset as well as the associated analytical code necessary to analyze this dataset. Additionally, we hope that our critique highlights that the goal of taphonomic research should not be 100% classification accuracy in experimental models. Instead, research should understand the past as precisely as possible while acknowledging that BSM shape can and will overlap, meaning models with realistic data will likely never achieve perfect discrimination. When ML models are constructed correctly, we can simultaneously assess the experimental precision of the model and the confidence levels for individual classifications of fossilized BSMs providing quantitative assessments of the probability that a specific action produced each mark.

## 4.3 Misclassification patterns in BSM identification

Our classification models reveal three trends about the shape of bone markings created by different actions. First, following qualitative observations (Behrensmeyer et al., 1986; Olsen and Shipman, 1988; Domínguez-Rodrigo et al., 2009), our model frequently misidentifies cut and trample marks for each other. Second, we show a similar trend of misclassifying carnivore tooth marks and percussion marks, which is also consistent with previous qualitative descriptions (Blumenschine and Selvaggio, 1988; Blumenschine et al., 1996; Galán et al., 2009). A third, and somewhat surprising, trend in our models is that a non-insignificant number of tooth marks are misclassified as cut and trample marks and vice versa. Below, we describe the relevance of these trends considering previous experimental work.

### 4.3.1 Cut & trample marks

Qualitatively, there is disagreement about what morphological criteria are diagnostic of trample marks. Trampling BSMs occur as

animals kick, walk on, or otherwise move a bone against sediment above, below, or along the ground (Olsen and Shipman, 1988). Generally, this process produces large patches of microscopic and easily identifiable abrasion marks (Domínguez-Rodrigo et al., 2009). However, in some instances, trampling moves bones against rocks or a surface that produces BSMs that macromorphologically mimic cut marks (Behrensmeyer et al., 1986; Domínguez-Rodrigo et al., 2009; Courtenay et al., 2020a).

Domínguez-Rodrigo et al. (2009) show that, compared to many cut marks, some but not all trample marks have sinuous groove trajectories without shoulder flaking. However, Domínguez-Rodrigo et al. (2017) also show that there is limited inter-observer agreement when qualitatively identifying these features. Furthermore, some criteria Domínguez-Rodrigo et al. (2009) use to define trample marks, such as shoulder flaking, are absent on fossils that have undergone weathering and other post-depositional processes. Further confusion surrounding what features are diagnostic of trampling is highlighted by descriptions in Behrensmeyer et al. (1986) and Olsen and Shipman (1988), who disagree whether trample marks have or lack internal microstriations.

Our quantitative assessment of trample and cut marks also reveals many overlapping morphological features between these mark types, as shown in Figure 4. Mean 3D depth and profile depth measurements are nearly equal between cut and trample marks (Table 2). Similarly, profile roughness as a proxy measurement for the presence and extent of microstriations is only slightly smaller in cut marks compared to trample marks (Table 2). Because these variables are measured on a microscopic scale, the quantitative measurements in the present study highlight nuances that were only previously qualitatively described, providing a more objective description of mark similarities.

Nonetheless, the same analysis demonstrates that there is some unique variation between cut and trample mark shape. These distinct measurements appear to reflect that trampling tends to produce broader marks than stone tool butchery. This is reflected by trample marks having an average radius almost four times

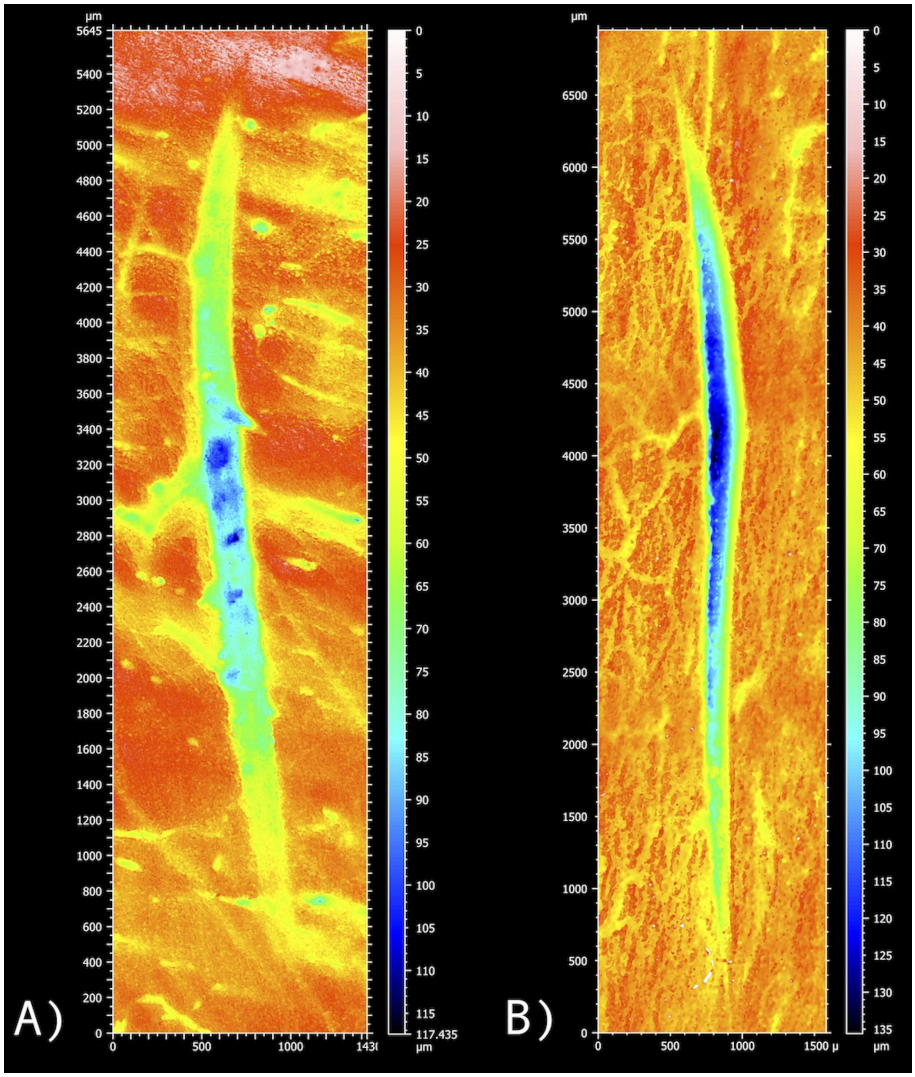


FIGURE 4
Example of a trample **(A)** and cut **(B)** mark with overlapping 3D characteristics, including length, width, and maximum depth.

greater than cut marks and almost double 3D and profile widths (Table 2). Despite these differences in cut and trample mark measurements, the standard deviations reported in Table 2 reveal that all measurement variables overlap to some degree, meaning that no one measurement variable can perfectly discriminate between these marks. Consequently, we consider a multivariate approach to capture the combined discriminatory power of multiple variables in identifying cut and trample marks.

The capacity for a multivariate approach to discriminate cut and trample marks is shown by three of the discriminatory models produced in the present study. First, we consider a QDA model and a RF model comparing cut, trample, percussion, and carnivore tooth marks, showing that even when jointly considering all measurement variables, cut and trample marks can overlap in shape. The extent of multivariate morphological overlap between cut and trample marks is shown by the QDA confusion matrix, where 104 out of the 228 misclassified marks occur between cut and trample marks, and the RF confusion matrix, with 107 out of the 244 misclassified marks (Tables 3, 4).

A follow-up RF model comparing only cut and trample marks reveals the multivariate features that help differentiate these two mark categories. This two-mark RF model was able to discriminate between cut and trample marks with approximately 79% accuracy. In general, the most important variables in this RF model, as depicted by a variable importance plot (VIP) (Figure 3), align with our univariate analysis of cut and trample mark characteristics, specifically that measurements associated with mark breadth and broadness are informative when separating these classes.

One surprising observation in the VIP of the RF model separating cut and trample marks is that maximum profile depth contributes significantly to this model's ability to discriminate between marks. When considered univariately, mean maximum profile depth is nearly equal between cut and trample marks. The importance of this depth variable is likely due to its covariation with another variable, or if it has a non-linear relationship that contributes to discrimination in this model. This observation underscores the broader utility of a multivariate approach in revealing hidden patterns that may be missed when considering only a singular variable.

In general, few quantitative BSM studies consider how trampling marks differ from butchery marks. One exception is Courtenay et al. (2020a), who use 3D microscopy, geometric morphometrics, and deep learning neural networks to distinguish trampling marks from cut marks with 100% accuracy. However, the methodological design of this study only considers intentionally created cut marks, while also suffering from data leakage problems caused by using a PCA and bootstrapping before splitting testing and training datasets. As such, the results reported in Courtenay et al. (2020a) likely confirm our findings that there are morphological differences between cut and trample marks. However, we remain skeptical that any method can discriminate these marks with 100% precision.

Other quantitative studies characterize the shape of trampling marks; however, they do not consider how trampling BSMs compare to cut marks (Courtenay et al., 2019c, 2020b) or they use primarily descriptive statistical tests that do not assess data or model discriminatory power (Souron et al., 2019). As such, the application of quantitative BSM modeling techniques for differentiating trampling marks from other taphonomic processes remains unclear.

The results of our study, alongside other qualitative and quantitative studies of trample and cut marks, indicate that cut and trample marks overlap in morphological characteristics. These similarities could easily lead to researchers misidentifying a fossilized trample mark as a cut mark, and vice versa.

### 4.3.2 Tooth & percussion marks

Qualitative shape characteristics of percussion marks are known to overlap with carnivore tooth marks (Blumenschine and Selvaggio, 1988, 1991; Blumenschine, 1995; Blumenschine et al., 1996; Galán et al., 2009). Both carnivore and percussive behaviors can create pit-shaped marks with similarly high breadth:depth ratios. Qualitative studies show that these marks often differ based on the presence or absence of internal microstriations (Blumenschine et al., 1996). However, Galán et al. (2009) show that when butchers use non-modified hammerstones, they create percussion marks lacking internal microstriations, which overlap morphologically with carnivore tooth marks.

The present quantitative assessment of tooth and percussion marks confirms qualitative observations that these mark types overlap morphologically. Specifically, we confirm previous qualitative results that the mean length and width of tooth marks tend to resemble percussion marks (Table 2), as shown in Figure 5. Additionally, our results show that both carnivore predation and percussion behaviors produce pit-shaped marks with similarly obtuse cross-sectional profiles (Table 2).

However, our results diverge from qualitative observations by showing that percussion marks are, on average, twice as deep as tooth marks (Table 2). These observations contrast previous descriptions of both mark types having similarly high breadth: depth ratios, as they have similar width values but differ in depth. This observation can likely be attributed to the high resolution of the confocal profilers used in this study, which can measure micrometer-scale differences in mark size and shape. Similarly, our technique shows that previously unconsidered features, such as mark volume and profile radius, are nearly twice as large in percussion marks compared to tooth marks (Table 2).

As in our analysis of cut and trample marks, this univariate analysis shows that no single measurement variable perfectly differentiates tooth and percussion marks. As such, we assess the discriminatory powers of our multivariate measurement dataset using three different classification models.

Confusion matrices produced by the four-mark QDA and RF models show that tooth marks are most frequently mistaken as percussion marks, and vice versa. In the QDA model, 27 of the 60 misclassified tooth marks are classified as percussion marks, and 17 of the 21 misclassified percussion marks are classified as tooth marks (Table 3). We observe a similar trend in the RF model, where 41 of the 72 misclassified tooth marks classify as percussion marks,

10.3389/fevo.2026.1681814

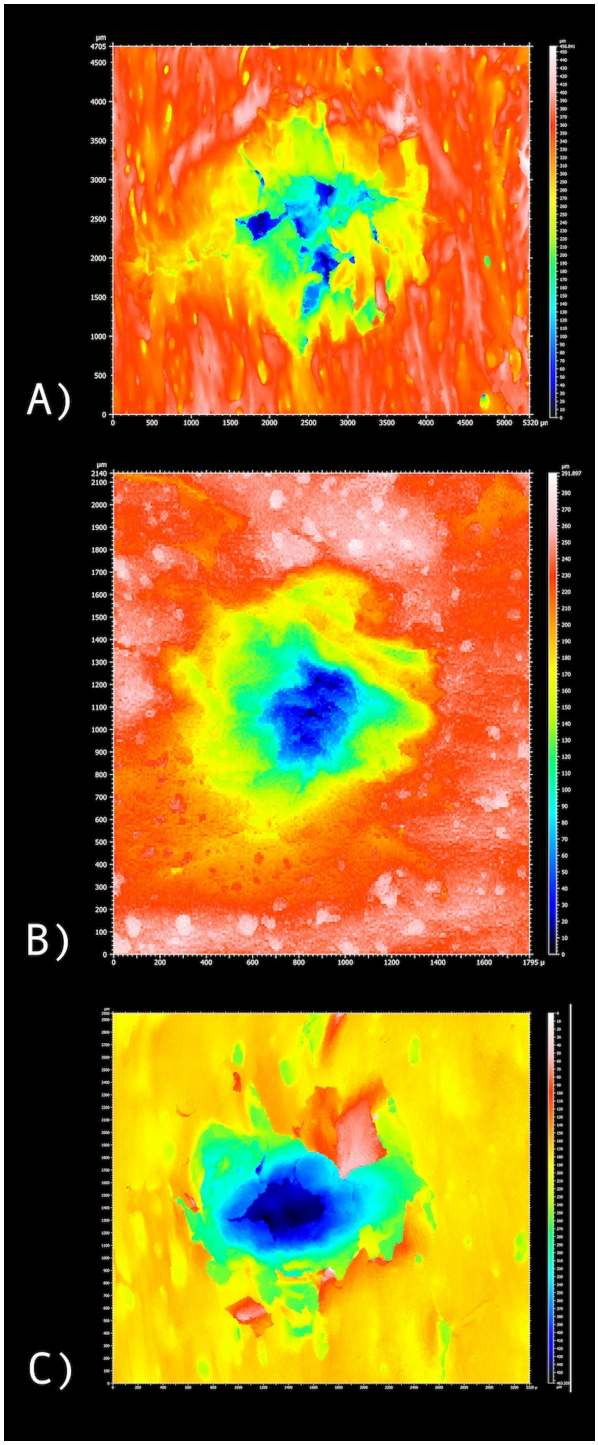


FIGURE 5
Example of a percussion **(A)**, tooth **(B)**, and trample **(C)** mark with overlapping 3D characteristics.

and 21 of the 22 misclassified percussion marks classify as tooth marks (Table 4).

We also generated a two-mark RF model comparing tooth and percussion marks to further investigate the multivariate classification powers of our dataset. This two-mark model was able to discriminate tooth and percussion marks with approximately 83% accuracy. A VIP of the RF model shows that maximum and mean depth measurements contribute most to the construction of this model, which aligns with our univariate measurement descriptions (Figure 3). Length and width measurements are less influential in our two-mark RF model, which agrees with previous qualitative and our univariate observations that these measurements tend to be similar.

In general, 3D studies of percussion marks have not received significant attention compared to cut and tooth marks. Yravedra et al. (2018) discriminate between carnivore tooth marks and percussion marks using a geometric morphometric and structured laser scanning approach with approximately 76% accuracy. This result is slightly lower than the 83% classification accuracy shown in our tooth and percussion RF model (Table 6).

In addition to providing a higher classification accuracy compared to Yravedra et al. (2018), the methodology described in the present study has several advantages. First, Yravedra et al. (2018) use a DAVID structured light-scanner, which has a very low reported maximum resolution of only 60,000 nm (Maté González et al., 2015) compared to the scanner resolution of 45 or 70 nm used in this study. Despite structured-light scanning methodologies being common in studies of BSMs (e.g., Arriaza et al., 2019; Courtenay et al., 2019d, 2019a; Maté-González et al., 2019), low-resolution data could easily impact mark classification when considering microscopic mark features. The issue is even more problematic when considering that in this study, the only inter-analyst test of replicability was in the geometric-morphometric landmarking procedure, not the scanning procedure that creates the mark. Furthermore, Yravedra et al. (2018) highlighted that their scanner is unable to capture "inconspicuous marks whose main morphological exterior and interior features could not be appreciated", suggesting that this method is not appropriate for studying many deep and/or wide percussion or tooth marks. The inconspicuous marks are key to accurately estimating hominin and carnivore involvement in the accumulation of fossil assemblages but are also the most difficult to distinguish from one another when using qualitative methods. Our high-resolution technique is capable of measuring these shallow and difficult to identify marks.

### 4.3.3 Trample and cut marks & tooth marks

An unanticipated trend in our quantitative BSM study is how frequently tooth marks misclassify as both cut and trample marks, and vice versa. Pante et al. (2017) demonstrate that cut and tooth marks can have overlapping morphological features that, when quantitatively measured in 3D, will lead to classificatory models occasionally misclassifying one mark as the other (approximate misclassification rate of 2.75%). Our results misclassify these marks with an approximate 7% misclassification rate, which exceeds previous misclassification estimations and could suggest that some fossilized BSMs classified as cut, tooth, or trample marks are misclassified.

The results of this study show that there is morphological overlap between cut, trample, and tooth marks. Twenty-four out of the 89 misclassified cut marks in the QDA model and 23 of the 93

misclassified cut marks in the RF model are classified as tooth marks (Tables 3, 4). Alternatively, of the 60 misclassified tooth marks in the QDA model, 11 misclassify as a cut mark, while 13 of the 72 misclassified tooth marks in the RF model misclassify as a cut mark (Tables 3, 4).

Previous qualitative studies primarily concern themselves with cut and tooth mark morphological variability. In general, these marks are considered relatively straightforward to distinguish. For example, Blumenschine et al. (1996) show that qualitative mark identification methods can reliably distinguish carnivore tooth marks and cut marks made by metal knives in experimental settings. However, Potts and Shipman (1981) show that fine tooth scratches by carnivores can easily be mistaken for cut marks in both microscopic and macroscopic features. In general, because of the presumed ease in distinguishing cut and tooth marks, the relationship between these marks is omitted from most quantitative studies of BSM morphology, with some researchers calling it unnecessary as the relationship is "relatively obvious and less informative" (Courtenay et al., 2019a) and "[does] not respond to any real archaeological questions" (Courtenay et al., 2019c).

We agree with the notion that qualitatively identifying cut marks that are extremely long, narrow, and with a clear V-shaped cross-sectional profile from some tooth scores or pits on bones that are extremely short, round, and have a more U-shaped profile is not problematic. However, many fossilized marks lack morphological clarity, making their classification as a tooth, cut, or other BSM problematic. For example, whether the 3.4-million-year-old marks on the Dikika fossils were produced by a hominin butcher or a different process has been the subject of intense debate for over a decade (McPherron et al., 2010; Domínguez-Rodrigo et al., 2012; Thompson et al., 2015), while some of the 2.5-million-year-old Bouri Hata marks have also been questioned (Sahle et al., 2017). These debates cloud our understanding of the origins of hominin carnivory and its role in the evolutionary history of humans.

Further examination of the misclassified tooth and cut marks in our QDA and RF models provides additional context for understanding why long-standing debates persist for some fossilized BSMs. In our QDA model, 16 of the 24 cut marks misclassified as tooth marks were produced by unmodified stone tools. Similarly, 13 of the 23 cut marks misclassified as tooth marks in the RF model were made by unmodified stones used as tools (Appendix B.1 & B.2). In their description of the 3.4-million-year-old Dikika marks, McPherron et al. (2010) hypothesized that early hominin butchers could have used naturally occurring and unmodified stones as tools to obtain animal tissue. This suggests cut marks made by unmodified stones could potentially be mistaken for crocodile tooth marks.

Our results also suggest there is potential for misclassification of crocodile tooth marks as cut marks. Specifically, of the tooth marks that were misclassified as cut marks, six of the 11 in the QDA model and six of the 13 in the RF model were produced by crocodiles (Appendix B.1, B.2). This point underscores the morphological overlap of BSMs produced by unmodified stones used as tools and crocodile teeth. Overall, the majority of crocodile tooth marks were correctly classified by both models (72% and 71% in the QDA and RF models, respectively), suggesting their potential for misclassification as cut marks on fossils is much lower after the emergence of stone tool technologies. These findings highlight the importance of applying more objective quantitative BSM modeling methods, as described in the present study, to analyze controversially identified BSMs.

The similarities between tooth marks and trample marks are more or less unexplored in both qualitative and quantitative experiments compared to trample and cut marks. As noted above, this is likely because trample marks are qualitatively described as cut mark mimics (Behrensmeyer et al., 1986; Olsen and Shipman, 1988). While the results of this study do support the qualitative similarities between cut and trample marks, it also shows that trample marks can share morphological features with tooth marks, as evidenced by the digital tooth and trample mark reconstructions in Figure 5. Some trample marks present width: length relationships as great or greater than tooth marks and have similarly shaped U-shaped cross-sectional profiles, as reflected in their broad opening angle and large radii (Table 2).

We note a similar observation of trample marks misclassified as tooth marks, with 16 of the 58 misclassified trample marks in the QDA model and 15 of the 57 misclassified trample marks in the RF model being classified as tooth marks (Tables 3, 4). Similarly, 22 of the 60 misclassified tooth marks in the QDA model were classified as trample marks, and 18 of the 72 misclassified tooth marks in the RF model classified as trample marks (Tables 3, 4).

Our observation that tooth marks can look like cut and trample marks and vice versa supports the data-centric approach we employ in this study. Compared to model-centric studies that focus on parameterizing and refining models, data-centric methods focus on improving the quantity and quality of the data by accepting redundancy or noise in the dataset (Jakubik et al., 2024). This approach, which is fast becoming a standard in many ML studies, allows for a broad understanding of the patterns in a dataset across all groups (e.g., our model comparing the four experimental BSM groups) before creating subset models to study relationships more in-depth (e.g., our models comparing cut and trample marks). In addition to finding and allowing for underlying and hidden relationships to be discovered in a dataset, data-centric ML methods do not necessarily focus on creating hyperparameterized models that may not apply to real-world scenarios (Zha et al., 2025).

While not the primary focus of any model developed in the present study, the observation that tooth marks, cut marks, and trample marks have overlapping morphological characteristics has important implications for using fossilized BSMs as a proxy for understanding hominin carnivory. Namely, it may suggest that qualitative observations of fossilized butchery BSMs previously considered "relatively obvious" and unworthy of further consideration may, in fact, be evidence of non-hominin-related activities, such as carnivore consumption.

## 5 Conclusion

Numerous studies over the past two decades highlight the potential of quantitative BSM identification methods for

discerning the specific actions that created marks on fossils (e.g., Bello and Soligo, 2008; Bello, 2011; Bello et al., 2011; Boschin and Crezzini, 2012; Maté González et al., 2015; Pante et al., 2017; Domínguez-Rodrigo and Baquedano, 2018; Otárola-Castillo et al., 2018, 2022; Yravedra et al., 2018; Courtenay et al., 2019a, 2020a, 2020b; Linares-Matás et al., 2019; Jiménez-García et al., 2020; Pobiner et al., 2023; Curran et al., 2025). However, these methods lack widespread adoption because most experimental procedures are unstandardized, which generates a discipline suffering from irreproducible experimental methodologies, data, and results (James and Thompson, 2015). In the present study, we describe the first open-source database of BSM measurements experimentally generated through simulated stone-tool butchery and percussion, carnivore feeding trials, and ungulate trampling using a quantitative BSM identification method shown to be replicable. Here, we also show how these data can be used to precisely identify what specific taphonomic action created a mark based on mark shape in experimental settings, which has applications for identifying similarly shaped BSMs on fossilized bones.

Our intention in publishing raw measurement values for the largest sample of experimentally generated BSMs to date is to facilitate scholarly collaborations and encourage the adoption of quantitative BSM modeling methods. Data generated by this study have the potential to assist researchers when analyzing morphologically ambiguous fossilized BSMs, improving the reliability and accuracy of our understanding of hominin carnivory.

While the results of this study are promising for discriminating actions based on fossilized BSM shape, we also recognize that our experimental BSM database is incomplete. No single experimental procedure can fully capture the vast array of possible BSM shapes produced by any taphonomic process. For example, the 411 cut marks included in the present database likely capture a sizable amount of morphological variation that could occur in all possible cut marks. However, it is possible that different stone tool technologies (e.g., blades) or raw materials, or even individual butchers could produce morphologically distinct BSMs not currently captured in our database. While it is important to acknowledge such methodological and experimental restrictions, this inherent limitation should not be viewed as a barrier to scientific progress, nor should it impede the use of experimental mark data to understand what effectors and actors (*sensu* Gifford-Gonzalez, 1991) created fossilized BSMs.

The current study marks the beginning of an ongoing project expanding and contributing additional marks to the BSM measurement database presented in this paper. Expanded versions of this BSM database will include larger samples of the mark types presented in this paper, as well as previously unstudied mark types, such as rodent gnawing, bone retoucher, and root etching. As this database grows, so will our understanding of the morphological variation that is possible for each BSM type.

While the primary aim of the present study is to introduce a working measurement database following the analytical protocols established by Pante et al. (2017), we recognize that other researchers may wish to analyze our experimental BSM database using alternative measurement methods. Accordingly, forthcoming publications will include the raw scanned files of our experimentally generated BSM database as openly available 3D coordinate XYZ text files.

An important aspect in presenting this dataset is that we do not intend for our methodology to supplant qualitative BSM identification methods, but instead, work in tandem with such methods. Fossilized marks that researchers agree are unambiguously created by a specific taphonomic action should not require further analysis. However, BSMs that are morphologically ambiguous or the subject of intense debate (e.g., Domínguez-Rodrigo and Barba, 2006; Blumenschine et al., 2007; McPherron et al., 2010; Domínguez-Rodrigo et al., 2012; Fariña et al., 2014; Thompson et al., 2015; Holen et al., 2017b; Ferrell, 2019; Holcomb et al., 2022) should be prioritized when modeling fossilized BSM morphology. Ultimately, we hope that this database can help clear up long-standing controversies about the origins of some marks on fossil bones that researchers continue to debate.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

## Author contributions

TK: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Writing – original draft, Writing – review & editing. AP: Conceptualization, Data curation, Investigation, Methodology, Writing – review & editing. TN: Data curation, Writing – review & editing. EO: Data curation, Methodology, Project administration, Writing – review & editing. IO: Data curation, Methodology, Project administration, Writing – review & editing. AT: Data curation, Investigation, Methodology, Project administration, Writing – review & editing. BP: Data curation, Resources, Writing – review & editing. MP: Conceptualization, Data curation, Investigation, Project administration, Resources, Software, Supervision, Writing – original draft, Writing – review & editing.

## Funding

The author(s) declared that financial support was received for this work and/or its publication. The profilometers used in this study were funded by the Department of Anthropology and Geography and the College of Liberal Arts, Colorado State University.

## Acknowledgments

Thank you to the following people who helped with data collection, curation, and experimental samples: Robert Blumenschine, Matthew Muttart, Jackson Njau, Jay Reti, and Rachel Winter. We also thank the Denver Zoo and Rist Canyon W.O.L.F. Sanctuary for permitting the use of their animals.

## Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo.2026.1681814/full#supplementary-material

## References

Abellán, N., Baquedano, E., and Domínguez-Rodrigo, M. (2022). High-accuracy in the classification of butchery cut marks and crocodile tooth marks using machine learning methods and computer vision algorithms. *Geobios* 72–73, 12–21. doi: 10.1016/j.geobios.2022.07.001

Andrews, P., and Cook, J. (1985). Natural modifications to bones in a temperate setting. *Man* 20, 675–691. doi: 10.2307/2802756

Aramendi, J., Arriaza, M. C., Yravedra, J., Maté-González, M. Á., Ortega, M. C., Courtenay, L. A., et al. (2019). Who ate OH80 (Olduvai Gorge, Tanzania)? A geometric-morphometric analysis of surface bone modifications of a *Paranthropus boisei* skeleton. *Quater. Int.* 517, 118–130. doi: 10.1016/j.quaint.2019.05.029

Arriaza, M. C., Aramendi, J., Courtenay, L. A., Maté-González, M. Á., Herranz-Rodrigo, D., González-Aguilera, D., et al. (2023). An evaluation of landmark-based methods to explore tooth score morphology: A case study on felids and hyenids. *Appl. Sci.* 13, 3864. doi: 10.3390/app13063864

Arriaza, M. C., Aramendi, J., Maté-González, M. Á., Yravedra, J., Baquedano, E., González-Aguilera, D., et al. (2019). Geometric-morphometric analysis of tooth pits and the identification of felid and hyenid agency in bone modification. *Quater. Int.* 517, 79–87. doi: 10.1016/j.quaint.2018.11.023

Arriaza, M. C., Yravedra, J., Domínguez-Rodrigo, M., Mate-González, M. Á., García Vargas, E., Palomeque-González, J. F., et al. (2017). On applications of micro-photogrammetry and geometric morphometrics to studies of tooth mark morphology: The modern Olduvai Carnivore Site (Tanzania). *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 488, 103–112. doi: 10.1016/j.palaeo.2017.01.036

Barreñada, L., Dhiman, P., Timmerman, D., Boulesteix, A.-L., and Van Calster, B. (2024). Understanding overfitting in random forest for probability estimation: a visualization and simulation study. *Diagn. Progn. Res.* 8, 14. doi: 10.1186/s41512-024-00177-1

Behrensmeyer, A. K., Gordon, K. D., and Yanagi, G. T. (1986). Trampling as a cause of bone surface damage and pseudo-cutmarks. *Nature* 319, 768–771. doi: 10.1038/319768a0

Bellat, M., Figueroa, J. D. O., Reeves, J. S., Taghizadeh-Mehrjardi, R., Tennie, C., and Scholten, T. (2025). Machine learning applications in archaeological practices: a review. *J. Comput. Appl. Archaeol.* 8. doi: 10.5334/jcaa.201

Bello, S. M. (2011). New results from the examination of cut-marks using three-dimensional imaging. *Develop. Quater. Sci.* 14, 249–262. doi: 10.1016/B978-0-444-53597-9.00013-3

Bello, S. M., and Soligo, C. (2008). A new method for the quantitative analysis of cutmark micromorphology. *J. Archaeol. Sci.* 35, 1542–1552. doi: 10.1016/j.jas.2007.10.018

Bello, S. M., Verveniotou, E., Cornish, L., and Parfitt, S. A. (2011). 3-dimensional microscope analysis of bone and tooth surface modifications: comparisons of fossil specimens and replicas. *Scanning* 33, 316–324. doi: 10.1002/sca.20248

Benito-Calvo, A., Arroyo, A., Sánchez-Romero, L., Pante, M., and de la Torre, I. (2018). Quantifying 3D micro-surface changes on experimental stones used to break bones and their implications for the analysis of Early Stone Age pounding tools. *Archaeometry* 60, 419–436. doi: 10.1111/arcm.12325

Bickler, S. H. (2021). Machine learning arrives in archaeology. *Adv. Archaeol. Pract.* 9, 186–191. doi: 10.1017/aap.2021.6

Blumenschine, R. J. (1988). An experimental model of the timing of hominid and carnivore influence on archaeological bone assemblages. *J. Archaeol. Sci.* 15, 483–502. doi: 10.1016/0305-4403(88)90078-7

Blumenschine, R. J. (1989). A landscape taphonomic model of the scale of prehistoric scavenging opportunities. *J. Hum. Evol.* 18, 345–371. doi: 10.1016/0047-2484(89)90036-5

Blumenschine, R. J. (1995). Percussion marks, tooth marks, and experimental determinations of the timing of hominid and carnivore access to long bones at FLK *Zinjanthropus*, Olduvai Gorge, Tanzania. *J. Hum. Evol.* 29, 21–51. doi: 10.1006/jhev.1995.1046

Blumenschine, R. J., Marean, C. W., and Capaldo, S. D. (1996). Blind tests of inter-analyst correspondence and accuracy in the identification of cut marks, percussion marks, and carnivore tooth marks on bone surfaces. *J. Archaeol. Sci.* 23, 493–507. doi: 10.1006/jasc.1996.0047

Blumenschine, R. J., and Pobiner, B. L. (2007). "Zooarchaeology and the ecology of Oldowan hominin carnivory," in *Evolution of the human diet* (New York: Oxford University Press, Inc.), 167–190.

Blumenschine, R. J., Prassack, K. A., Kreger, C. D., and Pante, M. C. (2007). Carnivore tooth-marks, microbial bioerosion, and the invalidation of Domínguez-Rodrigo and Barba's, (2006) test of Oldowan hominin scavenging behavior. *J. Hum. Evol.* 53, 420–426. doi: 10.1016/j.jhevol.2007.01.011

Blumenschine, R. J., and Selvaggio, M. M. (1988). Percussion marks on bone surfaces as a new diagnostic of hominid behaviour. *Nature* 333, 763–765. doi: 10.1038/333763a0

Blumenschine, R. J., and Selvaggio, M. M. (1991). On the marks of marrow bone processing by hammerstones and hyenas: their anatomical patterning and archaeological implications. *Cult. Beginnings.: Approaches. to Understanding. Early. Hominid. Life-ways. Afr. Savanna.* 19, 17e32.

Bonney, H. (2014). An investigation of the use of discriminant analysis for the classification of blade edge type from cut marks made by metal and bamboo blades. *Am. J. Phys. Anthropol.* 154, 575–584. doi: 10.1002/ajpa.22558

Boschin, F., and Crezzini, J. (2012). Morphometrical analysis on cut marks using a 3D digital microscope. *Int. J. Osteoarchaeol.* 22, 549–562. doi: 10.1002/oa.1272

Bourgeon, L., Burke, A., and Higham, T. (2017). Earliest human presence in North America dated to the last glacial maximum: new radiocarbon dates from Bluefish Caves, Canada. *PloS One* 12, e0169486. doi: 10.1371/journal.pone.0169486

Braun, D. R., Pante, M., and Archer, W. (2016). Cut marks on bone surfaces: influences on variation in the form of traces of ancient behaviour. *Interface Focus* 6, 20160006. doi: 10.1098/rsfs.2016.0006

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324

Bunn, H. T. (1981). Archaeological evidence for meat-eating by Plio-Pleistocene hominids from Koobi Fora and Olduvai Gorge. *Nature* 291, 574–577. doi: 10.1038/291574a0

Bunn, H. T. (2007). *Meat made us human. Evolution of the human diet: the known, the unknown, and the unknowable*. (New York: Oxford University Press, Inc.) 191–211.

Bunn, H. T., and Ezzo, J. A. (1993). Hunting and scavenging by Plio-Pleistocene hominids: Nutritional constraints, archaeological patterns, and behavioural implications. *J. Archaeol. Sci.* 20, 365–398. doi: 10.1006/jasc.1993.1023

Calder, J., Coil, R., Melton, J. A., Olver, P. J., Tostevin, G., and Yezzi-Woodley, K. (2022). Use and misuse of machine learning in anthropology. *IEEE BITS. Inf. Theory Magazine.* 2, 102–115. doi: 10.1109/MBITS.2022.3205143

Capaldo, S. D. (1997). Experimental determinations of carcass processing by Plio-Pleistocene hominids and carnivores at FLK 22 (*Zinjanthropus*), Olduvai Gorge, Tanzania. *J. Hum. Evol.* 33, 555–597. doi: 10.1006/jhev.1997.0150

Chen, C., Liaw, A., Breiman, L., and others (2004). *Using random forest to learn imbalanced data* Vol. 110 (Berkeley: University of California), 24.

Courtenay, L. A., Herranz-Rodrigo, D., González-Aguilera, D., and Yravedra, J. (2021). Developments in data science solutions for carnivore tooth pit classification. *Sci. Rep.* 11, 10209. doi: 10.1038/s41598-021-89518-4

Courtenay, L. A., Huguet, R., González-Aguilera, D., and Yravedra, J. (2020a). A hybrid geometric morphometric deep learning approach for cut and trampling mark classification. *Appl. Sci.* 10, 150. doi: 10.3390/app10010150

Courtenay, L. a., Huguet, R., and Yravedra, J. (2020b). Scratches and grazes: a detailed microscopic analysis of trampling phenomena. *J. Microsc.* 277, 107–117. doi: 10.1111/jmi.12873

Courtenay, L. A., Maté-González, M. Á., Aramendi, J., Yravedra, J., González-Aguilera, D., and Domínguez-Rodrigo, M. (2018). Testing accuracy in 2D and 3D geometric morphometric methods for cut mark identification and classification. *PeerJ* 6, e5133. doi: 10.7717/peerj.5133

Courtenay, L. A., Vanderesse, N., Doyon, L., and Souron, A. (2024). Deep learning-based computer vision is not yet the answer to taphonomic equifinality in bone surface modifications. *J. Comput. Appl. Archaeol.* 7, 388–411. doi: 10.5334/jcaa.145

Courtenay, L. A., Yravedra, J., Aramendi, J., Maté-González, M. Á., Martín-Perea, D. M., Uribelarrea, D., et al. (2019a). Cut marks and raw material exploitation in the lower pleistocene site of Bell's Korongo (BK, Olduvai Gorge, Tanzania): A geometric morphometric analysis. *Quater. Int.* 526, 155–168. doi: 10.1016/j.quaint.2019.06.018

Courtenay, L. A., Yravedra, J., Huguet, R., Aramendi, J., Maté-González, M. Á., González-Aguilera, D., et al. (2019b). Combining machine learning algorithms and geometric morphometrics: A study of carnivore tooth marks. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 522, 28–39. doi: 10.1016/j.palaeo.2019.03.007

Courtenay, L. A., Yravedra, J., Huguet, R., Ollé, A., Aramendi, J., Maté-González, M. Á., et al. (2019c). New taphonomic advances in 3D digital microscopy: A morphological characterisation of trampling marks. *Quater. Int.* 517, 55–66. doi: 10.1016/j.quaint.2018.12.019

Courtenay, L. A., Yravedra, J., Mate-González, M. Á., Aramendi, J., and González-Aguilera, D. (2019d). 3D analysis of cut marks using a new geometric morphometric methodological approach. *Archaeol. Anthropol. Sci.* 11, 651–665. doi: 10.1007/s12520-017-0554-x

Curran, S. C., Drăguşin, V., Pobiner, B., Pante, M., Hellstrom, J., Woodhead, J., et al. (2025). Hominin presence in Eurasia by at least 1.95 million years ago. *Nat. Commun.* 16, 836. doi: 10.1038/s41467-025-56154-9

de Heinzelin, J., Clark, J. D., White, T., Hart, W., Renne, P., WoldeGabriel, G., et al. (1999). Environment and behavior of 2.5-million-year-old Bouri hominids. *Science* 284, 625–629. doi: 10.1126/science.284.5414.625

de Juana, S., Galán, A. B., and Domínguez-Rodrigo, M. (2010). Taphonomic identification of cut marks made with lithic handaxes: an experimental study. *J. Archaeol. Sci.* 37, 1841–1850. doi: 10.1016/j.jas.2010.02.002

Domínguez-Rodrigo, M. (2019). Successful classification of experimental bone surface modifications (BSM) through machine learning algorithms: a solution to the controversial use of BSM in paleoanthropology? *Archaeol. Anthropol. Sci.* 11, 2711–2725. doi: 10.1007/s12520-018-0684-9

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

Domínguez-Rodrigo, M., and Baquedano, E. (2025). On bootstrapping, data overfitting and crocodiles: an additional comment to McPherron et al. (2022). *Archaeol. Anthropol. Sci.* 17, 62. doi: 10.1007/s12520-025-02183-w

Domínguez-Rodrigo, M., and Barba, R. (2006). New estimates of tooth mark and percussion mark frequencies at the FLK Zinj site: the carnivore-hominid-carnivore hypothesis falsified. *J. Hum. Evol.* 50, 170–194. doi: 10.1016/j.jhevol.2005.09.005

Domínguez-Rodrigo, M., Bunn, H. T., and Yravedra, J. (2014). A critical re-evaluation of bone surface modification models for inferring fossil hominin and carnivore interactions through a multivariate approach: Application to the FLK Zinj archaeofaunal assemblage (Olduvai Gorge, Tanzania). *Quater. Int.* 322–323, 32–43. doi: 10.1016/j.quaint.2013.09.042

Domínguez-Rodrigo, M., Cifuentes-Alcobendas, G., Jiménez-García, B., Abellán, N., Pizarro-Monzo, M., Organista, E., et al. (2020). Artificial intelligence provides greater accuracy in the classification of modern and ancient bone surface modifications. *Sci. Rep.* 10, 18862–18862. doi: 10.1038/s41598-020-75994-7

Domínguez-Rodrigo, M., Cifuentes-Alcobendas, G., Vegara-Riquelme, M., and Baquedano, E. (2025a). Reassessing deep learning (and meta-learning) computer vision as an efficient method to determine taphonomic agency in bone surface modifications. *Biol. Methods Protoc.* 10, bpaf057. doi: 10.1093/biomethods/bpaf057

Domínguez-Rodrigo, M., Courtenay, L. A., Cobo-Sánchez, L., Baquedano, E., and Mabulla, A. (2022). A case of hominin scavenging 1.84 million years ago from Olduvai Gorge (Tanzania). *Ann. New York. Acad. Sci.* 1510, 121–131. doi: 10.1111/nyas.14727

Domínguez-Rodrigo, M., de Juana, S., Galán, A. B., and Rodríguez, M. (2009). A new protocol to differentiate trampling marks from butchery cut marks. *J. Archaeol. Sci.* 36, 2643–2654. doi: 10.1016/j.jas.2009.07.017

Domínguez-Rodrigo, M., and Pickering, T. R. (2003). Early hominid hunting and scavenging: A zooarcheological review. *Evol. Anthropol.: Issues. News. Rev.* 12, 275–282. doi: 10.1002/evan.10119

Domínguez-Rodrigo, M., Pickering, T. R., and Bunn, H. T. (2012). Experimental study of cut marks made with rocks unmodified by human flaking and its bearing on claims of ∼3.4-million-year-old butchery evidence from Dikika, Ethiopia. *J. Archaeol. Sci.* 39, 205–214. doi: 10.1016/j.jas.2011.03.010

Domínguez-Rodrigo, M., Pizarro-Monzo, M., Cifuentes-Alcobendas, G., Vegara-Riquelme, M., Jiménez-García, B., and Baquedano, E. (2024). Computer vision enables taxon-specific identification of African carnivore tooth marks on bone. *Sci. Rep.* 14, 6881. doi: 10.1038/s41598-024-57015-z

Domínguez-Rodrigo, M., Saladié, P., Cáceres, I., Huguet, R., Yravedra, J., Rodríguez-Hidalgo, A., et al. (2017). Use and abuse of cut mark analyses: The Rorschach effect. *J. Archaeol. Sci.* 86, 14–23. doi: 10.1016/j.jas.2017.08.001

Domínguez-Rodrigo, M., Vegara-Riquelme, M., Palomeque-González, J., Jiménez-García, B., Cifuentes-Alcobendas, G., Pizarro-Monzo, M., et al. (2025b). Testing the reliability of geometric morphometric and computer vision methods to identify carnivore agency using Bi-Dimensional information. *Quater. Sci. Adv.* 17, 100268. doi: 10.1016/j.qsa.2025.100268

Fariña, R. A., Tambusso, P. S., Varela, L., Czerwonogora, A., Di Giacomo, M., Musso, M., et al. (2014). Arroyo del Vizcaíno, Uruguay: a fossil-rich 30-ka-old megafaunal locality with cut-marked bones. *Proc. R. Soc. B.: Biol. Sci.* 281, 20132211. doi: 10.1098/rspb.2013.2211

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern recogn. Lett.* 27, 861–874. doi: 10.1016/j.patrec.2005.10.010

Ferrell, P. M. (2019). The Cerutti Mastodon Site reinterpreted with reference to freeway construction plans and methods. *PaleoAmerica* 5, 1–7. doi: 10.1080/20555563.2019.1589663

Friendly, M. (2010). HE Plots for repeated measures designs. *J. Stat. Softw.* 37, 1–40. doi: 10.18637/jss.v037.i04

Galán, A. B., Rodríguez, M., de Juana, S., and Domínguez-Rodrigo, M. (2009). A new experimental study on percussion marks and notches and their bearing on the interpretation of hammerstone-broken faunal assemblages. *J. Archaeol. Sci.* 36, 776–784. doi: 10.1016/j.jas.2008.11.003

Gifford-Gonzalez, D. (1991). Bones are not enough: Analogues, knowledge, and interpretive strategies in zooarchaeology. *J. Anthropol. Archaeol.* 10, 215–254. doi: 10.1016/0278-4165(91)90014-O

Gümrükçü, M., and Pante, M. C. (2018). Assessing the effects of fluvial abrasion on bone surface modifications using high-resolution 3-D scanning. *J. Archaeol. Sci.: Rep.* 21, 208–221. doi: 10.1016/j.jasrep.2018.06.037

Halabaku, E., and Bytyçi, E. (2024). Overfitting in machine learning: a comparative analysis of decision trees and random forests. *IASC* 39, 987–1006. doi: 10.32604/iasc.2024.059429

Haynes, G. (2017). The cerutti mastodon. *PaleoAmerica* 3, 196–199. doi: 10.1080/20555563.2017.1330103

Holcomb, J. A., Mandel, R. D., Otárola-Castillo, E., Rademaker, K., Rosencrance, R. L., McDonough, K. N., et al. (2022). Does the evidence at Arroyo del Vizcaíno (Uruguay) support the claim of human occupation 30, 000 years ago? *PaleoAmerica* 8, 285–299. doi: 10.1080/20555563.2022.2135476

Holen, S., Deméré, T., Fisher, D., Fullagar, R., Paces, J., Jefferson, G., et al. (2017a). Broken bones and hammerstones at the Cerutti Mastodon site: A reply to Haynes (In PaleoAmerica: a journal of early human migration and dispersal). *PaleoAmerica* 4, 8–11. doi: 10.1080/20555563.2017.1396835

Holen, S. R., Deméré, T. A., Fisher, D. C., Fullagar, R., Paces, J. B., Jefferson, G. T., et al. (2017b). A 130, 000-year-old archaeological site in southern California, USA. *Nature* 544, 479–483. doi: 10.1038/nature22065

Jakubik, J., Vössing, M., Kühl, N., Walk, J., and Satzger, G. (2024). Data-centric artificial intelligence. *Bus. Inf. Syst. Eng.* 66, 507–515. doi: 10.1007/s12599-024-00857-8

James, E. C., and Thompson, J. C. (2015). On bad terms: Problems and solutions within zooarchaeological bone surface modification studies. *Environ. Archaeol.* 20, 89–103. doi: 10.1179/1749631414Y.0000000023

Jiang, Y., Cukic, B., and Menzies, T. (2008). Does transformation help. *Defects, (2008b)*. Available online at: http://menzies.us/pdf/08transform.pdf (Accessed April 11, 2025).

Jiménez-García, B., Aznarte, J., Abellán, N., Baquedano, E., and Domínguez-Rodrigo, M. (2020). Deep learning improves taphonomic resolution: high accuracy in differentiating tooth marks made by lions and jaguars. *J. R. Soc. Interface* 17, 20200446. doi: 10.1098/rsif.2020.0446

Keevil, T. (2018). Inferring Early Stone Age tool technology and raw material from cut mark micromorphology using high-resolution 3-D scanning with applications to Middle Bed II, Olduvai Gorge, Tanzania. Fort Collins, CO, USA: Colorado State University.

Khondoker, M., Dobson, R., Skirrow, C., Simmons, A., and Stahl, D. (2016). A comparison of machine learning methods for classification using simulation with multiple real data examples from mental health studies. *Stat. Methods Med. Res.* 25, 1804–1823. doi: 10.1177/0962280213502437

Korkmaz, S., Goksuluk, D., and Zararsiz, G. (2014). MVN: an R package for assessing multivariate normality. *R. J.* 6, 151–162. doi: 10.32614/RJ-2014-031

Kovarovic, K., Aiello, L. C., Cardini, A., and Lockwood, C. A. (2011). Discriminant function analyses in archaeology: are classification rates too good to be true? *J. Archaeol. Sci.* 38, 3006–3018. doi: 10.1016/j.jas.2011.06.028

Krasinski, K. E., and Blong, J. C. (2020). Unresolved questions about site formation, provenience, and the impact of natural processes on bone at the Bluefish Caves, Yukon Territory. *Arctic. Anthropol.* 57, 1–21. doi: 10.3368/aa.57.1.1

Liaw, A., and Wiener, M. (2002). Classification and regression by randomForest. *R. News* 2, 18–22.

Linares-Matás, G. J., Yravedra, J., Maté-González, M. Á., Courtenay, L. A., Aramendi, J., Cuartero, F., et al. (2019). A geometric-morphometric assessment of three-dimensional models of experimental cut-marks using flint and quartzite flakes and handaxes. *Quater. Int.* 517, 45–54. doi: 10.1016/j.quaint.2019.05.010

Litynski, M. L., and Pante, M. C. (2023). Experiments suggest rockfall an improbable cause for bone surface modification on 24, 000-year-old bone at Bluefish Caves, Canada. *J. Archaeol. Sci.* 160, 105860. doi: 10.1016/j.jas.2023.105860

Lones, M. A. (2021). How to avoid machine learning pitfalls: a guide for academic researchers. *arXiv. preprint. arXiv:2108.02497*. doi: 10.48550/arXiv.2108.02497

Mantovan, L., and Nanni, L. (2020). The computerization of archaeology: survey on artificial intelligence techniques. *SN. Comput. Sci.* 1, 267. doi: 10.1007/s42979-020-00286-w

Marginedas, F., Rodríguez-Hidalgo, A., and Saladié, P. (2023). Rodent gnawing over fresh, dry and thermo altered bones: an experimental study with archaeological implications at El Mirador Cave (Atapuerca, Spain). *Historical. Biol.* 35, 1470–1483. doi: 10.1080/08912963.2022.2098487

Maté-González, M. Á., Aramendi, J., González-Aguilera, D., and Yravedra, J. (2017). Statistical comparison between low-cost methods for 3D characterization of cut-marks on bones. *Remote Sens.* 9, 873. doi: 10.3390/rs9090873

Maté-González, M. Á., Courtenay, L. A., Aramendi, J., Yravedra, J., Mora, R., González-Aguilera, D., et al. (2019). Application of geometric morphometrics to the analysis of cut mark morphology on different bones of differently sized animals. Does size really matter? *Quater. Int.* 517, 33–44. doi: 10.1016/j.quaint.2019.01.021

Maté-González, M. Á., Estaca-Gómez, V., Aramendi, J., Sáez Blázquez, C., Rodríguez-Hernández, J., Yravedra Sainz de los Terreros, J., et al. (2023). Geometric morphometrics and machine learning models applied to the study of Late Iron Age cut marks from Central Spain. *Appl. Sci.* 13, 3967. doi: 10.3390/app13063967

Maté-González, M. Á., Palomeque-González, J. F., Yravedra, J., González-Aguilera, D., and Domínguez-Rodrigo, M. (2018). Micro-photogrammetric and morphometric differentiation of cut marks on bones using metal knives, quartzite, and flint flakes. *Archaeol. Anthropol. Sci.* 10, 805–816. doi: 10.1007/s12520-016-0401-5

Maté González, M. Á., Yravedra, J., González-Aguilera, D., Palomeque-González, J. F., and Domínguez-Rodrigo, M. (2015). Micro-photogrammetric characterization of cut marks on bones. *J. Archaeol. Sci.* 62, 128–142. doi: 10.1016/j.jas.2015.08.006

McLachlan, G. J. (2005). *Discriminant analysis and statistical pattern recognition* (Hoboken, New Jersey, USA: John Wiley & Sons).

McPherron, S. P., Alemseged, Z., Marean, C. W., Wynn, J. G., Reed, D., Geraads, D., et al. (2010). Evidence for stone-tool-assisted consumption of animal tissues before 3.39 million years ago at Dikika, Ethiopia. *Nature* 466, 857–860. doi: 10.1038/nature09248

McPherron, S., Archer, W., Otarola-Castillo, E., Torquato, M., and Keevil, T. (2022). Machine learning, bootstrapping, null models, and why we are still not 100% sure which bone surface modifications were made by crocodiles. *J. Hum. Evol.* 164, 103071. doi: 10.1016/j.jhevol.2021.103071

Moscovich, A., and Rosset, S. (2022). On the cross-validation bias due to unsupervised preprocessing. *J. R. Stat. Soc. Ser. B.: Stat. Method.* 84, 1474–1502. doi: 10.1111/rssb.12537

Muttart, M. V. (2017). Taxonomic distinctions in the 3D micromorphology of tooth marks with application to feeding traces from middle bed II, Olduvai Gorge, Tanzania. Fort Collins, CO, USA: Colorado State University.

Nikita, E., and Nikitas, P. (2020). Sex estimation: a comparison of techniques based on binary logistic, probit and cumulative probit regression, linear and quadratic discriminant analysis, neural networks, and naïve Bayes classification using ordinal variables. *Int. J. Legal. Med.* 134, 1213–1225. doi: 10.1007/s00414-019-02148-4

Njau, J. K., and Blumenschine, R. J. (2006). A diagnosis of crocodile feeding traces on larger mammal bone, with fossil examples from the Plio-Pleistocene Olduvai Basin, Tanzania. *J. Hum. Evol.* 50, 142–162. doi: 10.1016/j.jhevol.2005.08.008

Olsen, S. L., and Shipman, P. (1988). Surface modification on bone: Trampling versus butchery. *J. Archaeol. Sci.* 15, 535–553. doi: 10.1016/0305-4403(88)90081-7

Otárola-Castillo, E., Torquato, M. G., Hawkins, H. C., James, E., Harris, J. A., Marean, C. W., et al. (2018). Differentiating between cutting actions on bone using 3D geometric morphometrics and Bayesian analyses with implications to human evolution. *J. Archaeol. Sci.* 89, 56–67. doi: 10.1016/j.jas.2017.10.004

Otárola-Castillo, E. R., Torquato, M. G., Keevil, T. L., May, A., Coon, S., Stow, E. J., et al. (2022). A new approach to the quantitative analysis of bone surface modifications: the bowser road mastodon and implications for the data to understand human-megafauna interactions in north america. *J. Archaeol. Method. Theory* 30, 1028–1063. doi: 10.1007/s10816-022-09583-5

Pante, M. C., Blumenschine, R. J., Capaldo, S. D., and Scott, R. S. (2012). Validation of bone surface modification models for inferring fossil hominin and carnivore feeding interactions, with reapplication to FLK 22, Olduvai Gorge, Tanzania. *J. Hum. Evol.* 63, 395–407. doi: 10.1016/j.jhevol.2011.09.002

Pante, M. C., Muttart, M. V., Keevil, T. L., Blumenschine, R. J., Njau, J. K., and Merritt, S. R. (2017). A new high-resolution 3-D quantitative method for identifying bone surface modifications with implications for the Early Stone Age archaeological record. *J. Hum. Evol.* 102, 1–11. doi: 10.1016/j.jhevol.2016.10.002

Pante, M. C., Njau, J. K., Hensley-Marschand, B., Keevil, T. L., Martín-Ramos, C., Peters, R. F., et al. (2018). The carnivorous feeding behavior of early Homo at HWK EE, Bed II, Olduvai Gorge, Tanzania. *J. Hum. Evol.* 120, 215–235. doi: 10.1016/j.jhevol.2017.06.005

Pante, M. C., Scott, R. S., Blumenschine, R. J., and Capaldo, S. D. (2015). Revalidation of bone surface modification models for inferring fossil hominin and carnivore feeding interactions. *Quater. Int.* 355, 164–168. doi: 10.1016/j.quaint.2014.09.007

Parkinson, J. A. (2018). Revisiting the hunting-versus-scavenging debate at FLK Zinj: A GIS spatial analysis of bone surface modifications produced by hominins and carnivores in the FLK 22 assemblage, Olduvai Gorge, Tanzania. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 511, 29–51. doi: 10.1016/j.palaeo.2018.06.044

Pobiner, B. L. (2007). *Hominin-carnivore interactions: evidence from modern carnivore bone modification and Early Pleistocene archaeofaunas (Koobi Fora, Kenya; Olduvai Gorge, Tanzania)* (New Jersey: Rutgers University).

Pobiner, B. (2020). The zooarchaeology and paleoecology of early hominin scavenging. *Evol. Anthropol.: Issues. News. Rev.* 29, 68–82. doi: 10.1002/evan.21824

Pobiner, B., Pante, M., and Keevil, T. (2023). Early Pleistocene cut marked hominin fossil from Koobi Fora, Kenya. *Sci. Rep.* 13, 9896. doi: 10.1038/s41598-023-35702-7

Pobiner, B., Rogers, M. J., Monahan, C. M., and Harris, J. W. K. (2008). New evidence for hominin carcass processing strategies at 1.5Ma, Koobi Fora, Kenya. *J. Hum. Evol.* 55, 103–130. doi: 10.1016/j.jhevol.2008.02.001

Potts, R., and Shipman, P. (1981). Cutmarks made by stone tools on bones from Olduvai Gorge, Tanzania. *Nature* 291, 577–580. doi: 10.1038/291577a0

Probst, P., Wright, M. N., and Boulesteix, A.-L. (2019). Hyperparameters and tuning strategies for random forest. *WIREs. Data Min. Knowledge. Discov.* 9, e1301. doi: 10.1002/widm.1301

R Core Team (2024). R: A Language and environment for statistical computing. Available online at: http://www.rstudio.com/ (Accessed July 9, 2024).

Revelle, W. (2024). psych: procedures for psychological, psychometric, and personality research (Evanston, Illinois: Northwestern University). Available online at: https://CRAN.R-project.org/package=psych (Accessed July 10, 2025).

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., et al. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinf.* 12, 77. doi: 10.1186/1471-2105-12-77

Sahle, Y., El Zaatari, S., and White, T. D. (2017). Hominid butchers and biting crocodiles in the African Plio–Pleistocene. *Proc. Natl. Acad. Sci. U.S.A.* 114, 13164–13169. doi: 10.1073/pnas.1716317114

Selvaggio, M. M. (1994). Carnivore tooth marks and stone tool butchery marks on scavenged bones: archaeological implications. *J. Hum. Evol.* 27, 215–228. doi: 10.1006/jhev.1994.1043

Shipman, P., and Rose, J. (1983). Early hominid hunting, butchering, and carcass-processing behaviors: Approaches to the fossil record. *J. Anthropol. Archaeol.* 2, 57–98. doi: 10.1016/0278-4165(83)90008-9

Solberg, H. E. (1978). Discriminant analysis. *CRC. Crit. Rev. Clin. Lab. Sci.* 9, 209–242. doi: 10.3109/10408367809150920

Souron, A., Napias, A., Lavidalie, T., Santos, F., Ledevin, R., Castel, C., et al. (2019). A new geometric morphometrics-based shape and size analysis discriminating

anthropogenic and non-anthropogenic bone surface modifications of an experimental data set. *IMEKO. TC* 560–565.

Srivastava, S., Gupta, M. R., and Frigyik, B. A. (2007). Bayesian quadratic discriminant analysis. *J. Mach. Learn. Res.* 8, 1277–1305.

Thompson, J. C., McPherron, S. P., Bobe, R., Reed, D., Barr, W. A., Wynn, J. G., et al. (2015). Taphonomy of fossils from the hominin-bearing deposits at Dikika, Ethiopia. *J. Hum. Evol.* 86, 112–135. doi: 10.1016/j.jhevol.2015.06.013

Van Der Ploeg, T., Austin, P. C., and Steyerberg, E. W. (2014). Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. *BMC Med. Res. Methodol.* 14, 137. doi: 10.1186/1471-2288-14-137

Venables, W. N., and Ripley, B. D. (2002). *Modern applied statistics with S.*, fourth (New York: Springer). Available online at: https://www.stats.ox.ac.uk/pub/MASS4/ (Accessed April 2, 2025).

Wolpert, D. H., and Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE Trans. Evol. Comput.* 1, 67–82. doi: 10.1109/4235.585893

Yravedra, J., Aramendi, J., Maté-González, M. Á., Austin Courtenay, L., and González-Aguilera, D. (2018). Differentiating percussion pits and carnivore tooth pits using 3D reconstructions and geometric morphometrics. *PloS One* 13, e0194324. doi: 10.1371/journal.pone.0194324

Yravedra, J., Courtenay, L. A., Herranz-Rodrigo, D., Linares-Matás, G., Rodríguez-Alba, J. J., Estaca-Gómez, V., et al. (2022). Taphonomic characterisation of tooth marks of extinct Eurasian carnivores through geometric morphometrics. *Sci. Bull.* 67, 1644–1648. doi: 10.1016/j.scib.2022.07.017

Yravedra, J., García-Vargas, E., Maté-González, M. Á., Aramendi, J., Palomeque-González, J. F., Vallés-Iriso, J., et al. (2017). The use of Micro-Photogrammetry and Geometric Morphometrics for identifying carnivore agency in bone assemblages. *J. Archaeol. Sci.: Rep.* 14, 106–115. doi: 10.1016/j.jasrep.2017.05.043

Zha, D., Bhat, Z. P., Lai, K.-H., Yang, F., Jiang, Z., Zhong, S., et al. (2025). Data-centric artificial intelligence: A survey. *ACM Comput. Surv.* 57, 129:1–129:42. doi: 10.1145/3711118