

# Machine learning approaches to dietary classification from dental microtexture in primates

Received: 15 September 2025

Accepted: 31 March 2026

Published online: 28 April 2026

Cite this article as: Estebanz-Sánchez F., Kit K., Ibáñez Estevez J.J. *et al.* Machine learning approaches to dietary classification from dental microtexture in primates. *Sci Rep* (2026). <https://doi.org/10.1038/s41598-026-47350-8>

Ferran Estebanz-Sánchez, Kristina Kit, Juan José Ibáñez Estevez, David R. Insua, Simón Rodríguez Santana & Laura M. Martínez

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

ARTICLE IN PRESS

# Machine Learning Approaches to Dietary Classification from Dental Microtexture in Primates

Ferran Estebarez-Sánchez<sup>1</sup>, Kristina Kit<sup>2</sup>, Juan José Ibáñez Estevez<sup>3,6</sup>, David R. Insua<sup>2</sup>, Simón Rodríguez Santana<sup>4,\*</sup>, and Laura M. Martínez<sup>5,†</sup>

<sup>1</sup>Archaeology of Social Dynamics, IMF-CSIC, Egipcíaques 15, 08001, Barcelona, Spain

<sup>2</sup>DataLab, Institute of Mathematical Sciences, ICMAT-CSIC, 28049, Madrid, Spain

<sup>3</sup>Cultural Landscape Research Group. Milà i Fontanals Institution. IMF-CSIC, 08001, Barcelona, Spain

<sup>4</sup>ICAI Engineering School, Universidad Pontificia de Comillas - IIT, 28015, Madrid, Spain

<sup>5</sup>Departament de Biologia Evolutiva, Ecologia i CC Ambientals. Facultat de Biologia. Universitat de Barcelona, Institut d'Arqueologia de la Universitat de Barcelona, 08007, Barcelona, Spain

<sup>6</sup>Institución Milá y Fontanals de Investigación en Humanidades, IMF-CSIC, 08001, Barcelona, Spain

\*[srsantana@comillas.edu](mailto:srsantana@comillas.edu)

†[lmartinez@ub.edu](mailto:lmartinez@ub.edu)

## ABSTRACT

Dental microwear texture (DMT) analysis is a critical proxy for reconstructing the diets of extant and extinct mammals. While craniodental morphology reflects selective pressures across evolutionary timescales, microwear captures localized, short-term dietary signals over weeks to months. However, the high dimensionality of modern 3D surface microtexture data, often spanning disparate parameter sets (such as ISO standards and Scale-Sensitive Fractal Analysis, SSFA), complicates classification, particularly when working with limited paleontological datasets. To address this, we present a robust machine learning pipeline designed to automatically classify primate samples ( $N = 99$ ) across 6 dietary groups and 7 species. Our methodology leverages a nested Leave-One-Out Cross-Validation framework to evaluate multiple classifiers, including Multinomial Logistic Regression (MLR) with various regularization approaches, Naive Bayes, and tree-based ensemble algorithms (e.g., Random Forests, XGBoost). Our results demonstrate that Lasso-regularized MLR and Naive Bayes yield the highest predictive performance while enforcing strict feature selection to maintain interpretability. Crucially, models relying exclusively on ISO parameters consistently outperformed those using SSFA, as ISO variables better capture the microroughness localized mechanical abrasions generated by specific diets in our dataset. Furthermore, the integration of novel Fourier-based descriptors and isotropy variables significantly enhanced models' discriminating power. By providing a mathematically rigorous framework to isolate precise ecological signals from noisy, high-dimensional data, this approach enables more accurate and reproducible dietary classifications. Ultimately, refining these dietary reconstructions is essential for resolving broader questions regarding niche partitioning, species evolution, and paleoecological dynamics.

## 1 Introduction

The study of diet and feeding habits of primates, as well as other mammals, is crucial to understand their ecology and evolution [1, 2, 3, 4]. Dietary reconstruction is based on the analysis of different proxies tested in extant species with well-known dietary habits [5]. The selection of such proxies is far from trivial as they record diet across different temporal scales and can, therefore, provide contradictory palaeodietary information, resulting in different reconstructions [5, 6]. In this context, proxies based on the analysis of craniodental morphology provide information related to evolutionary processes (e.g. adaptation, speciation). In contrast, other proxies, such as 2D microwear (MW) or 3D microtexture (DMT) patterns, provide finer (palaeo)dietary information at seasonal or, even, weekly resolution [7, 8, 9]. This may explain, for example, the apparent contradiction between the craniodental morphology of some fossil species and their microwear data, e.g. in *Australopithecus anamensis* [10, 11] or in *Paranthropus* genus [12, 13]. Dental microwear and microtexture have thus become important proxies [5, 14, 15, 16, 17, 18] to infer the dietary habits of extant [19, 20, 21] and extinct primates and hominins [10, 11, 12, 13, 15] since they provide useful information within a relatively short time frame before the death of the individual, while it is a technique that can be effectively applied to fossils with no chronological limit.

Microwear/microtexture studies have different objectives, including (1): **Within-group variability analysis** (e.g. intraspecific, intrapopulation). These studies assess variations within a group (e.g. population, species), to identify how factors such as sex [22, 23, 24], age [21, 22, 24, 25, 26, 27], chronology or location [10, 12, 24, 28] may influence microwear and microtexture patterns. This approach is particularly useful when the aim is to characterize internal variability without necessarily

reconstructing a specific diet; (2) **Between-group variability of unknown diet groups** (e.g. interspecific, interpopulation, intergroup): This category includes comparisons among several groups whose diets are not well established. The objective is to identify relative differences in abrasion or texture between groups. This approach highlights ecological contrast (abrasive vs softer diets) between groups, even in the absence of a known dietary reference sample [29, 30, 31, 32]; (3) **Between-group variability of known diet groups** to determine whether microwear or microtexture analysis can significantly identify groups (e.g. populations or species) whose diets are known. For example, microwear and microtexture have been shown to differentiate various groups of recent hunter-gatherers and agriculturalists [20, 21], as well as cecopithecoid taxa with distinct feeding regimes [16]. This approach is crucial for building reference datasets, which subsequently underpin paleodietary interpretations. Reference microwear models both in Hominoidea and Cercopithecoidea [19, 33, 34] were developed using this strategy and have since been applied to the study of fossil hominins and other primates [10, 11, 12, 13, 17, 35]; (4) **Reconstruction of the palaeodiet of a group** (e.g. species, population) This approach directly compares both signatures of a fossil species or population with an appropriate reference sample of known diet. By evaluating their similarity researchers can infer the dietary regime of extinct individuals. This method is widely applied in palaeoanthropology and bioarchaeology and has proven effective in reconstructing the diets of fossil hominins, non-human primates, and past human populations [11, 15, 17, 22, 27, 36, 37, 38].

Numerous articles address more than one of the previously mentioned objectives, including intra and intergroup analysis of unknown dietary groups [39, 40], between-group variability of known diet groups and paleodietary reconstruction of a specific archaeological or paleontological group [36, 41], intraspecific analysis, reconstruction of the paleodiet of a specific group [10] or test the relation between microwear and isotopic signal [6, 17].

Systematic approaches based on statistical/machine learning techniques have been used in microwear/DMT studies, with the choice of the methodologies dependent on the number of variables. The buccal microwear pattern consists of 15 variables [22, 39], while occlusal (high-magnification) include fewer variables, typically between 3 and 6 [42, 43, 44, 45, 46, 47]. Therefore, the statistical strategy when analyzing buccal microwear patterns usually emphasizes reducing data dimensionality to represent the groups studied graphically and analyze their similarity. One of the most commonly used methods is Principal Component Analysis (PCA) [20, 26, 39, 48], which projects the data in directions resulting from the linear combination of the original variables [49]. This facilitates 2D plot representations of the original data in the directions of maximal information [50]. Alternatively, other studies have chosen Discriminant Analysis (DA) [10, 11, 13, 26] or Canonical Variate Analysis [15, 31, 51], which maximally separates the predefined groups of interest based on the data. In general, dimensionality reduction techniques are not typically used in occlusal microwear, at least for primate and hominin studies, the exception being [17] with *Theropithecus oswaldi*. However, in low-magnification occlusal studies with a similar number of variables, multidimensionality reduction techniques are commonly applied [52, 53, 54, 55, 56]. Ultimately, primate and hominin high-magnification occlusal microwear pattern variability is represented through simple biplots of pitting incidence vs. scratch width [42, 46, 47], pitting incidence vs. pit width [43] or number of scratches vs. number of pits [45].

The statistical/Machine Learning (ML) strategy for DMT studies also depends on the set of variables used (typically, scale-sensitive fractal analysis (SSFA) and dental surface texture analysis from the international standard (ISO 25178)) as each set size may vary among studies. Broadly, ML offers a data-driven framework capable of automatically identifying complex, high-dimensional patterns to classify specimens based on these multivariate textural signatures. Primate and hominin SSFA studies, conducted on both occlusal and buccal surfaces, typically rely on a limited number of parameters, most commonly between two and six [14, 57, 58, 59, 60, 61, 62, 63, 64]. With rare exceptions [58], multidimensionality reduction techniques are generally not applied in SSFA analyses; instead, palaeodietary differences are usually explored using bivariate plots based on the most discriminant variables, most often  $e_{pLsar}$  (Length-scale Anisotropy of the relief) versus  $A_{sfc}$  (Area-Scale Fractal Complexity) [14, 60, 63]. In contrast to SSFA studies, few ISO-based DMT articles in primates and hominins have been published to date, either in the occlusal [65, 66, 67] and the buccal surfaces [16, 17]. Nevertheless, the analysis of 3DST involves the use of a large number of ISO 25178 / ISO 12781 parameters [16, 17, 68], so that the use of dimensionality reduction techniques is more common [68], especially PCA [69, 70, 71, 72, 73], as well as DA [69, 70].

With respect to selecting the most relevant parameters in 3DST studies, a pioneering study on occlusal 3DST in primates [74] highlighted a subset of ISO parameters ( $S5v$ ,  $Sq$ ,  $Vm$ ,  $Spd$ ,  $Sha$ ,  $Sda$ ) as indicative of the interaction between food and enamel during mastication. However, only a limited number of variables were considered. Later studies expanded the set of ISO variables, increasing analytical complexity and leading to what has been called a “jungle of parameters” [75]. Discriminative power varies across studies and sample compositions, likely due to high inter-variable correlations [16, 17, 68]. Moreover, the criteria for defining a parameter as diet-discriminative differ between contributions [75] due to multiple factors (e.g. data sources, noise contamination and others). Thus, no universal set of optimal discriminative parameters has emerged. Still, some trends suggest that combining spatial and height parameters yields better discrimination power [76].

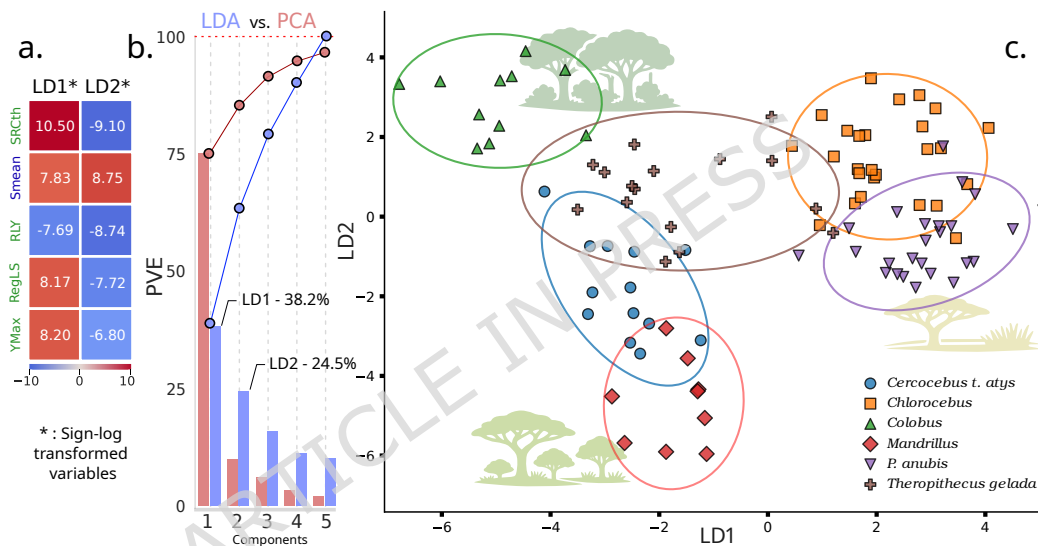
Our study develops a methodological pipeline for preprocessing and analysing DMT variables (3DST –ISO– or SSFA) alongside Fourier and Furrows variables (labeled `Other`) for palaeodietary reconstruction by performing groups and species classification, with a focus on model interpretability and the assessment of variable relevance [77]. Furrow and isotropy

parameters have been used recently to discriminate dietary regimes of mammals [65, 67, 78]. Specifically, we aim to contribute by (i) **proposing a framework** for identifying effective statistical and machine learning methods to classify primate species using buccal 3DST data, prioritising interpretability while incorporating ensemble approaches (e.g., random forests, boosting) for contrast; (ii) providing a reusable **code pipeline** for identifying key SSFA and ISO variables, facilitating future comparisons and data interpretation; and (iii) evaluating **variable importance** obtained in the constructed classification models to determine which features best discriminate microtexture patterns.

## 2 Results

### 2.1 Data preprocessing results

We applied the full data preprocessing pipeline to maximise the information available per group and species. For complete information on the dataset, including final variables selected, please refer to Section 4.3. Despite the strong non-normality and internal collinearity observed within several variable sets (see Section 4), most variables still displayed group and species-level structure when projected into low-dimensional space. The LDA visualization in Figure 1 (right) shows that, after preprocessing, LD directions contain a meaningful discriminative signal that distinguishes groups in this reduced dimensionality space. This motivated retaining all variables, rather than restricting the analysis to the subset passing normality or *variance inflation factor* (VIF, [79]) thresholds, since the full feature representation provides more informative class separation.



**Figure 1.** (Left - a) First 5 variables of the main two LDA components for group classification in preprocessed dataset. Values displayed in log-scale, with signs re-introduced afterwards. Variable names are color-coded, ISO in blue, SSFA in green. (Center - b) Proportion of variance explained (PVE) for both LDA (in blue) and PCA (in red) for the dataset. (Right - c) Data projection using the first two LD directions. Ellipses show Gaussian fits covering  $2\sigma$  for each label. Best viewed in color.

In the left panel (a) of Figure 1 we display top-5 variable loadings of the first two LD components (in sign-log-transformed scale). To compare to the usual decomposition using PCA, in the center panel (b) we show the proportion of variance explained (PVE) by the addition of each LDA (blue) and PCA (red) component. Even though PCA explains a larger portion of the variance in the initial components, since the LDA projections are constructed to maximize separability under the LDA assumptions [50] we see that the data clusters appear quite separate in the low-dimensionality figure, included in the right panel of the figure (c), including the corresponding Gaussian ellipses to illustrate group dispersion. This LDA projection reveals biologically interpretable structure: LD1 separates folivorous taxa such as *Colobus* (negative values) from open-habitat feeders like *Chlorocebus* and *P. anubis* (positive values), while *Theropithecus gelada* spans a broader range along this axis. LD2 differentiates *Mandrillus* and *Cercocebus atys*, both associated with hard or brittle dietary components, from the remaining groups, consistent with *Cercocebus* specializing in hard fruits in closed habitats and *Mandrillus* exploiting a diverse diet including fruits, leaves, and brittle seeds.

### 2.2 Classification models

We report performances of all models for the group-classification task in Table 1 (for the species classification results, see the SI). Rows correspond to the four variable sets (ISO, SSFA, Other –Fourier and Furrows– and All). For each method,

we display the best result obtained from either the VIF-filtered data or its non-filtered counterpart, so reported results in Table 1 include only the best one between them. We also evaluated each method using the original and standardized datasets, although these generally provided worse results than those presented here. For each metric, the highest-performing method is highlighted in bold red and the second highest in bold, with ties indicated when applicable. Given the small multiclass dataset (6 groups, 7 species), performance was estimated using LOOCV, and errors represent the standard errors of macro-averaged metrics. Micro-averaged accuracies are denoted  $\mu\text{-Acc.}$ , macro-averaged as  $M\text{-Acc.}$ . We also include  $\text{LDA}_C$  (LDA applied to the VIF-filtered datasets) for comparison [80]. For a complete definition of all models and metrics, please see Section 4.5.

We benchmarked results against two baselines: (i) random classifiers yield accuracies near 0.16 ( $\approx 1/6$ ) for groups and 0.17 ( $\approx 1/7$ ) for species; (ii) LDA, the standard in species classification, was applied after preprocessing, showing improved performance. Table 1 shows that VIF collinearity removal generally enhances accuracy in LDA, except for `Other`, where variable reduction may amplify noise. QDA did not outperform LDA, consistent with small, noisy datasets where LDA's lower variance offers greater robustness [81].

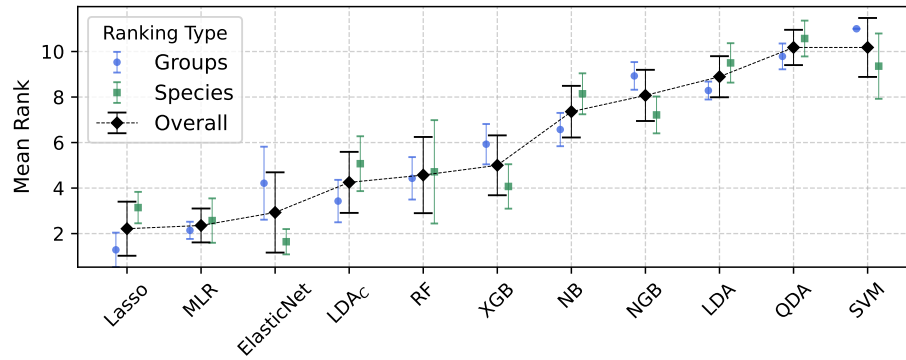
**Table 1.** Classification results for **Groups**. Each model tuned for optimal performance per variable set with (i) the *original*, preprocessed data without the *variance inflation factor* (VIF) filter for collinearity, and (ii) the *clean* (with  $c$  subscript for LDA), VIF-filtered and no collinear variables. Best performance in each case is highlighted in red and bold, second-best in bold only. For full description on each method (columns), and metric (rows) please refer to Section 4.5.

Metrics	LDA	$\text{LDA}_C$	QDA	MLR	Lasso	EN	NB	SVM	RF	XGB	NGB
$\mu\text{-Acc.}$	0.27	0.28	0.21	0.33	0.32	0.33	<b>0.34</b>	0.30	0.28	0.32	<b>0.34</b>
$M\text{-Acc.}$	$0.26 \pm 0.03$	$0.24 \pm 0.06$	$0.17 \pm 0.07$	<b><math>0.32 \pm 0.04</math></b>	$0.31 \pm 0.05$	$0.31 \pm 0.06$	<b><math>0.38 \pm 0.10</math></b>	$0.24 \pm 0.09$	$0.23 \pm 0.07$	$0.29 \pm 0.06$	$0.31 \pm 0.05$
I Prec.	$0.25 \pm 0.04$	$0.26 \pm 0.05$	$0.18 \pm 0.06$	$0.35 \pm 0.07$	$0.34 \pm 0.07$	$0.33 \pm 0.05$	<b><math>0.38 \pm 0.07</math></b>	$0.23 \pm 0.08$	$0.23 \pm 0.06$	$0.29 \pm 0.06$	<b><math>0.37 \pm 0.05</math></b>
S Rec.	$0.26 \pm 0.03$	$0.24 \pm 0.06$	$0.17 \pm 0.07$	<b><math>0.32 \pm 0.04</math></b>	$0.31 \pm 0.05$	$0.31 \pm 0.06$	<b><math>0.38 \pm 0.10</math></b>	$0.24 \pm 0.09$	$0.23 \pm 0.07$	$0.29 \pm 0.06$	$0.31 \pm 0.05$
O F1	$0.26 \pm 0.03$	$0.24 \pm 0.05$	$0.16 \pm 0.06$	<b><math>0.33 \pm 0.05</math></b>	<b><math>0.32 \pm 0.05</math></b>	<b><math>0.32 \pm 0.05</math></b>	<b><math>0.32 \pm 0.05</math></b>	$0.22 \pm 0.08$	$0.23 \pm 0.06$	$0.29 \pm 0.06$	<b><math>0.32 \pm 0.03</math></b>
C-k	0.11	0.10	0.04	0.17	0.15	0.16	0.22	0.11	0.09	0.16	0.17
AUC-ROC	0.60	0.67	0.51	0.71	0.71	0.73	0.68	0.71	0.67	0.69	0.61
$\mu\text{-Acc.}$	0.19	0.28	0.23	0.31	0.31	0.29	0.27	0.30	0.34	0.30	0.30
$M\text{-Acc.}$	$0.15 \pm 0.05$	$0.22 \pm 0.08$	$0.15 \pm 0.10$	$0.25 \pm 0.08$	<b><math>0.26 \pm 0.06</math></b>	$0.22 \pm 0.08$	<b><math>0.27 \pm 0.03</math></b>	$0.22 \pm 0.10$	$0.25 \pm 0.10$	$0.22 \pm 0.09$	$0.23 \pm 0.08$
S Prec.	$0.14 \pm 0.05$	$0.21 \pm 0.07$	$0.08 \pm 0.05$	$0.24 \pm 0.06$	<b><math>0.27 \pm 0.05</math></b>	$0.22 \pm 0.06$	<b><math>0.27 \pm 0.04</math></b>	$0.17 \pm 0.08$	$0.23 \pm 0.07$	$0.21 \pm 0.08$	$0.26 \pm 0.07$
S Rec.	$0.15 \pm 0.05$	$0.22 \pm 0.08$	$0.15 \pm 0.10$	$0.25 \pm 0.08$	<b><math>0.26 \pm 0.06</math></b>	$0.22 \pm 0.08$	<b><math>0.27 \pm 0.03</math></b>	$0.22 \pm 0.10$	$0.25 \pm 0.10$	$0.22 \pm 0.09$	$0.23 \pm 0.08$
F F1	$0.15 \pm 0.05$	$0.21 \pm 0.08$	$0.10 \pm 0.06$	$0.24 \pm 0.07$	<b><math>0.26 \pm 0.06</math></b>	$0.22 \pm 0.07$	<b><math>0.26 \pm 0.03</math></b>	$0.19 \pm 0.08$	$0.24 \pm 0.08$	$0.21 \pm 0.08$	$0.23 \pm 0.06$
A C-k	-0.01	0.09	-0.04	0.13	0.13	0.10	0.11	0.09	0.16	0.11	0.10
AUC-ROC	0.57	0.64	0.49	0.65	0.67	0.68	0.62	0.65	0.68	0.64	0.57
$\mu\text{-Acc.}$	<b>0.36</b>	0.30	0.22	0.33	0.34	0.32	0.23	0.35	0.26	0.24	<b>0.36</b>
$M\text{-Acc.}$	<b><math>0.30 \pm 0.07</math></b>	$0.23 \pm 0.10$	$0.16 \pm 0.07$	$0.26 \pm 0.09$	<b><math>0.28 \pm 0.07</math></b>	$0.24 \pm 0.09$	$0.18 \pm 0.07$	$0.24 \pm 0.13$	$0.21 \pm 0.07$	$0.17 \pm 0.08$	$0.26 \pm 0.11$
t Prec.	<b><math>0.34 \pm 0.05</math></b>	$0.22 \pm 0.08$	$0.14 \pm 0.05$	$0.25 \pm 0.08$	<b><math>0.31 \pm 0.05</math></b>	$0.24 \pm 0.07$	$0.18 \pm 0.07$	$0.16 \pm 0.08$	$0.23 \pm 0.06$	$0.13 \pm 0.06$	$0.25 \pm 0.08$
h Rec.	<b><math>0.30 \pm 0.07</math></b>	$0.23 \pm 0.10$	$0.16 \pm 0.07$	$0.26 \pm 0.09$	<b><math>0.28 \pm 0.07</math></b>	$0.24 \pm 0.09$	$0.18 \pm 0.07$	$0.24 \pm 0.13$	$0.21 \pm 0.07$	$0.17 \pm 0.08$	$0.26 \pm 0.11$
e F1	<b><math>0.31 \pm 0.06</math></b>	$0.22 \pm 0.08$	$0.15 \pm 0.06$	$0.25 \pm 0.08$	<b><math>0.28 \pm 0.06</math></b>	$0.23 \pm 0.08$	$0.18 \pm 0.07$	$0.18 \pm 0.09$	$0.21 \pm 0.06$	$0.15 \pm 0.07$	$0.24 \pm 0.09$
r C-k	0.20	0.12	0.01	0.16	0.17	0.14	0.06	0.14	0.07	0.02	0.18
AUC-ROC	0.67	0.67	0.53	0.68	0.67	0.67	0.64	0.65	0.62	0.54	0.62
$\mu\text{-Acc.}$	0.22	0.32	0.11	0.34	0.36	0.31	0.32	0.16	0.27	0.27	0.31
$M\text{-Acc.}$	$0.18 \pm 0.06$	$0.28 \pm 0.05$	$0.10 \pm 0.02$	$0.32 \pm 0.06$	<b><math>0.34 \pm 0.06</math></b>	$0.27 \pm 0.08$	<b><math>0.35 \pm 0.07</math></b>	$0.11 \pm 0.07$	$0.23 \pm 0.07$	$0.23 \pm 0.06$	$0.28 \pm 0.04$
A Prec.	$0.19 \pm 0.07$	$0.28 \pm 0.05$	$0.12 \pm 0.04$	<b><math>0.32 \pm 0.04</math></b>	<b><math>0.34 \pm 0.04</math></b>	$0.27 \pm 0.08$	$0.31 \pm 0.06$	$0.05 \pm 0.03$	$0.23 \pm 0.06$	$0.23 \pm 0.05$	$0.30 \pm 0.03$
l Rec.	$0.18 \pm 0.06$	$0.28 \pm 0.05$	$0.10 \pm 0.02$	$0.32 \pm 0.06$	<b><math>0.34 \pm 0.06</math></b>	$0.27 \pm 0.08$	<b><math>0.35 \pm 0.07</math></b>	$0.11 \pm 0.07$	$0.23 \pm 0.07$	$0.23 \pm 0.06$	$0.28 \pm 0.04$
l F1	$0.18 \pm 0.07$	$0.28 \pm 0.05$	$0.10 \pm 0.02$	<b><math>0.31 \pm 0.05</math></b>	<b><math>0.33 \pm 0.05</math></b>	$0.27 \pm 0.07$	$0.30 \pm 0.05$	$0.07 \pm 0.04$	$0.23 \pm 0.06$	$0.23 \pm 0.05$	$0.29 \pm 0.03$
C-k	0.06	0.15	-0.03	0.18	0.20	0.13	0.19	-0.13	0.08	0.08	0.13
AUC-ROC	0.50	0.67	0.46	0.68	0.68	0.72	0.67	0.28	0.71	0.67	0.63

It is worth noting that recent preceding publications relied primarily on LDA [80], which, even when aided by our improved preprocessing pipeline, still performs worse overall than most of the classifiers evaluated here (except for the `Other` variable set). This leads us to believe that LDA should only be used as a reference point for datasets similar to ours, and that other methods should be used for prediction on new samples. Note also that our evaluation here is intentionally stringent. By using LOOCV and reporting both macro- and micro-averaged metrics, we adopt a conservative scoring strategy that minimizes optimistic bias and class-imbalance effects. As a result, the reported metric values may appear modest, which is expected for a 6-class classification problem with fewer than 100 observations evaluated under such strict conditions. In this setting, each fold tests a single left-out individual, leading to particularly harsh penalties across all performance metrics.

To further illustrate these results, we ranked all models in Section 4.5 from best to worst for each dataset and metric in both the group- and species-level classification tasks. We then computed the mean rank and its standard error for each method across all metrics and datasets, as shown in Figure 2 (lower values indicate better performance). MLR [50] and its regularized variants (Lasso and Elastic Net) consistently ranked highest, reflecting their ability to handle multiclass problems with collinearity and noise through coefficient shrinkage. NB was competitive for `ISO` variables on non-VIF-filtered data, often leading on the original datasets, which suggests that VIF filtering may remove variables carrying independent predictive signal important for NB. LDA performed best with the `Other` variable set, benefitting from VIF filtering ( $\text{LDA}_C$  ranking 4th), consistent with its reliance on linear decision boundaries and low-variance estimation (for additional details see SI).

MLR models performed similarly with `ISO` variables alone and the full `All` set, suggesting that the inclusion of additional variables yields only marginal gains. Overall, `ISO` variables provided the strongest and most stable results, whereas `SSFA` and `Other` did not show a clear performance benefit. Consequently, `ISO` variables alone emerge as the most robust choice in this setting, supporting the majority of top-performing metrics in both classification tasks (see SI for further details and species-level results). Although these findings are not necessarily generalizable, they motivate further investigation using larger



**Figure 2.** Mean ranking across metrics for all methods for groups (blue dots), species (green squares) and overall (black diamonds). Lower is better. Best viewed in color.

and more diverse datasets.

### 2.3 Automated variable selection

Among all the methods included in our pipeline, MLR models seem to best suit our interests since they provide improved robustness and clear interpretation, with their regularized variants (*Lasso*, *Elastic Net*) being particularly apt to perform automated variable selection while maintaining high performance (see Table 1). The results in Figure 3 here correspond to Lasso MLR applied to the All dataset (73 variables across ISO, SSFA, and Other) without VIF filtering so that Lasso can select variables directly. Group-level results are presented here, while species-level results are included in the SI.

After training on the classification task, Lasso reduces the number of used variables from 73 to 32 for both group and species classification (for more details on the model's properties, please refer to Section 4.5). The selected sets are very similar: FLTV and Spk (groups) are replaced by S5v and Sz (species), all of which play minor roles. This stability is expected because species classification refines groups by subdividing one of them. Interpretation of these results is aided by the Lasso coefficients (Figure 3, top). Variables are ordered by importance (left = highest), and the sign of the re-scaled coefficients reflects whether higher values make more or less likely belonging to a given group.

The kernel density estimates (KDE, [50]) plots for each figure (Figure 3, bottom) help understand these choices by the model. For example, high Sal (an ISO autocorrelation measure) is characteristic of *Colobus* and aligns with its positive coefficients, whereas *Chlorocebus* and *T. gelada* show lower Sal. *T. gelada* also shows high Isotropy2 and Sci but low surface complexity. *Mandrillus* and *Cercocebus*, which process brittle foods, are distinctly non-isotropic, with high Smr1, S5p, and Smr2. *Papio* surfaces tend to be rounded and non-isotropic. Across models, variables such as DomWave, Smr2, Sal, Sci, and HAsfc81 consistently rank among the most informative features.

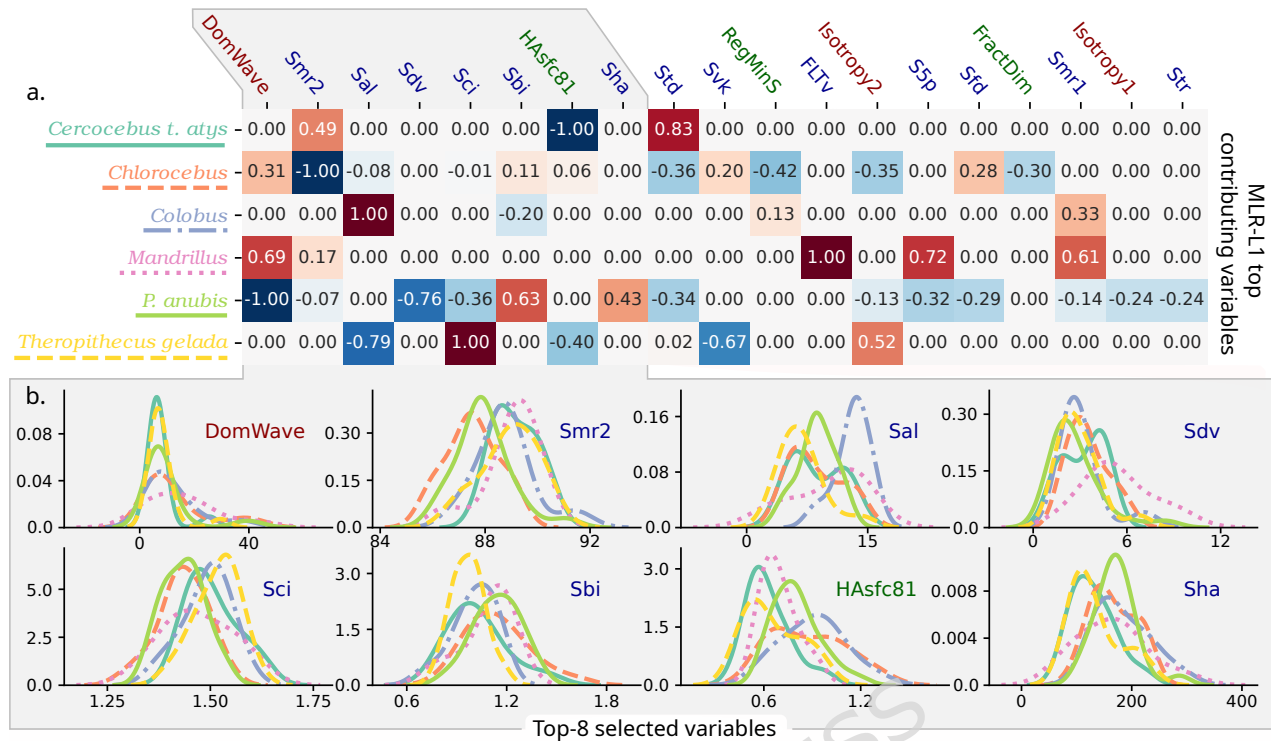
## 3 Discussion

### Reducing the number of parameters

Microtexture datasets are typically high-dimensional, especially when large ISO parameter sets are used, as in many recent buccal and occlusal studies [9, 16, 17, 82, 83, 84]. This dimensionality issue increases redundancy and complicates interpretation, underscoring the need for explicit variable selection. In contrast, SSFA relies on fewer descriptors and generally produces lower-dimensional representations (Section 1). In our dataset, correlation analyses showed that variables within both frameworks were either weakly related or strongly collinear, as showcased in Figure 5. Applying a collinearity filter (Section 2.1) reduced the ISO set to nine variables (Sku, Str, Std, Shv, Smr1, Sfd, Sci, Smean, Stdi), retaining complementary height, spatial, and volume descriptors as recommended previously [76]. These parameters have demonstrated sensitivity to buccal microtexture variation [16, 69, 85, 86].

After filtering, ISO variables accounted for nearly half of the retained predictors and yielded the strongest classification performance across metrics (Table 1). This indicates that, once redundancy is controlled, ISO descriptors capture most of the discriminative signal in our sample. Fourier-based variables (the Other set), such as isotropy and dominant wavelength, also contributed to group separation, as shown by the Lasso coefficients (Figure 3).

Although not the primary focus here, SSFA variables showed lower internal collinearity than ISO parameters, consistent with previous work [87]. Their more independent structure suggests that they may complement ISO descriptors, even if they contributed less to discrimination in the present analysis.



**Figure 3.** *Top (a):* Highest 19 Lasso MLR coefficients for group classification re-scaled group-wise. Variable names are color-coded, with ISO variables in blue, SSFA, in green, and Other, in red. *Bottom (b):* KDE plots for the top 8 variables. Groups are distinguished with colors and line types, reflected in the names and underlines for each group in the top-part of the figure. Best viewed in color.

### Specific versus generic level comparisons

Our study applied several statistical tests to characterize cercopithecoid species of known diet to suggest key statistical/ML tools to classify and discriminate species based on microtexture patterns. One relevant point to comment is that whether we consider only the ISO variables, the SSFA ones or both together, the performance of classifiers is better when working at the more genus (group) level due to there being one fewer class (with *Chlorocebus pygerythrus* and *Chlorocebus aethiops* grouped in a single genus category *Chlorocebus sp*) than at the species level (i.e., when both *Chlorocebus* species are treated independently; see SI with the species level analysis). In studies of occlusal SSFA, whether working at the genus or species level as a factor or grouping variable, significant differences were observed in extant primates ([34]); however, the authors suggested that in the absence of substantial differences between species of the same genus, the most pertinent analysis would be a generic-level comparison. A previous study [16] found that there were no significant statistical differences between both *Chlorocebus* species, with only the (Sal) variable showing a significant difference out of 29 of them. Our results here support that working at the genus level performs better than at the species level. This approach is particularly relevant in studies of fossil primates [88], where sample sizes are often limited. In such contexts, combining closely related taxa may help increase statistical power [34, 88]; however, this should only be done when there is reasonable evidence that the taxa included share a comparable dietary range. Therefore, grouping specimens at the genus level should not be considered a purely taxonomic decision but rather an analytical strategy that assumes similar dietary behaviour among the taxa included. Under these conditions, increasing the effective sample size may improve the robustness of paleodietary interpretations [88, 89].

### Comparison of different classifiers and types of variables

Another objective of this study was to assess the relative performance of different classifiers. For models using only ISO variables, the best results across most metrics were obtained with the NB classifier, followed by MLR and, less consistently, NGB. Although NB achieved the highest performance when restricted to ISO variables, our final analyses rely on Lasso-regularized MLR, as this model incorporates all variable sets simultaneously (fourth row in Table 1) and yields the best overall performance. Moreover, MLR models consistently ranked among the top-performing methods across metrics (Figure 2), a property that is particularly valuable in applied contexts where stability and interpretability across variable configurations are preferred over isolated performance peaks.

LDA was included to maintain comparability with previous analyses of the same sample [16, 17]. Despite the proposed preprocessing pipeline, however, LDA consistently underperformed relative to most other classifiers. This likely reflects the mismatch between LDA's assumptions (*e.g.*, class structure and covariance homogeneity) and the characteristics of our dataset, as well as the limited sample size per class, which makes the method sensitive to small variations even under LOO cross-validation. Performance improved after VIF filtering (LDA<sub>C</sub>), but remained inferior to NB and MLR across all metrics. Given these results, for this dataset (or others similar to it, particularly when using ISO variables), we suggest LDA should be considered a baseline reference rather than a competitive classifier, despite its traditional use in dental microtexture research. Given that many studies rely on DA or PCA for dimensionality reduction in similarly high-dimensional settings [71, 78, 85, 86, 90, 91, 92, 93, 94], it is plausible that alternative classifiers would also yield improved performance in those contexts. In this regard, our results support a broader use of MLR, which has only rarely been applied in occlusal microwear studies [95, 96].

For models using SSFA variables, NB and Lasso MLR yielded the best results, with Lasso MLR demonstrating the most consistent performance. Uniquely, our analysis incorporated all 31 available SSFA parameters rather than the traditional subset of 2–4 variables (*e.g.*, *Asfc*, *epLsar* [87]). While this precludes direct generalization to past studies, it highlights the untapped potential of exploiting broader SSFA feature sets.

However, despite previous analyses have relied on SSFA for palaeodietary reconstruction [60, 68, 74, 76, 87], our best overall classifiers relied exclusively on ISO parameters. This advantage likely arises because ISO variables capture microroughness, localized mechanical abrasions (*e.g.*, precise depths and volumes) tied directly to physical food-enamel interactions [74, 97, 98]. Conversely, SSFA measures broad, scale-invariant complexity, which may underestimate subtle geometric signatures required to differentiate closely related taxa. Finally, because ISO and SSFA describe the identical physical surface, combining them introduces severe multicollinearity and redundant noise rather than orthogonal discriminatory information, ultimately degrading the performance of sensitive models.

While combining ISO and SSFA variables might be expected to improve discrimination by providing complementary information [74, 99], our results do not support this assumption. Although previous studies have included both variable types, they have generally analysed them separately [74, 83, 84, 99, 100, 101], and only a few have combined them within a single statistical framework [9, 71, 80, 93, 102]. To our knowledge, none have systematically compared their joint discriminative power across taxa with different diets [68]. In this study, combining ISO and SSFA variables resulted in similar or poorer performance than using ISO variables alone at both the group and species levels. By contrast, variables related to isotropy and dominant wavelength (*Other* set) contributed meaningful discriminatory information, and this was the only variable set for which LDA achieved its best overall performance (Table 1). Although such variables have been applied to other materials (*e.g.*, lithic tools [103, 104]), they have not previously been incorporated into primate dental microtexture analyses. Therefore, we consider this finding warrants further investigation.

Beyond the choice of variables and algorithms, classification success varied notably depending on the specific ecology of the taxa analyzed (see the SI for further group-wise classification analysis). On the one hand, some groups are represented with fewer points than others, leading to variable statistical power for the different methods depending on each case. On the other hand, species possessing highly specialized and stereotyped diets naturally exhibit highly distinct and homogeneous microwear signatures, making them significantly easier for the algorithms to isolate in the feature space. In contrast, generalist or opportunistic species, whose diets fluctuate heavily with seasonal and regional availability, produce highly heterogeneous microtexture patterns. These broad dietary niches inherently result in overlapping statistical distributions in the feature space, thereby increasing the difficulty of algorithmic classification and driving down their specific accuracy metrics.

Finally, we emphasize that conclusions regarding variable selection must be interpreted cautiously given the limited sample size and the focus on buccal enamel surfaces. While ISO and SSFA variables may provide complementary information [74, 105], their combination within a single ML pipeline may introduce sufficient noise to offset potential gains. Consequently, trends identified here require validation using larger and more diverse samples, as well as occlusal surfaces, to assess whether (1) ISO variables consistently outperform SSFA in classification, (2) Fourier-based and furrow-related variables enhance ISO-based discrimination, and (3) ISO and SSFA parameters are better treated as complementary, but not directly combinable within the same analytical framework.

### **Conclusions, limitations and future directions**

Our findings indicate that, for buccal enamel surfaces in the cercopithecoid sample analysed, ISO texture parameters provide greater discriminatory power than SSFA variables. Across all evaluated classifiers, models relying exclusively on ISO variables consistently achieved higher or more stable performance than those based on SSFA parameters or on the combined variable sets. This suggests that ISO parameters may be more sensitive to the microtexture differences relevant for distinguishing dietary groups in this specific sample, successfully capturing fundamental aspects of surface texture and the direct mechanical interaction between food particles.

Despite these promising results, several limitations must be acknowledged. First, the dataset is limited in size and this

effect increases the influence of noise and individual-level variability, limiting the robustness of cross-validation estimates. The use of LOOCV allowed leveraging all available observations, but it also means that species with scarce representation may disproportionately affect classification results. Second, although our preprocessing pipeline reduced collinearity and improved the stability of the models, it cannot fully compensate for the inherent variability of microtexture data, nor for potential measurement noise introduced during scanning or data extraction. These factors could be improved upon by conducting similar studies in larger datasets.

These limitations directly inform several avenues for future research. Increasing sample size remains a priority, both to improve the stability of classification models and to enable more granular analyses (e.g., seasonality or population-level variation). Additionally, the promising behaviour of variables related to isotropy and dominant wavelength (here included in the Other set) suggests that their role should be explored further across taxa and surface types. Beyond enlarging the dataset, more expressive statistical frameworks e.g. Bayesian hierarchical modelling or other probabilistic ML methods could substantially enhance inference by explicitly modelling uncertainty and allowing information to be shared across individuals, species, or dietary groups. This would potentially improve parameter interpretability and stabilize estimates in data-limited settings. Such approaches may also help quantify the degree to which variability arises from individual- versus group-level factors, which is difficult to assess with classical methods. Further methodological work is needed to evaluate how these frameworks respond to different variable types, particularly in studies where dimensionality reduction is common. Finally, replicating this analysis on occlusal enamel surfaces, which differ functionally from buccal ones, would be crucial to determining whether the relative advantage of ISO variables generalizes beyond the current context.

Taken together, our study offers a comprehensive and reproducible pipeline for the preprocessing, selection, and statistical classification of microwear-derived dietary data. By systematically comparing a broad set of interpretable classification models, our framework clarifies which variables and variable sets contribute most strongly to separating dietary groups, providing a robust methodological foundation that can be readily extended to new datasets and to future applications in both extant and fossil dietary reconstruction.

## 4 Methods

### 4.1 Primate reference collection

This study is based on the open-access dataset of [16], which comprises 98 lower second molars (LM2) from seven African Cercopithecoidea species with distinct dietary adaptations (Table 2). Tooth selection follows the criteria described in [16] and related studies [15, 106], with a single modification: *Chlorocebus pygerythrus* and *Chlorocebus aethiops* are grouped into a single category due to their similar habitats and diets. Results obtained under the original species-level separation are reported in the supplementary information. The sample sizes listed in Table 2 correspond to the raw dataset, whereas the effective number of observations may differ after preprocessing, as illustrated in Figure 1. In particular, the application of the outlier filter may exclude some specimens when predefined criteria are met, resulting in smaller group sizes than those initially reported.

**Table 2.** Primate Cercopithecoidea sample and dietary information. \* = Category *Chlorocebus sp.* includes *Chlorocebus pygerythrus* and *Chlorocebus aethiops* individuals.

Species	Sample size	Region of origin	Habitat	Dietary category
<i>Colobus polykomus</i>	11	Liberia	Forest	Foliate-seed eater
<i>Mandrillus sphinx</i>	9	Cameroon and Gabon	Forest	Foliate-Frugivore-Seed eater
<i>Papio anubis</i>	25	Kenya	Savanna	Opportunistic omnivore
<i>Theropithecus gelada</i>	16	Ethiopia	Grasslands-highland plateaus	Grass eater
<i>Chlorocebus pygerythrus</i>	11	Tanzania	Savanna-woodland	Opportunistic omnivore
<i>Chlorocebus aethiops</i>	15	Kenya	Savanna	Opportunistic omnivore
* <i>Chlorocebus sp.</i>	26	East Africa	Savanna	Opportunistic omnivore
<i>Cercocebus atys</i>	13	Cameroon, Congo and Liberia	Forest	Hard object feeder

### 4.2 Image processing

Buccal enamel surfaces were scanned using a Sensofar PluNeox confocal microscope at IMF-CSIC, with a 20× objective (0.45 NA) following [16, 17]. As required for buccal microwear [10, 11, 15], teeth were positioned perpendicular to the light

beam at a 90° angle. Four non-overlapping 138 × 102µm sub-areas were selected from larger 650 × 550µm fields. Surfaces were leveled using least squares, had form removed with a 2nd-degree polynomial [16] and filtered (0.5% tail threshold and Gaussian 3 × 3). For each subarea we extracted 45 ISO and 20 SSFA variables with SensoMAP 7 (Digital Surf) from which we used a subset of 31 (further details on the variables selected, please refer to the SI).

SSFA [107] was designed to capture the multiscale nature of tooth–food interaction [87, 105] and is often considered the most effective tool for dietary reconstruction [74, 76]. Commonly used parameters include  $Asfc$  (complexity),  $Smc$ ,  $epLsar$  (anisotropy),  $HAsfc$  (heterogeneity), and  $Tfv$  [87]. Originally computed using Toothfrax and Sfrax [87, 108] software, these are now implemented in MountainsMap (Digital Surf), which includes more scale-sensitive variables and a corrected  $NewEpslar$  but excludes  $Tfv$  [108]. Importantly, outputs from different software are not directly comparable despite their similar discriminative power [108]. In this study, SSFA values were generated *ex novo*, avoiding compatibility issues.

ISO 25178-2/ISO 12781 parameters provide standardized indices describing surface texture [97, 98], and have been used in 3DST analyses to reflect food–enamel interactions [74]. They offer an alternative to SSFA for dietary analysis [16, 17, 69, 70, 74, 109], though they are not directly comparable [105]. Some studies have combined both types of variables [99, 109, 110, 111], yet no systematic comparison exists [68]. In pitheciids, combining both sets facilitated detection of subtle dietary differences [99].

Finally, in addition to SSFA and ISO variables, we include 8 variables (labeled `Other`) related to Fourier analysis (*e.g.*, isotropy, directionality, periodicity), and furrow-related features previously used in lithic use-wear [103, 111, 112, 113] and dietary inference [67, 78, 102].

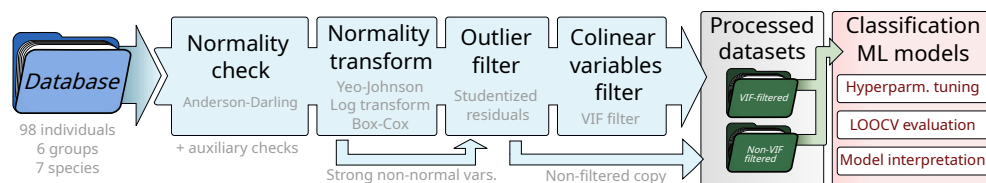
### 4.3 Data preprocessing

This study aims to develop a standardized data preprocessing pipeline for reliably identifying key variables in primate species classification. To ensure consistent input for classification models, the complete dataset underwent the preprocessing pipeline, schematically represented in Figure 4, consisting of the following steps:

1. *Checking for normality.* Since some models (*e.g.*, LDA) assume or benefit from normally distributed variables [80], we assessed normality using visual tools (boxplots, Q-Q plots) and the Anderson-Darling test ( $\alpha = 0.05$ ) [114], selected for its statistical power and suitability for our sample size.
2. *Transforming for normality.* Variables failing the normality test were transformed to approximate normality. While the log transform is standard in species classification [16], we used the Yeo-Johnson transformation [115] due to its effectiveness with zero and negative values. This generalizes Box-Cox and includes the log transform as a special case [116]. For a variable  $x$  and transformation parameter  $\lambda$  (estimated via maximum likelihood, [50, 115]), it is defined as:

$$T(x, \lambda) = \begin{cases} \frac{(x+1)^\lambda - 1}{\lambda}, & x \geq 0, \lambda \neq 0, \\ \log(x+1), & x \geq 0, \lambda = 0, \\ -\frac{(-x+1)^{2-\lambda} - 1}{2-\lambda}, & x < 0, \lambda \neq 2, \\ -\log(-x+1), & x < 0, \lambda = 2. \end{cases}$$

3. *Detecting and removing outliers.* Following [80], we flagged and removed outliers (characterized there as points with studentized residuals  $\geq 3$ ) since such extreme observations can disproportionately influence model fitting and reduce classification accuracy. This is particularly important in our leave-one-out measurement framework, where each point serves as a test case once, making outliers especially impactful on classification accuracy. Overall, only a small number of cases were affected, with only six points removed from the dataset.



**Figure 4.** Overview of the complete preprocessing and modeling pipeline. preprocessing steps are blue, with extra arrows indicating (1) the preservation of highly non-normal variables and (2) the non-VIF-filtered copy of the preprocessed dataset. Processed datasets are in green, while the schematic modelling and classification phase is in red. Best viewed in color.

4. *Dealing with colinear variables.* To address collinearity, which affects models like LDA, we computed the variance inflation factor (VIF, [79]) after standardization, removing variables with  $VIF \geq 2$  [117].

These steps form the core of the workflow summarized in Figure 4, which presents the transformation, filtering, and modeling stages in a unified scheme. The diagram emphasizes the sequential structure of the pipeline and illustrates how preprocessing decisions propagate into model training and interpretation. The preprocessing pipeline is highlighted in blue, with additional connections indicating that we retained variables exhibiting strong non-normal behavior that could not be adequately transformed in the second step. It also reflects our choice to use a version of the processed dataset without applying the VIF filter. These decisions result in two datasets, both of which are subsequently used in the classification tasks for the machine learning models whose results are presented.

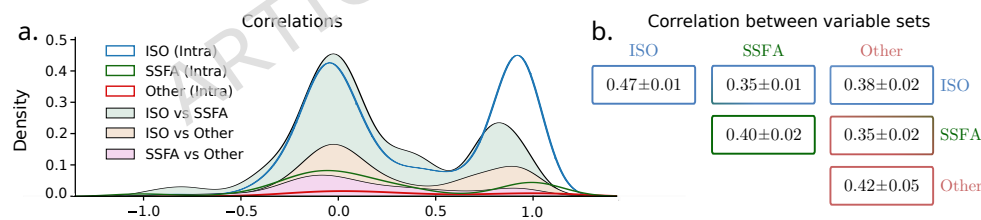
The previous steps are critical depending on the algorithm employed. Since not all classification methods are equally sensitive to collinearity, we retain two versions of the dataset: one with and one without the VIF filter. This allows the use of additional information from variables excluded by the collinearity filter when appropriate. Notably, no missing values were reported in this dataset, so no imputation was necessary. If missing data were present, their treatment would need to be evaluated based on their potential impact.

Our pipeline automates most of the preprocessing steps, though certain decisions still require expert judgment. The implementation is tailored to this dataset but can serve as a generalizable framework. For completeness, we store all datasets separately by variable set, as well as a merged version that combines all variables.

#### 4.4 Data analysis

To study the non-normality in the variables, we tested distributions with the Anderson-Darling test: only 12/45 ISO, 2/20 SSFA, and 2/8 Other variables (Fourier and Furrows) appeared normal. Log-transformations had little effect, while the Yeo-Johnson transformation notably improved results, leaving 3 ISO (Sal, Sdr, FLTq), 8 SSFA (YMax, SRCth, RegMinS, RegMaxS, RegLS, RLY, RegCo, SMaxComp), and 2 Other (DomWave, Isotropy2) variables non-normal. Thus, SSFA had the highest proportion of non-normal variables (40%), compared to ISO (7%) and Other (15%). Despite this, non-normal variables were predictive (Figure 3), so all were retained for optimal performance.

We also examined variable correlations both within and across variable sets to assess redundancy within sets and shared information between them. The Kernel Density Estimate (KDE, [50]) plots in Figure 5 reveal bimodal intra-set distributions, with most variables being either uncorrelated or strongly positively correlated. Cross-set correlations were generally centered near zero, indicating that the sets provide complementary information. Mean absolute correlations further confirmed that intra-set dependencies are stronger than inter-set dependencies. These findings underscore the value of combining variable sets to achieve a richer data representation and improve classification performance (see SI for details).



**Figure 5.** Variable analysis: (left - a) KDE of correlations for each variable set, both within itself (stroke-only curves) and with other variable sets (color-filled curves). (right - b) mean absolute correlation across variable sets. Best viewed in color.

Finally, we also studied potential outliers and collinearities. The studentized residuals filter removed 6 outliers, and the VIF filter identified collinearities within variable sets. After filtering, 9 ISO variables (Sku, Str, Std, Shv, Smrl, Sfd, Sci, Smean, and Stdi), 7 SSFA variables (FractDim, RegMinS, RegCo, CustHAsfc, MadAsfc, Eplsar, New\_Eplsar), and 4 Other variables (MaxDensF, Isotropy1, DomWave, Isotropy2) remained, indicating substantial co-dependence within each set. When applying collinearity filtering to the complete dataset, 13 variables remained: 7 from ISO (Sku, Sal, Std, Smrl, Sci, Smean, Stdi), 2 from SSFA (RegMinS, RegCo), and 4 from Other (Isotropy1, Amplitude, DomWave, New\_Eplsar). This suggests many SSFA variables provide limited additional information. Since model sensitivity to collinearity varies, we retained two dataset versions: one VIF-filtered and one unfiltered, applying all classification models to both. Results are reported for the best-performing configuration.

#### 4.5 Modeling and evaluation

Once the dataset has been preprocessed, we analyze it with the primary goal of correctly classifying primate *groups* (or *species*) based on dental textures as proxies for dietary habits. Since no single classifier can be universally best [118], our pipeline

includes several multi-class classifiers well documented in the literature, *e.g.*, [81], including *Extreme Gradient Boosting* (*XGBoost*, [119]) and *Natural Gradient Boosting* (*NGBoost*, [120]). Given our focus on interpretability, we exclude neural network-based models [121], which typically require larger datasets, extensive hyperparameter tuning, and high computational cost, without offering greater interpretability relevant to our objectives [50]. Specifically, the classifiers included in the pipeline, ordered by decreasing interpretability (and increasing flexibility), were:

- *Linear discriminant analysis* (**LDA**, [50]) introduced as the typical baseline in dental data-based classification, as in the originating study [16].
- *Quadratic discriminant analysis* (**QDA**, [50]) extends LDA by using quadratic separating functions and allowing different class covariance matrices. While it introduces more parameters than LDA, the formulation remains similar. Despite being an extension of LDA, QDA has not been widely applied in species classification [103, 122].
- *Naïve Bayes* (**NB**), serving as a baseline with minimal analyst intervention, constrained by the parametric form of the distributions selected and the strong structural assumption of conditional independence between variables for each separate class [50].
- *Multinomial logistic regression* (**MLR**) is a multiclass extension of logistic regression. We use it with regularization terms, specifically *Lasso* and *ElasticNet*, to simplify the model and promote variable selection [50]. The *Lasso* adds an  $\ell_1$  penalty, forcing some coefficients to zero and yielding a sparse, interpretable model. The *ElasticNet*, combining  $\ell_1$  and  $\ell_2$  penalties, is less sparse but better handles correlated predictors. Both methods can also be viewed as maximum-a-posteriori (MAP) estimates, offering a Bayesian perspective on sparsity.
- *Random Forest* (**RF**), another baseline requiring minimal analyst intervention drawing on the idea of decorrelating classification trees [123]. Besides, it requires more hyperparameter tuning than previous approaches.
- *Boosting methods* constitute another general family of methods showing excellent performance in various settings, being also robust to overfitting. Although we tested also *gradient boosting* [124] and *AdaBoost* [125], based on the results, we chose to report the metrics just for *XGBoost* (**XGB** [119]) and *NGBoost* (**NGB**, [120]), as they provided the best results within this family. However, they require an extensive hyperparameter tuning process.
- *Support vector machines* (**SVMs**, [50]) offer a robust and interpretable alternative, relying on support vectors. Originally designed for binary classification, we apply a *one-vs-rest* strategy for multiclass classification. This involves constructing  $n_{\text{classes}} \times (n_{\text{classes}} - 1)/2$  classifiers, which are then combined for the one-vs-rest approach. We tested *linear*, *RBF*, *polynomial* (degrees 2, 3, and 4), and *sigmoid* SVM kernels.

We used the implementation in `sklearn` [126] for almost all methods, except for *XGBoost* and *NGBoost* for which we employed the original implementations by their authors ([XGB code](#) and [NGB code](#)).

To better understand variable importance, we focus on MLR models, whose coefficients directly indicate the relevance of each variable to the model's prediction. Specifically, we analyze the fitted parameters from MLR models for both classification tasks (*groups* and *species*), where the sign of each coefficient shows the direction of association with a class, and its magnitude reflects the strength of that contribution. This allows identification of variables that consistently contribute across tasks or are uniquely important for specific classes, as detailed in Section 2.

Classifier performance was assessed using several *metrics* to provide a comprehensive evaluation. Given the manageable dataset size, performance measures were estimated via leave-one-out cross-validation (LOO-CV) on the entire dataset, training each model on all but one data point, evaluating on the held-out point, and averaging the results across all data points. This enables full utilisation of data and provides low-bias estimates of model performance, particularly generalisation error. Although LOO-CV may yield slightly lower performance estimates compared to fixed test sets, it offers a reliable and conservative evaluation by reducing variability and minimising overestimation due to arbitrary train-test splits. The chosen metrics were:

- *Accuracy* (Acc.): Defined as the ratio of correctly classified examples to the total number of instances in the dataset, also used for hyperparameter selection when necessary.
- *Recall* (Rec., *sensitivity*): Represents the fraction of instances correctly classified by the method (true positives divided by sum of true positives and false negatives). Specifically, it is the ratio of correctly classified primates in each class to all individuals in that class.
- *F1-score* (*F1*): A common classification metric, defined as the harmonic mean of precision and recall:  $F1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$ . Precision is the proportion of predicted positives that are correctly classified, and recall is the proportion of actual positives correctly identified. The F1-score provides a robust assessment of model performance, being high only when both precision and recall are high.

- *Cohen's Kappa* ( $C-\kappa$ ): A robust measure of the agreement between the classifier and actual labels beyond chance. It quantifies the proportion of predictions not explained by random guessing, defined as  $\kappa = (p_o - p_e)/(1 - p_e)$ , where  $p_o$  is the observed agreement (accuracy) and  $p_e$  is the expected agreement by chance, calculated in the multiclass case as  $p_e = N^{-2} \sum_{k=1}^K n_k p_k$ , with  $K$  categories,  $N$  observations,  $n_k$  the predicted proportion for class  $k$ , and  $p_k$  the actual proportion. The measure ranges from  $-1$  to  $+1$ , with higher values indicating greater agreement, zero indicating random guessing, and negative values worse than random guessing.
- *AUC-ROC* (Area Under the ROC curve) is a common metric for evaluating classification performance. In multiclass classification, it is computed using the *one-vs-rest* approach, where each class is treated as positive in turn, and the rest as negative. The overall AUC-ROC score is the average of these individual values, reflecting the classifier's ability to discriminate between classes. Scores close to 1 indicate better performance.

For more information on each metric, please see [50]. In all metrics we computed their *macro-average* values by calculating the metric for each class (group or species) and then averaging these across all classes. However, for Cohen's Kappa and AUC-ROC, we used the *micro-average* values. Micro-averaging aggregates contributions from all classes, considering the total true positives, false positives, and false negatives, which we found to be more stable due to the high variance caused by the small number of instances per class. For accuracies, we report both micro ( $\mu\text{-Acc.}$ ) and macro ( $M\text{-Acc.}$ ) averages to provide broader context for the pipeline's results.

To ensure unbiased estimates of model generalization and prevent data leakage, we employed a two-stage nested cross-validation strategy. Hyperparameter optimization was strictly confined to an inner loop utilizing 5-fold cross-validation, with accuracy serving as the primary metric to evaluate the parameter spaces for each method. For models with highly complex hyperparameter spaces (e.g., RF, XGB, NGB), we implemented randomized search cross-validation to efficiently navigate the combinations within a viable computational budget, ensuring fair comparisons across methods. Conversely, models such as LDA, QDA, and Naive Bayes required no tuning, while Multinomial Logistic Regression (MLR) involved evaluating specific regularization paths for the Lasso and ElasticNet penalties. The hyperparameter space for SVMs is detailed above, while in tree-based models optimization focused on critical architectural parameters, e.g. the number of estimators, maximum depth, minimum samples to split a node, and bootstrap settings. Once the optimal configurations were identified via this 5-fold tuning phase, the final predictive performance of each model was independently evaluated using Leave-One-Out Cross-Validation (LOOCV), thereby maximizing the training data utilized for our final anthropological assessments. This approach ensures fair evaluation across methods while maintaining accessibility, allowing the pipeline to run without specialized computational resources. It also reflects considerations of computational efficiency and energy consumption, while producing reliable performance estimates and supporting interpretability studies. For full details on the final models' parameters, please resort to the [code repository](#) (to be made public upon acceptance, code included with the submission).

## References

1. Meloro, C. *et al.* Chewing on the trees: Constraints and adaptation in the evolution of the primate mandible. *Evolution* **69**, 1690–1700, DOI: [10.1111/evo.12694](https://doi.org/10.1111/evo.12694) (2015).
2. DeCasien, A. R., Williams, S. A. & Higham, J. P. Primate brain size is predicted by diet but not sociality. *Nat. Ecol. & Evol.* **1**, 1–7, DOI: [10.1038/s41559-017-0112](https://doi.org/10.1038/s41559-017-0112) (2017).
3. Grabowski, M., Kopperud, B. T., Tsuboi, M. & Fansen, T. F. Both diet and sociality affect primate brain-size evolution. *Syst. Biol.* **72**, 404–418, DOI: [10.1093/sysbio/syac075](https://doi.org/10.1093/sysbio/syac075) (2023).
4. López-Aguirre, C., Lang, M. M. & Silcox, M. T. Diet drove brain and dental morphological coevolution in strepsirrhine primates. *Plos One* **17**, e0269041, DOI: [10.1371/journal.pone.0269041](https://doi.org/10.1371/journal.pone.0269041) (2022).
5. Davies, M. & Pineda-Munoz, S. The temporal scale of diet and dietary proxies. *Ecol. Evol.* **6**, 1883–1897, DOI: <https://doi.org/10.1002/ece3.2054> (2016).
6. Pérez-Pérez, A., Martínez, L. M., Gómez, M., Estebarez-Sánchez, F. & Romero, A. Correlations among dietary proxies in african fossil hominins: Dental buccal microwear, occlusal texture and  $^{13}\text{C}$  stable isotopes. *J. Archaeol. Sci. Reports* **22**, 384–391, DOI: [10.1016/j.jasrep.2018.03.013](https://doi.org/10.1016/j.jasrep.2018.03.013) (2018).
7. Teaford, M. F. & Oyen, O. J. In vitro and in vivo turnover in dental microwear. *Am. J. Phys. Anthropol.* **80**, 447–460, DOI: [10.1002/ajpa.1330800405](https://doi.org/10.1002/ajpa.1330800405) (1989).
8. Romero, A., Galbany, J., Juan, J. D. & Pérez-Pérez, A. Brief communication: Short- and long-term *in vivo* human buccal-microwear turnover. *Am. J. Phys. Anthropol.* **148**, 467–472, DOI: [10.1002/ajpa.22054](https://doi.org/10.1002/ajpa.22054) (2012).

9. Winkler, D. E. *et al.* The turnover of dental microwear texture: Testing the "last supper" effect in small mammals in a controlled feeding experiment. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **557**, DOI: [10.1016/j.palaeo.2020.109930](https://doi.org/10.1016/j.palaeo.2020.109930) (2022).
10. Estebarez, F., Martínez, L. M., Galbany, J., Turbón, D. & Pérez-Pérez, A. Testing hypotheses of dietary reconstruction from buccal dental microwear in *Australopithecus afarensis*. *J. Hum. Evol.* **57**, 739–750, DOI: [10.1016/j.jhevol.2009.07.007](https://doi.org/10.1016/j.jhevol.2009.07.007) (2009).
11. Estebarez, F., Galbany, J., Martínez, L. M., Turbón, D. & Pérez-Pérez, A. Buccal dental microwear analyses support greater specialization in consumption of hard foodstuff for *Australopithecus anamensis*. *J. Anthropol. Sci.* **90**, 163–185 (2012).
12. Ungar, P. S., Grine, F. E. & Teaford, M. F. Dental microwear and diet of the plio-pleistocene hominin *Paranthropus boisei*. *PLoS One* **3**, 10.1371, DOI: [10.1371/journal.pone.0002044](https://doi.org/10.1371/journal.pone.0002044) (2008).
13. Martínez, L. M., Estebarez-Sánchez, F., Galbany, J. & Pérez-Pérez, A. Testing dietary hypotheses of East African hominines using buccal dental microwear data. *PLOSOne* **11**, DOI: [10.1371/journal.pone.0165447](https://doi.org/10.1371/journal.pone.0165447) (2016).
14. Scott, R. *et al.* Dental microtexture analysis shows within-species diet variability in fossil hominins. *Nature* **436**, 693–695, DOI: [10.1038/nature03822](https://doi.org/10.1038/nature03822) (2005).
15. Martínez, L. M. *et al.* Buccal dental-microwear and feeding ecology of Early Pleistocene *Theropithecus oswaldi* from Cueva Victoria (Spain). *J. Hum. Evol.* **142**, 1–12, DOI: [10.1016/j.jhevol.2019.102736](https://doi.org/10.1016/j.jhevol.2019.102736) (2020).
16. Martínez, L. M. *et al.* Effectiveness of buccal dental-microwear texture in African Cercopithecoidea dietary discrimination. *Am. J. Biol. Anthropol.* **179**, 678–686, DOI: [10.1002/ajpa.24635](https://doi.org/10.1002/ajpa.24635) (2022).
17. Ramírez-Pedraza, I. *et al.* Multiproxy approach to reconstruct fossil primate feeding behavior: Case study for macaque from the plio-pleistocene site guefaït-4.2 (eastern morocco). *Front. Ecol. Evol.* **11**, 1–16, DOI: [10.3389/fevo.2023.1011208](https://doi.org/10.3389/fevo.2023.1011208) (2023).
18. Jiménez-Hidalgo, E., Estebarez-Sánchez, F., DeSantis, L. R., Krajcarz, M. T. & Meloro, C. Reconstructing paleodiets: Challenges and advances. *Front. Ecol. Evol.* DOI: [10.3389/fevo.2023.1267012](https://doi.org/10.3389/fevo.2023.1267012) (2023).
19. Galbany, J., Estebarez, F., Martínez, L. M. & Pérez-Pérez, A. Buccal dental microwear variability in extant african hominoidea: taxonomy versus ecology. *Primates* **50**, 221–230, DOI: [10.1007/s10329-009-0139-0](https://doi.org/10.1007/s10329-009-0139-0) (2009).
20. Romero, A., Ramírez-Rozzi, F. V., Juan, J. D. & Pérez-Pérez, A. Diet-related buccal dental microwear patterns in central african pygmy foragers and bantu-speaking farmer and pastoralist populations. *PLoS One* **8**, e84804, DOI: [doi:10.1371/journal.pone.0084804](https://doi.org/10.1371/journal.pone.0084804) (2013).
21. Schmidt, C. W. *et al.* Dental microwear texture analysis of homo sapiens sapiens: Foragers, farmers, and pastoralists. *Am. journal physical anthropology* **169**, 207–226 (2019).
22. Pérez-Pérez, A., de Castro, J. B. & Arsuaga, J. Nonocclusal Dental Microwear Analysis of 300,000-year-old *Homo heidelbergensis* teeth from Sima de los Huesos (Sierra de Atapuerca, Spain). *Am. J. Phys. Anthropol.* **108**, 433–457, DOI: [10.1002/\(SICI\)1096-8644\(199904\)108:4<433::AID-AJPA5>3.0.CO;2-5](https://doi.org/10.1002/(SICI)1096-8644(199904)108:4<433::AID-AJPA5>3.0.CO;2-5) (1999).
23. Jarosřová, I., Vávra, J., Kiřik, J. & Horáková, M. Buccal dental microwear of a barbarian population from Prague-Zličín—a study of the micration period in the Czech Republic. *Interdiscip. Archaeol. Nat. Sci. Archaeol.* **VII**, 55–70 (2016).
24. Bodoriková, S. *et al.* Trace elements and dental microwear as the indicators of diet strategies in individuals from the cemetery of Borovce (8<sup>th</sup>-12<sup>th</sup>, Pieřňany district, Slovakia). *Anthropologie* **62**, 329–347, DOI: [10.2670/anthro.19.09.06.1](https://doi.org/10.2670/anthro.19.09.06.1) (2019).
25. Pérez-Pérez, A., Lalueza, C. & Turbón, D. Intraindividual and intragroup variability of buccal tooth striation pattern. *Am. J. Phys. Anthropol.* **94**, 175–187, DOI: [10.1002/ajpa.1330940203](https://doi.org/10.1002/ajpa.1330940203) (1994).
26. Pinilla-Pérez, B., Romero, A. & Pérez-Pérez, A. Age-related variability in buccal dental-microwear in middle and upper pleistocene human populations. *Anthropol. Rev.* **74**, 25–37, DOI: [10.2478/v10044-010-0005-0](https://doi.org/10.2478/v10044-010-0005-0) (2011).

27. Jarošová, I. & Tvrđý, Z. Diet and diversity of early farmers in neolithic period (Ibk): Buccal dental microwear and stable isotopic analysis at Vedrovice (Czech Republic) and Nitra-Horné Krškany (Slovakia). *Anthropologie* **55**, 353–384 (2017).
28. Peterson, A., Abella, E. F., Grine, F. E., Teaford, M. F. & Ungar, P. S. Microwear textures of australopithecus africanus and paranthropus robustus molars in relation to paleoenvironment and diet. *J. Hum. Evol.* **119**, 42–63, DOI: [10.1016/j.jhevol.2018.02.004](https://doi.org/10.1016/j.jhevol.2018.02.004) (2018).
29. Salazar-García, D., Romero, A., García-Borja, P., Subirà, M. & Richards, M. A combined dietary approach using isotope and dental buccal-microwear analysis of human remains from the neolithic, roman and medieval periods from the archaeological site of Tossal de les Basses (Alicante, Spain). *J. Archaeol. Sci. Reports* **6**, 601–619, DOI: <http://dx.doi.org/10.1016/j.jasrep.2016.03.002> (2016).
30. Romero, A., Martínez-Ruiz, N. & de Juan, J. Non-occlusal dental microwear in a bronze-age human sample from East Spain. *Anthropologie* **XLII**, 65–69, DOI: <https://www.jstor.org/stable/26292675> (2004).
31. Pérez-Pérez, A. *et al.* The diet of the first europeans from atapuerca. *Sci. Reports* **7**, 44319, DOI: [DOI:10.1038/srep43319](https://doi.org/10.1038/srep43319) (2017).
32. Hernando, R. *et al.* Integrating buccal and occlusal dental microwear with isotope analyses for a complete paleodietary reconstruction of holocene populations from hungary. *Sci. Reports* **11**, DOI: [10.1038/s41598-021-86369-x](https://doi.org/10.1038/s41598-021-86369-x) (2021).
33. Galbany, J. & Pérez-Pérez, A. Buccal enamel microwear variability in *Cercopithecoidea* primates as a reflection of dietary habits in forested and open savanna environments. *Anthropologie* **42**, 13–20, DOI: <https://hdl.handle.net/2445/160222> (2004).
34. Scott, R., Teaford, M. & Ungar, P. Dental microwear texture and anthropoid diets. *Am. J. Phys. Anthropol.* **147**, 551–579, DOI: [10.1002/ajpa.2200](https://doi.org/10.1002/ajpa.2200) (2012).
35. Ungar, P. S., Grine, F. E., Teaford, M. F. & El Zaatari, S. Dental microwear and diets of african early homo. *J. human evolution* **50**, 78–95 (2006).
36. Ungar, P. & Teaford, M. Preliminary examination of non-occlusal dental microwear in anthropoids: Implications for the study of fossil primates. *Am. J. Phys. Anthropol.* **100**, 101–113, DOI: [10.1002/\(SICI\)1096-8644\(199605\)100:1%3C101::AID-AJPA10%3E3.0.CO;2-4](https://doi.org/10.1002/(SICI)1096-8644(199605)100:1%3C101::AID-AJPA10%3E3.0.CO;2-4) (1996).
37. Bodoriková, S. *et al.* Dietary reconstruction from trace element analysis and dental microwear in an early medieval population from Gáň (Galanta district, Slovakia). *Anthropol. Anzeiger. J. Biol. Clin. Anthropol.* **70**, 229–248, DOI: [10.1127/0003-5548/2013/0256](https://doi.org/10.1127/0003-5548/2013/0256) (2013).
38. García-González, R., Sánchez-Puente, Z., Arsuaga, J. L. & Carretero, J. M. Dietary inferences from dental microwear patterns in chalcolithic populations from the iberian peninsula: the case of El Portalón de Cueva Mayor (Sierra de Atapuerca, Burgos, Spain) and El Alto de la Huesera (Álava, Spain). *Archaeol. Anthropol. Sci.* **11**, 3811–3823, DOI: [10.1007/s12520-018-0711-x](https://doi.org/10.1007/s12520-018-0711-x) (2019).
39. Pérez-Pérez, A., Espurz, V., de Castro, J. M. B., de Lumley, M. A. & Turbón, D. Non-occlusal dental microwear variability in a sample of middle and late pleistocene human populations from europe and the near east. *J. Hum. Evol.* **44**, 497–513, DOI: [10.1016/S0047-2484\(03\)00030-7](https://doi.org/10.1016/S0047-2484(03)00030-7) (2003).
40. Hernando, R. *et al.* Inferring childhood dietary maturation using buccal and occlusal deciduous molar microwear: a case study from the recent prehistory of the iberian peninsula. *Archaeol. Anthropol. Sci.* **12**, 1–13, DOI: [10.1007/s12520-019-00997-z](https://doi.org/10.1007/s12520-019-00997-z) (2020).
41. Lalueza, C., Pérez-Pérez, A. & Turbón, D. Dietary inferences through buccal microwear analysis of middle and late pleistocene human fossil. *Am. J. Biol. Anthropol.* **100**, 367–387, DOI: [10.1002/\(SICI\)1096-8644\(199607\)100:3<3C367::AID-AJPA5>3E3.0.CO;2-R](https://doi.org/10.1002/(SICI)1096-8644(199607)100:3<3C367::AID-AJPA5>3E3.0.CO;2-R) (1996).
42. Teaford, M. F., Maas, M. C. & Simons, E. L. Dental microwear and microstructure in early oligocene primates from the Faym, Egypt: Implications for diet. *Am. J. Phys. Anthropol.* **101**, 527–543, DOI: [10.1002/\(SICI\)1096-8644\(199612\)101:4%3C527::AID-AJPA7%3E3.0.CO;2-S](https://doi.org/10.1002/(SICI)1096-8644(199612)101:4%3C527::AID-AJPA7%3E3.0.CO;2-S) (1996).
43. Daegling, D. J. & Grine, F. E. Terrestrial foraging and dental microwear in papio ursinus. *Primates* **40**, 559–572, DOI: [10.1007/BF02574831](https://doi.org/10.1007/BF02574831) (1999).

44. King, T., Aiello, L. C. & Andrews, P. Dental microwear of *Griphopithecus alpani*. *J. Hum. Evol.* **36**, 3–31, DOI: [10.1006/jhev.1998.0258](https://doi.org/10.1006/jhev.1998.0258) (1999).
45. Merceron, G., Blondel, C. & L De Bonis, L. V., GD Koufos. A new method of dental microwear analysis: application to extant primates and ouranopithecus macedoniensis (late miocene of greece). *Palaios* **20**, 551–561, DOI: [10.2110/palo.2004.p04-17](https://doi.org/10.2110/palo.2004.p04-17) (2005).
46. Grine, F. E., Ungar, P. S., Teaford, M. F. & El-Zaatari, S. Molar microwear in *praeanthropus afarensis*: Evidence for dietary stasis through time and under diverse paleoecological conditions. *J. Hum. Evol.* **51**, 297–319, DOI: [10.10520/EJC96571](https://doi.org/10.10520/EJC96571) (2006).
47. Grine, F. E., Ungar, P. S. & Teaford, M. F. Was the early pliocene hominin "australopithecus" anamensis a hard object feeder? *South Afr. J. Sci.* **102**, 301–310, DOI: [10.10520/EJC96571](https://doi.org/10.10520/EJC96571) (2006).
48. Nava, A. *et al.* Multipronged dental analyses reveal dietary differences in last foragers and first farmers at Grotta Continenza, central Italy (15,500–7000 bp). *Sci. Reports* **11**, 4261, DOI: [10.1038/s41598-021-82401-2](https://doi.org/10.1038/s41598-021-82401-2) (2021).
49. Vidal, R., Ma, Y. & Sastry, S. S. *Principal Component Analysis* (Springer, 2016).
50. Bishop, C. M. *Pattern Recognition and Machine Learning (Information Science and Statistics)* (Springer, 2007), 1 edn.
51. Petraru, O.-M., Groza, V.-M., Lobiuc, A., Bejenaru, L. & Popovici, M. Dental microwear as a diet indicator in the seventeenth-century human population from iasi city, romania. *Archaeol. Anthropol. Sci.* **12**, 184, DOI: [10.1007/s12520-020-01159-2](https://doi.org/10.1007/s12520-020-01159-2) (2020).
52. Merceron, G. & Madelaine, S. Molar microwear pattern and palaeoecology of ungulates from La Berbie (Dordogne, France): environment of neanderthals and modern human populations of the middle/upper palaeolithic. *Boreas* **35**, 272–278, DOI: [0.1080/03009480600584873](https://doi.org/10.1080/03009480600584873) (2006).
53. Rivals, F., Solounias, N. & Mithlacher, M. C. Evidence for geographic variation in the diets of late pleistocene and early holocene bison in north america, and differences from the diets of recent bison. *Quat. Res.* **68**, 338–346, DOI: [10.1016/j.yqres.2007.07.012](https://doi.org/10.1016/j.yqres.2007.07.012) (2007).
54. Rivals, F., Solounias, N. & Schaller, G. B. Diet of mongolian gazelles and tibetan antelopes from steppe habitats using premaxillary shape, tooth mesowear and microwear analyses. *Mammalian Biol.* **76**, 358–364, DOI: [10.1016/j.mambio.2011.01.005](https://doi.org/10.1016/j.mambio.2011.01.005) (2011).
55. Pappa, S., Schreve, D. C. & Rivals, F. The bear necessities: A new dental microwear database for the interpretation of palaeodiet in fossil ursidae. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **514**, 168–188, DOI: [10.1016/j.palaeo.2018.10.015](https://doi.org/10.1016/j.palaeo.2018.10.015) (2019).
56. Ramírez-Pedraza, I. *et al.* Microwear and isotopic analyses on cave bear remains from toll cave reveal both short-term and long-term dietary habits). *Sci. Reports* **9**, 1–16, DOI: [10.1038/s41598-019-42152-7](https://doi.org/10.1038/s41598-019-42152-7) (2019).
57. Ungar, P., Krueger, K. L., Blumenscine, R. J., Njau, J. & Scott, R. S. Dental microwear texture analysis of hominins recovered by the olduvai landscape paleoanthropology project, 1995-2007. *J. Hum. Evol.* **63**, 429–437, DOI: [10.1016/j.jhev.2011.04.006](https://doi.org/10.1016/j.jhev.2011.04.006) (2012).
58. Delezene, L. K. *et al.* Premolar microwear and tooth use in *Australopithecus afarensis*. *Journal of Hum. Evol.* **65**, 282–293, DOI: [10.1016/j.jhev.2013.06.001](https://doi.org/10.1016/j.jhev.2013.06.001) (2013).
59. Aliaga-Martínez, A., Romero, A., Galbany, J., Hernández-Aguilar, R. A. & Pérez-Pérez, A. Buccal dental microwear texture and catarrhine diets. *Am. J. Phys. Anthropol.* **163**, 462–473, DOI: [10.1002/ajpa.23219](https://doi.org/10.1002/ajpa.23219) (2017).
60. Ungar, P. S., Scott, J. R. & Steininger, C. M. Dental microwear differences between eastern and southern african fossil bovids and hominins. *South Afr. J. Sci.* **112**, 1–5, DOI: [10.17159/sajs.2016/20150393](https://doi.org/10.17159/sajs.2016/20150393) (2016).
61. Martin, F., Plastiras, C.-A., Souron, A. & Boissier, J.-R. Dietary niches of terrestrial cercopithecines from the plio-pleistocene shungura formation, ethiopia: evidence from dental microwear texture analysis. *Sci. Reports* **8**, 14502, DOI: [10.1038/s41598-018-32092-z](https://doi.org/10.1038/s41598-018-32092-z) (2018).

62. Schmidt, C. W. *et al.* Dental microwear texture analysis of homo sapiens sapiens: foragers, farmers, and pastoralists. *Am. J. Phys. Anthropol.* **169**, 207–226, DOI: [10.1002/ajpa.23815](https://doi.org/10.1002/ajpa.23815) (2019).
63. Williams, F. L. *et al.* Dietary reconstruction of Spy I using dental microwear texture analysis. *Comptes Rendus Palevol* **18**, 1083–1094, DOI: [10.1016/j.crpv.2019.06.004](https://doi.org/10.1016/j.crpv.2019.06.004) (2019).
64. Williams, F. L. *et al.* Reconstructing the diet of Kůlna 1 from the Moravian karst (Czech Republic). *J. Paleolithic Archaeol.* **4**, 1–23, DOI: [10.1007/s41982-021-00099-0](https://doi.org/10.1007/s41982-021-00099-0) (2021).
65. Schulz-Kornas, E., Stuhlträger, J., Clauss, M., Wittig, R. M. & Kupczik, K. Dust affects chewing efficiency and tooth wear in forest dwelling western chimpanzees (pan troglodytes verus). *Am. J. Phys. Anthropol.* **169**, 66–77, DOI: [10.1002/ajpa.23808](https://doi.org/10.1002/ajpa.23808) (2019).
66. Merceron, G. *et al.* Further away with dental microwear analysis: Food resource partitioning among plio-pleistocene monkeys from the shungura formation, ethiopia. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **572**, 110414, DOI: [10.1016/j.palaeo.2021.110414](https://doi.org/10.1016/j.palaeo.2021.110414) (2021).
67. Stuhlträger, J. *et al.* Dental wear patterns reveal dietary ecology and season of death in a historical chimpanzee population. *PLoS One* **16**, e0251309, DOI: [10.1371/journal.pone.0251309](https://doi.org/10.1371/journal.pone.0251309) (2021).
68. DeSantis, L. Dental microwear textures: Reconstructing diets of fossil mammals. *Surf. Topogr. Metrol. Prop* **4**, DOI: [10.1088/2051-672X/4/2/023002](https://doi.org/10.1088/2051-672X/4/2/023002) (2016).
69. Purnell, M. A., Crumpton, N., Gill, P., Jones, G. & Rayfield, E. J. Within-guild dietary discrimination from 3-d textural analysis of tooth microwear in insectivorous mammals. *J. Zool.* **291**, 249–257, DOI: [10.1111/jzo.12068](https://doi.org/10.1111/jzo.12068) (2013).
70. Purnell, M. A. & Darras, L. P. 3d tooth microwear texture analysis in fishes as a test of dietary hypotheses of durophagy. *Surf. Topogr. Metrol. Prop.* **4**, 014006, DOI: [0.1088/2051-672X/4/1/014006](https://doi.org/10.1088/2051-672X/4/1/014006) (2016).
71. Winkler, D. E. *et al.* Dental microwear texture gradients in guinea pigs reveal that material properties of the diet affect chewing behaviour. *J. Exp. Biol.* **224**, DOI: [10.1242/jeb.242446](https://doi.org/10.1242/jeb.242446) (2021).
72. Miyamoto, K., Kubo, M. O. & Yokohata, Y. The dental microwear texture of wild boars from japan reflects inter- and intra-population feeding preferences. *Front. Ecol. Evol.* **10**, 110414, DOI: [1576460.3389/fevo.2022.9](https://doi.org/10.3389/fevo.2022.9) (2022).
73. Sato, K., Sato, T. & Kubo, M. Reconstructing diets of hunted sika deer from torihama shell midden site (ca. 6,000 years ago) by dental microwear texture analysis. *Front. Ecol. Evol.* **10**, DOI: [10.3389/fevo.2022.957038](https://doi.org/10.3389/fevo.2022.957038) (2022).
74. Calandra, I., Schulz, E., Pinnov, M., Khron, S. & Kaiser, T. M. Teasing apart the contributions of hard dietary items on 3d dental microtextures in primates. *J. Hum. Evol.* **63**, 85–98, DOI: [10.1016/j.jhevol.2012.05.001](https://doi.org/10.1016/j.jhevol.2012.05.001) (2012).
75. Francisco, A., Brunetière, N. & Merceron, G. Gathering and analyzing surface parameters for diet identification purposes. *Technologies* **6**, 1–21, DOI: [10.3390/technologies6030075](https://doi.org/10.3390/technologies6030075) (2018).
76. Francisco, A., Blonde, C., Brunetière, N., Ramdarshan, A. & Merceron, G. Enamel surface topography analysis for diet discrimination. a methodology to enhance and select discriminative parameters. *Surf. Topogr. Metrol. Prop.* **6**, 015002, DOI: [10.1088/2051-672X/aa9dd3](https://doi.org/10.1088/2051-672X/aa9dd3) (2018).
77. Miller, T. Explanation in artificial intelligence: Insights from the social sciences. *Artif. Intell.* **267**, 1–38, DOI: [10.1016/j.artint.2018.07.007](https://doi.org/10.1016/j.artint.2018.07.007) (2019).
78. Schulz, E., Calandra, I. & Kaiser, T. Feeding ecology and chewing mechanics in hoofed mammals: 3d tribology of enamel wear. *Wear* **300**, 169–179, DOI: [doi:10.1016/j.wear.2013.01.115](https://doi.org/10.1016/j.wear.2013.01.115) (2013).
79. Hair, J. F. *Multivariate data analysis* (Prentice Hall, 2009).
80. Martin, L. F. *et al.* Dental microwear texture analysis correlations in guinea pigs (cavia porcellus) and sheep (ovis aries) suggest that dental microwear texture signal consistency is species-specific. *Front. Ecol. Evol.* **10**, DOI: [10.3389/fevo.2022.958576](https://doi.org/10.3389/fevo.2022.958576) (2022).
81. James, G., Witten, D., Hastie, T., Tibshirani, R. & Taylor, J. *Statistical Learning*, 15–67 (Springer International Publishing, Cham, 2023).

82. Aiba, K., Miura, S. & Kubo, M. O. Dental microwear texture analysis in two ruminants, Japanese serow (*Capricornis capripus*) and sika deer (*Cervus nippon*), from central Japan. *Mammal Study* **44**, 193–192, DOI: [10.3106/ms2018-00](https://doi.org/10.3106/ms2018-00) (2019).
83. Winkler, D. E., E. S.-K., Kaiser, T. & Tütken, T. Dental microwear texture reflects dietary tendencies in extant lepidosaurs despite their limited use of oral food processing. *Proc. Royal Soc. Lond. B* **286**, 20190544, DOI: [10.1098/rspb.2019.0544](https://doi.org/10.1098/rspb.2019.0544) (2019).
84. Schulz-Kornas, E. *et al.* Everything matters: Molar microwear texture in goats (*Capra aegagrus hircus*) fed diets of different abrasiveness. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **552**, 66–77, DOI: [10.1016/j.palaeo.2020.109783](https://doi.org/10.1016/j.palaeo.2020.109783) (2020).
85. Purnell, M. A., Seehausen, O. & Galis, F. Quantitative three-dimensional microtextural analyses of tooth wear as a tool for dietary discrimination in fishes. *Interface* **9**, 2225–2233, DOI: [10.1098/rsif.2012.0140](https://doi.org/10.1098/rsif.2012.0140) (2012).
86. Bestwick, J., Unwin, D. M., Butler, R. J. & Purnell, M. A. Dietary diversity and evolution of the earliest flying vertebrates revealed by dental microwear texture analysis. *Nat. Commun.* **11**, 5293, DOI: [10.1038/s41467-020-19022-2](https://doi.org/10.1038/s41467-020-19022-2) (2021).
87. Scott, R. S. *et al.* Dental microwear texture analysis: technical considerations. *J. Hum. Evol.* **51**, 339–349, DOI: [10.1016/j.jhevol.2006.04.006](https://doi.org/10.1016/j.jhevol.2006.04.006) (2006).
88. Shearer, B. M. *et al.* Dental microwear profilometry of African non-cercopithecoid catarrhines of the early Miocene. *J. Hum. Evol.* **78**, 33–43, DOI: [10.1016/j.jhevol.2014.08.011](https://doi.org/10.1016/j.jhevol.2014.08.011) (2015).
89. Martínez, L. M., Roig, A. E. D., Alrousan, M. & Estebananz-Sánchez, F. On sample size and buccal enamel preservation in dental microwear. *AWRANA Congr.* (2022).
90. Kaiser, T. M., Clauss, M. & Schulz-Kornas, E. A set of hypotheses on tribology of mammalian herbivore teeth. *Surf. Topogr.: Metrol. Prop.* **4**, DOI: [10.1088/2051-672X/4/1/014003](https://doi.org/10.1088/2051-672X/4/1/014003) (2016).
91. Mihlbachler, M. C., Rusnack, F. & Beatty, B. L. Experimental approaches to assess the effect of composition of abrasives in the cause of dental microwear. *R. Soc. Open Sci.* 1–38, DOI: [10.1098/rsos.211549](https://doi.org/10.1098/rsos.211549) (2022).
92. Kubo, M. O. & Fujita, M. Diets of Pleistocene insular dwarf deer revealed by dental microwear texture analysis. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **20**, 110098, DOI: [10.1016/j.palaeo.2020.110098](https://doi.org/10.1016/j.palaeo.2020.110098) (2021).
93. Ungar, P. S. *et al.* Incisor microwear of arctic rodents as a proxy for microhabitat preference. *Mammalian Biol.* **101**, 133–152, DOI: [10.1007/s42991-021-00138-x](https://doi.org/10.1007/s42991-021-00138-x) (2021).
94. Sergio, J.-M., Gourichon, L., Muñiz, J. & Ibáñez, J. J. Discriminating dietary behaviour between wild and domestic goats using dental microwear texture: first results from a modern reference set and early Neolithic goat exploitation in the southern Levant. *J. Archaeol. Sci.* **155**, 105779, DOI: [10.1016/j.jas.2023.105779](https://doi.org/10.1016/j.jas.2023.105779) (2023).
95. Gurtov, A. *Dental Microwear Texture Analysis of Early Pleistocene Hominin Prey: A Test of Foraging Seasonality*. Ph.D. thesis, UNIVERSITY OF WISCONSIN – MADISON (2016).
96. Jiménez-Manchón, S., Émilie Blaise & Gardeisen, A. Exploring low-magnification dental microwear of domestic ungulates: Qualitative observations to infer palaeodiets. *Quat. Int.* **557**, 12–22, DOI: [10.1016/j.quaint.2020.02.021](https://doi.org/10.1016/j.quaint.2020.02.021) (2020).
97. Pagani, L., Qi, Q., Jing, X. & Scott, P. J. Towards a new definition of areal surface texture parameters on freeform surface. *Measurement* **109**, 281–291, DOI: [10.1016/j.measurement.2017.05.028](https://doi.org/10.1016/j.measurement.2017.05.028) (2017).
98. Marinello, F. & Pezzuolo, A. Application of ISO 25178 standard for multiscale 3D parametric assessment of surface topographies. *IOP Conf. Series: Earth Environ. Sci.* **275**, 012011, DOI: [10.1088/1755-1315/275/1/012011](https://doi.org/10.1088/1755-1315/275/1/012011) (2019).
99. Ragni, A. J., Teaford, M. F. & Ungar, P. S. A molar microwear texture analysis of pitheciid primates. *Am. J. Primatol.* **79**, e22697, DOI: [10.1002/ajp.22697](https://doi.org/10.1002/ajp.22697) (2017).
100. Winkler, D. E. *et al.* Indications for a dietary change in the extinct Bovid genus *Myotragus* (Plio-Holocene, Mallorca, Spain). *Geobios* **46**, 143–150, DOI: [10.1016/j.geobios.2012.10.010](https://doi.org/10.1016/j.geobios.2012.10.010) (2013).
101. Martin, L. F. *et al.* Dental wear at macro- and microscopic scale in rabbits fed diets of different abrasiveness: A pilot investigation. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **556**, DOI: [10.1016/j.palaeo.2020.109886](https://doi.org/10.1016/j.palaeo.2020.109886) (2020).

102. Arman, S. D., Gully, G. A. & Prideaux, G. J. Dietary breadth in kangaroos facilitated resilience to quaternary climatic variations. *Science* **387**, 167–171, DOI: [10.1126/science.adq4340](https://doi.org/10.1126/science.adq4340) (2025).
103. Ibáñez, J. J., Lazuen, T. & González-Urquijo, J. Identifying experimental tool use through confocal microscopy. *J. Archaeol. Method Theory* **26**, 1176–1215, DOI: [10.1007/s10816-018-9408-9](https://doi.org/10.1007/s10816-018-9408-9) (2019).
104. Ibáñez, J. J. & Mazzucco, N. Quantitative use-wear analysis of stone tools: Measuring how the intensity of use affects the identification of the worked material. *PLoS One* **16**, e0257266, DOI: [10.1371/journal.pone.0257266](https://doi.org/10.1371/journal.pone.0257266) (2021).
105. Ungar, P. & Evans, A. Exposing the past: surface topography and texture of paleontological and archeological remains. *Surf. Topogr. Metrol. Prop.* **4**, 040302, DOI: [10.1088/2051-672X/4/4/040302](https://doi.org/10.1088/2051-672X/4/4/040302) (2016).
106. Gordon, K. A study of microwear on chimpanzee molars: Implications for dental microwear analysis. *Am. J. Phys. Anthropol.* **59**, 195–215, DOI: [10.1002/ajpa.1330590208](https://doi.org/10.1002/ajpa.1330590208) (1982).
107. Ungar, P. S., Brown, C. A., Bergstrom, T. S. & Walker, A. Quantification of dental microwear by tandem scanning confocal microscopy and scale-sensitive fractal analyses. *Scanning* **25**, 185–193, DOI: [10.1002/sca.4950250405](https://doi.org/10.1002/sca.4950250405) (2003).
108. Calandra, I. *et al.* Surface texture analysis in toothfrax and mountainsmap® ssfa module: Different software packages, different results? *Peer Community Journal: section Archaeol.* **2**, DOI: [doi.org/10.24072/pci.archaeo.100024](https://doi.org/10.24072/pci.archaeo.100024) (2022).
109. Schulz, E., Calandra, I. & Kaiser, T. M. Applying tribology to teeth of hoofed mammals. *Scanning* **32**, 162–182, DOI: [10.1002/sca.20181](https://doi.org/10.1002/sca.20181) (2010).
110. Merceron, G. *et al.* Dental microwear and controlled food testing on sheep: The trident project. *Biosurface Biotribology* **3**, 174–561, DOI: [10.1016/j.bsbt.2017.12.005](https://doi.org/10.1016/j.bsbt.2017.12.005) (2017).
111. Arman, S. D., Prowse, T. A., Couzens, A. M., Ungar, P. S. & Prideaux, G. J. Incorporating intraspecific variation into dental microwear texture analysis. *Interface* **153**, 20180957, DOI: [10.1098/rsif.2018.0957](https://doi.org/10.1098/rsif.2018.0957) (2019).
112. Ibáñez, J. J., González-Urquijo, J. E. & Gibaja, J. Discriminating wild vs domestic cereal harvesting micropolish through laser confocal microscopy. *J. Archaeol. Sci.* **48**, 96–103, DOI: [10.1016/j.jas.2013.10.012](https://doi.org/10.1016/j.jas.2013.10.012) (2014).
113. Ibáñez, J. J., Anderson, P. C., González-Urquijo, J. & Gibaja, J. Cereal cultivation and domestication as shown by microtexture analysis of sickle gloss through confocal microscopy. *J. Archaeol. Sci.* **73**, 62–81, DOI: [10.1016/j.jas.2016.07.011](https://doi.org/10.1016/j.jas.2016.07.011) (2016).
114. Anderson, T. W. & Darling, D. A. A test of goodness of fit. *J. Am. statistical association* **49**, 765–769 (1954).
115. Yeo, I.-K. & Johnson, R. A. A new family of power transformations to improve normality or symmetry. *Biometrika* **87**, 954–959, DOI: [10.1093/biomet/87.4.954](https://doi.org/10.1093/biomet/87.4.954) (2000).
116. Kuhn, M. & Johnson, K. *Feature Engineering and Selection: A Practical Approach for Predictive Models* (Chapman and Hall/CRC, 2019).
117. Stine, R. A. Graphical interpretation of Variance Inflation Factors. *The Am. Stat.* **49**, 53–56, DOI: [10.1080/00031305.1995.10476113](https://doi.org/10.1080/00031305.1995.10476113) (1995).
118. Wolpert, D. H. The lack of a priori distinctions between learning algorithms. *Neural computation* **8**, 1341–1390, DOI: [10.1162/neco.1996.8.7.1341](https://doi.org/10.1162/neco.1996.8.7.1341) (1996).
119. Chen, T. & Guestrin, C. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 785–794, DOI: [10.1145/2939672.2939785](https://doi.org/10.1145/2939672.2939785) (2016).
120. Duan, T. *et al.* NGBoost: Natural Gradient Boosting for Probabilistic Prediction. In *International Conference on Machine Learning*, 2690–2700, DOI: [10.48550/arXiv.1910.03225](https://doi.org/10.48550/arXiv.1910.03225) (PMLR, 2020).
121. Gallego, V. & Rios Insua, D. Current advances in neural networks. *Annu. Rev. Stat. Its Appl.* **9**, 197–222, DOI: <https://dx.doi.org/10.1146/annurev-statistics-040220-112019> (2022).
122. Jiménez-Manchón, S. *et al.* Comparative analysis of confocal microscopy objective magnifications on dental microwear texture analysis. implications for dietary reconstruction in caprines. *J. Archaeol. Sci. Reports* **58**, 104716, DOI: [10.1016/j.jasrep.2024.104716](https://doi.org/10.1016/j.jasrep.2024.104716) (2024).

123. Biau, G. & Scornet, E. A random forest guided tour. *Test* **25**, 197–227 (2016).
124. Natekin, A. & Knoll, A. Gradient boosting machines, a tutorial. *Front. neurorobotics* **7**, 21 (2013).
125. Freund, Y. & Schapire, R. E. A decision-theoretic generalization of on-line learning and an application to boosting. *J. computer system sciences* **55**, 119–139 (1997).
126. Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

## Author Contributions

All authors contributed to the study conception and design. Material preparation and data collection were performed by LM, FES and JJIE, whereas analysis was performed by SRS, KK and DRI. The first draft of the manuscript was written by FES, LMM, SRS and DRI, and all authors commented on later versions of the manuscript. All authors read and approved the final manuscript.

## Additional Information

**Competing Interests** The authors declare no competing interests. All data generated or analysed during this study are included in this published article (and its supplementary information files).

**Acknowledgements** The authors are grateful to all the curators and technical personnel of the different institutions where the specimens were molded.

**Data Availability** The datasets analysed during the current study are publicly available as part of previously published research on extant primates. All relevant datasets, along with the code used to conduct the analyses, are included in the Supplementary Information files and are also available in the code repository (*to be released upon time of acceptance*). Regional collaborators contributed from the outset, ensuring local ecological and methodological perspectives were incorporated.

**Funding** This research was supported by Ministerio de Ciencia e Innovación, Grant Numbers: PDC2021-121613-I00, PID2021-124662OB-I00, funded by MCIN/AEI/10.13039/501100011033 and the European Union “NextGeneration EU/ORTR”, PID2020-112963GB-I00 and PID2023-148818NB-I00 by ERDF A way of making Europe, by the European Union.