

Automated multi-view silhouette generation and morphometric clustering of isolated theropod teeth using machine learning

CAROLINA S. MARQUES^{1*}
EMMANUEL DUFOURQ²
SORAIA PEREIRA^{1,3,4}

VANDA F. SANTOS^{5,6,7}
ELISABETE MALAFAIA^{5,6,8}

1. Centro de Estatística e Aplicações (CEAUL), Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal.
2. Department of Computer Science, University of Northern Iowa, Cedar Falls, United States of America.
3. Centro de Matemática (CMAT), Universidade do Minho, Guimarães, Portugal.
4. Departamento de Matemática, Universidade do Minho, Guimarães, Portugal.
5. Instituto Dom Luiz (IDL), Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal.
6. Departamento de Ciências da Terra e Energia, Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal.
7. Paleolbérica Research Group, University of Alcalá, Spain.
8. Museu Nacional de História Natural e da Ciência (MUHNAC), Universidade de Lisboa, Lisboa, Portugal.

*Corresponding author

Recibido: 30 de mayo 2026 - Aceptado: 22 de julio 2026 - Publicado: 27 de octubre 2026

Para citar este artículo: Carolina S. Marques, Emmanuel Dufourq, Soraia Pereira, Vanda F. Santos, & Elisabete Malafaia (2026). Automated multi-view silhouette generation and morphometric clustering of isolated theropod teeth using machine learning. *Publicación Electrónica de la Asociación Paleontológica Argentina* 26(3): 15–43.

Link a este artículo: <http://dx.doi.org/10.5710/PEAPA.22.07.2026.582>

©2026 Marques, Dufourq, Pereira, Santos, Malafaia

AUTOMATED MULTI-VIEW SILHOUETTE GENERATION AND MORPHOMETRIC CLUSTERING OF ISOLATED THEROPOD TEETH USING MACHINE LEARNING

CAROLINA S. MARQUES^{1*}, EMMANUEL DUFOURQ², SORAIA PEREIRA^{1,3,4}, VANDA F. SANTOS^{5,6,7}, AND ELISABETE MALAFAIA^{5,6,8}

¹Centro de Estatística e Aplicações (CEAUL), Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal. csmarques@ciencias.ulisboa.pt; soraia.pereira@math.uminho.pt

²Department of Computer Science, University of Northern Iowa, Cedar Falls, United States of America. emmanuel@cs.uni.edu

³Centro de Matemática (CMAT), Universidade do Minho, Guimarães, Portugal.

⁴Departamento de Matemática, Universidade do Minho, Guimarães, Portugal.

⁵Instituto Dom Luiz (IDL), Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal. vafsantos@ciencias.ulisboa.pt; efmalafaia@ciencias.ulisboa.pt

⁶Departamento de Ciências da Terra e Energia, Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal.

⁷Paleolbérica Research Group, University of Alcalá, Spain.

⁸Museu Nacional de História Natural e da Ciência (MUHNAC), Universidade de Lisboa, Lisboa, Portugal.

*Corresponding author

 CSM: <https://orcid.org/0000-0002-5936-9342>; ED: <https://orcid.org/0000-0002-6986-3408>; SP: <https://orcid.org/0000-0002-7336-1320>; VFS: <https://orcid.org/0000-0001-5842-7857>; EM: <https://orcid.org/0000-0003-4894-4257>

Abstract. Isolated theropod teeth are abundant in mesozoic vertebrate assemblages and often taxonomically informative, but their description still relies on manual measurements and expert-defined characters. We present an automated computer-vision workflow that converts raw multi-view photographs of isolated theropod teeth into standardized binary silhouettes in labial, lingual, mesial, distal, and basal cross-sectional views. A segmentation model initialized from dinosaur-track image binarization was used and then finetuned on tooth-like shapes. The model was applied to 74 theropod teeth (10 for testing) from Guimarota, a Late Jurassic locality in Portugal, containing one of the most important assemblages of isolated small theropod teeth in Europe. From the silhouettes, we extracted width, height, and contour-angle descriptors from whole crown and upper, middle, and lower crown regions. Dimensionality reduction and density-based clustering (HDBSCAN) were used to explore the measurement space, and cluster assignments were interpreted with random forest feature importance and decision trees. The analysis used 64 independent teeth represented by 1280 derived view-region records; the most influential variables included height and width measurements. HDBSCAN identified five clusters and eleven unassigned specimens, suggesting morphological heterogeneity. Cluster-assigned silhouettes were then used to train an EfficientNet model on 53 assigned teeth. As an exploratory internal-consistency check, the model recovered cluster-derived classes with high mean performance under repeated stratified cross-validation, although with some fold-to-fold variation, especially in the least represented cluster. Gradient-weighted Class Activation Mapping (Grad-CAM) highlighted anatomically meaningful silhouette regions, including apices, lateral contours, basal margins, and cross-sectional perimeter. This workflow provides a scalable framework for reproducible theropod tooth morphometrics.

Key words. Guimarota Theropoda isolated teeth. Semantic segmentation. Transfer learning. HDBSCAN. UMAP dimensionality reduction. Random forest. Convolutional neural network.

Resumen. GENERACIÓN AUTOMATIZADA DE SILUETAS EN MÚLTIPLES VISTAS Y AGRUPAMIENTO MORFOMÉTRICO DE DIENTES AISLADOS DE TERÓPODOS MEDIANTE APRENDIZAJE AUTOMÁTICO. Los dientes aislados de terópodos son abundantes en los registros de vertebrados mesozoicos y pueden ser taxonómicamente informativos, pero su estudio depende de mediciones manuales. Presentamos un flujo de trabajo automatizado de visión por computadora que convierte fotografías multivista de dientes aislados en siluetas binarias estandarizadas. Usamos una arquitectura de segmentación de imágenes por aprendizaje profundo. El modelo se aplicó a 74 dientes aislados de terópodos (diez para prueba) de Guimarota, una localidad clásica del Jurásico Superior en Portugal, que ha proporcionado uno de los conjuntos más importantes de dientes aislados de pequeños terópodos en Europa. A partir de las siluetas se extrajeron ancho, alto y descriptores angulares del contorno completo de los dientes y de las regiones superior, media e inferior de la corona. El análisis se basó en 64 dientes independientes representados por 1280 registros de regiones de vista derivados. La reducción de dimensionalidad y el agrupamiento basado en densidad identificaron cinco agrupamientos y once especímenes no asignados, lo que sugiere heterogeneidad morfológica. Las asignaciones se interpretaron mediante importancia de variables de bosque aleatorio y árboles de decisión. Como comprobación exploratoria de consistencia interna un modelo EfficientNet multirrama recuperó las clases derivadas de los agrupamientos con alta precisión media en validación cruzada estratificada repetida, aunque con variación entre pliegues, especialmente en el agrupamiento menos representado. Las visualizaciones Grad-CAM destacaron regiones anatómicamente relevantes, como ápices, contornos laterales, márgenes basales y perímetros transversales. Este flujo de trabajo proporciona un marco escalable para la morfometría dental reproducible de terópodos.

Palabras clave. Dientes aislados de terópodos de Guimarota. Segmentación semántica. Aprendizaje por transferencia. HDBSCAN. Reducción de dimensionalidad UMAP. Bosque aleatorio. Red neuronal convolucional.

INTRODUCTION

Isolated theropod teeth are among the most frequently recovered dinosaur remains in many Mesozoic vertebrate microfossil assemblages. Their abundance makes them especially valuable where skeletal remains are rare, fragmentary, or taxonomically ambiguous. The Upper Jurassic Guimarota coal mine near Leiria, Portugal, is a classic example where hundreds of isolated small theropod teeth have been described, representing several morphotypes. These morphotypes include forms compared with the early-branching avialan *Archaeopteryx* von Meyer, 1861 (see, e.g., Louchart & Pouech, 2017, for a discussion of the interpretation of these specimens), *Compsognathus* Wagner, 1859, and other small-bodied coelurosaurians, as well as more early-branching forms, including allosaurian and ceratosaurian theropods (e.g., Zinke, 1998; Rauhut et al., 2000).

Theropod tooth studies have long relied on standardized anatomical terminology and morphometric variables such as crown height, crown-base length, crown-base width, curvature, position and extent of the carinae, denticle density, crown compression and elongation ratios. Hendrickx et al. (2015) formalized terminology for theropod teeth and emphasized that isolated teeth preserve taxonomically important anatomical information, even when skeletal association is absent. Quantitative studies have shown that multivariate analyses of isolated theropod teeth can recover meaningful ecological and taxonomic signals, including differences among small theropod tooth morphotypes and temporal patterns of faunal turnover (Larson & Currie, 2013). Similar approaches have been applied to Late Jurassic theropod tooth assemblages using multivariate and cladistic methods, reinforcing the utility of isolated teeth for reconstructing faunal diversity (Gerke & Wings, 2016).

Despite this progress, most tooth morphometric workflows remain completely manual and measured differently by researchers (e.g., Hendrickx et al., 2019). Researchers commonly trace outlines, identify landmarks, orient specimens, and measure features using interactive image software. This introduces several limitations: measurements are time-consuming, observer-dependent, and difficult to replicate across large collections; small differences in image orientation or contrast can alter outline-based measurements; and multi-view information is rarely integrated into

a single computational pipeline. These limitations are particularly important for isolated teeth, where taxonomic information may be encoded in subtle differences among labial, lingual, mesial, distal, and basal cross-sectional (hereafter referred to as 'cross-sectional') profiles.

Machine learning and deep learning are increasingly used in paleontology to reduce such bottlenecks. Recent reviews note that AI applications in paleontology now include fossil classification, segmentation, prediction, and image-based identification, although manual workflows still dominate many areas of the field (Yu et al., 2024). Transfer learning and image-classification examples in fossil or dental material include bovid tooth classification (Domínguez-Rodrigo et al., 2023), fossil shark tooth identification (Barucci et al., 2024), and isolated theropod tooth classification from morphometric datasets (Wills et al., 2021; Marques et al., 2025). Deep learning has been applied to fossil CT segmentation, including dinosaur fossil segmentation with U-Net and modified DeepLabv3+ architectures (Yu et al., 2022). Large-scale fossil image classification has also been demonstrated using transfer learning on a Fossil Image Dataset of more than 415,000 images, with reported test accuracy around 90% (Liu et al., 2023). These studies show that computer vision can support paleontological classification and segmentation, but few methods directly target automated multi-view measurement of theropod teeth.

The segmentation of theropod tooth images presents a specific challenge because crown margins, denticles, and other diagnostic structures can be narrow, irregular, or difficult to distinguish from the background depending on the viewpoint. Preserving these boundaries is important to retain morphological information for subsequent analyses. DeepLabv3+ is well suited to this problem because it is designed to preserve semantic information while recovering sharper object boundaries (Chen et al., 2018).

Transfer learning refers to the practice of starting from a model that has already learned general visual features from a large image dataset, and then adapting it to a more specific task. This approach is appropriate because paleontological image datasets are often small, whereas simulated data can provide large numbers of controlled training examples. Pan and Yang (2010) describe precisely this situa-

tion, where knowledge from a source domain can improve performance in a related target domain where labelled data are limited.

Here we present a methodology for AI-generated multi-view silhouettes to automate morphometric feature extraction from isolated theropod teeth (Fig. 1). Our study has five main aims: first, to generate standardized silhouettes in binarized form from multiple tooth-viewpoint photographs; second, to extract reproducible measurements and angular descriptors from the silhouettes; third, to explore morphometric structure using Uniform Manifold Approximation and Projection (UMAP; McInnes et al., 2018), and Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN; McInnes et al., 2017); fourth, to use random forests to identify which automatically extracted variables contribute most strongly to cluster separation; and fifth, to evaluate, as an internal

consistency check, whether HDBSCAN-derived morphometric groups can be recovered from five-view tooth images using a multi-branch EfficientNet model (Tan & Le, 2019), and interpreted using Gradient-weighted Class Activation Mapping (Grad-CAM; Selvaraju et al., 2017) visualizations. The approach is novel because it links simulated-data transfer learning, DeepLabv3+ segmentation, multi-view theropod tooth silhouettes, automated measurement extraction, unsupervised clustering, and supervised image classification.

MATERIALS AND METHODS

Case-study material and image organization

The case-study dataset comprises 74 isolated theropod tooth crowns from the Guimarota coal-mine assemblage near Leiria, Portugal, recovered from Kimmeridgian deposits of the Alcobaça Formation (Zinke, 1998; Fürsich et al., 2022),

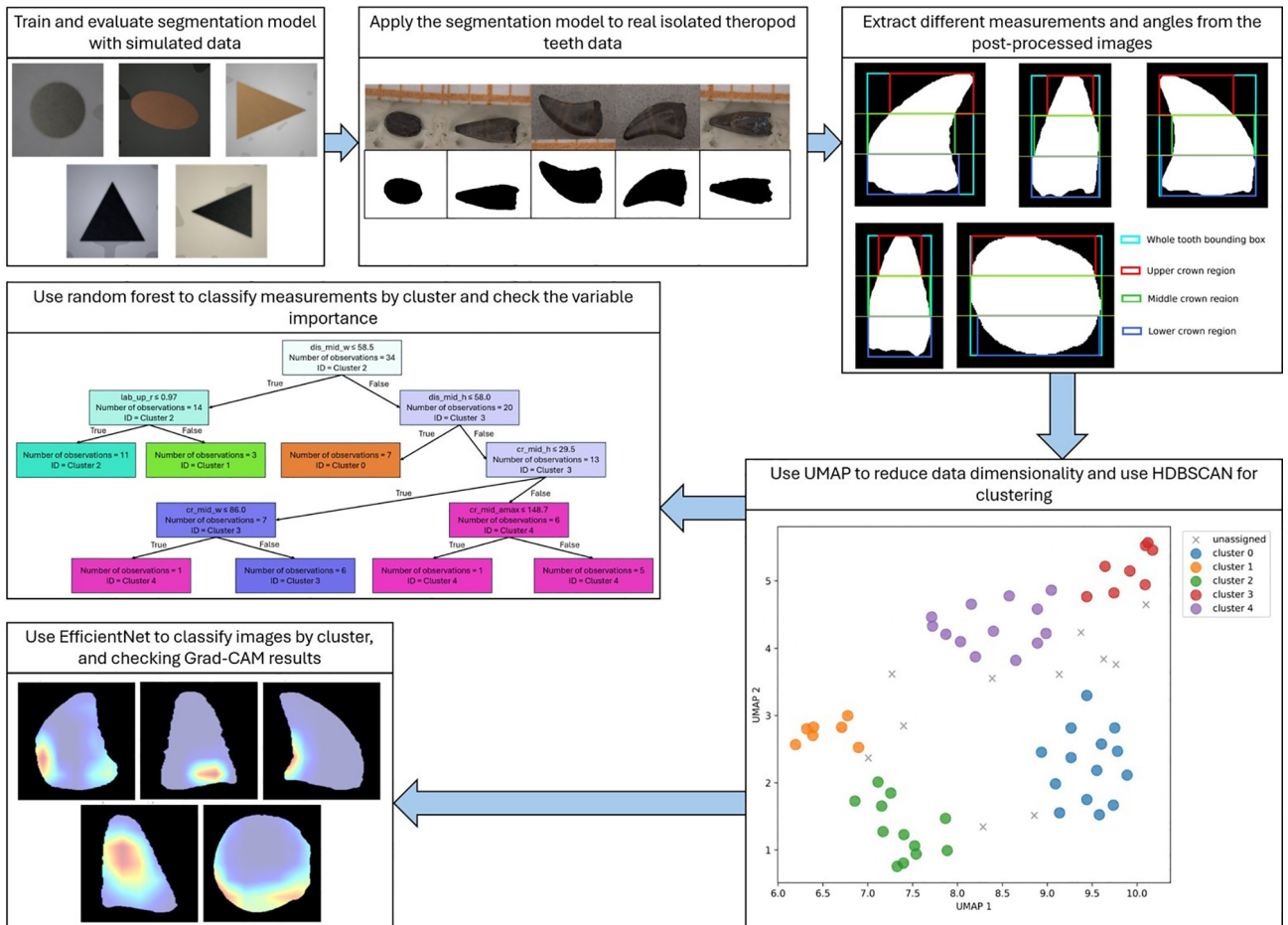


Figure 1. Workflow used for automated multi-view theropod tooth silhouette generation, morphometric clustering, and classifier interpretation.

where 10 teeth are left out from the whole analysis and used as a test set. Guimarota is a microvertebrate-rich Upper Jurassic locality in which dinosaur remains are dominated by isolated teeth rather than associated cranial or postcranial material (Zinke, 1998; Martin & Krebs, 2000; Rauhut et al., 2000). Previous work on the Guimarota theropod teeth recognized a taxonomically diverse assemblage of small and medium-sized theropods, including teeth attributed to *Compsognathus*, allosaurid and ceratosaurid morphotypes, *Richardoestesia* Currie, Rigby & Sloan, 1990, and *Paronychodon* Cope, 1876, as well as morphotypes tentatively referred to dromaeosaurids, troodontids, and tyrannosauroids (Zinke, 1998; Rauhut et al., 2000). Even though previous studies have identified a high diversity of theropod tooth morphotypes, these were primarily based on morphological comparisons, and the review of this collection based on more recent methodologies, such as discriminant analysis or machine learning, has never been done. Because isolated theropod teeth can be difficult to identify below broad clade or morphotype level, the present sample is treated primarily as a morphometric/morphotype dataset rather than as a set of definitive species-level occurrences (Larson & Currie, 2013; Hendrickx et al., 2015; Hendrickx et al., 2020).

Institutional abbreviations. MG, Museu Geológico do Laboratório Nacional de Energia e Geologia, Lisboa, Portugal.

Simulated training data and transfer learning

The segmentation model was based on DeepLabv3+. The training strategy followed a staged transfer learning design. First, the model was initialized from an available dinosaur-track binarization workflow, in which photographs were converted into binarized masks separating the track impression (white) from the background (black). This pretraining stage was selected because the task is directly analogous to converting paleontological photographs into standardized black-and-white masks. Starting from a model already trained to produce binarized paleontological images reduced the need for the network to learn the basic foreground-background representation required for binary segmentation from scratch. As an additional comparison, we also evaluated a DeepLabv3+ model initialized with conventional ImageNet-pretrained weights. However, this alternative produced inferior segmentation results for our

dataset, likely because ImageNet (Deng et al., 2009) pretraining emphasizes natural-image features rather than the binary foreground-background separation required for this task. We therefore adopted the dinosaur-track pretraining workflow as the final initialization strategy. This choice should be interpreted as a task-matched practical decision rather than as evidence that dinosaur-track pretraining is universally superior to conventional ImageNet or other semantic-segmentation pretraining approaches. Second, the network was finetuned on a synthetic dataset generated in Python using geometric drawing functions, primarily triangles, circles, and ellipses, with randomized size, orientation, aspect ratio, background noise, blur, colour variation, and local deformation. These elementary geometric shapes were used as simplified training objects rather than realistic tooth replicas because the target images consist of isolated theropod teeth photographed against relatively simple backgrounds. Under these conditions, the key segmentation task is mainly foreground-background separation and silhouette recovery rather than recognition of complex anatomical texture. Triangular forms approximate sharp crown profiles, circular/elliptical forms approximate basal or cross-sections, and noise simulates photographic artifacts, background texture, and imperfect preservation. The output retained was a binarized silhouette for future measurement extraction. The binarized masks were treated as the analytical product because they allowed deterministic contour extraction, region subdivision, and pixel-based morphometrics.

Segmentation post-processing

The predicted masks for the 74 teeth were post-processed before measurement. The 74 teeth include 10 teeth used for testing the trained models. In this context, binarization means converting each photograph into a two-colour mask in which the tooth silhouette is separated from the background. The final mask is therefore an analytical simplification rather than a copy of the original photograph. The binarization procedure uses Gaussian blurring, Otsu thresholding (Otsu, 1979) in both possible polarities, morphological opening and closing, and largest-component selection while ignoring components touching the image border. Gaussian blurring smooths small pixel-level noise

before thresholding, Otsu thresholding automatically selects an intensity cutoff, and padding adds a fixed border around the cropped tooth mask to reduce edge effects during measurement. This design allowed the workflow to handle either dark-on-light or light-on-dark silhouettes and to reject background or frame artifacts. The images were then manually checked for missing parts of the teeth and background parts that were selected as teeth.

The post-processed binarized images were treated as the reference masks for evaluating the predicted masks generated from real tooth images. Agreement was quantified at the view level using intersection over union (IoU; Jaccard, 1912) and Dice coefficient as overlap-based metrics (Dice, 1945), and mean symmetric boundary error, 95th-percentile Hausdorff distance, and boundary precision, recall, and F1-score with a 2-pixel tolerance as boundary-based metrics (Taha & Hanbury, 2015; Karimi & Salcudean, 2020; Cheng et al., 2021). This evaluation assessed how closely the segmentation model predictions matched the final silhouettes used for downstream morphometric extraction.

After the masks' evaluation, each selected mask was cropped to its bounding box to reduce irrelevant background and padding and rotation were applied to ensure a standardized analytical region.

Automatic measurement extraction

To standardize the extraction of tooth measurements, each specimen was represented using five predefined views: labial, mesial, lingual, distal, and cross-sectional. These orientations follow standard dental terminology commonly used in descriptions of isolated theropod teeth, where crown morphology is interpreted relative to the labiolingual and mesiodistal axes (Smith & Dodson, 2003; Hendrickx et al., 2015). Using these views allowed the same measurement procedure to be applied consistently across specimens and ensured that shape information was captured from anatomically meaningful perspectives. In addition to the whole-tooth mask, each tooth was divided into upper, middle, and lower regions. The three regions were defined geometrically by dividing the cropped silhouette bounding box into equal thirds along the apicobasal axis; they should therefore be interpreted as

standardized image regions rather than homologous anatomical landmarks. This regional subdivision was introduced in the present study to capture variation along the apicobasal axis, because theropod tooth shape, curvature, and other diagnostic structures may vary between the crown base, mid-crown, and apical portion. Three regions were chosen as a compromise between capturing apical, mid-crown, and basal variation and avoiding excessive fragmentation of the small dataset. Broken, incomplete, recurved, or obliquely oriented crowns may affect these geometric regions, and this potential source of variation is considered when interpreting regional descriptors. Thus, while the five views follow established anatomical terminology, the regional masks represent a study-specific framework designed to quantify how tooth morphology changes across different parts of the crown. For each mask, the whole-tooth bounding box was measured first to provide an overall estimate of tooth size and proportions. Measurements were then extracted from the upper, middle, and lower regional masks, allowing both global and localized aspects of tooth shape to be represented in the morphometric dataset. This method produced a standardized set of view-region observations for subsequent clustering and classification analyses.

The final feature matrix included measurements from the labial, lingual, mesial, distal, and cross-sectional views, and from the whole, upper, middle, and lower crown regions. For each view-region combination, six descriptors were retained: width, height, height-to-width ratio, mean contour angle, minimum contour angle, and maximum contour angle. Width and height were included to quantify crown dimensions, while the height-to-width ratio was included to summarize relative elongation or compression, comparable to the use of dimensional and ratio-based variables in previous theropod tooth morphometric studies (Smith et al., 2005; Marques et al., 2025). Contour angles were computed by approximating the mask contour as a polygon and calculating internal angles at successive contour vertices. The polygon simplification parameter was set as a fraction of contour arc length, with Ramer–Douglas–Peucker algorithm (Ramer, 1972; Douglas & Peucker, 1973), with epsilon set to 0.008, which increases angular stability while retaining detailed outline geometry.

The contour-angle descriptors were included as image-derived summaries of outline geometry, intended to quantify average and extreme changes in crown profile across each anatomical region. This was motivated by the recognized taxonomic relevance of crown profile, curvature, and cross-sectional morphology in theropod dental studies (Smith et al., 2005; Hendrickx et al., 2019). Together, these six descriptors provided a compact, interpretable, and repeatable representation of crown size, proportional shape, and contour morphology. Feature names encoded the anatomical view, crown region, and measurement type. Thus, variables such as *lab_wh_w*, *lin_mid_r*, *dis_low_amean*, and *cr_up_amin* represented, respectively, width, height-to-width ratio, mean contour angle, and minimum contour angle measured from specific view-region combinations. The resulting measurement matrix contained 64 teeth × 5 views × 4 regions = 1280 records. At specimen level, these records were reshaped into a 64 × 120 feature matrix. Because the height-to-width ratio is algebraically derived from height and width, the three variables are not statistically independent. We retained the ratio because it is biologically interpretable and comparable with previous tooth morphometric studies, but this redundancy is acknowledged when interpreting UMAP structure and feature-importance rankings.

Data preprocessing

The automatically extracted measurements contained repeated observations for each tooth across anatomical views and regions. These data were transformed into a tooth-level feature matrix, with one row per tooth and one column per derived measurement. This matrix contained 64 independent specimens and 120 predictors. Missing measurements are a recognized issue in theropod tooth morphometric datasets, and previous studies have addressed them using mean or column-average substitution, group-mean supplementation, exclusion of incomplete specimens or variables, and explicit evaluation of missing-data effects on classification performance (Gerke & Wings, 2016; Serrano-Martínez et al., 2016; Wills et al., 2021; Hendrickx et al., 2024; Marques et al., 2025). In the present study, there were no missing measurements, as all the measurements were obtained automatically from teeth

silhouettes. The measurements matrix was then standardized so that variables expressed in different units, such as pixel distances, ratios, and angular measurements, contributed comparably to the ordination and clustering analyses. The method for the standardization used in this study was the standard scaler, which transforms the data so that the mean is 0 and the standard deviation (SD) is 1 (Sola & Sevilla, 1997). It uses the Z-score normalization, where the standardized value is given by

$$Z = \frac{(x-\mu)}{\sigma} \quad (1)$$

where x is the original value, μ the mean value of the variable, and σ is the SD of the variable.

Dimensionality reduction, unsupervised clustering and sensitivity analysis

Dimensionality reduction was performed using UMAP to represent the multivariate measurement space in two dimensions for visualization and density-based exploratory clustering. Clustering was performed on the two-dimensional UMAP embedding using HDBSCAN. HDBSCAN was selected because it does not require the number of clusters to be specified in advance and can classify specimens with weak density-based support as noise or unassigned. This is appropriate for fossil morphometric datasets, where some specimens may occupy intermediate or low-density regions of morphospace. Specimens that were not assigned to a dense group were therefore treated as unassigned rather than being forced into artificial clusters. Clustering was applied to the UMAP embedding rather than directly to the original measurement matrix. The original feature space contained many correlated variables derived from repeated view–region combinations, which can make density-based clustering sensitive to noise, redundant descriptors, and differences in variable scale. UMAP provided a lower-dimensional representation of the dominant multivariate structure of the dataset while preserving local neighbourhood relationships among specimens. This representation is well suited to HDBSCAN because the algorithm identifies clusters from local density structure and therefore benefits from an embedding in which specimens with similar morphometric profiles are positioned close together.

However, because UMAP is itself a nonlinear dimensionality reduction method, the resulting embedding may distort global distances and local densities relative to the original measurement space. For this reason, the UMAP–HDBSCAN clusters were treated as an exploratory morphospace-based solution rather than as a definitive classification. The stability of the inferred groups was therefore evaluated using the sensitivity analyzes described below, in which UMAP neighbourhood size, UMAP minimum distance, random seed, HDBSCAN minimum cluster size, and HDBSCAN minimum samples were systematically varied.

The baseline analysis used for the stability study had the following parameter combination: UMAP $n_neighbors = 10$, UMAP $min_dist = 0.01$, random seed 42, HDBSCAN $min_cluster_size = 6$, and HDBSCAN $min_samples = 3$. The cluster assignments obtained from this baseline run were then used as the comparison point for the sensitivity analysis. In other words, each alternative parameter combination was compared with the baseline clustering to assess how much the inferred cluster assignments changed when UMAP or HDBSCAN parameters were varied.

To evaluate the robustness of the clustering solution to algorithmic parameter choices, a sensitivity analysis was performed across a full grid of UMAP and HDBSCAN parameters. The tested UMAP values were $n_neighbors = 5, 10, 15$, $min_dist = 0.00, 0.01, 0.05$, and random seeds 0, 42, 123. The tested HDBSCAN values were $min_cluster_size = 4, 6, 8$ and $min_samples = 1, 3, 5$. This produced 243 parameter combinations. For each combination, UMAP dimensionality reduction and HDBSCAN clustering were recalculated from the same imputed and standardized feature matrix.

For every clustering solution, the number of recovered clusters, the number and percentage of unassigned specimens, the minimum, median, and maximum cluster size, the percentage represented by the largest cluster, and the mean and median HDBSCAN membership probabilities were recorded. Cluster-size distributions were also saved for each run. Internal density-based cluster validity was evaluated using the Density-Based Clustering Validation index (DBCV; Moulavi et al., 2014), calculated from the UMAP embedding and HDBSCAN labels using Euclidean

distances. Higher DBCV values indicate stronger density separation among the recovered clusters. DBCV was calculated for each sensitivity run and included in the full and summarized sensitivity tables.

To quantify the stability of clustering assignments relative to the baseline solution, each sensitivity run was compared with the baseline labels using the adjusted Rand index and normalized mutual information. These metrics measure the similarity between alternative cluster assignments and the reference clustering while accounting for differences in cluster numbering.

Cluster interpretation

To analyze the density structure underlying the cluster solution, a HDBSCAN condensed-tree was obtained. To identify the measurements that distinguished each cluster, cluster-profile analyzes were performed using the imputed feature matrix in the original measurement scale. For each HDBSCAN cluster, the mean value of every feature was calculated and compared with the global mean across all specimens. The difference between the cluster mean and the global mean was expressed as a z-score

$$Z = \frac{(\bar{x}_{cluster} - \bar{x}_{global})}{s_{global}} \quad (2)$$

where $\bar{x}_{cluster}$ is the cluster mean for a given feature, $\bar{x}_{global}$ is the global mean for that feature, and s_{global} is the global standard deviation. Positive z-scores indicate features that are larger or more strongly expressed in a given cluster relative to the full sample, whereas negative z-scores indicate features that are reduced relative to the full sample. For each cluster, features were ranked by absolute z-score. This information provided an anatomical interpretation of the clusters by identifying which views, regions, dimensions, ratios, or angular descriptors contributed most strongly to each group.

For each HDBSCAN cluster, the centroid was calculated in UMAP space as the mean position of all specimens assigned to that cluster. Euclidean distances from each specimen to the cluster centroid were then computed, and the specimen closest to the centroid was selected as the representative tooth for that cluster.

Random forest classification was used as a supervised

interpretability step to evaluate which automatically extracted measurements best explained the unsupervised HDBSCAN groups. Random forests are appropriate here because they can model nonlinear feature interactions and provide feature-importance rankings and information on the decision process (Breiman, 2001). This model was not used to define the clusters; rather, it was trained after clustering to identify the variables that most strongly separated the HDBSCAN-defined groups. Only specimens assigned to HDBSCAN clusters were included in this analysis, while unassigned specimens were excluded. The random forest analysis therefore used 53 clustered specimens and 120 predictors. The corresponding feature matrix was extracted from the original measurement data and median-imputed. Since random forest models do not require standardized predictors, the classifier was trained on the imputed measurements in their original scale. This allowed decision thresholds in individual trees to be interpreted directly in pixels, ratios, or degrees.

A random forest classifier with 200 trees was trained using the HDBSCAN cluster labels as target classes. The number of trees was fixed at 200 as a standard baseline choice to obtain stable ensemble predictions while maintaining reasonable computational efficiency. The model used balanced class weights, and a fixed random seed of 42. Model performance was evaluated using repeated stratified k-fold cross-validation. In this procedure, the dataset was repeatedly divided into training and validation subsets while preserving the relative number of specimens from each cluster in every split. Up to five folds were used, but the number of folds was automatically limited by the smallest cluster size, because each fold must contain at least one specimen from every class. The cross-validation was repeated three times with different fold partitions, producing multiple independent train-validation evaluations. Model performance was summarized using accuracy, weighted precision, weighted recall, weighted F1-score, and a normalized confusion matrix. Metrics were reported as mean $\pm$ standard deviation across folds, with approximate 95% confidence intervals. Row-normalized training and validation confusion matrices were obtained for each fold and summarized as mean $\pm$ standard deviation matrices across all folds. Cross-validation was used only to estimate model

performance. After cross-validation, a final model was trained on all clustered specimens solely to generate the final feature-importance ranking, for interpretation and future application only, not for estimating cross-validation performance.

Feature importance was obtained by using impurity-based feature importance, which was extracted from each fitted random forest and summarized across folds as mean $\pm$ standard deviation. For each predictor, the importance value was calculated as the normalized total reduction in node impurity attributed to that variable across all trees in the ensemble. The resulting scores were ranked in descending order to identify the variables that contributed most strongly to the classification of specimens into the HDBSCAN-derived morphometric groups. This analysis was used as an interpretability step to determine which automatically extracted measurements, anatomical views, and crown regions were most influential in distinguishing the unsupervised clusters.

Cluster-based multi-view EfficientNet classification

A multi-branch EfficientNet classifier was used as an image-based consistency check to assess whether the HDBSCAN-derived morphometric classes could be recovered from the binarized multi-view tooth silhouettes. The supervised model used only teeth assigned to HDBSCAN clusters. The unassigned HDBSCAN class (-1) was excluded, and any tooth lacking one or more of the five required views was skipped. This produced an image dataset of 53 teeth distributed across five cluster-derived classes. The images were then grouped into classes with the following labels: cluster 0 = class 0, cluster 1 = class 1, cluster 2 = class 2, cluster 3 = class 3, and cluster 4 = class 4. The class distribution was 14, 6, 12, 8, and 13 teeth for classes 0, 1, 2, 3, and 4, respectively.

The classification model was implemented as a multi-branch EfficientNet-BO architecture, based on the EfficientNet family proposed by Tan and Le (2019). EfficientNet-BO was selected as the backbone because it provides a favorable balance between classification performance and computational efficiency through compound scaling of network depth, width, and input resolution. In addition, the availability of ImageNet pretrained EfficientNet-BO weights

enabled transfer learning from a compact and well-established convolutional backbone, which was suitable for processing multiple anatomical views while limiting computational cost.

The model used five separate EfficientNet branches, one for each view. Each branch was initialized with ImageNet pretrained weights and adapted to grayscale silhouettes by replacing the first RGB convolution with a one-channel convolution initialized from the mean of the original RGB weights. Each view-specific branch produced a feature vector, and the five branch outputs were concatenated before classification. The classifier consisted of dropout, a fully connected layer with 512 units, a ReLU activation (Nair & Hinton, 2010), a second dropout layer, and a final linear layer with five output classes. Input images were resized to 224 x 224 pixels. During training, conservative geometric data augmentation was applied, including random rotation up to 5°, translation up to 3%, and scaling between 0.95 and 1.05. These ranges were selected to simulate minor variations in tooth orientation, positioning, and image scale that may occur during image acquisition or preprocessing, while preserving the anatomical class label and avoiding unrealistic deformation of the dental structures. Validation images were resized and normalized without augmentation. Training used a weighted cross-entropy loss to account for class imbalance and the AdamW optimizer (Loshchilov & Hutter, 2019) with a learning rate of $1e^{-4}$ and weight decay of $1e^{-4}$. The model was trained for 13 epochs with a batch size of four. Additional trial runs with more training epochs did not improve validation and training performance, so the 13-epoch setting was retained to avoid unnecessary training and reduce the risk of overfitting. The previously mentioned hyper-parameters and augmentation values were selected based on visual inspection and preliminary training runs, so the supervised results are reported as exploratory until confirmed with an independent test set. Cross-validation was used only to estimate model performance.

To address the limitations of a single small train-test split, EfficientNet performance was evaluated using repeated stratified k-fold cross-validation rather than one fixed validation set. The number of folds was set to the smaller of five or the minimum class size, and the cross-validation was repeated three times using a fixed random

seed. For each fold, the model was trained from a new initialization and evaluated on the held-out validation subset. Loss, accuracy, weighted and macro precision, weighted and macro recall, and weighted and macro F1-score were calculated for each fold. Per-class precision, recall, F1-score, and support were also recorded. Final performance was reported as mean $\pm$ standard deviation across folds, with approximate 95% confidence intervals. Training and validation confusion matrices were obtained for each fold and summarized as mean $\pm$ standard deviation matrices. Training curves were also summarized across folds as mean $\pm$ standard deviation for accuracy and loss. Because the predicted labels are derived from the same morphometric clustering workflow, this supervised image-based analysis was interpreted as a consistency check rather than as independent validation of biological or taxonomic classes.

View-specific Grad-CAM visualization

After cross-validation, a final multi-branch EfficientNet model was trained on all clustered teeth. This final model was not used to estimate predictive performance; it was used only for Grad-CAM visualization and for subsequent application to additional unseen specimens. Grad-CAM was used to improve interpretability by highlighting the image regions for each anatomical view that most influenced the EfficientNet model classification decisions. For each view-specific branch, the target layer was the final convolutional feature block of the corresponding EfficientNet branch. Gradients of the predicted-class score were backpropagated to this layer, spatial weights were computed by averaging gradients across the feature-map dimensions, and the weighted feature maps were summed and passed through a rectified linear activation to produce a class-discriminative activation map. The map was resized to the 224 x 224 input size and normalized. Because the input images were black-background/white-tooth masks, Grad-CAM maps were restricted to the tooth silhouette before visualization. A tooth mask was generated from the normalized input image, all activation values outside the mask were set to zero, and the resulting view-specific heatmap was overlaid on the original silhouette. Thus, the Grad-CAM images show the regions of each tooth view

that contributed most strongly to the predicted HDBSCAN-derived class.

Exploratory prediction and classifier concordance on unseen teeth

As an additional exploratory consistency test, both trained classifiers were applied to a set of ten previously unseen teeth. The multi-branch EfficientNet model classified each specimen from the five binarized anatomical views, whereas the random forest classifier used the corresponding tooth-level morphometric feature matrix generated with the same preprocessing and pivoting steps used for the training data. For each specimen, predicted cluster classes and class probabilities were obtained for both models and compared to assess concordance between the image-based and morphometric classifiers. Because independent taxonomic labels were unavailable for these teeth, this analysis was interpreted as an external prediction and model-agreement exercise rather than as a formal accuracy test.

Implementation information and computation information

All analyzes were implemented in Python 3.10.19 (Python Software Foundation, 2022). The DeepLabv3+ segmentation model was obtained using the `segmentation_models_pytorch` library (Iakubovskii, 2019). Image reading, thresholding, contour detection, connected-component filtering, and geometric measurements used the library OpenCV (Bradski, 2000); additional image operations used PIL/Pillow (Murray et al., 2024) and scikit-image (van der Walt et al., 2014). Machine learning analyzes were performed with scikit-learn (Pedregosa et al., 2011), including median imputation, feature standardization, supervised modelling, training/validation splitting, class weighting, and model evaluation, and random forest feature-importance analysis; nonlinear dimensionality reduction used umap-learn (McInnes et al., 2018); and density-based clustering used hdbscan (McInnes et al., 2017). Visualizations were produced using Matplotlib and Seaborn. For the measurement-based analyzes, the median-imputed information was obtained with SimpleImputer from scikit-learn. The multi-branch EfficientNet model, data loaders, training

procedure, and Grad-CAM implementation were built in PyTorch (Paszke et al., 2019) and torchvision. Training and inference runtimes were recorded during the same execution of the multi-view EfficientNet-B0 workflow. The analysis used a CPU on a Windows 10 system equipped with an Intel64 Family 6 Model 154 Stepping 3 GenuineIntel processor, with 20 CPU cores, and 31.71 GB of system RAM. Training time was measured as wall-clock time for the full 13-epoch training procedure, including the validation set evaluation performed at the end of each epoch. Inference time was measured on the held-out validation set at tooth level, where each tooth consisted of five anatomical-view images. Two inference timings were recorded: a forward-pass time, representing model prediction after the five views had been loaded and moved to the computing device, and an end-to-end inference time, including image loading, preprocessing, and prediction.

RESULTS

Simulated training examples and segmentation results

The simulated training dataset (Fig. 2A) produced a broad range of simple tooth-shape silhouettes. Triangular shapes generated sharp apical profiles comparable to labial or mesial tooth outlines, whereas circular and elliptical shapes approximated cross-sectional views. Noise, blur, and colour variation increased the diversity of image conditions and reduced the gap between clean synthetic outlines and real specimen photographs. In the final segmentation stage, the model generated binarized masks from the noisy images (Fig. 2B). The binarized outputs were visually simpler but analytically superior because they allowed deterministic contour extraction.

Segmentation performance was evaluated by comparing the predicted binarized masks generated from the real tooth images with the corresponding post-processed black-and-white masks used for downstream morphometric extraction. Overall agreement was high, with a mean IoU of 0.90 ± 0.21 and a mean Dice coefficient of 0.93 ± 0.18 . Median values were higher than the means, with a median IoU of 0.98 and a median Dice coefficient of 0.990, indicating that most predicted masks closely matched the final post-processed silhouettes. Boundary agreement

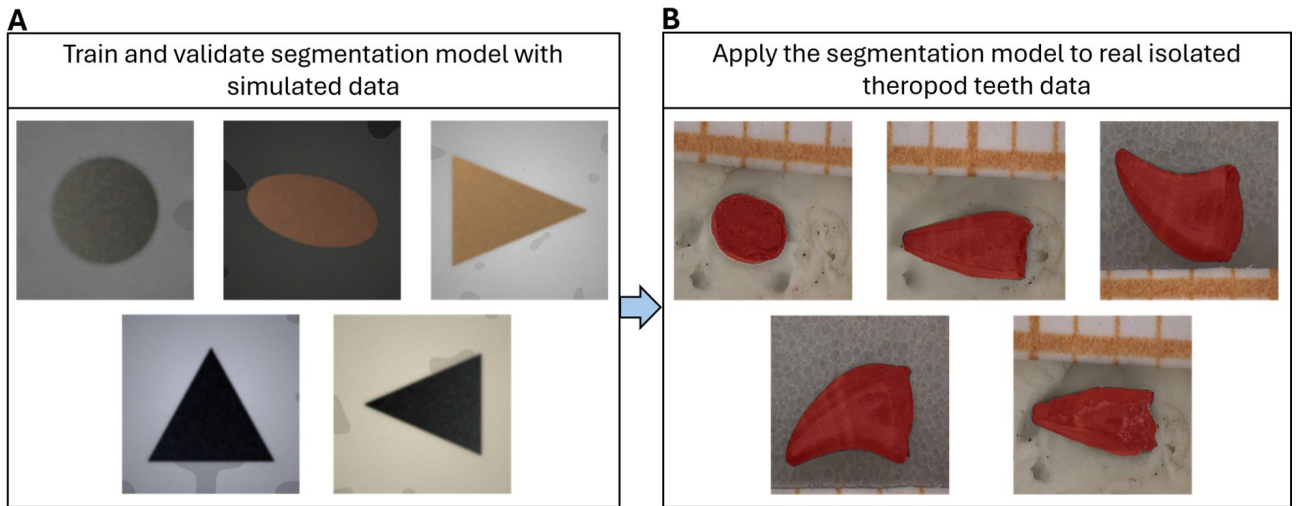


Figure 2. A, Simulated training images. B, Guimarota theropod tooth (MG27793_12) image and DeepLabv3+ outputs (cross, mesial, labial, lingual, and distal). For each viewpoint, it shows the original image, overlaid with the binary mask (in red) obtained by the model.

was also generally strong, with a mean symmetric boundary error of 3.50 ± 9.09 pixels and a median boundary error of 0.50 pixels. The 95th-percentile Hausdorff distance had a mean of 10.50 ± 21.24 pixels and a median of 1.88 pixels. Using a 2-pixel boundary tolerance, the mean boundary precision, recall, and F1-score were 0.85 ± 0.25 , 0.82 ± 0.27 , and 0.85 ± 0.24 , respectively. The lower minimum values and larger standard deviations indicate that a small number of specimens or views showed poorer agreement, but the high median overlap and boundary scores demonstrate that the segmentation workflow usually recovered the tooth silhouettes with high fidelity.

View-level segmentation consistency was high across the five anatomical views. IoU values were highest for the labial and lingual views, with mean IoU values of 0.93 ± 0.21 and 0.92 ± 0.23 , respectively, followed by mesial (0.89 ± 0.24), distal (0.88 ± 0.23), and cross-sectional views (0.87 ± 0.28). Dice coefficients showed the same pattern, with mean values of 0.96 ± 0.15 for labial, 0.95 ± 0.17 for lingual, 0.92 ± 0.20 for mesial, 0.91 ± 0.19 for distal, and 0.89 ± 0.27 for cross-sectional views. Median IoU and Dice values were very high in all views, ranging from 0.95 to 1.00 for IoU and from 0.98 to 1.00 for Dice, indicating that most view-level predictions closely matched the final post-processed silhouettes. Boundary errors were generally low, with median symmetric boundary errors between 0.00

and 1.37 pixels and median 95th-percentile Hausdorff distances between 0.00 and 4.62 pixels. The larger standard deviations and occasional low minimum values indicate that a small number of specimens or views showed poorer agreement, but the overall view-level results support the robustness of the segmentation and post-processing workflow used for morphometric extraction.

Measurement extraction

The final measurement table contained 1280 derived view-region records from 64 independent isolated theropod teeth. Each tooth was represented by 20 view-region combinations: five views and four regions (Fig. 3). These records increased feature richness but did not increase the number of independent biological specimens.

Whole-view summary statistics showed clear differences among anatomical views. Labial and lingual views had the largest mean widths, whereas mesial and distal views had higher mean height-to-width ratios, consistent with narrower profiles in mesial and distal views. The cross view was lower and broader, as expected for a transverse or basal outline (Tab. 1).

Sensitivity analysis for UMAP-HDBSCAN clustering

The sensitivity analysis confirmed that the baseline UMAP-HDBSCAN parameter combination recovered a

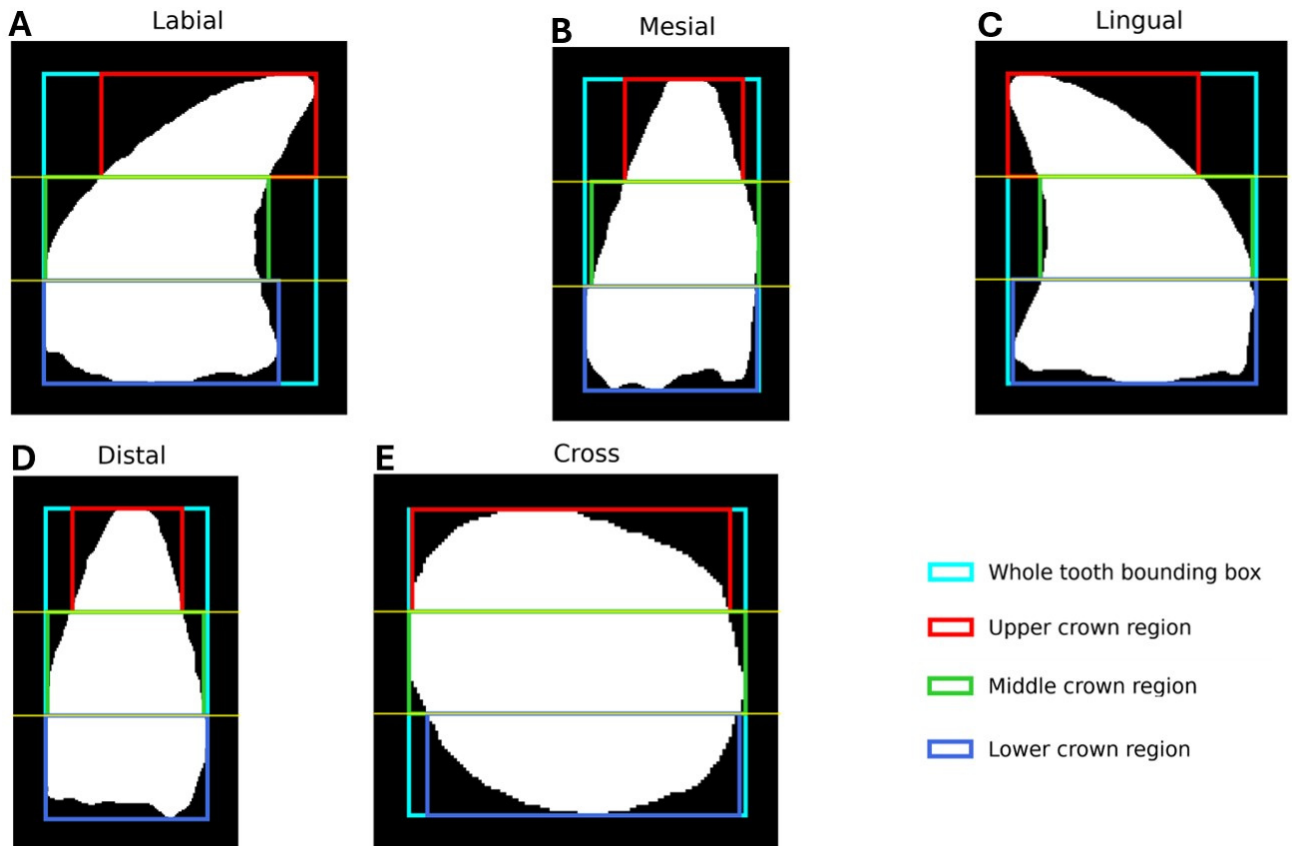


Figure 3. Example automated measurement overlay (MG27793_12). Binary masks with whole-tooth bounding box, upper/middle/lower crown regions, regional width and height, height-to-width ratio, and mean contour angle. A, labial view. B, mesial view. C, lingual view. D, distal view, E, cross-sectional view.

TABLE 1. Summary results for the different tooth views

View	Mean width px	SD	Mean height px	SD	Mean H:W	SD	Mean contour angle	SD
cross	114.73	36.16	79.23	30.99	0.72	0.15	148.01	4.61
distal	82.17	37.30	172.56	40.08	2.33	0.79	138.43	5.31
labial	129.16	42.12	171.97	40.24	1.41	0.42	139.84	5.46
lingual	128.44	40.62	168.78	42.45	1.40	0.42	140.11	4.24
mesial	83.72	40.81	172.78	39.95	2.31	0.77	138.22	5.82

five-cluster solution, with 11 of 64 specimens classified as unassigned noise and a mean HDBSCAN membership probability of 0.972. Across the three random seeds tested for the baseline parameter combination, the

number of recovered clusters remained constant at five, with an unassigned proportion of $17.19 \pm 1.56\%$, a mean membership probability of 0.957 ± 0.017 , and a DBCV score of 0.446 ± 0.051 . Agreement with the baseline run was

moderate to high across seeds, with a mean adjusted Rand index of 0.753 and a minimum adjusted Rand index of 0.605.

Across the full sensitivity grid, DBCV values ranged from 0.017 to 0.707, indicating substantial variation in internal density-based cluster validity among parameter combinations. The highest DBCV values were obtained under parameter settings that generally recovered fewer clusters, especially combinations with smaller UMAP neighbourhood size and larger HDBSCAN minimum cluster size. Thus, density-based validation favoured a somewhat coarser three- to four-cluster structure, whereas the five-cluster solution was most stable in the region surrounding the baseline parameters. These results indicate that the baseline clustering captures a reproducible structure, but that the precise number of clusters remains partly sensitive to the balance between cluster granularity and density separation.

Baseline UMAP-HDBSCAN clustering of multi-view tooth measurements

The measurement-based morphometric matrix produced a structured distribution of the 64 Guimarota theropod teeth in UMAP space. HDBSCAN identified five morphometric clusters and a small set of unassigned specimens (Fig. 4).

Of the 64 teeth analyzed, 53 were assigned (see Table 2 for the distribution of the teeth for further analysis) to clusters and 11 were classified as unassigned low-density specimens.

Cluster 0 was the largest group, containing 14 specimens, followed by Cluster 4 with 13 specimens, Cluster 2 with 12 specimens, Cluster 3 with eight specimens, and Cluster 1 with six specimens (Fig. 5). The unassigned specimens represented 17.2% of the total sample, indicating that most specimens occupied relatively dense regions of the UMAP morphospace, while a minority remained morphologically intermediate or low-density.

The clusters were not randomly distributed in the embedding. Cluster 1 occupied the left side of the morphospace, Cluster 2 formed a lower-left group, Cluster 4 occupied the upper-left to central area, Cluster 0 was concentrated mainly in the lower-right region, and Cluster 3

formed an upper-right group. The unassigned specimens occurred between and around these dense regions. This distribution suggests that the automatically extracted measurements captured morphological differences among the teeth rather than producing a single continuous undifferentiated cloud.

The HDBSCAN condensed tree showed five selected stable branches corresponding to the clusters recovered in the UMAP morphospace. The selected clusters persisted across different density levels, supporting the interpre-

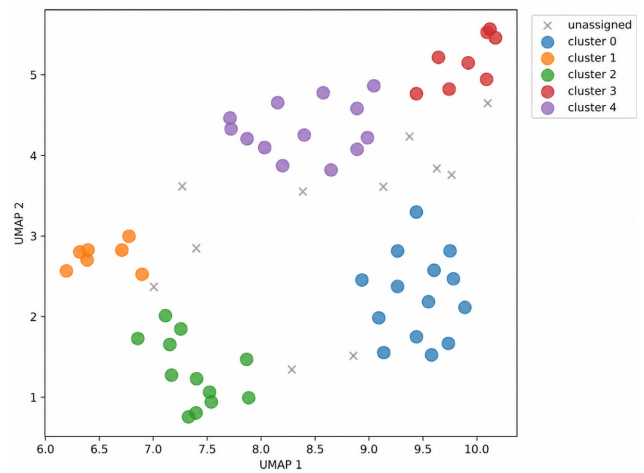


Figure 4. Uniform Manifold Approximation and Projection (UMAP)-Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) clustering of Guimarota theropod tooth measurements. The UMAP projection shows the distribution of specimens based on automatically extracted multi-view measurements. Colours indicate HDBSCAN clusters, and unassigned specimens are shown separately.

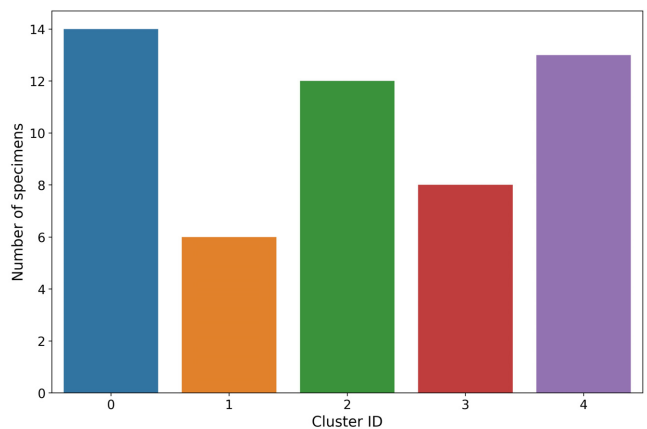


Figure 5. Number of teeth assigned to each Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) class. Cluster sizes show that most teeth were assigned to one of five morphometric groups, with only 11 unassigned specimens.

TABLE 2. Representative isolated theropod teeth, and their distance to the centroid of the cluster in UMAP representation.

Cluster	Representative tooth	Distance to UMAP centroid	Cluster size
0	MG27797D56	0.11	14
1	MG27786_3	0.12	6
2	MG27748	0.090	12
3	MG27767	0.038	8
4	MG27784_3	0.074	13

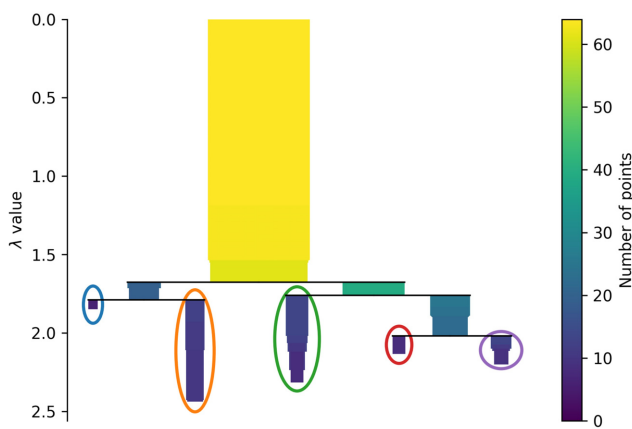


Figure 6. Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) condensed tree showing selected clusters. The plot summarizes the hierarchical density structure used by HDBSCAN to extract clusters. Coloured branches indicate the selected clusters, while branch height reflects density level, expressed as λ . Larger and more persistent branches correspond to more stable morphometric groups.

tation that the UMAP-HDBSCAN parameterization recovered high-density morphometric groups rather than only arbitrary visual subdivisions (Fig. 6).

Representative teeth (Tab. 2; Figs. S1-S5, Supplementary Information) were selected as the specimens closest to each cluster centroid in UMAP space. These representatives provide reference specimens for visual comparison among the groups.

Cluster-specific morphometric profiles and cluster-specific variability

Cluster-profile analysis identified the measurements that most strongly distinguished each group from the global sample mean. These profiles were calculated as z-scores of the cluster mean relative to the global mean,

allowing positive and negative deviations to be interpreted anatomically.

Cluster 0 was mainly characterized by negative deviations in lingual, labial, distal, and cross-sectional ratio and angular variables (Fig. 7). The strongest reduced variables included the different lingual ratios, cross-sectional whole minimum angle, lingual upper and middle ratios, lingual whole ratio, distal lower, middle, and upper ratios, and distal upper ratio. Lingual height variables were also reduced. This combination suggests that Cluster 0 includes teeth with relatively lower height-to-width proportions and reduced lingual height expression compared with the full sample.

Cluster 1 was distinguished almost entirely by positive height-to-width ratios across several views and regions (Fig. 8). The strongest variables included distal middle and lower ratios, distal whole ratio, distal upper ratio, mesial lower, upper, and middle ratios, lingual upper and middle

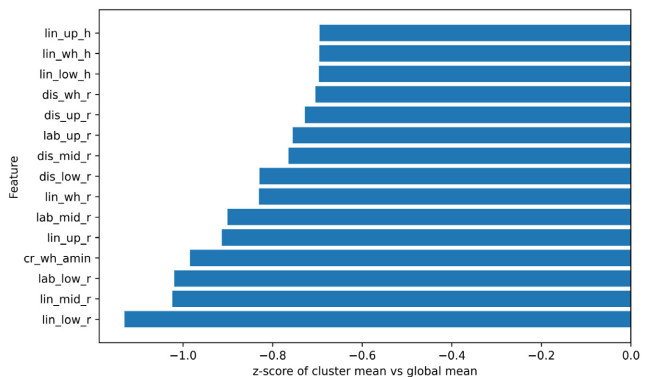


Figure 7. Cluster 0 profile. The most distinctive measurements for Cluster 0 show reduced height-to-width ratios, lingual height variables, and selected angular descriptors.

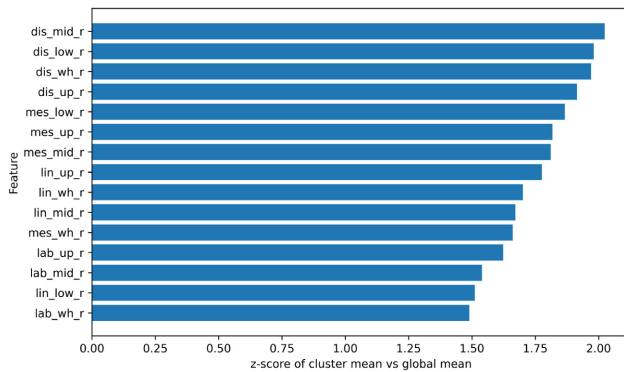


Figure 8. Cluster 1 profile. Cluster 1 is characterized by high height-to-width ratios across distal, mesial, lingual, and labial views, suggesting relatively elongate or narrow tooth silhouettes.

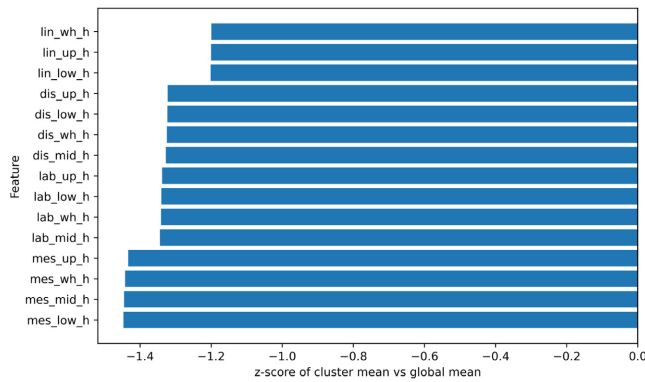


Figure 9. Cluster 2 profile. Cluster 2 is distinguished by reduced height measurements across mesial, labial, distal, and lingual views.

ratios, lingual whole ratio, labial upper middle, and lower ratios, and labial whole ratio. This indicates that Cluster 1 represents a relatively slender morphotype, with taller or narrower outlines across multiple anatomical views.

Cluster 2 showed predominantly negative deviations in height variables across several views (Fig. 9). The most reduced variables included mesial lower and middle height, mesial whole height, labial middle, and lower height, distal middle height, labial upper height, distal whole height, distal lower, and upper height, lingual lower and upper height, and lingual whole height. This profile indicates that Cluster 2 represents a group of relatively low-height silhouettes across labial, mesial, distal, and lingual views.

Cluster 3 showed a different profile, with positive deviations in width variables and negative deviations in several angular descriptors (Fig. 10). The strongest positive variables included mesial whole width, mesial upper and

middle width, lingual whole width, distal whole width, lingual upper width, and labial whole width. In contrast, reduced variables included mesial and distal minimum-angle measurements and mesial mean-angle measurements. This combination suggests relatively wide silhouettes with more acute or distinctive local contour geometry.

Cluster 4 was distinguished by strong positive deviations in cross-sectional and lingual variables (Fig. 11). The highest values included cross-sectional lower and middle height, cross-sectional whole height, cross-sectional upper height, cross-sectional lower, middle, and upper width, cross-sectional whole width, lingual lower, middle, and upper height, lingual whole height, cross-sectional whole ratio, and cross-sectional middle ratio. This profile suggests a group with relatively robust cross-sectional silhouettes and increased lingual height expression.

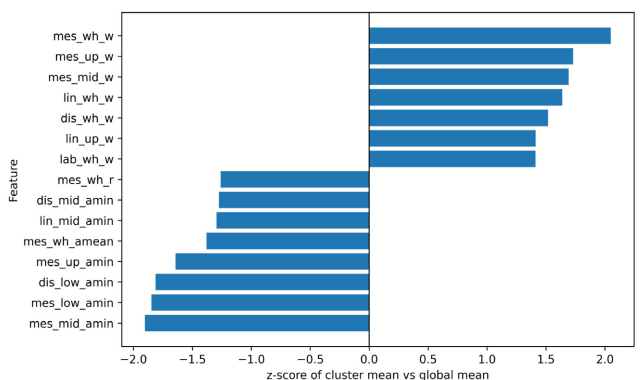


Figure 10. Cluster 3 profile. Cluster 3 is characterized by reduced distal and lingual heights and reduced labial and lingual height-to-width ratios.

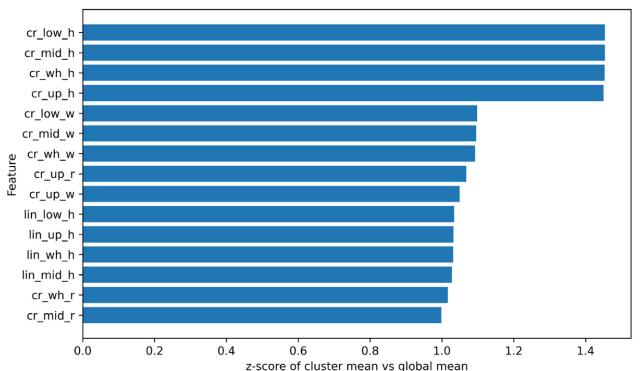


Figure 11. Cluster 4 profile. Cluster 4 is distinguished by increased cross-sectional height and width measurements and by positive lingual height variables.

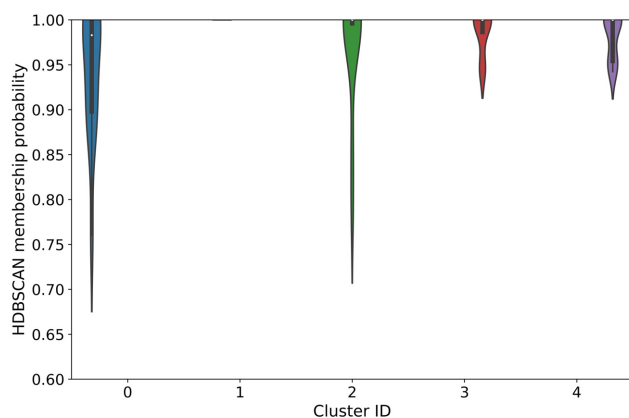


Figure 12. Violin plots showing the distribution of Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) cluster membership probabilities for each identified tooth cluster (clusters 0–4). The y-axis represents the probability assigned by HDBSCAN to each specimen belonging to its respective cluster, ranging from 0 to 1. Wider sections of each violin indicate a higher density of specimens with similar probabilities. Internal boxplots summarize the median and interquartile range of the distributions. Most specimens exhibit high membership probabilities (>0.90), indicating that the identified morphotypes are well supported and internally coherent in feature space. Clusters 0 and 2 display slightly broader distributions and lower minimum probabilities, suggesting comparatively higher morphological variability or weaker cluster boundaries relative to the other groups.

Membership probability distributions for each cluster indicate generally high clustering confidence across all groups, with most specimens exhibiting probabilities close to 1.0 (Fig. 12). Clusters 3 and 4 showed the highest assignment stability, characterized by narrow violin distributions concentrated near maximal probability values, suggesting strong internal morphological consistency. Clusters 0 and 2 exhibited broader distributions and slightly lower probability tails, with some specimens presenting probabilities below 0.80, indicating greater morphological heterogeneity or partial overlap with neighbouring clusters in multidimensional feature space.

Random forest interpretation of HDBSCAN clusters

Random forest classification was used to evaluate which automatically extracted measurements of the 53 teeth best explained the HDBSCAN-derived clusters. Only specimens assigned to HDBSCAN clusters were included, and unassigned specimens were excluded. The final random forest dataset contained 53 clustered teeth. Cluster 0 contained 14 teeth, Cluster 1 contained 6 teeth, Cluster 2

contained 12 teeth, Cluster 3 contained 8 teeth, and Cluster 4 contained 13 teeth. Instead of using a single train-validation split, model performance was evaluated using repeated stratified k-fold cross-validation. Five stratified folds were used and repeated three times, producing 15 train-validation evaluations. Stratification preserved the relative representation of the five HDBSCAN-derived clusters in each fold.

Across folds, the mean training support was approximately 11.2 teeth for Cluster 0, 4.8 for Cluster 1, 9.6 for Cluster 2, 6.4 for Cluster 3, and 10.4 for Cluster 4 (Fig. S6, Supplementary Information). The corresponding mean validation supports were approximately 2.8, 1.2, 2.4, 1.6, and 2.6 teeth, respectively (Fig. S6, Supplementary Information). These values reflect the expected 80/20 distribution produced by five-fold stratified cross-validation and show that Cluster 1 was consistently the least represented class in both training and validation partitions.

The random forest recovered the HDBSCAN-derived classes very strongly from the morphometric feature matrix. Training performance achieved values of 1.00 across folds, with the row-normalized training confusion matrix showing 1.00 ± 0.00 on the diagonal (Fig. S7A, Supplementary Information). Validation performance remained high (Fig. S7B, Supplementary Information), although some misclassifications were observed. The row-normalized validation confusion matrix showed that Cluster 0 was correctly classified in 0.93 ± 0.14 of validation cases, with the remaining 0.07 ± 0.14 assigned to Cluster 4. Cluster 1 showed the lowest and most variable validation performance, with 0.63 ± 0.48 correctly classified and 0.37 ± 0.48 assigned to Cluster 2. Cluster 2 was recovered without validation misclassification, with a diagonal value of 1.00 ± 0.00 . Cluster 3 was correctly classified in 0.87 ± 0.30 of validation cases, with 0.13 ± 0.30 assigned to Cluster 4. Cluster 4 was also recovered without validation misclassification, with a diagonal value of 1.00 ± 0.00 .

The weighted and macro-averaged metrics confirmed strong overall classification performance, with values of 1.00 for all the training metrics (Fig. S8, Supplementary Information). Validation weighted precision, recall, and F1-score were approximately 0.90, 0.92, and 0.90, respectively (Fig. S8, Supplementary Information). Validation macro pre-

cision, recall, and F1-score were slightly lower, approximately 0.87, 0.89, and 0.87, respectively (Fig. S8, Supplementary Information), reflecting the influence of the smaller Cluster 1. All per-class training metrics presented values of 1.00 (Fig. S9, Supplementary Information). Per-class validation metrics showed the same pattern as weighted and macro-averaged metrics: Cluster 0 had high precision, recall, and F1-score, Cluster 2 and Cluster 4 were recovered particularly strongly, and Cluster 3 also showed high performance (Fig. S10, Supplementary Information). Cluster 1 had the lowest precision, recall, and F1-score, with large standard deviations across folds, consistent with its small sample size and partial overlap with Cluster 2 in the validation confusion matrix.

The plotted decision tree, representing one estimator from the random forest, shows that the first split was based on distal middle width ($\text{dis_mid_w} \leq 58.5$), followed by additional splits involving labial upper ratio, distal middle height, cross-sectional middle height, cross-sectional middle width, and cross-sectional middle maximum angle. This example tree shows that the classifier separated the morphometric clusters using a combination of distal, labial, and cross-sectional measurements rather than a single measurement or anatomical view (Fig. 13).

The random forest feature-importance analysis showed

that the highest-ranking variables were dominated by height and width measurements from the distal and cross-sectional views, with additional contributions from mesial and lingual measurements (Fig. 14). The most important variable was distal whole height (dis_wh_h), followed by cross-sectional middle height (cr_mid_h), cross-sectional whole height (cr_wh_h), distal middle width (dis_mid_w), distal whole width (dis_wh_w), cross-sectional lower height (cr_low_h), mesial middle height (mes_mid_h), distal lower width (dis_low_w), mesial whole width (mes_wh_w), and cross-sectional upper height (cr_up_h). Additional high-ranking variables included distal middle height, lingual upper width, mesial lower height, lingual whole width, mesial whole height, distal lower ratio, lingual upper height, mesial middle minimum angle, distal lower height, and mesial upper height. Multi-view EfficientNet classification of cluster-derived morphometric groups.

After removing HDBSCAN unassigned teeth, and retaining teeth represented by all five required views, five cluster-derived classes were used for supervised image classification. The final EfficientNet dataset contained 53 teeth. Cluster 0 contained 14 teeth, Cluster 1 contained 6 teeth, Cluster 2 contained 12 teeth, Cluster 3 contained 8 teeth, and Cluster 4 contained 13 teeth. Instead of using a single train-validation split, model performance was

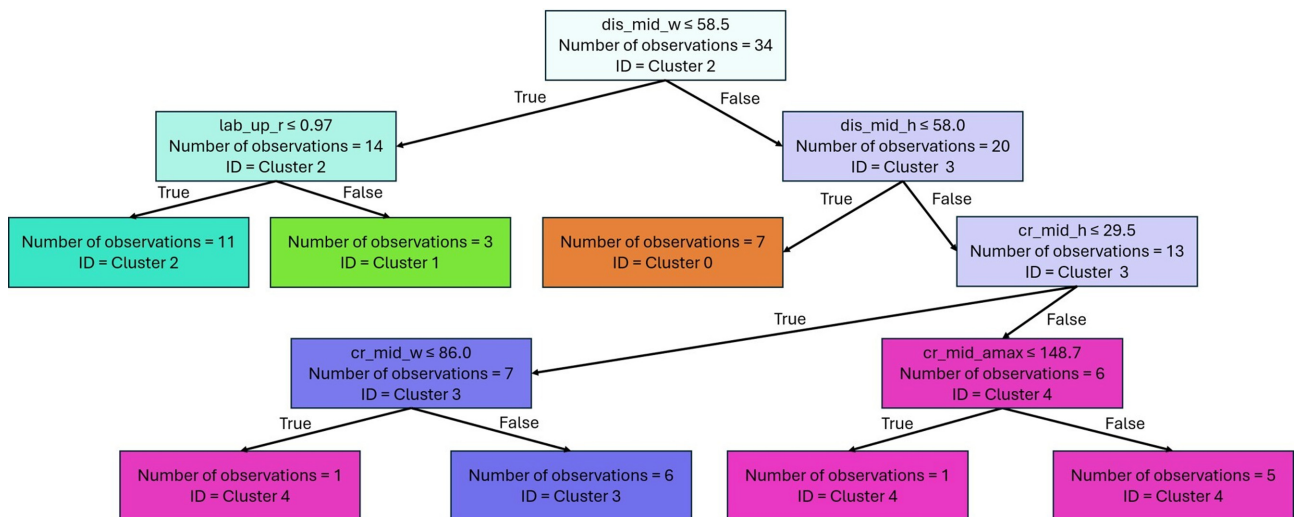


Figure 13. Example decision tree from the random forest classifier. The displayed tree shows one estimator from the random forest and illustrates representative decision rules used to classify teeth into Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN)-derived clusters. The tree should be interpreted as an illustrative example because the full random forest combines predictions across many decision trees.

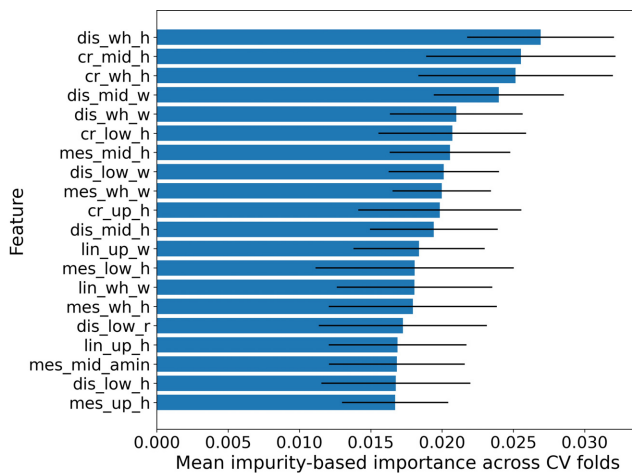


Figure 14. Mean and standard deviation for the random forest feature importance across the cross-validation folds. The plot shows the 20 variables with the highest mean impurity-based importance scores in the random forest classifier used to separate the HDBSCAN-derived morphometric clusters. Higher values indicate a greater contribution to classification across the forest. Because these values can be affected by correlated or high-variance predictors, they should be interpreted as exploratory indicators of cluster separation rather than as direct measures of biological importance. The standard deviation is seen on the black horizontal lines.

evaluated using repeated stratified k-fold cross-validation. Five stratified folds were used and repeated three times, producing 15 train-validation evaluations. Stratification preserved the relative representation of the five cluster-derived classes in each fold. Across folds, the mean training support was approximately 11.2 teeth for Cluster 0, 4.8 for Cluster 1, 9.6 for Cluster 2, 6.4 for Cluster 3, and 10.4 for Cluster 4 (Fig. S11, Supplementary Information). The corresponding mean validation supports were approximately

2.8, 1.2, 2.4, 1.6, and 2.6 teeth, respectively (Fig. S11, Supplementary Information).

The multi-branch EfficientNet model showed consistent improvement across repeated cross-validation. Mean training accuracy (Fig. 15A) increased from approximately 0.39 in epoch 1 to 0.85 by epoch 5, 0.95 by epoch 7, and approached 1.00 from approximately epoch 9 onward. Mean validation accuracy (Fig. 15A) also increased rapidly during the first epochs, from approximately 0.39 in epoch 1 to 0.71 by epoch 5, 0.82 by epoch 7, and a maximum of approximately 0.91 around epoch 9. After this point, validation accuracy remained high but fluctuated slightly among folds, ending at approximately 0.89 by the final epoch. This pattern indicates strong recovery of the cluster-derived classes from the five-view silhouettes, while also showing that validation performance was not identical across repeated folds. Loss curves (Fig. 15B) showed the same general pattern. Mean training loss decreased steadily from approximately 1.58 in epoch 1 to less than 0.10 after epoch 9, reaching very low values by the final epochs. Mean validation loss decreased from approximately 1.59 in epoch 1 to approximately 0.53 by epoch 8 and then stabilized between approximately 0.42 and 0.47 during the final epochs. The divergence between the near-zero training loss and the higher validation loss after the middle epochs suggests that the model fitted the training folds very closely, while validation performance remained high but more variable. This pattern is expected given the small number of teeth per class and the repeated use of small validation folds.

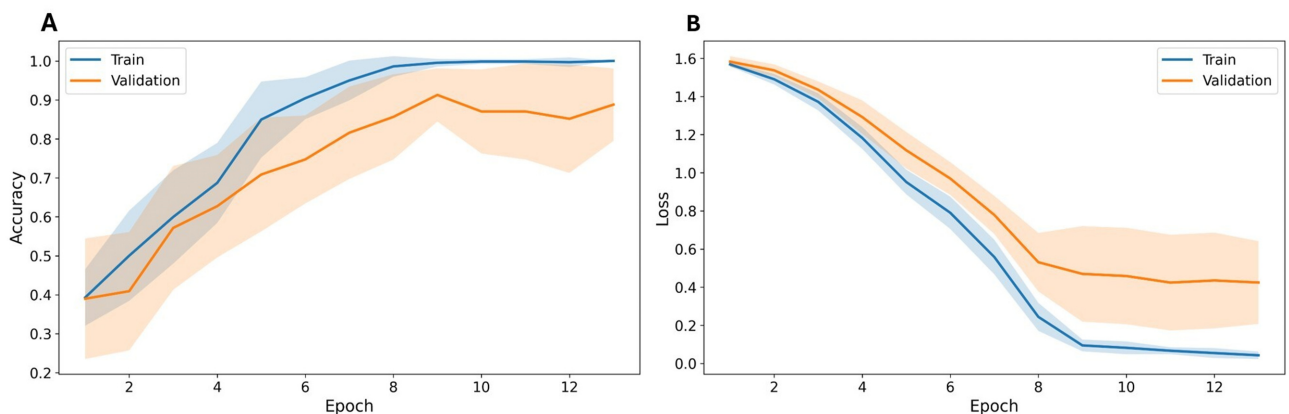


Figure 15. Multi-branch EfficientNet training curves across repeated stratified cross-validation folds. A, accuracy curves. B, loss curves. Lines show the mean across folds and shaded bands show $\pm$ standard deviation (SD).

The mean training confusion matrix (Fig. S12A, Supplementary Information) showed complete separation of the five cluster-derived classes, with all diagonal values equal to 1.00. In validation (Fig. S12B, Supplementary Information), most predictions remained on the diagonal, but occasional misclassifications occurred among morphologically similar clusters. Cluster 0 was correctly classified in 0.90 ± 0.18 of validation cases, with minor confusion mainly with Cluster 2 and Cluster 4. Cluster 1 showed the lowest diagonal value, 0.73 ± 0.42 , with misclassifications primarily into Cluster 2 and, less frequently, Cluster 4. Cluster 2 was correctly classified in 0.92 ± 0.17 of cases, Cluster 3 in 0.87 ± 0.30 , and Cluster 4 in 0.93 ± 0.14 .

Training performance for both weighted and macro-averaged metrics and per-class metrics presented values close to 1.00 for all classes (Fig. S13 and Fig. S14, Supplementary Information). Mean validation weighted precision, recall, and F1-score were approximately 0.90, 0.89, and 0.88, respectively. Macro-averaged values were slightly lower, with mean macro precision of approximately 0.89, macro recall of approximately 0.87, and macro F1-score of approximately 0.86. The difference between weighted and macro scores reflects the small and uneven class sizes, particularly the lower stability of Cluster 1.

Validation performance remained high but varied among classes (Fig. S15, Supplementary Information). Cluster 0 showed the strongest validation recovery, with mean precision of approximately 0.96, recall of 0.90, and F1-score of 0.91. Cluster 2 also performed strongly, with precision of approximately 0.85, recall of 0.92, and F1-score of 0.87. Cluster 4 showed similar performance, with precision of approximately 0.90, recall of 0.93, and F1-score of 0.90 (Fig. S15, Supplementary Information). Cluster 3 had validation precision of approximately 0.93, recall of 0.87, and F1-score of 0.89. Cluster 1 was the most variable class, with validation precision of approximately 0.80, recall of 0.73, and F1-score of 0.75. This lower and more variable performance is consistent with the small size of Cluster 1, which contained only six teeth.

View-specific Grad-CAM results

Grad-CAM visualizations were summarized using representative validation examples from each of the five

cluster-derived EfficientNet classes. Because the performance estimates were obtained from repeated cross-validation, the Grad-CAM results should be interpreted as a model-interpretation analysis rather than as an independent validation of taxonomic identity. The panels show the five anatomical viewpoints used by the model for each specimen, with class-discriminative activation overlaid on the black-background/white-tooth masks. This cluster-level presentation makes it possible to compare how the multi-branch EfficientNet used different parts of the tooth silhouette across the morphometric groups.

Across all five clusters, activation was not distributed uniformly across the masks. Instead, the strongest responses were concentrated on silhouette regions corresponding to anatomically interpretable areas, including the apical region, basal margin, lateral contour, longitudinal crown axis, and cross-sectional perimeter. Because the maps were generated separately for each view-specific branch, the results also show that no single anatomical view alone accounted for all of the predictions.

In Cluster 0 (Fig. 16), the Grad-CAM maps emphasized basal and lateral parts of the crown in several lateral views, with additional activation along the lower part of the cross-sectional outline. This pattern is consistent with a class recognized from broad silhouette geometry and marginal shape rather than from a single point of the tooth.

In Cluster 1 (Fig. 16), activations were stronger along lateral contour zones and central to mid-crown portions of the silhouette, with the cross-sectional view showing a more localized response on one side of the outline. This suggests that the model used elongate-outline geometry, localized contour curvature, and cross-sectional asymmetry to recover this cluster-derived class.

In Cluster 2 (Fig. 16), the strongest responses were distributed across basal margins, apical or lateral transitions, and peripheral cross-sectional regions. These activations are consistent with the measurement-based interpretation that this class was distinguished by regional height, width, and contour information expressed across multiple views.

Cluster 3 (Fig. 16) showed a different Grad-CAM pattern. The strongest responses occurred near the apex in one lateral view and along the long axis of the crown in the narrow mesial or distal views. The cross-sectional view also

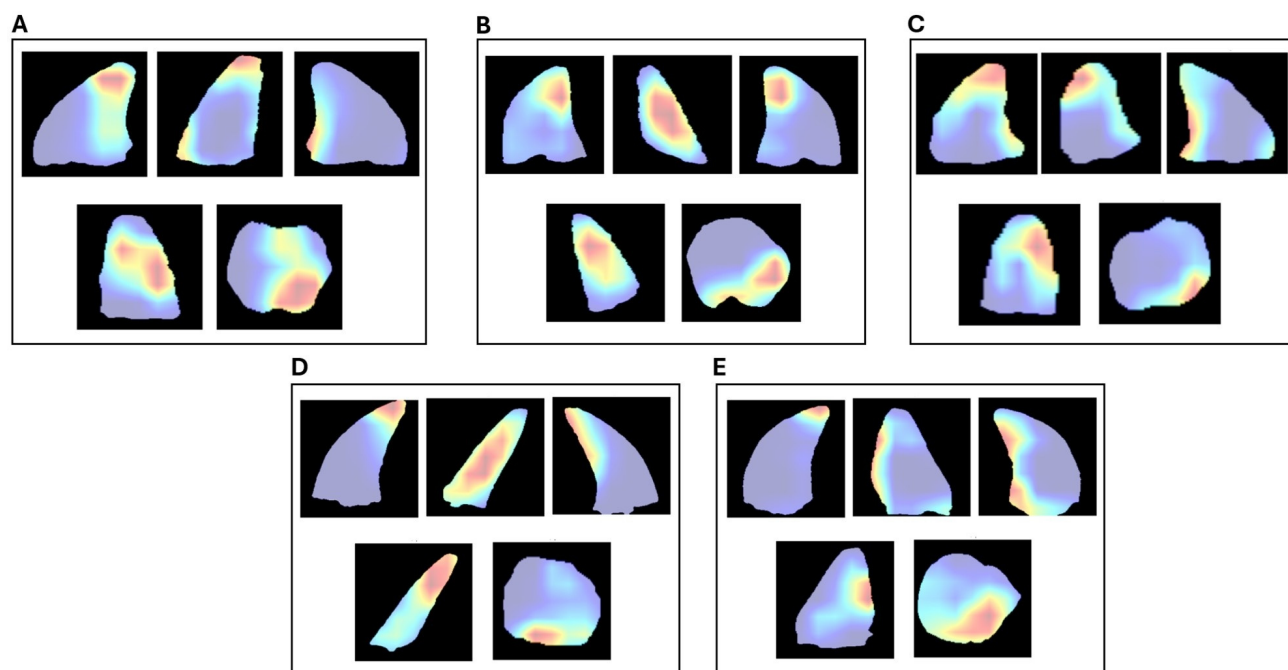


Figure 16. Gradient-weighted Class Activation Mapping (Grad-CAM) summaries for example teeth for each cluster. The five view-specific overlays show activation concentrated mainly on basal and lateral crown regions and on the lower perimeter of the cross-sectional silhouette. A, Cluster 0 is represented by tooth MG27752. B, Cluster 1 is represented by tooth MG27808D118. C, Cluster 2 is represented by tooth MG27712. D, Cluster 3 is represented by tooth MG27768. E, Cluster 4 is represented by tooth MG27793_5.

showed activation along part of the lower lateral perimeter. This supports the interpretation that Cluster 3 is visually recovered through apical shape, longitudinal crown geometry, and lateral outline curvature.

In Cluster 4 (Fig. 16), the Grad-CAM maps emphasized basal and lateral contour regions in several views and the lower perimeter of the cross-sectional outline. This pattern agrees with the cluster-profile and random forest results, which indicated that cross-sectional height and width measurements and lingual height variables contributed strongly to the separation of this cluster.

The Grad-CAM results show convergence between the measurement-based and image-based components of this study. The cluster profiles and Random Forest feature-importance results identified view-specific widths, heights, ratios, angular descriptors, and cross-sectional measurements as important variables, while Grad-CAM showed that the EfficientNet focused on the silhouette regions where those measurements are anatomically expressed. The Grad-CAM maps therefore support the interpretation that the EfficientNet recovered the cluster-derived morphome-

tric groups from tooth shape rather than from arbitrary image regions.

Predicting cluster ID for unseen isolated theropod teeth

The trained multi-branch EfficientNet and random forest classifiers were applied to ten unseen teeth to evaluate whether newly processed specimens could be assigned to the previously defined cluster-derived morphometric groups. The two classifiers produced concordant predictions for seven of the ten specimens (Tab. 3). Both models assigned MG27669, MG27676, and MG27707 to Cluster 2, MG27682, MG27809_2, MG27809_3, and MG27809_4 to Cluster 4. Disagreements occurred for three specimens: MG27683 was assigned to Cluster 4 by EfficientNet but to Cluster 1 by random forest; MG27703 was assigned to Cluster 0 by EfficientNet but to Cluster 4 by random forest; and MG27705 was assigned to Cluster 2 by EfficientNet but to Cluster 1 by random forest.

EfficientNet prediction probabilities ranged from 0.431 to 0.997, whereas random forest probabilities ranged from

TABLE 3. Probabilities correspond to the predicted cluster and the associated probability for each model and each test tooth. Agreement indicates whether both classifiers assigned the specimen to the same cluster.

Specimen	EfficientNet prediction	EfficientNet probability	Random forest prediction	Random forest probability	Agreement
MG27669	Cluster 2	0.90	Cluster 2	0.46	Yes
MG27676	Cluster 2	0.56	Cluster 2	0.58	Yes
MG27682	Cluster 4	0.89	Cluster 4	0.49	Yes
MG27683	Cluster 4	0.43	Cluster 1	0.64	No
MG27703	Cluster 0	0.90	Cluster 4	0.37	No
MG27705	Cluster 2	0.51	Cluster 1	0.34	No
MG27707	Cluster 2	0.96	Cluster 2	0.79	Yes
MG27809_2	Cluster 4	1.00	Cluster 4	0.89	Yes
MG27809_3	Cluster 4	0.99	Cluster 4	0.81	Yes
MG27809_4	Cluster 4	1.00	Cluster 4	0.85	Yes

0.340 to 0.885. The highest-confidence concordant predictions were observed for MG27809_2, MG27809_3, and MG27809_4, which were assigned to Cluster 4 by both classifiers, with EfficientNet probabilities above 0.99 and random forest probabilities between 0.79 and 0.885. The discordant cases generally had lower or more ambiguous probabilities in at least one model, suggesting that these specimens may lie closer to boundaries between cluster-derived morphometric groups.

Overall, the agreement in seven of ten unseen specimens supports broad consistency between the silhouette-based EfficientNet classifier and the measurement-based random forest classifier. However, the three discordant predictions indicate that unseen specimens should not be treated as definitive assignments. These results are best interpreted as exploratory cluster-derived predictions and as a model-concordance check, rather than as independent taxonomic identifications or formal external validation.

Grad-CAM visualizations for the unseen teeth showed that the EfficientNet predictions were mainly driven by localized regions of the binarized silhouettes rather than by the entire tooth outline uniformly (Fig. S16, Supplementary Information). Activation was concentrated along anatomically informative areas, including apical regions, lateral crown margins, basal edges, and the transverse/cross-section

perimeter. High-confidence predictions, particularly those assigned to Cluster 4, showed more consistent activation patterns across views, whereas lower-confidence predictions displayed more dispersed or view-specific activation. These visualizations indicate that the model used biologically relevant silhouette features when assigning unseen specimens to the cluster-derived classes, although the highlighted regions should be interpreted as model-attention patterns rather than direct anatomical diagnoses.

DISCUSSION

Morphometric structure of the Guimarota theropod tooth sample

The UMAP-HDBSCAN analysis recovered five morphometric groups among the selected 64 Guimarota theropod teeth, with only eleven specimens remaining unassigned. This result indicates that the automated multi-view measurement workflow captured structured variation in the dataset. The clusters occupy distinct regions of morphospace and differ in their dominant measurements, suggesting that the method is sensitive to variation in tooth outline, height, width, proportions, and contour geometry.

The clustering should not be interpreted directly as a taxonomic classification. The morphology of isolated theropod teeth generally shows a high degree of convergence

among taxa and it can also vary due to tooth position, ontogeny, wear, breakage, preservation, and imaging orientation. Therefore, the clusters are best interpreted as morphometric groups rather than taxon-level identifications. Nevertheless, the recovery of multiple dense groups supports the usefulness of automated silhouette-based morphometrics for grouping isolated theropod tooth assemblages and identifying specimens that deserve closer anatomical comparison.

Compared with traditional theropod tooth morphometric studies such as Gerke and Wings (2016) and Ősi et al. (2025), which rely on manually defined measurements and/or wear variables, the present workflow is not intended to replace anatomical interpretation. Its contribution is to automate standardized multi-view silhouette extraction and provide a reproducible exploratory screen for morphological variation before detailed taxonomic assessment.

Anatomical interpretation of the clusters

The cluster profiles suggest that the groups differ in biologically interpretable ways. Cluster 1 is the clearest example of a proportion-based group. Its most distinctive variables are height-to-width ratios from distal, mesial, lingual, and labial views. This pattern suggests relatively elongate or narrow crowns. In contrast, Cluster 0 is characterized by reduced ratios and reduced lingual height variables, suggesting lower or broader outlines in several views.

Cluster 2 is characterized mainly by reduced height variables across mesial, labial, distal, and lingual views. This group may therefore represent lower-height silhouettes, although this interpretation depends on whether all images were standardized to the same physical scale. If images were not scale-calibrated, some height-related signals could reflect differences in image size, magnification, segmentation extent, or preservation rather than biological size alone.

Clusters 3 and 4 are not identical. Cluster 3 combines positive width variables across mesial, lingual, distal, and labial views with negative angular descriptors, suggesting wider silhouettes with distinctive local contour geometry. Cluster 4 shows strong positive deviations in cross-sectional height and width measurements and positive

lingual height variables, suggesting a group with a more robust cross-sectional profile. These cluster interpretations should be compared visually using representative specimens before assigning biological meaning.

Importance of multi-view measurements

The results show that no single view dominates all cluster distinctions. Mesial, distal, lingual, labial, and cross-sectional variables all appear among the most distinctive cluster-profile features or the random forest importance rankings. This supports the central premise of the study: isolated theropod teeth are better represented by multi-view information than by a single labial or lingual outline alone.

The cross-section view was particularly important in Cluster 4 and in the random forest model, where cross-sectional middle and lower height ranked among the important predictors. This suggests that apical outline information may carry useful morphometric signals. Similarly, distal and mesial measurements were important in separating several clusters, highlighting the value of including narrow mesial and distal views rather than relying only on lateral profiles.

Usefulness of cluster profiles and Random Forest interpretation

The cluster-profile plots provide the most direct anatomical interpretation of the HDBSCAN groups. Because they compare each cluster mean with the global mean, they show whether a group is characterized by increased or decreased widths, heights, ratios, or angular values. This information is especially useful for paleontological interpretation because it translates unsupervised clusters into recognizable morphological differences.

The random forest analysis provides a complementary interpretation. It shows which variables are most effective for reproducing the HDBSCAN labels in a supervised framework. The high importance of variables such as distal whole height, cross-sectional middle and lower height, mesial upper height, distal lower width, labial middle height, and cross-sectional whole height indicates that linear measurements across multiple views are central to separating the groups. However, random forest importance should be

interpreted cautiously because correlated variables can share importance, and impurity-based importance can favour variables with greater variability. Therefore, this information should be treated as a supporting interpretability analysis rather than as the primary evidence for cluster identity.

The example decision tree presents how some decision rules separate clusters, but it represents only one tree from the forest. It should not be interpreted as the definitive classification rule. The full forest combines a total of 200 trees, and the feature-importance information provides a more stable summary of the variables contributing to classification.

EfficientNet recovery of cluster-derived morphotypes and Grad-CAM interpretability

The EfficientNet step adds an internal consistency check of visual recoverability. The Convolutional Neural Network (CNN) was not trained to identify external taxonomic labels; it was trained to recover morphometric clusters derived from the UMAP-HDBSCAN workflow. The validation performance in the held-out split therefore suggests that the five-view black-background/white-tooth images contained sufficient silhouette information to reproduce the cluster-derived morphometric groups in this dataset. This result supports the internal consistency of the workflow: the groups recovered from quantitative measurements also correspond to patterns that can be learned directly from tooth silhouettes.

The Grad-CAM results are important because they make the image classifier more interpretable. Rather than relying only on accuracy, the cluster-level Grad-CAM summaries show that the model used silhouette and contour regions corresponding to anatomically plausible areas, including apical shape, lateral contour, basal margins, longitudinal crown geometry, and cross-sectional perimeter. The representative examples from Clusters 0-4 indicate that different classes were associated with different combinations of activated regions: basal and lateral contour regions in Clusters 0 and 4, localized lateral and mid-crown contour signals in Cluster 1, cross transitions and peripheral cross-sectional regions in Cluster 2, and apical plus long-axis crown geometry in Cluster 3. These regions correspond closely to the

types of variables that appear in the cluster profiles and random forest feature importances.

Implications for automated theropod tooth analysis

These results demonstrate that AI-generated and post-processed silhouettes can be converted into quantitative descriptors capable of separating isolated theropod teeth into interpretable morphometric groups. The approach is useful because it reduces manual measurement effort, allows consistent extraction of many view-region variables, and provides reproducible outputs that can be analyzed through cluster profiles and representative specimens.

For the Guimarota site, this workflow provides a way to organize a morphologically diverse isolated-tooth assemblage before detailed taxonomic study. Representative specimens such as MG27797D56, MG27786_3, MG27748, MG27767, and MG27784_3 can be used as references for comparing cluster morphology. These specimens should be examined visually together with the original images and silhouettes to determine whether the automatically recovered groups correspond to recognizable tooth morphotypes.

The EfficientNet and Grad-CAM results strengthen this implication because they show that the clusters are numerical groupings in a measurement table and visually recoverable from the multi-view silhouettes. The cluster-level Grad-CAM summaries further indicate which silhouette regions were most influential for each representative class, making the approach useful for screening, organizing, and prioritizing fossil tooth specimens for later detailed anatomical and taxonomic evaluation.

Why simulated data are useful

The use of simulated training data is a key advantage. Paleontological datasets often suffer from small sample sizes, inconsistent imaging, and limited pixel-level annotations. Simulated silhouettes allow the model to learn basic object-background separation before being exposed to more realistic tooth-shape outlines. Although triangles and circles are simplified relative to real teeth, they provide controllable endmembers for sharp crown profiles and rounded cross-sections. Noise augmentation further helps the model tolerate photographic variation.

This approach does not replace real annotated tooth masks. Rather, it provides a pretraining and finetuning strategy that reduces the amount of manually labelled tooth data required. The strategy is consistent with transfer learning principles, in which knowledge from a broader or synthetic source domain is adapted to a smaller target domain (Pan & Yang, 2010).

Value of multi-view analysis for isolated theropod teeth studies

The proposed workflow addresses a practical gap in theropod tooth morphometrics: the lack of an automated, reproducible method for turning multi-view images into standardized silhouettes and then into quantitative descriptors. Traditional theropod tooth studies depend on detailed anatomical description and measurement, and these remain essential. However, automated silhouettes can accelerate the repetitive parts of the workflow and make it easier to compare large assemblages under consistent measurement rules.

For the Guimarota fossil record, where isolated theropod teeth are abundant and morphologically diverse, automated measurement is especially useful. The original Guimarota theropod tooth assemblage includes numerous small specimens and multiple morphotypes, making it difficult to summarize diversity using a small number of manual measurements (Zinke, 1998). A multi-view silhouette methodology captures crown height, width and regional shape, angular contour structure, and differences among labial, lingual, mesial, distal, and cross-section views.

Performance and limitations of the cluster-derived multi-view classifier

The supervised EfficientNet analysis should be interpreted as a cluster-recognition experiment rather than as a direct taxonomic classification. The five output classes were derived from HDBSCAN clusters. As a result, the EfficientNet model was trained and evaluated on 53 teeth distributed across five cluster-derived classes. This is important because model performance reflects the ability to recover previously defined groups, not necessarily the ability to identify theropod taxa independently.

Although the image dataset contains multiple views per

tooth, the number of independent biological specimens is smaller than the number of images. Each tooth contributes with five correlated images, one for each viewpoint. For this reason, model evaluation was performed at the tooth level, ensuring that all views of a given tooth remained within the same training or validation partition. This design reduces the risk of data leakage that would occur if different views of the same specimen were distributed across training and validation sets.

The small number of independent specimens remains an important limitation. Even though repeated stratified cross-validation provides a more robust estimate than a single train-validation split, each validation fold still contains only a small number of teeth distributed across five cluster-derived classes. As a result, class-level performance metrics are sensitive to the classification of very few specimens, especially for the least represented clusters. Cluster 1, for example, contains only six teeth, meaning that its representation in any validation fold is necessarily limited. The higher variability observed for this cluster should therefore be interpreted in light of its small sample size rather than as evidence of a clearly weaker biological signal. The reported classifier performance should be treated as an exploratory internal consistency assessment of the Guimarota multi-view dataset rather than as a definitive estimate of generalization to other theropod tooth assemblages. This limitation is especially relevant when comparing the present study with other paleontological machine-learning studies, because many such studies use larger numbers of independent specimens, images, or measurements. Direct comparison is therefore not straightforward. The present dataset is structured around repeated anatomical views of the same tooth, rather than fully independent image observations, and the effective biological sample size is the number of teeth, not the number of images.

At the same time, the dataset should not simply be regarded as an “easy” classification problem. The classes are not taxonomic labels imposed *a priori*, but cluster-derived groups based on the structure of the dataset. Some classes may share similar size ranges or external outlines, in which case successful classification would suggest that the model is using combinations of view-specific features rather than size alone. The multi-branch architecture is

relevant in this study because each anatomical view provides different information, where labial and lingual views emphasize crown outline and curvature, mesial and distal views capture profile shape and carinal geometry, and cross-sections summarize crown geometry in transverse view. Therefore, correct classification of similarly sized teeth supports the interpretation that the model is learning multi-view shape information rather than only separating large and small specimens.

The Grad-CAM visualizations provide a useful qualitative check on this interpretation, since activation maps were generated separately for each EfficientNet branch and restricted to the tooth silhouette, they help identify which regions of each anatomical view contributed most strongly to the predicted cluster. These visualizations indicate whether the model focused on biologically meaningful regions, such as the crown apex, basal outline, carinae, or cross-sectional shape, rather than on background artefacts. However, Grad-CAM should be interpreted cautiously, because it helps visualize model attention, but it does not by itself prove that the classifier used taxonomically diagnostic characters. The model results are therefore best understood as evidence that the selected Guimarota teeth contain visually separable multi-view morphotypes, while the biological and taxonomic meaning of those morphotypes requires independent anatomical interpretation.

Exploratory prediction and classifier concordance on unseen teeth

The application of the trained classifiers to previously unseen teeth provided an additional exploratory test of how transferable the workflow is to newly processed specimens. Because the EfficientNet and random forest models use different representations of the same teeth, their comparison is informative. EfficientNet classifies specimens directly from the five binarized anatomical views, whereas the random forest uses the derived morphometric feature matrix. Agreement between the two models therefore suggests that the image-based and measurement-based components of the workflow capture overlapping aspects of tooth shape.

At the same time, predictions for unseen specimens should be interpreted cautiously. Cluster-derived assign-

ments represent placement within a learned morphometric structure, not taxonomic identifications. Lower prediction probabilities or disagreement between classifiers may indicate that a specimen falls near the boundary between previously defined clusters, occupies an intermediate region of morphospace, or expresses shape features that are represented differently by the image-based and measurement-based models. Such cases may also suggest that a tooth is not well represented by the original cluster structure and could belong to an unsampled or underrepresented morphometric group.

The EfficientNet model can use distributed information from complete multi-view silhouettes, including local outline curvature, apical shape, basal geometry, and cross-sectional perimeter. The random forest, in contrast, is restricted to predefined measurements such as view-specific heights, widths, ratios, and angular descriptors. Therefore, discordant or low-confidence assignments may highlight specimens for which the full silhouette contains information not fully captured by the extracted measurements, or cases in which specific measured variables strongly influence the morphometric classification.

This analysis remains exploratory because the unseen teeth lack independent taxonomic labels. The predictions cannot be used to estimate external accuracy or validate the biological identity of the clusters. Instead, they demonstrate that new specimens can be passed through the same segmentation, measurement, classification, and interpretation pipeline. The key value of this step is to identify whether new teeth are confidently placed within the existing morphometric groups or whether they show uncertainty that may reflect boundary positions, unusual morphology, or potentially unsampled regions of theropod tooth morphospace.

Limitations and next steps

Several limitations should be considered. First, the clusters were generated from image-derived measurements, so they depend on segmentation quality, image orientation, and scale consistency. Since the input images were not standardized to the same physical scale, size-related variables such as height and width may partly reflect imaging differences. Future analyses should include scale calibration

or size-normalized shape variables to separate true size differences from proportional shape differences.

Second, HDBSCAN identifies dense regions in morphospace but does not prove that these groups correspond to biological taxa. The clusters should be validated against expert morphotype assignments, traditional tooth measurements, preservation categories, and, where possible, independent anatomical characters. Third, the random forest was used as an interpretability tool for existing clusters, not as an independent predictive classifier. Therefore, its performance and feature importance should be presented as exploratory. Fourth, the EfficientNet model was trained on labels derived from unsupervised clustering rather than on independent taxonomic assignments. Therefore, its classification performance should not be interpreted as taxonomic validation. Instead, it evaluates whether the HDBSCAN-derived morphometric groups are visually recoverable from the multi-view tooth silhouettes. Although repeated stratified cross-validation provides a more conservative evaluation than a single train-validation split, the number of independent teeth remains small and each validation fold contains only a few specimens per class, especially for the least represented cluster. Consequently, fold-level and class-level performance metrics are sensitive to the classification of individual specimens. The EfficientNet results should therefore be interpreted as an exploratory internal consistency check for the present Guimarota dataset, rather than as a final estimate of generalization to independent material.

Future work should focus on six improvements. First, collect more theropod teeth images from Guimarota, create a curated manually annotated validation set of images in all five views, and apply this methodology to the full collection. Second, expand the synthetic generator to produce more realistic theropod tooth silhouettes, including recurvature, basal constriction, carina-like edges, broken apices, and taphonomic noise. Third, compare silhouette-based clusters with traditional theropod tooth measurements and expert morphotype assignments. Fourth, integrate 2D measurements, raw silhouette embeddings, and 3D visual-hull descriptors into a combined multimodal clustering framework. Fifth, include information from the denticles as they have been used in other studies as an important

feature to distinguish taxonomic groups. Sixth, apply this methodology to other theropod teeth from different locations.

CONCLUSIONS

This study presents a reproducible AI-assisted methodology intended to be reproducible through shared code, data, and image inputs for multi-view theropod tooth silhouette generation, measurement extraction, clustering, and interpretation. Using DeepLabv3+ transfer learning from simulated data, the method produces binarized silhouettes that can be analyzed automatically.

Applied to 64 theropod teeth from the Guimarota locality, with 10 additional teeth reserved as an exploratory test set, the method successfully converted multi-view tooth silhouettes into a specimen-level morphometric matrix incorporating labial, lingual, mesial, distal, and cross-sectional information. UMAP ordination combined with HDBSCAN clustering recovered five morphometric groups and 11 unassigned specimens, indicating that the automated measurements captured structured variation within the assemblage while also preserving low-density or ambiguous specimens. Cluster-profile analyzes showed that the groups differed in interpretable anatomical variables, including view-specific width, height, height-to-width ratios, and contour-angle measurements. Random Forest feature importance further indicated that variables from multiple views contributed to group separation, supporting the value of a multi-view approach rather than reliance on a single tooth outline.

The cluster-derived image folders provided a second supervised step. After exclusion of unassigned specimens, 53 teeth with five views were used to train a multi-branch EfficientNet model. The model recovered the five HDBSCAN-derived morphometric classes with high mean validation accuracy under repeated stratified cross-validation, rather than relying on a single held-out validation split. Across 15 train-validation evaluations, validation accuracy was approximately 0.89, indicating that the cluster-derived classes were generally visually recoverable from the multi-view tooth silhouettes, although performance varied among folds and was less stable for the least represented cluster. Cluster-level view-specific Grad-CAM summaries showed

that predictions were influenced by the tooth silhouette itself, especially the apex, lateral margins, basal regions, longitudinal crown outline, and cross-sectional perimeter. The representative Grad-CAM examples showed that each cluster-derived class could be associated with a different combination of activated silhouette regions. Within this exploratory validation design, these results suggest that the morphometric groups are statistically identifiable from measurements, and visually recoverable from standardized multi-view silhouette images.

As an additional exploratory test, the trained multi-branch EfficientNet and Random Forest classifiers were applied to the 10 previously unseen teeth excluded from the clustering and cross-validation analyzes. The two classifiers produced concordant cluster predictions for seven of the 10 specimens, indicating substantial agreement between the image-based and morphometric approaches when applied to new material. Because independent taxonomic or morphotype labels were unavailable for these specimens, this test was interpreted as an external prediction and classifier-concordance exercise rather than as a formal assessment of classification accuracy.

The results demonstrate that automated silhouette-based measurements can provide a useful exploratory tool for grouping isolated theropod teeth into morphometric groups. These groups should not be interpreted directly as taxonomic assignments, but they provide a reproducible basis for identifying representative specimens, comparing morphometric groups, and selecting teeth for more detailed anatomical or taxonomic study. The approach is particularly useful for assemblages such as that from the Guimarota fossil site, where isolated teeth are abundant and potentially represent multiple sources of variation, including taxonomy, tooth position, size, preservation, and imaging orientation.

Future work should focus on validating the recovered clusters against expert morphotype assignments, traditional theropod tooth measurements, and scale-calibrated images. The sensitivity analysis shows that the five-cluster solution is reproducible around the baseline parameter region, but that coarser three- to four-cluster solutions can receive higher DBCV scores under other parameter combi-

nations. Additional improvements should include larger manually annotated validation datasets, more realistic simulated training data, expanded sensitivity tests for UMAP and HDBSCAN parameters, and integration of two-dimensional silhouette descriptors with optional three-dimensional visual-hull reconstructions. Despite these limitations, the proposed workflow represents a step toward more automated, reproducible, and scalable quantitative analysis of isolated theropod teeth.

DATA AND CODE AVAILABILITY

The source code, trained models, final measurement tables, image files used for the EfficientNet workflow, and other relevant results are available in Zenodo (DOI: 10.5281/zenodo.21412439). Researchers requiring access to the original photographs for reproducibility or comparative purposes should contact the corresponding author.

FUNDING

This work is funded by national funds through FCT – Fundação para a Ciência e a Tecnologia, I.P., under the doctoral scholarship FCT/CEAUL, UI/BD/154258/2022 (<https://doi.org/10.54499/UI/BD/154258/2022>), the individual contract CEECIND/01770/2018 (<https://doi.org/10.54499/CEECIND/01770/2018/CP1534/CT0004>), CEAUL Research Unit, UID/00006/2025 (<https://doi.org/10.54499/UID/00006/2025>), and by the European Union – NextGenerationEU through the project UID/PRR/00006/2025 (<https://doi.org/10.54499/UID/PRR/00006/2025>), and FCT, I.P./MCTES (PT) through national funds (PIDDAC): LA/P/0068/2020 (<https://doi.org/10.54499/LA/P/0068/2020>), UID/50019/2025 – (<https://doi.org/10.54499/UID/50019/2025>), and by the European Union – NextGenerationEU under projects UID/PRR/50019/2025 (<https://doi.org/10.54499/UID/PRR/50019/2025>) and UID/PRR2/50019/2025 (<https://doi.org/10.54499/UID/PRR2/50019/2025>).

ACKNOWLEDGMENTS

The authors would like to thank Jorge Sequeira and Rúben Dias from Museu Geológico of the Laboratório Nacional de Energia e Geologia (LNEG), Roberto Keller and Judite Alves from Museu Nacional de História Natural e da Ciência (MUHNAC) for access to specimens and equipment. The authors would also like to thank the editors of this volume for considering this contribution.

REFERENCES

- Barucci, A., Ciacchi, G., Liò, P., Azevedo, T., Di Cencio, A., Merella, M., Bianucci, G., Bosio, G., Casati, S., & Collareta, A. (2024). An explainable convolutional neural network approach to fossil shark tooth identification. *Bollettino della Società Paleontologica Italiana*, 63(3), 215–227. <https://doi.org/10.4435/BSPI.2024.15>
- Bradski, G. (2000). The OpenCV library. *Dr. Dobbs's Journal of Software Tools*, 25(11), 120–125.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
- Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In V. Ferrari, M. Hebert, C. Sminchisescu, & Y. Weiss (Eds.), *Computer vision – ECCV 2018* (pp. 833–851). Springer. https://doi.org/10.1007/978-3-030-01234-2_49
- Cope, E. D. (1876). Descriptions of some vertebrate remains from the Fort Union beds of Montana. *Proceedings of the Academy of Natural Sciences of Philadelphia*, 28, 248–261.
- Currie, P. J., Rigby, J. K., Jr., & Sloan, R. E. (1990). Theropod teeth from the Judith River Formation of southern Alberta, Canada. In K. Carpenter & P. J. Currie (Eds.), *Dinosaur systematics: Approaches and perspectives* (pp. 107–126). Cambridge University Press. <https://doi.org/10.1017/CBO9780511608377.011>
- Cheng, B., Girshick, R., Dollár, P., Berg, A. C., & Kirillov, A. (2021). Boundary IoU: Improving object-centric image segmentation evaluation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 15334–15342). Nashville, USA. <https://doi.org/10.1109/CVPR46437.2021.01508>
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 248–255). Nashville, USA. <https://doi.org/10.1109/CVPR.2009.5206848>
- Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology*, 26, 297–302. <https://doi.org/10.2307/1932409>
- Domínguez-Rodrigo, M., Brophy, J., Mathews, G. J., Pizarro-Monzo, M., & Baquedano, E. (2023). African bovid tribe classification using transfer learning and computer vision. *Annals of the New York Academy of Sciences*, 1530(1), 152–160. <https://doi.org/10.1111/nyas.15067>
- Douglas, D. H., & Peucker, T. K. (1973). Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *The Canadian Cartographer*, 10(2), 112–122. <https://doi.org/10.3138/FM57-6770-U75U-7727>
- Fürsich, F. T., Schneider, S., Werner, W., Lopez-Mir, B., & Pierce, C. S. (2022). Life at the continental-marine interface: Palaeoenvironments and biota of the Alcobaça Formation, Late Jurassic, central Portugal, with a formal definition of the unit appended. *Palaeobiodiversity and Palaeoenvironments*, 102, 1–52. <https://doi.org/10.1007/s12549-021-00496-x>
- Gerke, O., & Wings, O. (2016). Multivariate and cladistic analyses of isolated teeth reveal sympatry of theropod dinosaurs in the Late Jurassic of northern Germany. *PLoS ONE*, 11, e0158334. <https://doi.org/10.1371/journal.pone.0158334>
- Hendrickx, C., Mateus, O., & Araújo, R. (2015). A proposed terminology of theropod teeth (Dinosauria, Saurischia). *Journal of Vertebrate Paleontology*, 35, e982797. <https://doi.org/10.1080/02724634.2015.982797>
- Hendrickx, C., Tschopp, E., & Ezcurra, M. D. (2020). Taxonomic identification of isolated theropod teeth: The case of the shed tooth crown associated with *Aerosteon* (Theropoda: Megaraptora) and the dentition of Abelisauridae. *Cretaceous Research*, 108, 104312. <https://doi.org/10.1016/j.cretres.2019.104312>
- Hendrickx, C., Mateus, O., Araújo, R., & Choiniere, J. (2019). The distribution of dental features in non-avian theropod dinosaurs: Taxonomic potential, degree of homoplasy, and major evolutionary trends. *Palaeontologia Electronica*, 22(3), 22.3.74, 1–110. <https://doi.org/10.26879/820>
- Hendrickx, C., Trapman, T. H., Wills, S., Holwerda, F. M., Stein, K. H. W., Rauhut, O. W. M., Melzer, R. R., Van Woensel, J., & Reumer, J. W. F. (2024). A combined approach to identify isolated theropod teeth from the Cenomanian Kem Kem Group of Morocco: Cladistic, discriminant, and machine learning analyses. *Journal of Vertebrate Paleontology*, e2311791. <https://doi.org/10.1080/02724634.2024.2311791>
- Iakubovskii, P. (2019). Segmentation Models PyTorch [Computer software]. GitHub. https://github.com/qubvel/segmentation_models.pytorch
- Jaccard, P. (1912). The distribution of the flora in the alpine zone. *New Phytologist*, 11, 37–50. <https://doi.org/10.1111/j.1469-8137.1912.tb05611.x>
- Karimi, D., & Salcudean, S. E. (2020). Reducing the Hausdorff distance in medical image segmentation with convolutional neural networks. *IEEE Transactions on Medical Imaging*, 39, 499–513. <https://doi.org/10.1109/TMI.2019.2930068>
- Larson, D. W., & Currie, P. J. (2013). Multivariate analyses of small theropod dinosaur teeth and implications for paleoecological turnover through time. *PLoS ONE*, 8, e54329. <https://doi.org/10.1371/journal.pone.0054329>
- Liu, X., Jiang, S., Wu, R., & Shu, W. (2023). Automatic taxonomic identification based on the Fossil Image Dataset (>415,000 images) and deep convolutional neural networks. *Paleobiology*, 49, 1–22. <https://doi.org/10.1017/pab.2022.14>
- Loshchilov, I., & Hutter, F. (2019). Decoupled weight decay regularization. *International Conference on Learning Representations*. <https://doi.org/10.48550/arXiv.1711.05101>
- Louchart, A., & Pouech, J. (2017). A tooth of Archaeopterygidae (Aves) from the Lower Cretaceous of France extends the spatial and temporal occurrence of the earliest birds. *Cretaceous Research*, 73, 40–46. <https://doi.org/10.1016/j.cretres.2017.01.004>
- Marques, C. S., Dufourq, E., Pereira, S., Santos, V. F., & Malafaia, E. (2025). Enhancing the classification of isolated theropod teeth using machine learning: A comparative study. *PeerJ*, 13, e19116. <https://doi.org/10.7717/peerj.19116>
- Martin, T., & Krebs, B. (2000). Guimarota: A Jurassic ecosystem. Verlag Dr. Friedrich Pfeil. *Geological Magazine*, 138(5), 625–626. <https://doi.org/10.1017/S0016756801316084>
- McInnes, L., Healy, J., & Astels, S. (2017). hdbscan: Hierarchical density based clustering. *Journal of Open Source Software*, 2, 205. <https://doi.org/10.21105/joss.00205>
- McInnes, L., Healy, J., Saul, N., & Großberger, L. (2018). UMAP: Uniform manifold approximation and projection. *Journal of Open Source Software*, 3, 861. <https://doi.org/10.21105/joss.00861>
- Moulavi, D., Jaskowiak, P. A., Campello, R. J. G. B., Zimek, A., & Sander, J. (2014). Density-based clustering validation. In Proceedings of the 2014 SIAM International Conference on Data Mining (pp. 839–847). Society for Industrial and Applied Mathematics. <https://doi.org/10.1137/1.9781611973440.96>

- Murray, A., van Kemenade, H., Wiredfool, Clark, J. A., Karpinsky, A., Baranovič, O., Yay295, Gohlke, C., Dufresne, J., Brett, M., DWesl, Schmidt, D., Kopachev, K., Houghton, A., REDxEYE, Keith-Magee, R., Mani, S., Landey, S., Koskela, A., ... Lahd, R. (2024). *Pillow [Computer software]*. Zenodo. <https://doi.org/10.5281/zenodo.19371132>
- Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted Boltzmann machines. In *Proceedings of the 27th International Conference on Machine Learning* (pp. 807–814). <https://doi.org/10.5555/3104322.3104425>
- Ősi, A., Kolláti, R., & Nagy, A. L. (2025). Morphological and dental wear pattern analysis of Upper Cretaceous theropod teeth from central Europe. *Palaeobiodiversity and Palaeoenvironments*, 105, 499–515. <https://doi.org/10.1007/s12549-025-00657-2>
- Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1), 62–66. <https://doi.org/10.1109/TSMC.1979.4310076>
- Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22, 1345–1359. <https://doi.org/10.1109/TKDE.2009.191>
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., & Chintala, S. (2019). PyTorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems*, 32, 8024–8035. <https://doi.org/10.48550/arXiv.1912.01703>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Lin, Z., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830. <https://doi.org/10.5555/1953048.2078195>
- Python Software Foundation. (2022). Python (Version 3.10.19) [Computer software]. <https://www.python.org/>
- Ramer, U. (1972). An iterative procedure for the polygonal approximation of plane curves. *Computer Graphics and Image Processing*, 1(3), 244–256. [https://doi.org/10.1016/S0146-664X\(72\)80017-0](https://doi.org/10.1016/S0146-664X(72)80017-0)
- Rauhut, O., Martin, T., & Krebs, B. (2000). The dinosaur fauna from the Guimarota mine. In T. Martin & B. Krebs (Eds.), *Guimarota: A Jurassic ecosystem* (pp. 75–82). Verlag Dr. Friedrich Pfeil.
- Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. *Proceedings of the IEEE International Conference on Computer Vision*, 618–626. <https://doi.org/10.1109/ICCV.2017.74>
- Serrano-Martínez, A., Vidal, D., Sciscio, L., Ortega, F., & Knoll, F. (2016). Isolated theropod teeth from the Middle Jurassic of Niger and the early dental evolution of Spinosauridae. *Acta Palaeontologica Polonica*, 61(2), 403–415. <https://doi.org/10.4202/app.00101.2>
- Smith, J. B., & Dodson, P. (2003). A proposal for a standard terminology of anatomical notation and orientation in fossil vertebrate dentitions. *Journal of Vertebrate Paleontology*, 23(1), 1–12. [https://doi.org/10.1671/0272-4634\(2003\)23\[1:APFAST\]2.0.CO;2](https://doi.org/10.1671/0272-4634(2003)23[1:APFAST]2.0.CO;2)
- Smith, J. B., Vann, D. R., & Dodson, P. (2005). Dental morphology and variation in theropod dinosaurs: Implications for the taxonomic identification of isolated teeth. *The Anatomical Record Part A*, 285A, 699–736. <https://doi.org/10.1002/ar.a.20206>
- Sola, J., & Sevilla, J. (1997). Importance of input data normalization for the application of neural networks to complex industrial problems. *IEEE Transactions on Nuclear Science*, 44(3), 1464–1468. <https://doi.org/10.1109/23.589532>
- Taha, A. A., & Hanbury, A. (2015). Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool. *BMC Medical Imaging*, 15, 29. <https://doi.org/10.1186/s12880-015-0068-x>
- Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In *Proceedings of the 36th International Conference on Machine Learning* (Vol. 97, pp. 6105–6114). PMLR. <https://doi.org/10.48550/arXiv.1905.11946>
- van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., Goullart, E., & Yu, T. (2014). Scikit-image: Image processing in Python. *PeerJ*, 2, e453. <https://doi.org/10.7717/peerj.453>
- von Meyer, H. (1861). *Archaeopteryx lithographica* (Vogel-Feder) und *Pterodactylus* von Solenhofen. *Neues Jahrbuch für Mineralogie, Geognosie, Geologie und Petrefakten-Kunde*, 1861, 678–679.
- Wagner, J. A. (1859). Über einige im lithographischen Schiefer neu aufgefundene Schildkröten und Saurier. *Gelehrte Anzeigen der Bayerischen Akademie der Wissenschaften*, 49, 553.
- Wills, S., Underwood, C. J., & Barrett, P. M. (2021). Learning to see the wood for the trees: Machine learning, decision trees, and the classification of isolated theropod teeth. *Palaeontology*, 64(1), 75–99. <https://doi.org/10.1111/pala.12512>
- Yu, C., Qin, F., Li, Y., Qin, Z., & Norell, M. (2022). CT segmentation of dinosaur fossils by deep learning. *Frontiers in Earth Science*, 9, 805271. <https://doi.org/10.3389/feart.2021.805271>
- Yu, C., Qin, F., Watanabe, A., Yao, W., Li, Y., Qin, Z., Liu, Y., Wang, H., Jiangzuo, Q., Hsiang, A., Ma, C., Rayfield, E. J., Benton, M., & Xu, X. (2024). Artificial intelligence in paleontology. *Earth-Science Reviews*, 252, 104765. <https://doi.org/10.1016/j.earscirev.2024.104765>
- Zinke, J. (1998). Small theropod teeth from the Upper Jurassic coal mine of Guimarota, Portugal. *Paläontologische Zeitschrift*, 72, 179–189. <https://doi.org/10.1007/BF02987825>

doi: 10.5710/PEAPA.22.07.2026.582

Recibido: 30 de mayo 2026

Aceptado: 22 de julio 2026

Publicado: 27 de octubre 2026

