

Combining machine learning algorithms and geometric morphometrics: A study of carnivore tooth marks

Lloyd A. Courtenay^{a,b,c,*}, José Yravedra^{c,d}, Rosa Huguet^{b,a,e}, Julia Aramendi^{c,f}, Miguel Ángel Maté-González^g, Diego González-Aguilera^g, Mari Carmen Arriaza^{h,i}

^a Àrea de Prehistòria, Universitat Rovira i Virgili (URV), Avinguda de Catalunya 35, 43002 Tarragona, Spain

^b Institut de Paleoeologia Humana i Evolució Social (IPHES), Zona Educacional, Campus Sescelades URV (Edifici W3) E3, 43700 Tarragona, Spain

^c Department of Prehistory, Complutense University, Prof. Aranguren s/n, 28040 Madrid, Spain

^d C. A. I. Archaeometry and Archaeological Analysis, Complutense University, Professor Aranguren s/n, 28040 Madrid, Spain

^e Unit Associated to CSIC, Departamento de Paleobiología, Museo de Ciencias Naturales, calle José Gutiérrez Abascal, s/n, 28006 Madrid, Spain

^f IDEA (Institute of Evolution in Africa), Covarrubias 36, 28010 Madrid, Spain

^g Department of Cartography and Terrain Engineering, Polytechnic School of Avila, University of Salamanca, Hornos Caleros 50, 05003 Avila, Spain

^h School of Geography, Archaeology and Environmental Studies, University of the Witwatersrand, Private Bag 3, 2050, South Africa

ⁱ Centre of Excellence in Paleosciences, University of the Witwatersrand, Johannesburg, South Africa

ARTICLE INFO

Keywords:

Artificial Intelligence
Bone Surface modifications
Statistics
Taphonomy

ABSTRACT

Since the 1980s an intense scientific debate has revolved around the hunting capacities of early hominin populations and the behavioral patterns of carnivores sharing the same ecosystem, and thus competing for the same resources. This debate, commonly known as the hunter-scavenger debate, fostered the emergence of a new research line into the Bone Surface Modifications (BSMs) produced by both taphonomic agents. Throughout the following 20 years, multiple studies concerning the action of carnivores have been developed, with a particular focus on the oldest archaeological sites in East Africa. Recent technological advances applied to taphonomy have provided new insight into carnivore BSMs. A newly developed part of this work relies on Geometric Morphometrics (GMM) studies aimed at discerning carnivore agency through the morphologic characterization of tooth scores and pits. GMM studies have produced promising results, however methodological limitations are still present. This paper presents the first combined application of Machine Learning (ML) algorithms and GMM to the analysis of carnivore tooth marks, generating classification rates of 100% between carnivore species in some cases.

1. Introduction

A fundamental line of research in human evolution consists in understanding early hominin subsistence patterns and the interactions with other paleo-communities in the paleoecosystem. Studying such a complex ecological niche ca. 3 to 1 Ma has fueled a wide range of debates regarding early human behavior, mostly focused on both East and South African sites. The famous *osteodontokeratic* culture was associated with *Australopiths* remains from the Makapansgat (South Africa) Member 3 deposit in the mid-20th century (Dart, 1957). Subsequently, Brain's (1967, 1969) neotaphonomic research showed that bone assemblages produced after human butchery and carnivore ravaging was very similar to that found in the Makapansgat assemblage. More importantly, the presence of carnivore activity was identified at the Cradle of Humankind sites in South Africa (Kromdraai, Swartkrans and

Sterkfontein) on hominin remains suggesting that *Australopiths* carcasses were accumulated by carnivores (Brain, 1981). From this perspective *Australopiths* were consequently depicted as prey rather than predators (Brain, 1981).

Similarly, questions regarding the subsistence patterns of early hominins became a heated source of discussion towards the end of the 20th century based on the study of East African archaeological sites such as Koobi Fora or Olduvai Gorge (Binford, 1981; Bunn, 1981). Throughout the 1980s and 1990s, the hunter-scavenger debate confronted the distinction of different taphonomic agents involved in bone modification, so that discerning between anthropogenic and carnivore activity became a key practice in archaeological research. Part of this development focused greatly on the ecology of different carnivore species and ethnoarchaeological studies, in which taphonomy played a pivotal role (Bunn, 1981; Blumenschine, 1988; Blumenschine and

* Corresponding author at: Àrea de Prehistòria, Universitat Rovira i Virgili (URV), Avinguda de Catalunya 35, 43002 Tarragona, Spain.

E-mail address: ladc1995@gmail.com (L.A. Courtenay).

<https://doi.org/10.1016/j.palaeo.2019.03.007>

Received 5 January 2019; Received in revised form 5 March 2019; Accepted 5 March 2019

Available online 07 March 2019

0031-0182/ © 2019 Elsevier B.V. All rights reserved.

Marean, 1993; Selvaggio, 1994; Blumenschine, 1995; Capaldo, 1997; Stanford and Bunn, 2001; Domínguez-Rodrigo, 2002; Pickering and Bunn, 2007; Domínguez-Rodrigo et al., 2007), expanding on previous studies addressing this matter (Dawkins, 1863; Schaller and Lowther, 1969; Sutcliffe, 1970; Houtson, 1979).

An important amount of taphonomic studies have tried to characterize the different types of osteological accumulations produced by carnivores, mainly based on the Bone Surface Modifications (BSMs) produced. Through this, the action of hyenas (Sutcliffe, 1970; Brain, 1981; Cruz-Uribe, 1991; Domínguez-Rodrigo, 2001; Pickering, 2002; Egeland et al., 2008; Domínguez-Rodrigo and Pickering, 2010; Domínguez-Rodrigo et al., 2015), lions (Domínguez-Rodrigo, 1999; Domínguez-Rodrigo et al., 2012; Gidna et al., 2013, 2014; Arriaza et al., 2016), wolves (Haynes, 1983a, b; Willey and Snyder, 1989; Stiner, 2004; Yravedra et al., 2011, 2012) and leopards (Brain, 1981; Kerbis-Peterhans, 1990; Ruiters and Berger, 2000; Pickering et al., 2011) have been widely studied. Additionally, studies of jaguars have also received attention, yet to a lesser degree (Martín, 2008; Domínguez-Rodrigo et al., 2015).

Multiple criteria have been analyzed to determine the agents involved in bone modification at archaeological sites. These include the type of prey (Brain, 1981; Haynes, 1982), skeletal part profiles (Marean and Frey, 1997; Bartram and Marean, 1999; Pickering et al., 2003; Faith et al., 2007), prey mortality profiles (Stiner, 1994; Bunn and Pickering, 2010), and tooth marks (Selvaggio and Wilder, 2001; Domínguez-Rodrigo and Piqueras, 2003; Delaney-Rivera et al., 2009; Andrés et al., 2012; Domínguez-Rodrigo et al., 2012). Tooth marks can further be divided into different categories, produced through different masticatory actions during carcass consumption. Among these, the two most frequent types are pits and scores. While pits present a circular morphology created by the direct imprint of the carnivore's tooth, a score is a shallow elongated groove across the surface of the bone with an equally rounded base (Binford, 1981; Blumenschine, 1995).

Initial studies into these two BSMs focused primarily on biometric dimensional properties (Selvaggio and Wilder, 2001; Domínguez-Rodrigo and Piqueras, 2003). While using 95% confidence intervals have been able to reduce a great percentage of equifinality present through studies of this type (Andrés et al., 2012), overlapping of samples is still large and, in many cases, little information serves to accurately build hypotheses. On the other hand, other authors have studied the properties of bone fracture patterns, differentiating carnivore agency from anthropogenic activities via fracture plane angles (Capaldo and Blumenschine, 1994; Pickering et al., 2005; Pickering and Egeland, 2006; Alcántara et al., 2006; Galán et al., 2009; Moclán and Domínguez-Rodrigo, 2018). Development of new variables such as the taphotypes have successfully differentiated between spotted hyenas and lions, alongside different types of felids (Domínguez-Rodrigo et al., 2015), nevertheless, the identification of different carnivore taxa involved in bone modification remains controversial.

The application of Geometric Morphometrics (GMM) to taphonomy has transformed the analysis of both anthropogenic (Maté-González et al., 2015, 2016, 2017a; Courtenay et al., 2017), and carnivore BSMs, including tooth scores (Yravedra et al., 2017), and tooth pits (Aramendi et al., 2017a, Arriaza et al., 2018). Despite the apparent success of these studies, certain aspects still need to be improved. Aramendi et al. (2017a), Arriaza et al. (2017, 2018) and Yravedra et al. (2017) were able to determine the action of some carnivore species based on the morphology of their traces up to certain degree, not reaching 100% of certainty. Additionally, while the tooth marks of hyenas and lions were distinguishable, wolves and jaguars are much harder to define. The limitations imposed by sample size and the number of carnivore species included in the analysis might have also affected the resolution obtained with these methods.

The pioneering introduction of Machine Learning (ML) algorithms into archaeological research, however, offers an excellent opportunity to analyse BSMs with a higher precision. The first introduction of ML

methods to the study of carnivore activity, based on the skeletal part representation observed in modern carnivore bone accumulations, provided 100% classification rates in some cases (Arriaza and Domínguez-Rodrigo, 2016). Equally remarkable results have been presented for the analysis of anthropogenic BSMs by means of ML techniques (Domínguez-Rodrigo, 2018; Domínguez-Rodrigo and Baquedano, 2018), as well as fracture patterns (Moclán and Domínguez-Rodrigo, Under Revision).

2. Materials and methods

In this paper we present the first combined effort to join GMM analysis and ML techniques to discern carnivore agency based on tooth mark morphology, inspired by implementations of advanced statistical techniques in taphonomy. For the purpose of this study we have reassessed the GMM models proposed by Aramendi et al. (2017a) and Yravedra et al. (2017), implementing ML algorithms for the processing of morphometric data.

2.1. Tooth mark samples

A total of 89 carnivore tooth pits and 127 scores on adult horse long bones were compared. All bones contained flesh when presented to the individuals. These included tooth marks generated by spotted hyenas (pits = 21, scores = 33), jaguars (pits = 20, scores = 34) and lions (pits = 24, scores = 30) in a controlled setting, in the Cabárceno Nature Park in Cantabria (Spain), were studied. All the marks were produced by multiple (minimum 2) adult individuals for each species. Tooth marks produced by wolves (pits = 24) were obtained from Cabárceno Park, as well as the natural wolf sites in mount Campelo, near Sobrado Dos Montxes, Galicia (scores = 30). In the case of wolves, the number of wolves ranged between 10 and 15 wild individuals.

Only tooth marks on long bone shafts (namely tibiae and radii) were selected for two main reasons. Firstly, diaphyses are denser than epiphyses, therefore teeth tend to penetrate cortical layers less. Secondly, diaphyses present a higher survival rate, thus the use of a sample based on shafts would be a more useful framework for future archaeo-paleontological analogies. Bone epiphyses are usually more susceptible to being destroyed during carnivorous feeding (e.g. furrowing) or taphonomically affected through fluvial agents (Lyman, 1994).

Pits and scores were selected on the basis of their preservation and general conditions. Inconspicuous or superficial tooth marks and those that present a bad cortical preservation or some type of post-depositional alteration were excluded from the analysis.

For more details about the lion sample see Gidna et al. (2013), for the hyena and jaguar samples see Domínguez-Rodrigo et al. (2015), and for wolf pit and score samples see Yravedra et al. (2011, 2012, 2017).

2.2. Virtual reconstruction of marks

Micro-Photogrammetry and computer vision techniques were used to create high-resolution 3D models of the pits and scores (Fig. 1). Precise metrical models were generated using images taken with oblique photography using a CANON EOS 700D with a 60-mm macro lens (Table S1) and following the specified protocol explained in Maté-González et al. (2015). The camera was self-calibrated to simultaneously compute the interior and exterior camera parameters. For data collection, a total of 9–10 photos were taken for each mark. The number of photos varies depending on the geometry of the bone and the shape of the mark (Aramendi et al., 2017a; Yravedra et al., 2017). Once the photographs had been taken, they were processed so as to generate a 3D model for each mark with the open-source photogrammetric reconstruction software GRAPHOS (inteGRAted PHOtogrammetric Suite) (González-Aguilera et al., 2016).

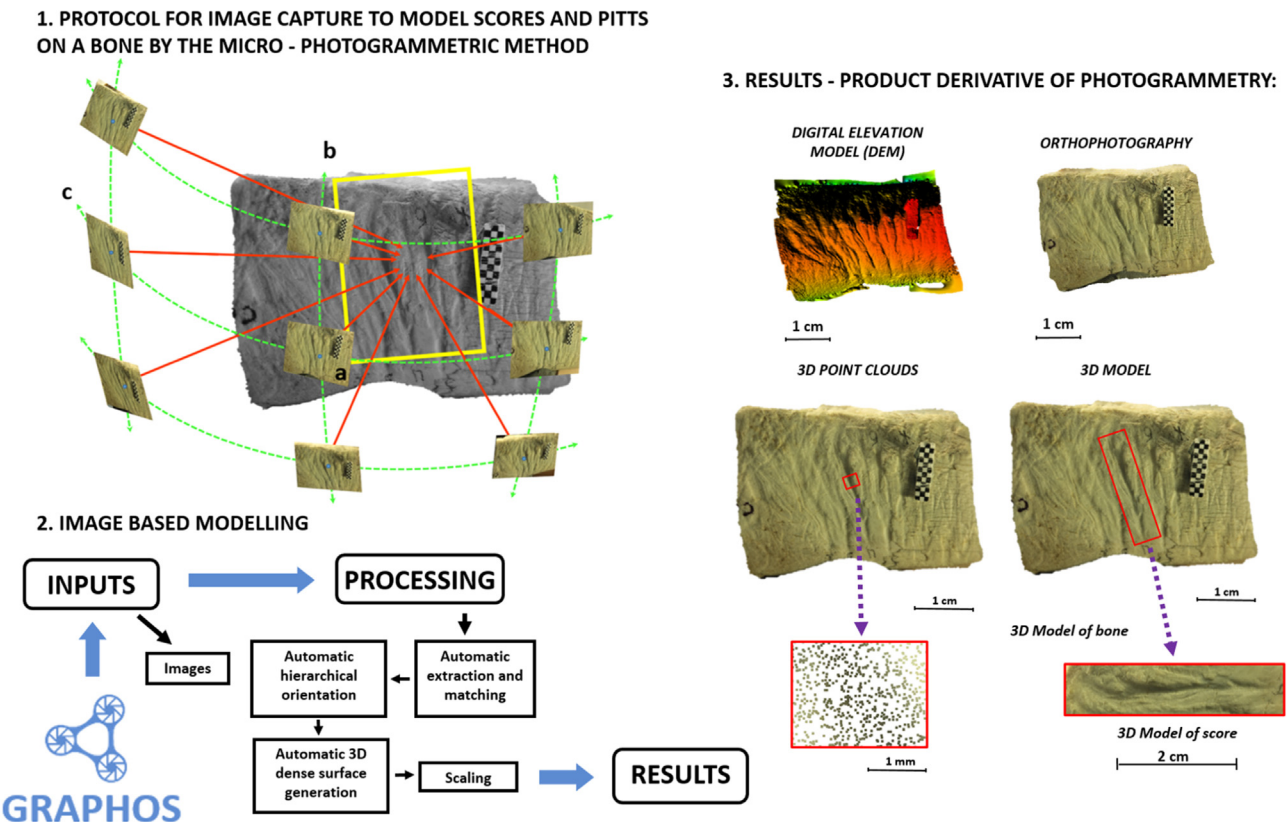


Fig. 1. Protocol for Image Capture and 3D reconstruction for each tooth mark prior to processing (Maté-González et al., 2015; Yravedra et al., 2017).

2.3. Geometric morphometrics

Geometric morphometric analyses of the tooth marks were based on two different models. A 7 2D homologous landmarks, as described by Yravedra et al. (2017), was used on tooth scores while tooth pits were studied using 17 3D homologous landmarks, as described by Aramendi et al. (2017a). Both models can be consulted in Table 1. All statistical

analyses were carried out in the free software R (www.rproject.org, Core-Team, 2015). All R packages and libraries with their respected bibliography can be consulted in Supplemental Appendix 1.

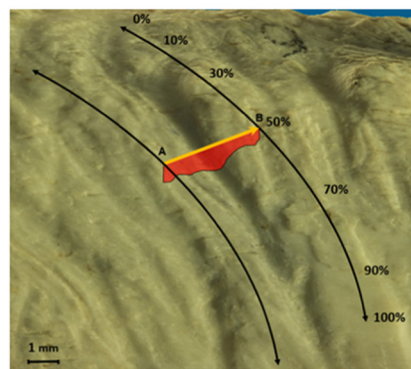
Tooth score cross sections were obtained through importing the 3D digital reconstruction of each tooth mark into a Global Mapper software (Fig. 2), where marks could be treated, extracting sections at mid length (between 30% and 70% of the marks' length) (Maté-González et al.,

Table 1
Geometric morphometric landmark models used for comparisons between taxa. Including Yravedra et al.'s (2017) 2D 7-landmark model and Aramendi et al. (2017a) 3D 17-landmark model.

| Model | N° | Landmark | Description |
|-------|----|---------------------|--|
| 2D | 1 | Left shoulder | Upper limit of the left side of the mark |
| | 2 | Midpoint left wall | Halfway between the bottom of the mark and the left shoulder |
| | 3 | Base left wall | Point where the left wall and bottom of the mark meet |
| | 4 | Depth | Deepest point of the mark |
| | 5 | Base right wall | Point where the right wall and bottom of the mark meet |
| | 6 | Midpoint right wall | Halfway between the bottom of the mark and the right shoulder |
| | 7 | Right shoulder | Upper limit of the right side of the mark |
| 3D | 1 | Length A | Upper limit of the longitudinal axis |
| | 2 | Length B | Lower limit of the longitudinal axis |
| | 3 | Width A | Left limit of the breadth axis |
| | 4 | Width B | Right limit of the breadth axis |
| | 5 | Depth | Most centered lowest point of the pit |
| | 6 | Left upper half A | Point at the first third between the upper limit of the long axis and the left limit of the breadth axis |
| | 7 | Left upper half B | Point at the second third between the upper limit of the long axis and the left limit of the breadth axis |
| | 8 | Left lower half A | Point at the first third between the left limit of the breadth axis and the lower limit of the long axis |
| | 9 | Left lower half B | Point at the second third between the left limit of the breadth axis and the lower limit of the long axis |
| | 10 | Right upper half A | Point at the first third between the upper limit of the long axis and the right limit of the breadth axis |
| | 11 | Right upper half B | Point at the second third between the upper limit of the long axis and the right limit of the breadth axis |
| | 12 | Right lower half A | Point at the first third between the right limit of the breadth axis and the lower limit of the long axis |
| | 13 | Right lower half B | Point at the second third between the right limit of the breadth axis and the lower limit of the long axis |
| | 14 | Interior length A | Upper inflection point on the longitudinal axis |
| | 15 | Interior length B | Lower inflection point on the longitudinal axis |
| | 16 | Interior width A | Left inflection point on the breadth axis |
| | 17 | Interior width B | Right inflection point on the breadth axis |

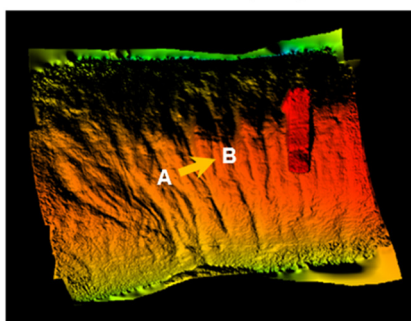
SCORE

3D MODEL



Side view

DIGITAL ELEVATION MODEL (DEM)



Orthogonal view

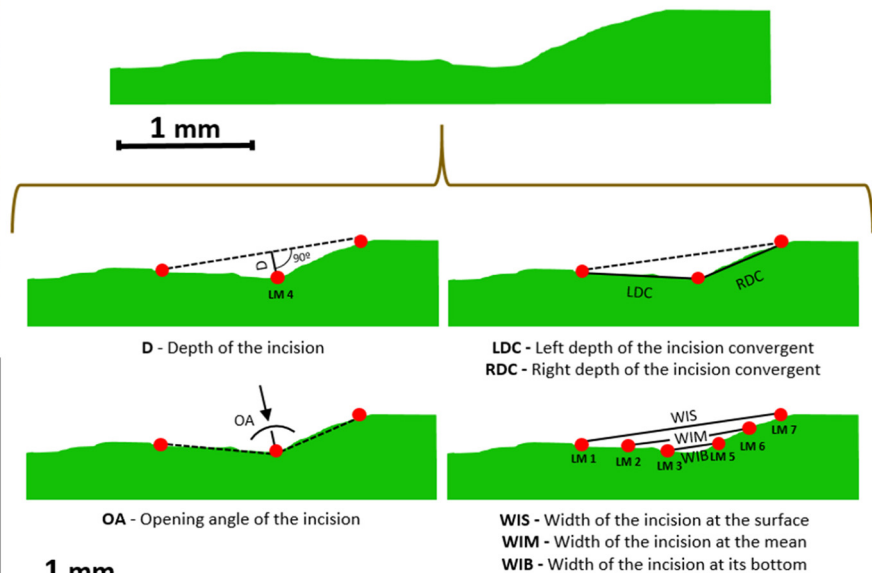


Fig. 2. Location of the seven landmarks in the 2D Geometric Morphometric model for tooth score cross-sections (Yravedra et al., 2017).

2015; Arriaza et al., 2017; Yravedra et al., 2017). These profiles were then extracted and imported into the free software tpsDig2 (v.2.1.7) where landmark data was collected.

3D digital reconstructions of each tooth pit were exported into the Avizo software (Visualization Sciences Group, USA) for landmarking (Fig. 3). Files containing landmark coordinates were then edited and imported into R where a full Procrustes fit and an orthogonal tangent projection (Dryden and Mardia, 1998) were used to normalise the data for further multivariate statistical analyses. This technique, commonly referred to as Generalized Procrustes Analysis (GPA), is used to standardize the form information through the application of superimposition procedures including translation, rotation and scaling. Any remaining differences are exposed through patterns of variation and covariation that can be assessed through several statistical tests (Slice, 2001; Rohlf, 1999). Principal component analyses (PCA) in shape and form space on the Procrustes superimposed landmarks were performed to reduce the large sets of variables to fewer dimensions that represent most of overall variance (Bookstein, 1991; Mitteroecker and Gunz, 2009). The PC scores calculated by the PCAs were then extracted and used for ML training.

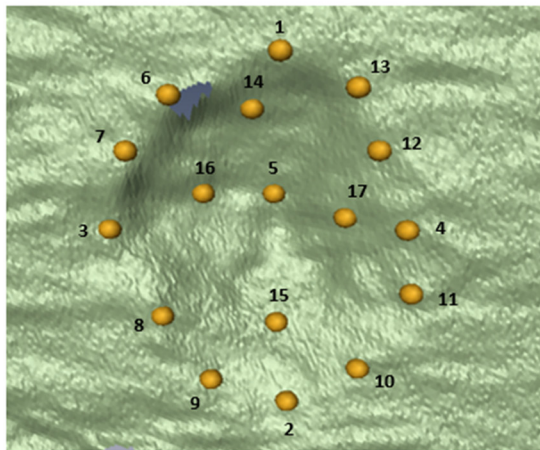
2.4. Machine learning algorithms

Recent applications of ML algorithms in archaeology have been described as powerful classifiers (Domínguez-Rodrigo, 2018). ML provides a treatment of raw data using structured techniques that assign meaning to the data (Lantz, 2013). Much like in humans, the learning process consists in data retrieval (input), translation (abstraction) and a final output where the abstract data is used to form the basis behind the action (generalization) (Lantz, 2013). In this case, ML is particularly useful in handling large amounts of data and efficiently providing a

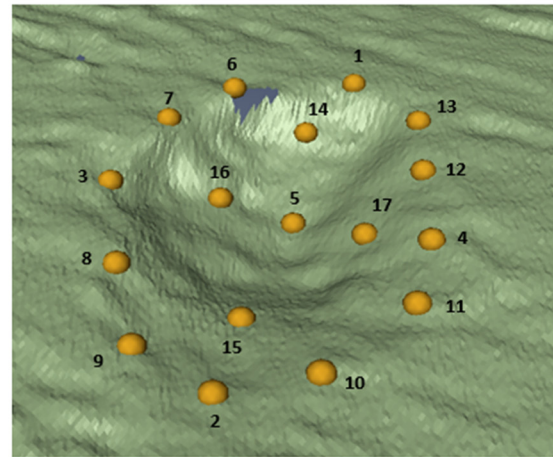
means of handling this information; through classification.

To ensure the accuracy of training models, the samples were bootstrapped 1000 times. To train each algorithm and generate our classification models, 70% of the sample was separated. Validation was then carried out on the remaining 30% of the sample. This is a standard process in testing predictive model efficiency. Model evaluation is performed through resampling techniques that can be used to estimate performance. This can be carried out using multiple methods; in this case we preferred the method based on subsampling and submodeling through k fold cross validation ($k = 10$). One main advantage of ML algorithms is the possibility of tuning the parameters in order to improve performance. This can either be done manually via altering values such as γ and cost , in the case of SVM models, or they can be tuned using additional lines of R coding (Supplemental Appendix). The parameters achieved through tuning are then applied in the models, thus improving performance.

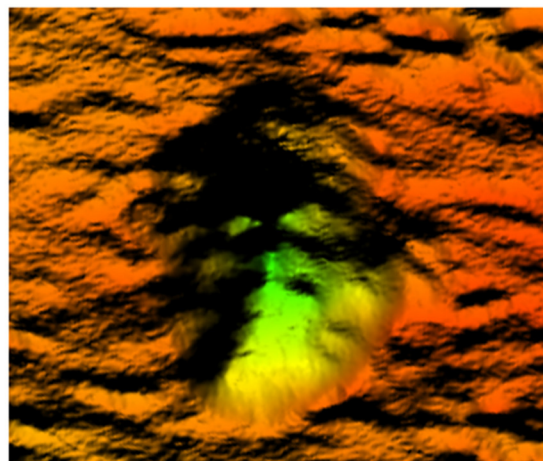
A random selection of 9 common ML algorithms was selected. Much like in the case of GMM statistical analysis of variance, PC scores produced through PCA were used for ML analysis. Comparison of ML algorithm performance was then carried out considering Kappa, sensitivity, specificity and balanced accuracy values through confusion matrix tables. The Kappa statistic adjusts accuracy through considering the possibility of a correct prediction by chance alone (Lantz, 2013). These values are presented as a value between -1 and 1 , with any value above 0.8 considered as a powerful predictive model. Sensitivity and Specificity tests combine different evaluations of Type I and Type II statistical errors in proportion with the rest of the calculated confusion matrix. Calculation of these ratios is a common practice in medical statistics, whereby sensitivity defines the likelihood of correctly diagnosing an illness while specificity is defined by the likelihood of correctly identifying the healthiness of an individual (Fawcett, 2006).

PIT**3D MODEL**

Orthogonal view



Side view

**DIGITAL ELEVATION
MODEL (DEM)**

4 mm

Fig. 3. Location of the seventeen landmarks in the 3D Geometric Morphometric model for tooth pits Aramendi et al. (2017a).

These can also be referred to as the true positive and true negative rates, thus defining the likeliness of misclassification (low number = high misclassification). A combination of these results can provide an additional balanced accuracy value that defines the power of the model (Lantz, 2013).

The 9 ML algorithms selected are the following:

Neural Network (NNET). NNET is inspired by human brain activities, using a computational model that mimics brain patterns. This method uses an interconnected group of nodes, involving input, connected weights, processing elements and a resulting output (Lantz, 2013; Wei and Chiu, 2015). Nodes are built through multiple regression methods. Training of the neural networks consists in the adjustment of weights through successive layers of nodes (Günther and Fritsch, 2010). The input data fed to the neural layers (perceptrons) is transformed via specific nonlinear sigmoidal functions. The parameters of these functions are usually optimised to minimize SQR (Sum of Square Residuals). These parameters exhibit a tendency to overfit the training data set. To avoid this, weight decay is used to reduce the model error for a given value of lambda (Domínguez-Rodrigo, 2018). This λ parameter must be specified

together with the number of hidden units.

Support Vector Machines (SVM). SVM maps input data into high dimensional feature space defined by the *kernel* function (Lantz, 2013). This is based on finding the optimum hyperplane that separates the training data by the maximum margin (Wei and Chiu, 2015). SVM algorithms are powerful classifiers in cases where distribution is non-linear (Cortes and Vapnik, 1995). SVM models are tuned through thresholds in order to define residuals that consequently contribute to a more robust fit. Tuning parameters consist in *cost*, which is used to penalize models with large residuals, and a *loss* function that determines the degree of overfitting training models.

Partial Least Squares Discriminant Analysis (PLSDA). This algorithm classifies class through predictor combinations that optimally separate classes (Mevik and Wehrens, 2007). The tuning components of PLSDA are found through latent variables (components) that maximise classification accuracy. Once these components are identified, the PLSDA model is retrained to ensure accuracy in the final model. PLSDA in many cases has been identified as more powerful than standard Linear Discriminant Analysis (LDA) methods.

K-Nearest Neighbour (KNN). This unsupervised learning algorithm

Table 2

Comparison of different ML algorithm performance used for discerning carnivore agency (excluding wolves) through Yravedra et al.'s (2017) 2D 7-landmark model in Shape space.

| Model | Model performance | | 95% confidence interval | | Hyena | | Jaguar | | Lion | |
|-------|-------------------|----------|-------------------------|-------|-------|-------|--------|-------|-------|-------|
| | Kappa | Accuracy | Lower | Upper | Spec. | Sens. | Spec. | Sens. | Spec. | Sens. |
| NNET | 0.54 | 0.69 | 0.63 | 0.74 | 1 | 0.50 | 0.90 | 1 | 0.95 | 1 |
| PLSDA | 0.79 | 0.86 | 0.82 | 0.90 | 0.90 | 0.83 | 1 | 0.80 | 0.99 | 0.97 |
| NB | 0.82 | 0.88 | 0.84 | 0.91 | 0.98 | 0.77 | 0.88 | 0.90 | 0.97 | 0.96 |
| CTREE | 0.9 | 0.93 | 0.90 | 0.96 | 0.94 | 0.95 | 0.99 | 0.87 | 0.95 | 1 |
| MDA | 0.92 | 0.95 | 0.91 | 0.97 | 0.98 | 0.87 | 0.94 | 0.96 | 1 | 1 |
| KNN | 0.99 | 0.99 | 0.98 | 1 | 0.99 | 1 | 1 | 0.98 | 1 | 1 |
| RF | 1 | 1 | 0.99 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| C5.0 | 1 | 1 | 0.99 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| SVM | 1 | 1 | 0.99 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

classifies unlabelled data by assigning them the class of the most similar labelled examples. This algorithm works well in samples with many variables and performs well when there are well-defined labelled sets. The main advantage of this algorithm is that it is easy to train. KNN identifies k cases in the sample as the nearest in similarity. Unlabelled cases are subsequently assigned by similarity. To predict the location of testing data in the predictor space, different k models are tested and compared to an error/accuracy parameter. To overcome the bias-variance trade-off an intermediate k value is usually selected. Larger k values tend to reduce the bias of variance but small patterns may go unnoticed (Domínguez-Rodrigo, 2018).

Mixture Discriminant Analysis (MDA). Initially conceived as an extension of LDA, MDA is built upon class-specific distributions combined into a single multivariate distribution. This is done by creating a per-class mixture (Kuhn and Johnson, 2013), consisting in the separation of class-specific means from class-specific covariance structures. MDA can be tuned using parameters in the number of distributions per class or sub-classes.

Naive Bayes (NB). NB algorithms apply Bayes' theorem through estimation of class probabilities and likelihoods on observed predictions. This results in dynamic estimates of posterior probabilities of classes. The conditional probability is then used for model classification. NB assumes that all predictors are independent and uses a nonparametric density modelling process.

Decision tress using the C5.0 algorithm (C5.0). This algorithm implements decision trees, employing ML techniques to ensure model accuracy. Unlike standard decision tress, the C5.0 algorithm tuning methods can reach comparable complexities to NN or SVM models. This model works through recursive partitioning of data. The C5.0 algorithm employs k -fold cross validation, dividing the sets into k subsets during the training and testing phases. Variance estimation decreases as k increases.

Conditional Inference Tress (CTREE). CTREES are similar to standard decision tress in as much as they recursively partition data through univariate splits on dependent variables. The main difference presented in this algorithm is that CTREES adapt significance test procedures in order to select the variables that maximise prediction accuracy (Wei and Chiu, 2015).

Random Forest (RF). RF is a more robust version of CTREE. Much like CTREE, the RF algorithm uses a small random number of the data set variables, instead of all the variables. Each selection produces an independent tree. The random variable selection is performed through bootstrap aggregation, known as bagging, splitting the data into multiple data sets for testing. These observations are referred to as out-of-bag (OOB) observations. RF produces estimates of how many iterations are needed to minimize the OOB error. After selecting a number of trees, the algorithm averages the results and produces a robust classification method. This avoids overfitting of results to data, as is more common in standard decision and

regression trees.

More information on ML models used in taphonomy can be consulted in Domínguez-Rodrigo (2018).

2.5. Fine tuning GMM models

Common malpractice in GMM studies is the assumption that the first PC and CV scores are the most important for sample differentiation. Theoretically, the first few PC scores represent the highest percentage of sample variance and covariance, however, in some cases analytical scores with lower percentiles are useful to identify deviations from the majority (Albrecht, 1992). In the case of tooth scores, the first 10 PC scores were successful in most statistical applications, however the PC scores of Aramendi et al.'s (2017a) 3D 17-landmark model proved to be problematic due to the high number of PC scores generated through PCA (51 in total), and their small percentage of represented variance.

One of the advantages of ML techniques is the ability to evaluate model performance through examining the importance of each variable. In this case we used each model evaluation of variable importance to observe the classification power for each PC score. As expected, some PC scores are more important for classification than others. Reduction of the number of PC scores used can prune statistical noise while avoiding overfitting. Once the ML algorithms had identified the top performing PC scores for agent classification, the models were retuned using only these PC scores.

3. Results

3.1. ML differentiation of tooth scores

ML algorithms based on the information extracted from PC scores successfully construct powerful classification models (Tables S2 and S3). ML models in both shape and form space are relatively effective, though results tend to be more accurate after removing wolves from the analysis (Table 2). Additionally, score classification model rates are higher and more precise when using pure shape information, rather than when including centroid size in form space. At least 3 ML algorithms are able to reach 100% classification rates, with C5.0, RF and SVM being the best performing models for discerning carnivore agency via score cross-section morphology (Table 2). Nevertheless, sensitivity, specificity and accuracy values are high in most cases (Fig. 4), and ML performance is much more powerful than traditional GMM LDA models (Tables 3, 4, S4 and S5).

3.2. ML differentiation of tooth pits

Much like the case of tooth scores, 3D tooth pit models repeatedly produce up to 100% classification rates (Fig. 5), with SVM, RF, C5.0, KNN, MDA and PLSDA providing the highest degree of precision

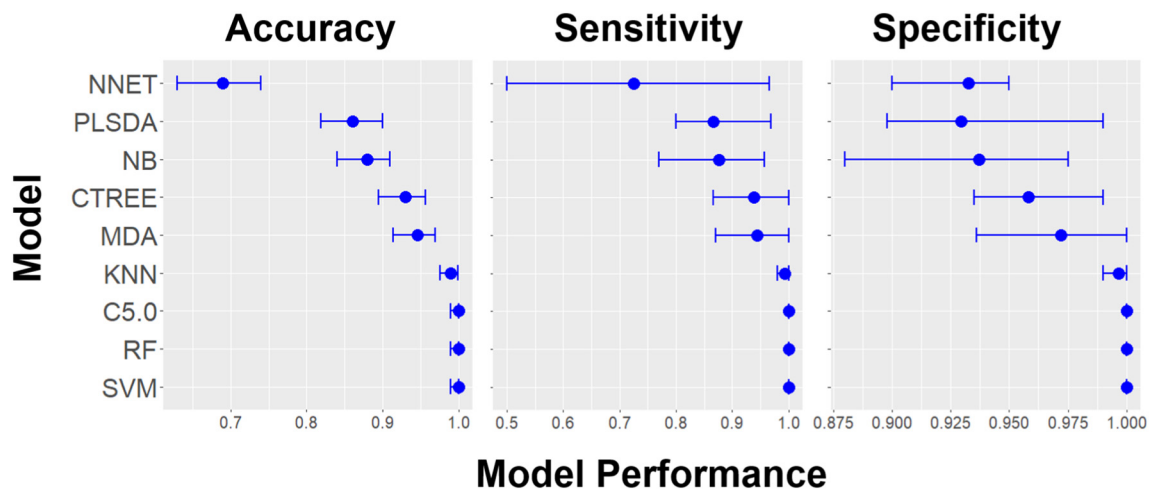


Fig. 4. Model performance data for machine learning algorithms when processing and classifying tooth score morphology.

Table 3

LDA performance comparing Lion, Hyena, Jaguar and Wolf tooth marks in both shape and form space for both GMM models. Specificity and sensitivity values are averaged, for full details consult Table S1.

| | Performance | | 95% confidence interval | | | |
|----------|-------------|----------|-------------------------|-------|-------|-------|
| | Kappa | Accuracy | Lower | Upper | Spec. | Sens. |
| 2D shape | 0.32 | 0.50 | 0.41 | 0.59 | 0.73 | 0.60 |
| 2D form | 0.35 | 0.51 | 0.42 | 0.60 | 0.84 | 0.51 |
| 3D shape | 0.13 | 0.35 | 0.25 | 0.46 | 0.78 | 0.34 |
| 3D form | 0.16 | 0.37 | 0.27 | 0.48 | 0.79 | 0.37 |

Table 4

LDA performance comparing only Lion, Hyena and Jaguar tooth marks. Scores are studied in shape space while pits are studied in form space. Specificity and sensitivity values are averaged, for full details consult Table S2.

| | Model performance | | 95% confidence interval | | | |
|--------|-------------------|----------|-------------------------|-------|-------|-------|
| | Kappa | Accuracy | Lower | Upper | Spec. | Sens. |
| Scores | 0.46 | 0.64 | 0.54 | 0.73 | 0.81 | 0.48 |
| Pits | 0.14 | 0.43 | 0.31 | 0.56 | 0.70 | 0.52 |

(Table 5). But contrary to 2D cross-section profiles, higher classification models are generated when combining shape and centroid size in form space (Tables S6 and S7).

The analysis including all 4 carnivore species (spotted hyenas, lions, jaguars and wolves) based on Aramendi et al.'s (2017a) model provides relatively low classification results (Tables S6 and S7). For this study, 4 consequent tuning tests were carried out on PC scores using ML algorithms. The first test includes all PC scores; the second model tuning includes the top 10 PC scores identified through test 1; the third tuning takes in the top 15 PC scores; and the final tuning procedure includes the top 20 PC scores. All tables can be consulted in Table S8. Model performance was then re-evaluated for each testing cycle (Fig. 6). According to the results, the inclusion of at least the 15 first PC scores obtains the best performance accuracy for discerning carnivore agency through tooth pit morphology.

Nevertheless, here again, the exclusion of wolf tooth pits from the analysis generates a much more powerful model without the need for PC tuning (Table 5, Fig. 6).

4. Discussion

Initial GMM studies of both scores and pits originally produced significant differences between most carnivore species (Arriaza et al., 2017; Yravedra et al., 2017; Aramendi et al., 2017a). Such differences are reflected in significant p values ($p < 0.05$) obtained via MANOVA

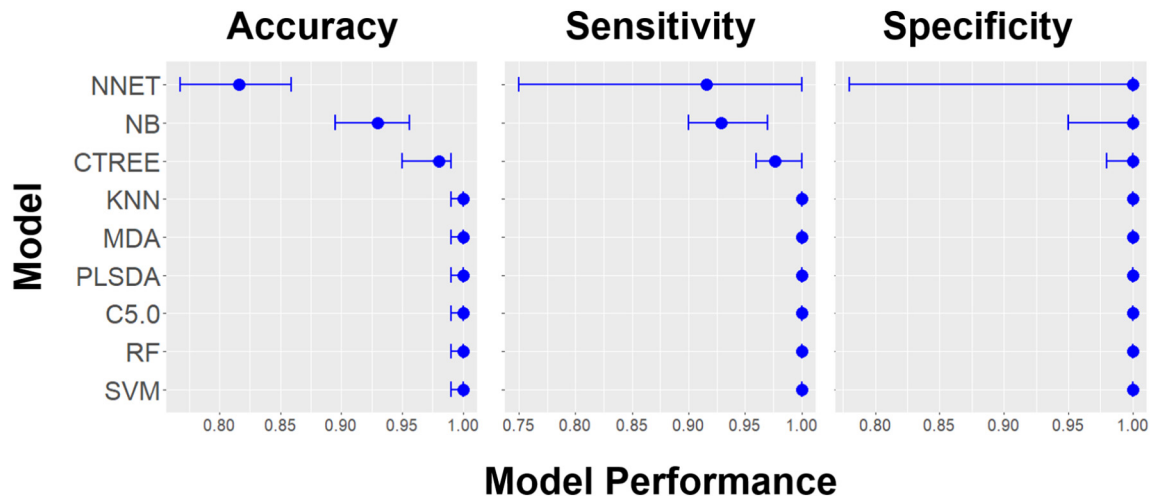


Fig. 5. Model performance data for machine learning algorithms when processing and classifying tooth pit morphology.

Table 5

Comparison of different ML algorithm performance used for discerning carnivore agency (excluding wolves) through [Aramendi et al. \(2017a\)](#) 3D 17-landmark model in Form space.

| Model | Model performance | | 95% confidence interval | | Hyena | | Jaguar | | Lion | |
|-------|-------------------|----------|-------------------------|-------|-------|-------|--------|-------|-------|-------|
| | Kappa | Accuracy | Lower | Upper | Spec. | Sens. | Spec. | Sens. | Spec. | Sens. |
| NNET | 0.73 | 0.82 | 0.77 | 0.86 | 1 | 0.75 | 0.78 | 1 | 0.82 | 1 |
| NB | 0.90 | 0.93 | 0.90 | 0.96 | 0.95 | 0.90 | 1 | 0.92 | 0.95 | 0.97 |
| CTREE | 0.97 | 0.98 | 0.95 | 0.99 | 1 | 1 | 0.98 | 0.96 | 0.98 | 0.97 |
| PLSDA | 1 | 1 | 0.99 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| MDA | 1 | 1 | 0.99 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| KNN | 1 | 1 | 0.99 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| RF | 1 | 1 | 0.99 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| C5.0 | 1 | 1 | 0.99 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| SVM | 1 | 1 | 0.99 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

tests and a relatively clear separation of groups in the graphs and the distances computed after Canonical Variance Analysis (CVA). Regardless, in these original studies, PCAs still presented significant overlapping in the majority of cases. LDA results were usually able to achieve relatively high classification rates ([Aramendi et al., 2017a](#); [Arriaza et al., 2017](#); [Yravedra et al., 2017](#)), however the efficiency and performance of LDA as classification method is not as powerful as desirable ([Tables 3 and 4](#)). Nevertheless, the exclusion of wolves from the analysis appears to clear up some confusion ([Table 2](#)), as previously

observed in other studies ([Yravedra et al., 2018](#)). In sum, GMM results *alone* show that the 2D analysis of tooth scores produce better results when performed in shape space, while the 3D analysis of tooth pits is more accurate when using form data ([Fig. 7](#)).

In this study a combined use of ML algorithms and taphonomic GMM models has been able to provide up to 100% classification when discerning between carnivore agencies. NNET and NB are generally the worst performing models, while SVM, C5.0 and RF consistently perform much better reaching exceptional results (Kappa = 1, Accuracy = 1).

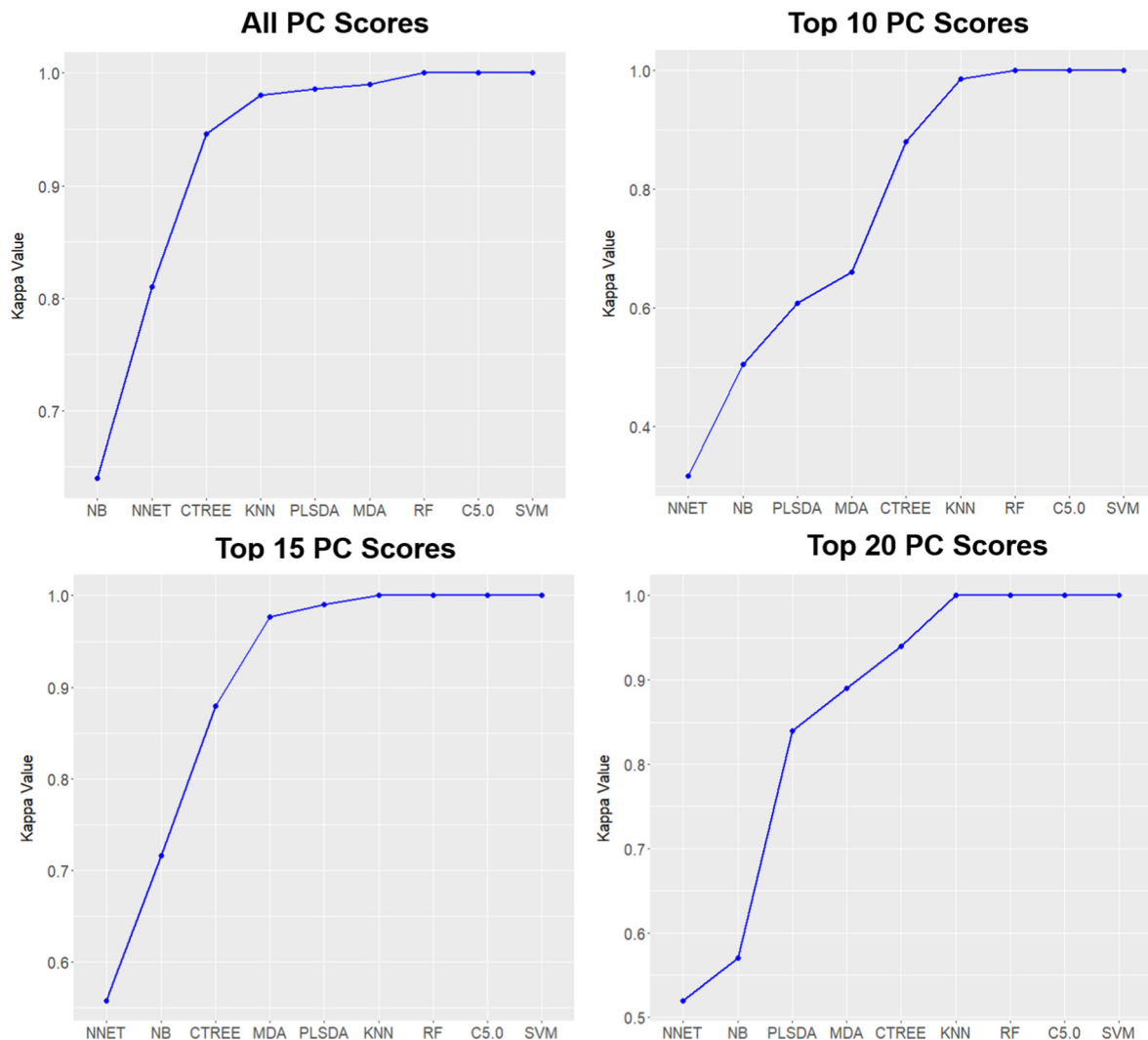


Fig. 6. Kappa value curves for each machine learning algorithm after each stage of the PC score tuning of [Aramendi et al. \(2017a\)](#) 3D seventeen landmark model.

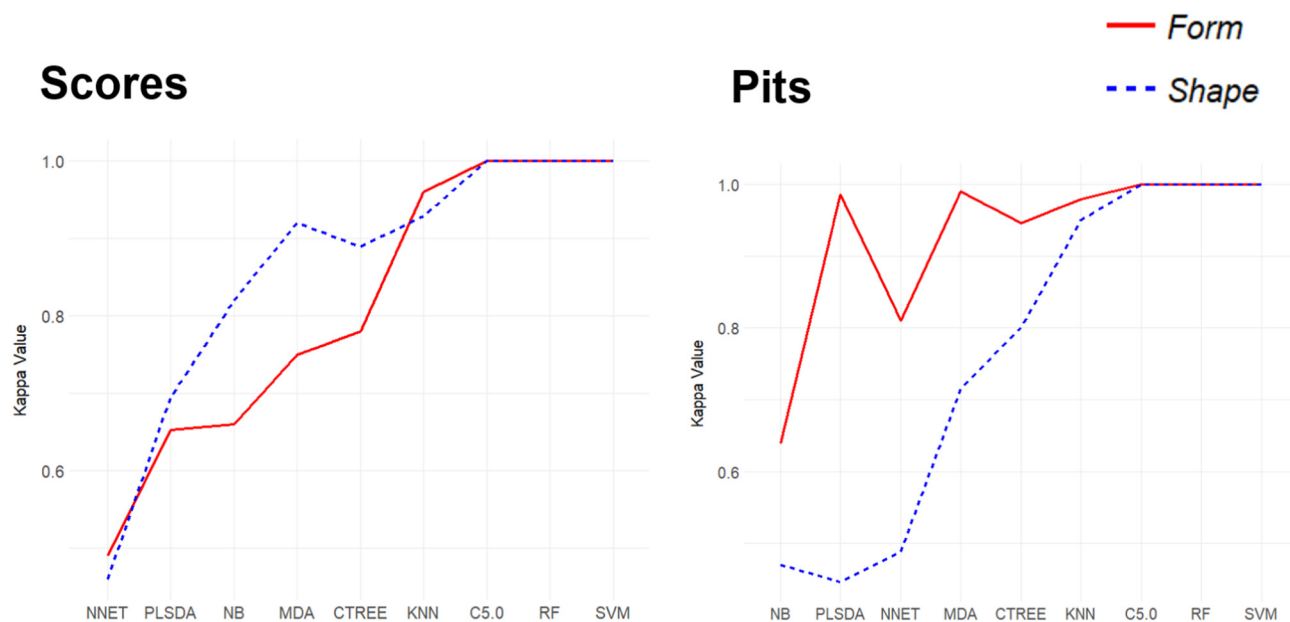


Fig. 7. Kappa value curves for machine learning algorithm performance when processing both pits and scores in shape space and form space.

CTREE, PLSDA, MDA and KNN are inconsistent, only providing high classification rates in some cases. As seen in the tuning of the 3D 17-landmark model, however, future use of ML and GMM based methods in classifying unknown taphonomic traces should take into account the possible effect that PC score selection could have on the quality of ML based classifications. Some algorithms, for example, were unable to provide 100% classification prior to tuning. The process of tuning, however, was able to highlight the potential of ML based methods on multiple accounts.

The poor performance of some models can be argued by the complexity of their internal architecture. In the case of NNET, if a more developed model is constructed using optimised hyperparameters, such as those typical of Deep Learning (Chollet and Alaire, 2017), these refined Deep Neural Networks are more likely to produce higher classification results. The optimisation of a Deep Learning model that confronts GMM data, therefore, would be a useful development for the future.

In general, ML based differentiation of carnivore agency through

Aramendi et al.'s (2017a) 3D 17-landmark model provides the best results (Fig. 8). Nevertheless, Yravedra et al.'s (2017) 2D 7 landmark model is still capable of producing powerful classification models. Most models achieve Kappa values above the margin of 0.8, thus, they can be considered powerful predictive models (Lantz, 2013). According to our results, tooth scores should be preferably analyzed on shape data by means of at least RF, C5.0 and SVM algorithms. In the case of tooth pits, the analysis should be rather performed in form space, however, care should be taken when dealing with a large number of PC scores. Tuning of models is highly recommended when dealing with great overlapping samples in PCAs.

These new statistical applications may be able to provide a deeper understanding of hominin-carnivore interactions, as considered to be a critical component in human evolutionary studies. The importance of understanding hominin-carnivore reactions is two-fold. Firstly, because carnivores may be responsible for bone accumulations and taphonomists must be able to differentiate between anthropogenic and carnivore-accumulated bone assemblages. Secondly, because some key

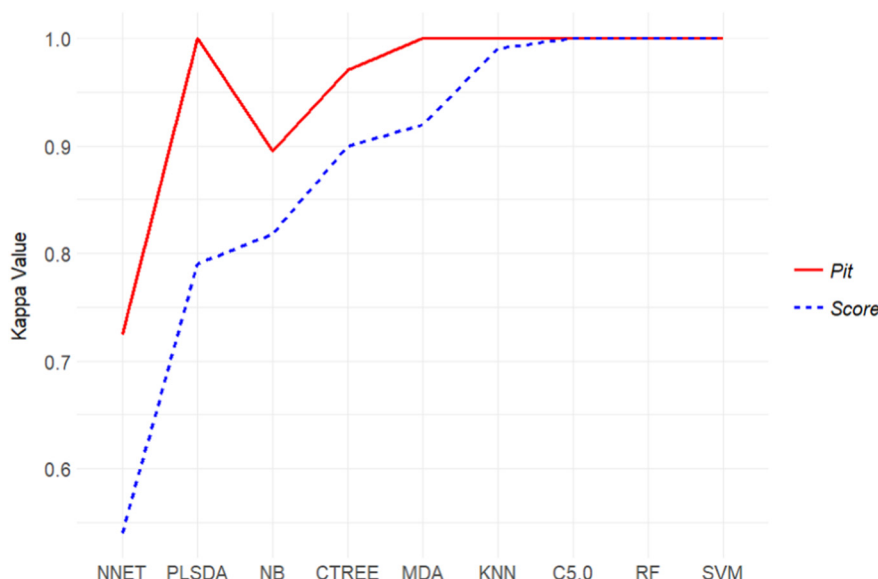


Fig. 8. Final kappa value curves for machine learning algorithm performance when processing pits in form space and scores in shape space.

paleoanthropological sites have been considered palimpsests in which both humans and carnivores may have played an important role in bone modification, such as those from Olduvai Gorge (Domínguez-Rodrigo et al., 2007). The improvement of classical BSM variables is therefore of great importance to help discerning the accumulating agents of bone assemblages at archaeological sites.

From this perspective, the application of GMM and micro-photogrammetric techniques in studying BSMs have helped to avoid some shortcomings of previous taphonomic variables, such as the equifinality arisen from the study of tooth mark frequencies and dimensions (Blumenschine, 1988; Selvaggio and Wilder, 2001; Domínguez-Rodrigo and Piqueras, 2003; Andrés et al., 2012). While bidimensional studies of tooth marks have been successful in distinguishing different sized carnivores, within these groups of large and small animals, little differentiation is possible (Andrés et al., 2012). This equifinality is hard to overcome relying on these variables alone, and their sole use is not recommended. Combined with tooth mark frequencies, skeletal profiles, fracture patterns and taphotypes, better results have been achieved (Domínguez-Rodrigo et al., 2007; Pineda et al., 2015; Aramendi et al., 2017b; Saladié et al., 2017). This highlights the need for a combined use of multiple criteria when analysing an assemblage.

Recent applications of ML based techniques in the study of modern carnivore sites have efficiently addressed some of these questions (Arriaza and Domínguez-Rodrigo, 2016), considering multiple traditional variables. Alongside these approaches, very recently these algorithms have been applied to multiple BSMs qualitative criteria (Domínguez-Rodrigo, 2018; Domínguez-Rodrigo and Baquedano, 2018)

However, the introduction of ML in archaeology occurred much later than in other sciences (Michalski et al., 1983). In other fields, the development of ML algorithms as part of the field of artificial intelligence have provided a much more efficient means of handling large amounts of raw data, using a more structured means of processing these data sets (Lantz, 2013; Paterson and Gibson, 2017). An additional advantage presented by ML is its versatility, using tuning functions to reach higher degrees of accuracy based on performance (Wolpert, 1996). With the inclusion of Computer Vision and Deep Learning techniques (Chollet and Alaire, 2017; Paterson and Gibson, 2017), the possibilities artificial intelligence may provide for future archaeological studies are ever-growing.

A certain degree of caution is required when assuming ML is the absolute solution to eliminating equifinality, however. A problematic concept that requires further confrontation lies in data collection, and the objectivity behind obtaining such information. The development of GMM over the years has come to provide increasingly more objective means of processing features of shape and form, however the most reliable and efficient means of obtaining the 3D models used is still under investigation (Maté-González et al., 2017b, c; Courtenay et al., 2017, Under review). Likewise, comparisons between different landmark models are still pending further investigation. Nevertheless, as presented here, combining both GMM and ML approaches is still able to provide interesting results. From this perspective the benefits of a hybrid ML and GMM approach to taphonomy could produce an important development in BSM analysis.

5. Conclusions

This paper presents the first combination of advanced statistical analysis using GMM and ML methods to discern carnivore agency through different types of tooth marks. The future of ML and GMM based methods is promising. Our results provide the first combined inclusion of both statistical techniques, achieving 100% classification rates in multiple models. So far, this technique has only been applied to determine carnivore agency, but can be expanded to include other GMM models.

The possibility provided by the combination of ML and GMM based techniques could be of utmost relevance when applied to Lower

Pleistocene sites, where the interaction of several modifying agents is still subject of discussion. Though the comparative sample still needs to be increased, adding more carnivore species and tooth marks, the methodological approach described throughout this paper opens up the door to new work lines in taphonomy. Furthermore, the development of this line of research could also shed light into certain debates regarding early human behaviour.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.palaeo.2019.03.007>.

Acknowledgements

Firstly, we would like to thank all the staff and members of both the IPHES and the Rovira i Virgili University. Along these lines, we would also like to acknowledge the staff at the Parque de la Naturaleza de Cabárceno and Santiago Borrigan for providing the samples used in our experiments. Finally we would like to thank Abel Moclán for his advice and comments during the process of writing this paper.

Additionally, we would like to thank the TIDOP Group from the Department of Cartographic and Land Engineering of the Higher Polytechnics School of Avila, University of Salamanca, for the use of their tools and facilities. We want to recognize the technical support provided by C.A.I. Arqueometry and Archaeological Analysis from Complutense University which has been very useful in carrying out the present paper. J.A. would like to thank Fundación La Caixa and the Spanish Education, Culture and Sports Ministry (FPU15/04585) for funding her postgraduate education program. The support of the DST-NRF Centre of Excellence in Palaeosciences (CoE-Pal) towards this research is also acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the CoE. Finally, this work was developed within the general framework of the Spanish MINECO-FEDER project CGL2015-65387-C3-1-P, the Catalan AGAUR project 2017-SGR-1040, and the URV project 2017-PFR-URV-B2-91.

References

- Albrecht, G.H., 1992. Assessing the affinities of fossils using canonical variates and generalized distances. *Hum. Evol.* 7 (4), 49–69.
- Alcántara, V.G., Egido, R.B., del Pino, J.M.B., Ruiz, A.B.C., Vidal, A.I.E., Aparicio, Á.F., Calleja, S.H., Jiménez, A.I., González, M.M., Gil, M.P., Tello, V.P., Calvo, J.R., Yravedra, J., Vidal, A.S., Domínguez-Rodrigo, M., 2006. Determinación de procesos de fractura sobre huesos frescos: un sistema de análisis de los ángulos de los planos de fracturación como discriminador de agentes bióticos. *Trab. Prehist.* 63 (1), 37–45.
- Andrés, M., Gidna, A.O., Yravedra, J., Domínguez-Rodrigo, J., 2012. A Study of Dimensional differences of tooth marks (pits and scores) on bones modified by small and large carnivores. *Journal of Archaeological Anthropological Sciences.* 4, 209–219.
- Aramendi, J., Maté-González, M.Á., Yravedra, J., Ortega, M.C., Arriaza, M.C., González-Aguilera, D., Baquedano, E., Domínguez-Rodrigo, M., 2017a. Discerning carnivore agency through the three-dimensional study of tooth pits: Revisiting crocodile feeding behaviour at FLK-Zinj and FLK NN3 (Olduvai Gorge, Tanzania). *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 488, 93–102.
- Aramendi, J., Uribealrrea, D., Arriaza, M.C., Arráiz, H., Barboni, D., Yravedra, J., Cruz Ortega, M., Gidna, A., Mabulla, A., Baquedano, E., Domínguez-Rodrigo, M., 2017b. The Paleocology and Taphonomy of AMK (Bed I, Olduvai Gorge) and its Contributions to the Understanding of the “Zinj” Paleolandscape, *Paleogeography, Paleoclimatology. Paleocology.* 488, 35–49.
- Arriaza, M.C., Domínguez-Rodrigo, M., 2016. When Felids and Hominins Ruled at Olduvai Gorge: a Machine Learning Analysis of the Skeletal Profiles of the Non-Anthropogenic Bed I Sites. *Quat. Sci. Rev.* 139, 43–52.
- Arriaza, M.C., Domínguez-Rodrigo, M., Yravedra, J., Baquedano, E., 2016. Lions as bone accumulators? Paleontological and ecological implications of a modern bone assemblage from Olduvai Gorge. *PLoS One* 11 (5), e0153797.
- Arriaza, M.C., Yravedra, J., Domínguez-Rodrigo, M., Maté-González, M.Á., Vargas, E.G., Palomeque-González, J.F., Aramendi, J., González-Aguilera, D., Baquedano, E., 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.
- Arriaza, M.C., Aramendi, J., Maté-González, M.Á., Yravedra, J., Baquedano, E., González-Aguilera, D., Domínguez-Rodrigo, M., 2018. Geometric morphometric analysis of tooth pits and identification of felid and hyenid agency in Bone modification. *Quat. Int.* <https://doi.org/10.1016/j.quaint.2018.11.023>.
- Bartram, L.E., Marean, C.W., 1999. Explaining the “Klasies Pattern”: Kua

- Ethnoarchaeology, the Die Kelders Middle Stone Age Archaeofauna. long bone fragmentation and carnivore ravaging, *Journal of Archaeological Science*. 26, 9–29.
- Binford, L.R., 1981. *Bones: Ancient Men and Modern Myths*. Academic Press.
- Blumenshine, R., 1988. An experimental model of the timing of hominid and carnivore influence on archaeological bone assemblages. *J. Archaeol. Sci.* 15, 483–502.
- Blumenshine, R., 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 (1), 21–51.
- Blumenshine, R., Marean, C.W., 1993. Long bone marrow yields of Some African ungulates. *J. Archaeol. Sci.* 20, 555–587.
- Bookstein, F., 1991. *Morphometric Tools for Landmark Data: Geometry and Biology*. In: New York. Press, Cambridge University.
- Brain, C.K., 1967. Hottentot food remains and their bearing on the interpretation of fossil bone assemblages. *Scientific Papers of the Namib Desert Research Station* 32, 1–11.
- Brain, C.K., 1969. The contribution of Namib Desert Hottentots to an understanding of Australopithecine bone accumulations. *Scientific Papers of the Namib Desert Research Station* 39, 13–22.
- Brain, C.K., 1981. The hunter or the hunted. An Introduction to African Cave Taphonomy. The University of Chicago press, Chicago.
- Bunn, H.T., 1981. Archaeological evidence for meat eating by Plio-Pleistocene hominids from Koobi Fora and Olduvai Gorge. *Nature*. 291, 574–577.
- Bunn, H.T., Pickering, T.R., 2010. Methodological recommendations for ungulate mortality analyses in paleoanthropology. *Quat. Res.* 74, 388–394.
- Capaldo, S., 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.
- Capaldo, S.D., Blumenshine, R.J., 1994. A quantitative diagnosis of notches made by hammerstone percussion and carnivore gnawing in bovid long bones. *Am. Antiq.* 59, 724–748.
- Chollet, F., Allaire, J.J., 2017. *Deep Learning with R*. Manning Publications, New York.
- Core-Team, 2015. A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. <https://www.Rproject.org/>, 2015. Accessed the 14th of February, 2018.
- Cortes, C., Vapnik, V., 1995. Support-Vector Networks. *Mach. Learn.* 20, 273–297.
- Courtenay, L.A., Yravedra, J., Maté-González, M.Á., Aramendi, J., González-Aguilera, D., 2017. 3D analysis of cut marks using a new geometric morphometric methodological approach. *Journal of Archaeological and Anthropological Sciences*. <https://doi.org/10.1007/s12520-017-0554-x>.
- Courtenay, L.A., Yravedra, J., Huguete, R., Ollé, A., Aramendi, J., Maté-González, M.Á., González-Aguilera, D., 2019. New taphonomic advances in 3D digital microscopy: a morphological characterisation of trampling marks. *Quat. Int.* <https://doi.org/10.1016/j.quaint.2018.12.019>. (under review).
- Cruz-Urbe, K., 1991. Distinguishing hyena from hominid bone accumulations. *J. Field Archaeol.* 18 (4), 467–486.
- Dart, R.A., 1957. The Osteodontokeratic culture of *Australopithecus africanus*. *Transvaal Museum Memoirs* 10, 1–105.
- Dawkins, W.B., 1863. Wookey Hole hyaena den. *Proceedings of the Somersetshire Archaeology and Natural History Society*. 11 (2), 197–219.
- Delaney-Rivera, C., Plummer, T.W., Hodgson, J.A., Forrest, F., Hertel, F., Oliver, J.S., 2009. Pits and pitfalls: taxonomic variability and patterning in tooth mark dimensions. *J. Archaeol. Sci.* 36, 2597–2608.
- Domínguez-Rodrigo, M., 1999. Flesh availability and bone modifications in carcasses consumed by lions: Palaeoecological relevance in hominid foraging patterns. *Palaeo* 2189. *Palaeogeography, Palaeoclimatology, Palaeoecology*. 149, 373–388.
- Domínguez-Rodrigo, M., 2001. A study of carnivore competition in riparian and open habitats of modern Savannas and its implications for hominid behavioral modelling. *J. Hum. Evol.* 40, 77–98.
- Domínguez-Rodrigo, M., 2002. Hunting and scavenging by early humans: the state of the debate. *J. World Prehist.* 16, 1–54.
- Domínguez-Rodrigo, M., 2018. Successful Classification of Experimental Bone Surface Modifications (BSM) through Machine Learning. A Solution to the Controversial Use of BSM in Paleanthropology, Archaeological and Anthropological Sciences, Algorithms. <https://doi.org/10.1007/s12520-018-0684-9>.
- Domínguez-Rodrigo, M., Baquedano, E., 2018. Distinguishing butchery cut marks from crocodile bite marks through machine learning methods. *Sci. Rep.* <https://doi.org/10.1038/s41598-018-24071-1>.
- Domínguez-Rodrigo, M., Pickering, T.R., 2010. Un Estudio Tafonómico Multivariante de las Acumulaciones de Fauna de Hienidos (*Crocota crocuta*) y Félidos (*Panthera pardus*). 1º Reunión de Científicos Sobre Cubiles de Hiena (y otros grandes carnívoros) en los Yacimientos Arqueológicos de la Península Ibérica 45–60.
- Domínguez-Rodrigo, M., Piqueras, A., 2003. The use of tooth pits to identify carnivore taxa in tooth-marked Archaeofaunas and their relevance to reconstruct hominid carcass processing behaviours. *J. Archaeol. Sci.* 30 (11), 1385–1391.
- Domínguez-Rodrigo, M., Barba, R., Egeland, C.P., 2007. *Deconstructing Olduvai*. Springer, The Netherlands.
- Domínguez-Rodrigo, M., Gidna, A.O., Yravedra, J., Musiba, C., 2012. A comparative neo-taphonomic study of felids, hyaenids and canids: an analogical framework based on long bone modification patterns. *Journal of Taphonomy*. 10 (3), 147–164.
- Domínguez-Rodrigo, M., Yravedra, J., Oganista, E., Gidna, A., Fourvel, J.B., Baquedano, E., 2015. A new methodology approach to the taphonomic study of paleontological and archaeological faunal assemblages: a preliminary case study from Olduvai Gorge (Tanzania). *J. Archaeol. Sci.* 59, 35–53.
- Dryden, I.L., Mardia, K.V., 1998. *Statistical Shape Analysis*. John Wiley & Sons, Chichester.
- Egeland, A.G., Egeland, C.P., Bunn, H.T., 2008. Taphonomic analysis of a modern spotted hyena (*Crocota crocuta*) den from Nairobi, Kenya. *Journal of Taphonomy*. 6 (3–4), 275–299.
- Faith, J.T., Marean, C.W., Behrensmeier, A.K., 2007. Carnivore competition, bone destruction, and bone density. *J. Archaeol. Sci.* 34 (12), 2025–2034.
- Fawcett, T., 2006. An Introduction to ROC analysis. *Pattern Recogn. Lett.* 27, 861–874.
- Galán, A.B., Rodríguez, M., de Juana, S., 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.
- Gidna, A., Yravedra, J., Domínguez-Rodrigo, M., 2013. A cautionary note on the use of captive carnivores to model wild predator behavior: a comparison of bone modification patterns on long bones by captive and wild lions. *J. Archaeol. Sci.* 40, 1903–1910.
- Gidna, A.O., Kisui, B., Mabulla, A., Musiba, C., Domínguez-Rodrigo, M., 2014. An ecological neo-taphonomic study of carcass consumption by lions in Tarangire National Park (Tanzania) and its relevance for human evolutionary biology. *Quat. Int.* 322–323, 167–180.
- González-Aguilera, D., López Fernández, L., Rodríguez-González, P., Guerrero, D., Hernandez-Lopez, D., Remondino, F., Menna, F., Nocerino, E., Toschi, I., Ballabeni, A., Gaiani, M., 2016. Development of an all-purpose free photogrammetric tool. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Conference.
- Günther, F., Fritsch, S., 2010. neuralnet. training of neural networks, *The R Journal*. 2 (1), 30–38.
- Haynes, G., 1982. Utilization and Skeletal Disturbances of North American Prey Carcasses. *Artic* 35, 266–281.
- Haynes, G., 1983a. Frequencies of spiral and green-bone fractures on ungulate limb bones in modern surface assemblages. *Am. Antiq.* 48, 102–114.
- Haynes, G., 1983b. A guide for differentiating mammalian carnivore taxa responsible for gnaw damage to herbivore limb bones. *Paleobiology*. 9, 164–172.
- Houtson D.C., 1979. The adaptations of scavengers in A. R. E. Sinclair and M. Narton Griffiths eds 263–286.
- Kerbis-Peterhans, J.C., 1990. The Role of Porcupines, Leopards and Hyaenas in Ungulate Carcass Dispersal: Implications for Paleoanthropology. University of Chicago, Chicago, Department of Anthropology.
- Kuhn, M., Johnson, K., 2013. *Applied Predictive Modeling*. Springer, London.
- Lantz, B., 2013. *Machine Learning with R*. Packt Publishing Ltd., Birmingham.
- Lyman, R.L., 1994. *Vertebrate Taphonomy*. Cambridge University Press.
- Marean, C.W., Frey, C.J., 1997. Animal bones from caves to cities: reverse utility curves as methodological artifacts. *Am. Antiq.* 62 (4), 698–711.
- Martin, F.M., 2008. Bone crunching felids at the end of the Pleistocene in Fuego-Patagonia, Chile. *Journal of Taphonomy* 6 (3–4), 337–372.
- Maté-González, M.Á., Yravedra, J., González-Aguilera, D., Palomeque-González, J.F., Domínguez-Rodrigo, M., 2015. Microphotogrammetric characterization of cut marks on bones. *J. Archaeol. Sci.* 62, 128–142.
- Maté-González, M.Á., Palomeque-González, J.F., Yravedra, J., González-Aguilera, D., Domínguez-Rodrigo, M., 2016. Micro-photogrammetric and morphometric differentiation of cut marks on bones using metal knives, quartzite and flint flakes. *Journal of Archaeological and Anthropological Science*. <https://doi.org/10.1007/s12520-016-0401-5>.
- Maté-González, M.Á., Yravedra, J., Martín-Perea, D., Palomeque-González, J., San-Juan-Blazquez, M., Estaca-Gómez, V., Uribealarea, D., Álvarez-Alonso, D., Cuartero, F., González-Aguilera, D., Domínguez-Rodrigo, M., 2017a. Flint and quartzite: distinguishing raw material through bone cut marks. *Archaeometry*. <https://doi.org/10.1111/arc.12327>.
- Maté-González, M.Á., Aramendi, J., Yravedra, J., Blasco, R., Rosell, J., González-Aguilera, D., Domínguez-Rodrigo, M., 2017b. Assessment of statistical agreement of three techniques for the study of cut marks: 3D Digital Microscope, Laser Scanning Confocal Microscopy and Micro-Photogrammetry. *J. Microsc.* 267 (3), 356–370. <https://doi.org/10.1111/jmi.12575>.
- Maté-González, M.Á., Aramendi, J., Yravedra, J., González-Aguilera, D., 2017c. Statistical comparison between low-cost methods for 3D characterization of cut-marks on bones. *Remote Sens.* <https://doi.org/10.3390/rs9090873>.
- Mevik, B.H., Wehrens, R., 2007. The pls package: principal component and partial least squares regression in R. *J. Stat. Softw.* 18 (2), 1–23.
- Michalski, R.S., Carbonell, J.G., Mitchell, T.M., 1983. *Machine Learning: An Artificial Intelligence Approach*, Volume I. Morgan Kaufman Publishers Inc., USA.
- Mitteroecker, P., Gunz, P., 2009. Advances in geometric morphometrics. *Evol. Biol.* 36, 235–247.
- Moclán, A., Domínguez-Rodrigo, M., 2018. An experimental study of the patterned nature of anthropogenic bone breakage and its impact on bone surface modification frequencies. *J. Archaeol. Sci.* 96, 1–13.
- Moclán, A., Domínguez-Rodrigo, M., 2019. Classifying Bone Breakage Patterns: an Experimental Analysis of Fracture Planes to Discern Between Hominin and Carnivore Activity using Machine Learning (ML) Algorithms (accepted).
- Paterson, J., Gibson, A., 2017. *Deep Learning: A Practitioner's Approach*. O'Reilly Media Inc, California.
- Pickering, T.R., 2002. Reconsideration of Criteria for Differentiating Faunal Assemblages Accumulated by Hyenas and Hominids. *Int. J. Osteoarchaeol.* 12, 127–141.
- Pickering, T.R., Bunn, H.T., 2007. The endurance running hypothesis and hunting and scavenging in Savanna Woodlands. *J. Hum. Evol.* 53, 434–438.
- Pickering, T.R., Egeland, C.P., 2006. Experimental patterns of hammerstone percussion damage on bones: implications for inferences of carcass processing by humans. *J. Archaeol. Sci.* 33, 459–469.
- Pickering, T.R., Marean, C.W., Domínguez-Rodrigo, M., 2003. Importance of Limb Bone Shaft fragments in zooarchaeology: a response to “on in situ attrition and vertebrate body part profiles” (2002), by M. C. Stiner, *Journal of Archaeological Science*. 30 (11), 1469–1482.

- Pickering, T.R., Domínguez-Rodrigo, M., Egeland, C.P., Brain, C.K., 2005. The contribution of limb bone fracture patterns to reconstructing early hominid behaviour at Swartkrans cave (South Africa): archaeological application of a new analytical method. *Int. J. Osteoarchaeol.* 15, 247–260.
- Pickering, T.R., Heaton, J.L., Zwodeski, S.E., Kuman, K., 2011. Taphonomy of bones from baboons killed and eaten by wild leopards in Mapungubwe National Park, South Africa. *Journal of Taphonomy* 9 (2), 117–159.
- Pineda, A., Saladié, P., Huguer, R., Cáceres, I., Rosas, A., García-Tabernero, A., Estalrich, A., Mosquera, M., Ollé, A., Vallverdú, J., 2015. Coexistence among large predators during the Lower Paleolithic at the Site of La Mina (Barranc de la Boella, Tarragona, Spain). *Quat. Int.* 388, 177–187.
- Rohlf, F.J., 1999. Shape statistics: procrustes superimpositions and tangent spaces. *J. Classif.* 16 (2), 197–223.
- Ruiter, D.J., Berger, L.R., 2000. Leopards as taphonomic agents in dolomitic caves – implications for bone accumulations in the hominid-bearing deposits of South Africa. *J. Archaeol. Sci.* 27 (8), 665–684.
- Saladié, P., Fernández, P., Rodríguez-Hidalgo, A., Huguet, R., Pineda, A., Cáceres, I., Marín, J., Vallverdú, J., Carbonell, E., 2017. The TD63.3 Faunal Assemblage of the Gran Dolina site (Atapuerca, Spain): a Late Early Pleistocene Hyena Den. DOI: <https://doi.org/10.1080/08912963.2017.1384476>.
- Schaller, G.B., Lowther, G.R., 1969. The relevance of carnivore behavior to the study of early hominids. *South W. Journal of Anthropology* 25, 307–341.
- Selvaggio, M.M., 1994. Identifying the Timing and Sequence of Hominid and Carnivore Involvement with Plio-Pleistocene Bone Assemblages from Carnivore Tooth Marks and Stone Tool Butchery Marks on Bone Surfaces. PhD Thesis. New Brunswick: Rutgers University.
- Selvaggio, M.M., Wilder, J., 2001. Identifying the involvement of multiple carnivore taxon with archaeological bone assemblages. *J. Archaeol. Sci.* 28, 465–470.
- Slice, D.E., 2001. Landmark coordinates aligned by procrustes analysis do not lie in Kendall's shape space. *Syst. Biol.* 50 (1), 141–149.
- Stanford, C.B., Bunn, H.T., 2001. *Meat Eating and Human Evolution*. Oxford University Press, New York.
- Stiner, M., 1994. *Honor Among Thieves: A Zooarchaeological Study of Neandertal Ecology*. Princeton University press, Princeton.
- Stiner, M.C., 2004. Comparative ecology and taphonomy of spotted hyenas, humans, and wolves in Pleistocene Italy. *Rev. Paleobiologie* 23, 771–785.
- Sutcliffe, A.J., 1970. Spotted hyena: Crusher, gnawer, digester and collector of bones. *Nature* 227, 110–1113.
- Wei, Y., Chiu, D., 2015. *Machine Learning with R Cookbook*. Packt Publishing, Birmingham.
- Wiley, P., Snyder, L.M., 1989. Canid modification of human remains: implications for time-since-death estimations. *J. Forensic Sci.* 34, 894–901.
- Wolpert, D.H., 1996. The existence of a priori distinctions between learning algorithms. *Neural Comput.* 8, 1391–1420.
- Yravedra, J., Lagos, L., Bárcena, F., 2011. A taphonomic study of wild wolf (*Canis lupus*) modification of horse bones in northwestern Spain. *Journal of Taphonomy*. 9 (1), 37–65.
- Yravedra, J., Lagos, L., Bárcena, F., 2012. The wild wolf (*Canis lupus*) as a dispersal agent of animal carcasses in Northwestern Spain. *J. Taphon.* 10, 219–238.
- Yravedra, J., García-Vargas, E., Maté-González, M.Á., Aramendi, J., Palomeque-González, J.F., Vallés-Iriso, J., Matesanz-Vicente, J., González-Aguilera, D., Domínguez-Rodrigo, M., 2017. The use of micro-photogrammetry and geometric morphometrics for identifying carnivore agency in bone assemblages. *J. Archaeol. Sci. Rep.* 14, 106–115.
- Yravedra, J., Aramendi, J., Maté-González, M.Á., Courtenay, L.A., González-Aguilera, D., 2018. Differentiating percussion pits and carnivore tooth pits using 3D reconstruction and geometric morphometrics. *PLoS One*. <https://doi.org/10.1371/journal.pone.0194324>.