Application of artificially intelligent systems for the identification of discrete fossiliferous levels

David M. Martín-Perea[1,2,3], Lloyd A. Courtenay[4], M. Soledad Domingo[5] and Jorge Morales[1]

[1] Palaeobiology Department, Museo Nacional de Ciencias Naturales - CSIC, Madrid, Spain
[2] Geodynamics, Stratigraphy and Palaeontology Department, Universidad Complutense de Madrid, Madrid, Spain
[3] Institute of Evolution in Africa, Madrid, Spain
[4] Department of Cartographic and Land Engineering, Higher Polytechnic School of Avila, University of Salamanca, Avila, Spain
[5] Sciences, Social Sciences and Mathematics Department, Universidad Complutense de Madrid, Madrid, Spain

ABSTRACT

The separation of discrete fossiliferous levels within an archaeological or paleontological site with no clear stratigraphic horizons has historically been carried out using qualitative approaches, relying on two-dimensional transversal and longitudinal projection planes. Analyses of this type, however, can often be conditioned by subjectivity based on the perspective of the analyst. This study presents a novel use of Machine Learning algorithms for pattern recognition techniques in the automated separation and identification of fossiliferous levels. This approach can be divided into three main steps including: (1) unsupervised Machine Learning for density based clustering (2) expert-in-the-loop Collaborative Intelligence Learning for the integration of geological data followed by (3) supervised learning for the final fine-tuning of fossiliferous level models. For evaluation of these techniques, this method was tested in two Late Miocene sites of the Batallones Butte paleontological complex (Madrid, Spain). Here we show Machine Learning analyses to be a valuable tool for the processing of spatial data in an efficient and quantitative manner, successfully identifying the presence of discrete fossiliferous levels in both Batallones-3 and Batallones-10. Three discrete fossiliferous levels have been identified in Batallones-3, whereas another three have been differentiated in Batallones-10.

INTRODUCTION

The Batallones Butte, located 30 km to the south of Madrid (Spain, Fig. 1A) and 1 km to the east of Valdemoro (Fig. 1B), is home to nine Late Miocene paleontological sites (named Batallones-1, Batallones-2 and so on). These sites hypothetically correspond to hourglass-shaped cavities with upper openings (Fig. 1C), formed as a consequence of pseudokarstic processes (*Pozo et al., 2004*; *Calvo et al., 2013*), where mammals became
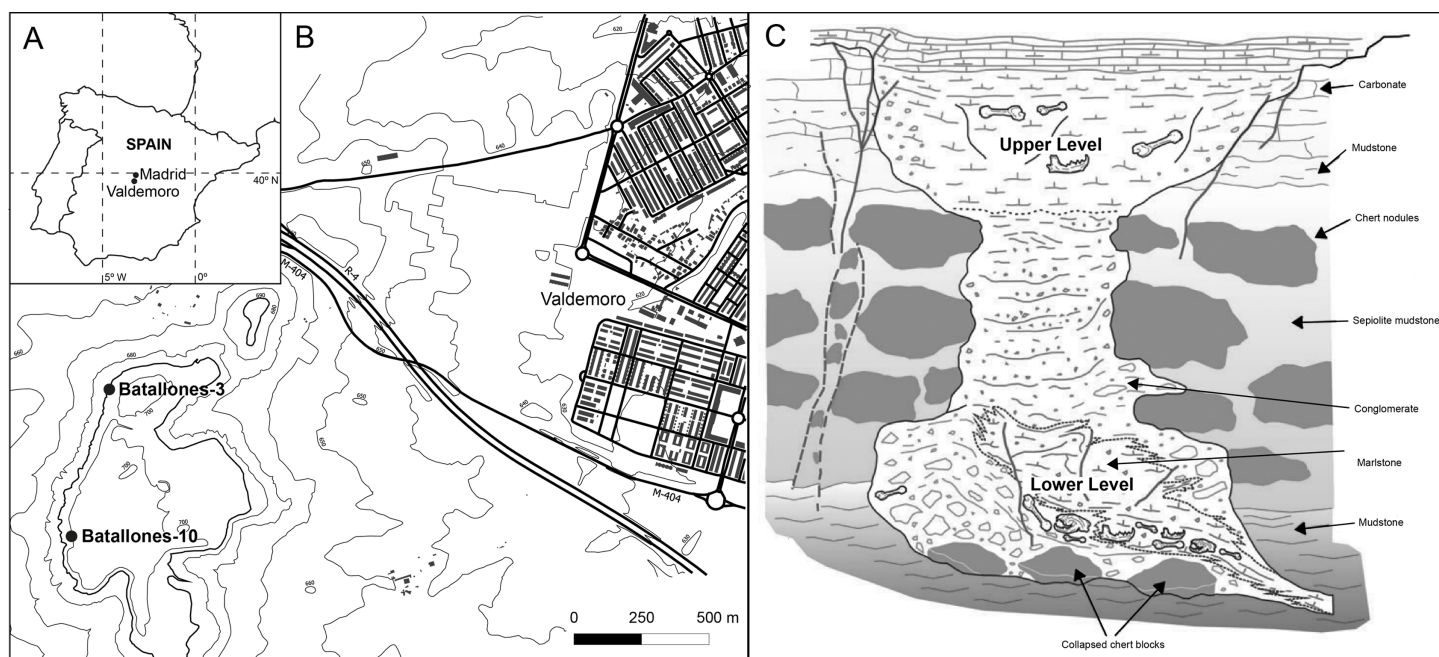
Figure 1 Geographical and geological background. (A) Map of Spain showing the location of the cities of Madrid and Valdemoro. (B) Detailed map of the situation of the Batallones Butte sites Batallones-3 and Batallones-10. (C) Hypothetical geomorphology and geology of Batallones Butte cavities, with an herbivore-dominated upper part and a carnivore-dominated lower cavity (modified from *Calvo et al. (2013)*).

Full-size 🖼 DOI: 10.7717/peerj.8767/fig-1

trapped (*Domingo et al., 2011*, *2012*, *2013a*). The upper part of this hour-glass structure is formed by deposits where mammalian herbivore bones predominate, whereas the lower part of the structure is overwhelmingly dominated by carnivoran remains (more than 90% of the fossils). Both assemblages contain abundant, diverse and well-preserved remains (*Domingo et al., 2013a*). These cavities were located in a landscape composed by woodland with patches of wooded grassland (*Domingo et al., 2013b*, *2016*)

The Batallones Butte sites have a late Vallesian age (ca. 9.1 Ma; early Late Miocene), based on the faunal association (*Morales et al., 1992*, *2004*; *Domingo, Alberdi & Azanza, 2007*; *Morales et al., 2008*; *López-Antoñanzas et al., 2010*; *Gómez Cano, Hernández Fernández & Álvarez-Sierra, 2011*). However, micromammal studies have shown that site formation was not synchronic between sites, with Batallones-10 being older than Batallones-1, which in turn is older than Batallones-3 (*López-Antoñanzas et al., 2010*).

The decimeter to centimeter-scale separation of discrete fossiliferous levels within the same lithostratigraphic unit is common in Cenozoic sites (*Canals, Vallverdú & Carbonell, 2003*; *Uribelarrea et al., 2014*; *Sañudo, Blasco & Fernández Peris, 2016*; *Gravina et al., 2018*; *Martín-Perea et al., 2019*), but tend to be qualitative, using two-dimensional transversal and longitudinal projection planes made in continuous strips (*Canals, Vallverdú & Carbonell, 2003*). Using similar methods, *Martín Escorza & Morales (2005)* preliminarily inferred discrete fossiliferous levels at Batallones-1 using spatial data from the first 3 years of the excavation (2,273 remains, 17.19% of the total Batallones-1 excavated sample).

Palaeoecological, palaeoenvironmental and taphonomical studies carried out at Batallones Butte (*Antón & Morales, 2000*; *Domingo et al., 2011*, *2013a*, *2013b*, *2016*) have treated the upper and lower part of the hourglass-shaped cavities separately based on their different depth and their different taxonomical composition (herbivores vs. carnivores, respectively). Due to the challenges posed by the structure of the deposits that filled the Batallones Butte cavities, such as lateral facies changes, deformations, depositional asymmetries, collapse structures, slickensides and/or local tilting typical of cave deposits (*Calvo et al., 2013*), geological sub-levels have not been identified within each of these parts. However, the question still remains on whether or not discrete fossiliferous levels are found embedded in the apparently geologically continuous deposits.

The pioneering introduction of Artificially Intelligent Algorithms (AIAs) in fields of archaeology and paleoanthropology has revolutionized numerous sub-disciplines such as those related with genetic sequencing (*Mondal, Bertranpetit & Lao, 2019*), site and object detection (*Anemone, Emerson & Conroy, 2011*; *Conroy et al., 2012*; *Emerson & Anemone, 2012*; *Emerson et al., 2015*; *Benhabiles & Tabia, 2016*; *Block et al., 2016*; *Wills, Choiniere & Barrett, 2018*; *Anemone & Conroy, 2018*; *Caspari & Crespo, 2019*; *Verschoof-van der Vaart & Lambers, 2019*), physical anthropology (*Bewes et al., 2019*), biomechanics (*Püschel et al., 2018*) restoration (*Derech, Tal & Shimshoni, 2018*; *Hermoza & Sipiarn, 2018*), as well as taphonomy (*Arriaza & Domínguez-Rodrigo, 2016*; *Domínguez-Rodrigo, 2019*; *Egeland et al., 2018*; *Byeon et al., 2019*; *Courtenay et al., 2019*; *Moclán, Domínguez-Rodrigo & Yravedra, 2019*).

AIAs, including those trained through Machine Learning and Deep Learning (ML & DL) techniques, present numerous possibilities for the processing of highly complex and noisy data sets (*Bishop, 2006*). Combined with their versatility, in many cases ML and DL approaches have outperformed human experts in a multitude of specialized tasks. Here we present a means of utilizing AIAs for the detection of patterns in 3D spatial data, establishing a method to objectively and efficiently detect discrete levels that human experts may have mistakenly overlooked. This study includes both unsupervised and supervised ML, joined through a hybrid Human-AI collaborative approach in order to establish a final model that can be used to describe the palaeontostratigraphic nature of each site.

The main goal of this study is to establish whether or not ML analysis can be used to quantitatively identify discrete fossiliferous levels in paleontological and archaeological sites based on the study of the spatial distribution of fossils. Specifically, this method has been applied and tested in two Batallones Butte sites (Batallones-3 and Batallones-10) in order to determine if the deposits are homogeneous or heterogeneous with discrete fossiliferous levels.

MATERIALS AND METHODS

The site of Batallones-3 has undergone systematic field excavations between the years 2001 and 2017, whereas excavations at Batallones-10 started in 2007 and are ongoing. Batallones-3 corresponds to the carnivoran-rich lower part of the hour-glass structure previously mentioned. It is not clear whether an upper part ever existed for this fossil site,

but, if it existed, it was destroyed by slope-erosion processes (this site is located in the slope of the Batallones Butte). In turn, Batallones-10 corresponds to the herbivore-rich upper part of the hour-glass structure and, if present, will have a lower part, yet to be excavated. At both sites, standard fossil vertebrate excavation protocols were followed in the extraction of paleontological remains (*Eberth, Rogers & Fiorillo, 2007*). On-site documentation of excavated remains has primarily consisted of taxonomical and anatomical identification as well as in depth documentation of the spatial distribution of remains. Spatial data obtained thus include standard x, y and z coordinates according to their position within a grid, as well as the trend and plunge of elongated fossil remains. Additional data, for future taphonomical studies, were also collected concerning the degree of articulation between anatomical elements, the overall preservation of the remains and the restoration techniques used.

For the purpose of this study, the data used for both supervised and unsupervised ML applications consisted solely in the spatial distribution of remains, including their x, y and z coordinates for in-depth 3D analysis. Manageable slices were extracted, similar to the approach proposed by *Canals, Vallverdú & Carbonell (2003)*, 2 for each site across the entire length of the excavated area. In the case of Batallones-10, one 2-m-wide slice was taken across the x and one across the y axis, searching for representative areas with the greatest object densities to capture a global vision of the entire site. In the case of Batallones-3, located in a domically-shaped cave with a debris cone in the middle under an inferred opening, two 2-m-wide slices were obtained along the site's x axis, one on each side of the debris cone.

Once slices were selected, the coordinates of each of the remains were imported into the free R software, x64 v.3.5.1 (www.rproject.org, *R Development Core Team, 2018*), for further analysis.

Unsupervised machine learning

For initial detection of hidden patterns among the levels of Batallones Butte site, an unsupervised density based clustering algorithm was used. Unsupervised ML algorithms are highly efficient AIAs that are exceptional for their use in tasks such as pattern recognition, anomaly detection, noise and dimensionality reduction, feature engineering as well as generative modeling (*Patel, 2019*). While unsupervised learning performs poorly in specifically defined tasks, their greatest advantage can be seen through their flexibility when applied to versatile datasets for general feature extraction. A popular application of unsupervised algorithms lies in clustering tasks, a component of pattern recognition that is, considered useful for processing and searching for hidden trends in highly noisy datasets.

A non-parametric Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm from the "fpc" R package was used for pattern recognition in 3D spatial data. Proposed as a means of overcoming clustering issues where groupings of samples are not straightforward (*Ester et al., 1996*), DBSCAN is highly efficient at detecting patterns in arbitrary and noisily distributed samples (*Sander et al., 1998*; *Schubert et al., 2017*).

DBSCAN performs through establishing areas of significant point densities, thus providing an efficient means of finding non-parametric patterns among noisy data sets while easily detecting outliers in low-density regions. This is especially useful when used for pattern recognition in spatial studies, where traditional clustering models such as partitioning and hierarchical algorithms tend to focus groupings around a centroid or subsamples in a convex manner. This frequently creates conflict in real-life circumstances.

DBSCAN has proven efficient when applied to 2D, 3D and/or higher dimensional feature spaces (*Ester et al., 1996*; *Sander et al., 1998*; *Schubert et al., 2017*). The algorithm works by separating points within the cluster (i.e., *core points*) from those on the border of the cluster (i.e., *border points*). The differentiation between the two is established mathematically depending on the reachability of neighboring points (q) from a core point (p). To define this, DBSCAN requires two main hyperparameters, the ε value defining the neighborhood (a.k.a. Eps) of a point and the minimum number of points (MinPts) that are required to form the cluster (i.e., the density of points). From here DBSCAN defines the neighborhood of p within dataset D as $N_{Eps}(p) = \{q \in D \mid dist(p, q) \leq \varepsilon\}$, adjusting for border points and the exclusion of noise through $p \in N_{Eps}(q)$ and $|N_{Eps}(q)| \geq$ MinPts (*Ester et al., 1996*). The definition of *density-reachable* points for clustering is thus established if the points can be linked in a chain, whereby $p_1, \ldots, p_n = q, p_n = p$ ensuring p_{i+1} is directly reachable through density from p_i. Through this definition, border points may not be density reachable if they do not fulfill the core point condition $|N_{Eps}(q)| \geq$ MinPts. The final components that define point density for clustering are established through *density-connected* points, whereby a point o is density-reachable from both p and q, essentially connecting both border and cluster points.

Once density has been established within the sample, the assigning of these points into cluster C requires the fulfillment of conditions; (1) $\forall\ p, q$: if $p \in C$ and q fulfill the definition of being density reachable, then $q \in C$, while (2) $\forall\ p, q \in C$: p must be density-connected to q. These defined conditions are named *Maximality* and *Connectivity*, respectively (*Ester et al., 1996*). In accordance with the aforementioned conditions, the final definition of noise can be logically established through $\{p \in D \mid \forall\ i: p \notin C_i\}$.

DBSCAN was thus trained on the entire dataset in an unsupervised manner, establishing an average MinPts value between 3 and 5. These values were intuitively chosen depending on slices with tightly packed densities (MinPts $\approx$ 5) and slices with a lower number of remains (MinPts $\approx$ 3). This abides by general rules recommended by numerous authors (*Sander et al., 1998*; *Schubert et al., 2017*), who define an approximate optimum MinPts = 2 $\times$ n$^\circ$ dimensions in the dataset which is then adjusted depending on the complexity of point distributions and the analyst's knowledge of the domain under consideration (*Sander et al., 1998*; *Schubert et al., 2017*). ε values were established in accordance with both the correspondent MinPts values and the slices under study. For optimization of this hyperparameter, k-distance graphs were plotted according to the nearest neighbor (*Ester et al., 1996*), employing the "elbow" technique to find a ε value according to both the MinPts parameter and the actual dataset (*Thorndike, 1953*; *Patel, 2019*; *Satopa et al., 2017*). The use of this heuristic, as provided by the "dbscan" R package, can then be adjusted to find the smallest ε value possible that can define a final model for

training. DBSCAN then establishes clusters within the dataset, separating noise into a separate "0" cluster and creating convex hulls through the use of border points q.

Collaborative intelligence learning

Considering the currently unquantifiable nature of numerous geological features that ML algorithms may be unable to detect, a human-in-the-loop collaborative strategy was proposed for this study (*Kamar, 2016*; *Holzinger, 2016*; *Dellermann et al., 2019*). The objectives of including this hybrid intelligence strategy meant that human interaction could complement the strengths of the pattern detection algorithms (*Simard et al., 2017*; *Dellermann et al., 2019*), whereby creating a bridge between both unsupervised and supervised techniques for model creation.

The Expert-in-the-Loop (EitL) approach adopted here consisted in the revision of the clustered dataset by considering possible underlying geological components that could be generating noise undetectable by the model. Components of this nature could easily include disconformities, geological faults, erosive surfaces, uneven sedimentation or even sedimentation of geological elements such as large boulders that could be separating point-densities that would essentially belong to the same layer.

The strategies employed consequently used EitL (in this case geologist-in-the-loop) interaction to manually correct clustered patterns, assessing which of the clustered groups were geologically separated and assigning these clusters to a new labeled layer. This was performed as objectively as possible through complimenting paleontological data routinely collected from the site during excavation and evaluating the nature of each of DBSCAN's groupings. EitL thus benefits through *Machine Teaching* (MT) in as much as the expert knowledge is used for troubleshooting and debugging (*Dellermann et al., 2019*; also known as a *sense-making* approach). Moreover, the amount of human input for models was monitored in a *collective* manner, using numerous *domain experts* to ensure accuracy (*Dellermann et al., 2019*). In cases where clusters seemed dubious or could not be clearly defined by either the EitL participants or the MT algorithm, these were stripped of their labels and included in the noise (cluster "0") group for objective classification by the trained ML algorithms in the supervised phase of the system. Once revised by the EitL participants, each cluster group was assigned a new label that could be used to train supervised algorithms in the following part of the workflow.

Supervised learning

Once spatial data points had been passed through the DBSCAN algorithm for initial pattern detection and then corrected by geologist-in-the-loop interventions, supervised algorithms were employed to define the final fossiliferous level models and classify those points in cluster group "0".

Considering the amount of data available, no bootstrapping procedures were deemed necessary prior to supervised training. Each algorithm was thus trained using 70:30% [*train:test*] splits using k-fold cross-validation ($k = 10$) in order to ensure the model could efficiently adjust its weights. While some authors propose the use of Spatial Cross-Validation (SCV) for studies regarding geographic data (*Miller, 2004*;

Brenning, 2012; *Pohjankukka et al., 2017*; *Lovelace, Nowosad & Muenchow, 2019*), these algorithms were not seen to produce any significant change to the quality of results. Additionally, hyperparameter optimization was performed using a random search loop function programed in R; establishing the best hyperparameter values for each model using a combination of random values until finding the optimum settings (*Bergstra & Bengio, 2012*). These loop algorithms ran for 50 iterations and were then extrapolated and used for the final classification models.

Two primary supervised ML algorithms were used for the final fine tuning of the fossiliferous level models.

- **Support Vector Machines (SVM)**. SVMs map out input vectors into a non-linear high dimensional feature space, using hyperplanes to calculate the degree of separation between samples (*Cortes & Vapnik, 1995*). In order to overcome traditional limitations imposed by linearity, a *kernel* function is used to define the feature space (*Bishop, 2006*). The constructed hyperplane can thus be used as a discriminant classifier decision surface which uses maximized margin or decision boundaries to reduce chances of overfitting (*Cortes & Vapnik, 1995*; *Bishop, 2006*). The consequent hyperplane can then be used to plot the maximized margin separations between samples and can provide a visual means of dividing strata, in an efficient and objectively computed way. For SVM applications the "e1071" R package was used.

- **Random Forest (RF)**. RF can be described as a robust and highly complex form of decision tree which has proven useful in the past for the processing of spatio-temporal data (*Hengl et al., 2018*). The RF algorithms use small random numbers of data set variables rather than the whole dataset, constructing an independent decision tree with each subsampling (*Breiman et al., 1984*; *Breiman, 2001*). The random variable selection is additionally performed using bootstrap aggregation, using a technique frequently referred to as out-of-bag observations. RF is then able to calculate the number of iterations needed to minimize the out-of-bag error. Once the number of trees has been selected, the algorithm averages results to produce a robust classification model (*Breiman, 2001*). For RF applications the "caret" R package was used.

Model evaluations were performed following standardized ML protocol, evaluating the Kappa (κ), Sensitivity, Specificity and Balanced Accuracy values obtained through confusion matrices (*Kuhn & Johnson, 2013*; *Lantz, 2013*). Sensitivity, Specificity and Balanced Accuracy are values derived from evaluations of Type I and Type II statistical errors in proportion with the rest of the calculated confusion matrix. These values are presented as numbers between 0 (poor) and 1 (high performing classifiers). κ values are a further statistical adjustment of the accuracy metric that measure model agreement relative to what would be expected by chance (*Kuhn, 2008*). These values are similarly presented as numbers between 0 and 1 although negative values can (although rarely) occur. Further evaluation of this statistic uses 0.8 as the threshold between poor and powerful classification models. Finally, loss metrics were recorded to evaluate the

predictive power of each model. For this, the Mean Squared Error (MSE) metric was employed:

$$\frac{1}{n}\sum_{i=1}^{n}(t_i - p_i)^2$$

taking into consideration the target (t_i) and predicted (p_i) values for each classification. MSE values are further interpreted considering AIAs of this type are trained to reduce the error produced in CV sets, therefore the smaller the MSE value the more powerful the predictive model. Model evaluation was performed using the "caret" R package.

Fine tuning of fossiliferous levels

Once the optimal models were defined, these were used to classify any of the points separated as noise by DBSCAN as well as the clusters that the EitL participants could not objectively define as a part of either fossiliferous level (so as to avoid subjectivity). In order to empirically class any point as a member of a fossiliferous level, the trained SVM and RF algorithms were used to predict the class label of each point. A threshold of 80% security was established as acceptable. Any points that could not be classed by the model with over an 80% predictive decision boundary were thus rejected and therefore not classified into any level in the final fossiliferous levels models.

RESULTS

Unsupervised machine learning clustering

DBSCAN clustering proved highly efficient at quickly processing each slice, grouping fossil remains into abundant clusters in both Batallones-3 (Fig. 2) and Batallones-10 (Fig. 3). DBSCAN results in most cases present very noisy profiles, with the detection of large numbers of tightly packed clusters. This is a result of the irregular density patterns detected in the data and probably results from numerous geological and paleontological disconformities that the algorithm is unable to quantify in some areas. Nevertheless, profiles become less noisy towards the extremities of each slice, helping the differentiation between levels, identifying at least three different fossiliferous levels.

On average, DBSCAN identified 166 points (5.6%) as noise in Batallones-10 and 62 (7.76%) in Batallones-3, with the highest levels of noise appearing in Batallones-10 x axis slice ($n = 141$) and Batallones-3 right slice ($n = 56$). For Batallones-3 right slice, this can be attributed to certain areas of the profile having little separation between the layers (Fig. 2C). Nevertheless, excluding noise, most clustered groups seem to follow a trend or pattern that highlights a separation between different fossiliferous levels (see "Collaborative Intelligence Learning").

Collaborative intelligence learning

While DBSCAN results produced a large number of clusters, detailed analysis through the EitL process identified numerous groupings that belonged to the same fossiliferous level. After careful evaluation, each of these separations could be recognized as products of minute or large disconformities. These can be attributed to the cavity's asymmetrical
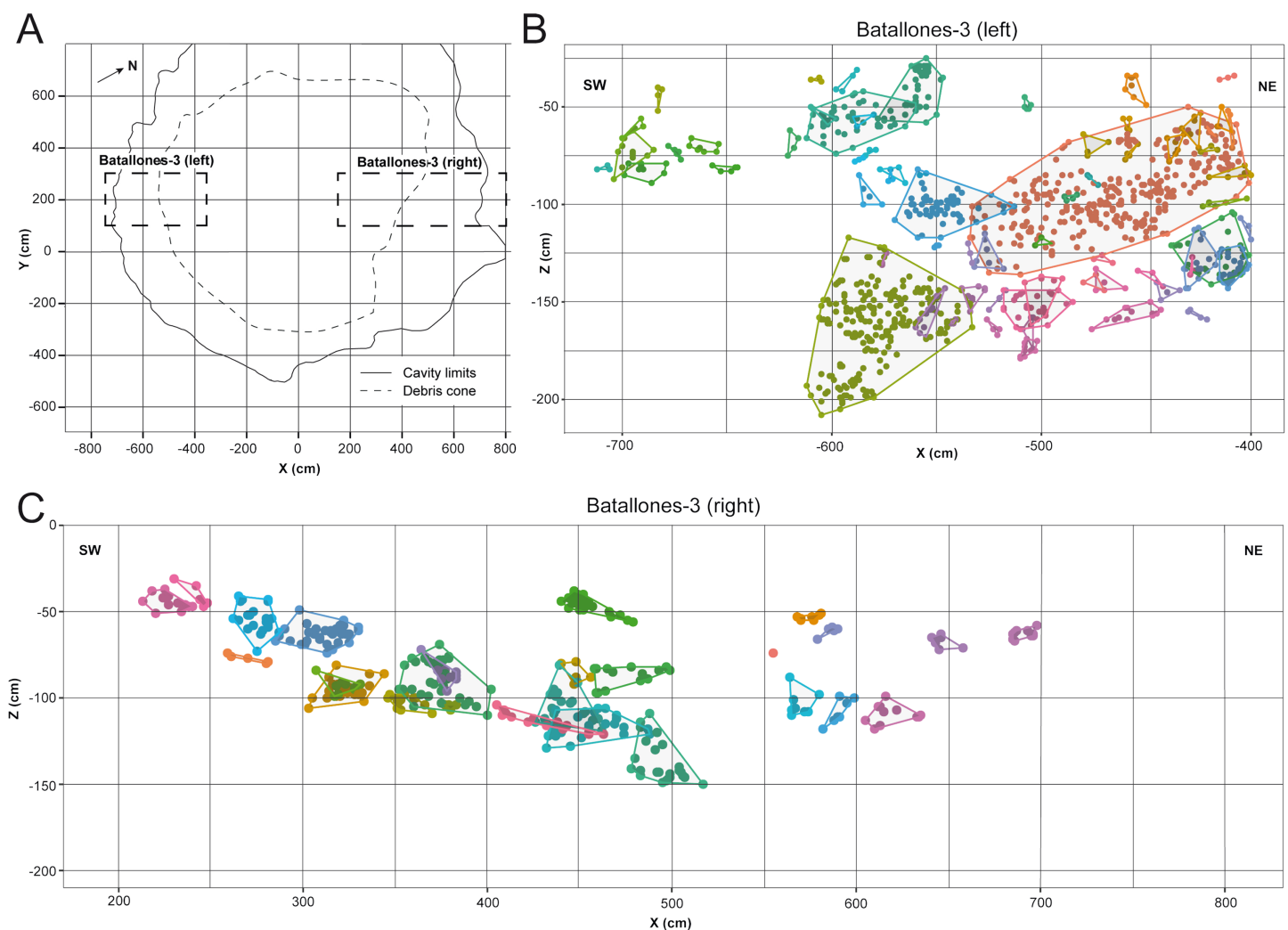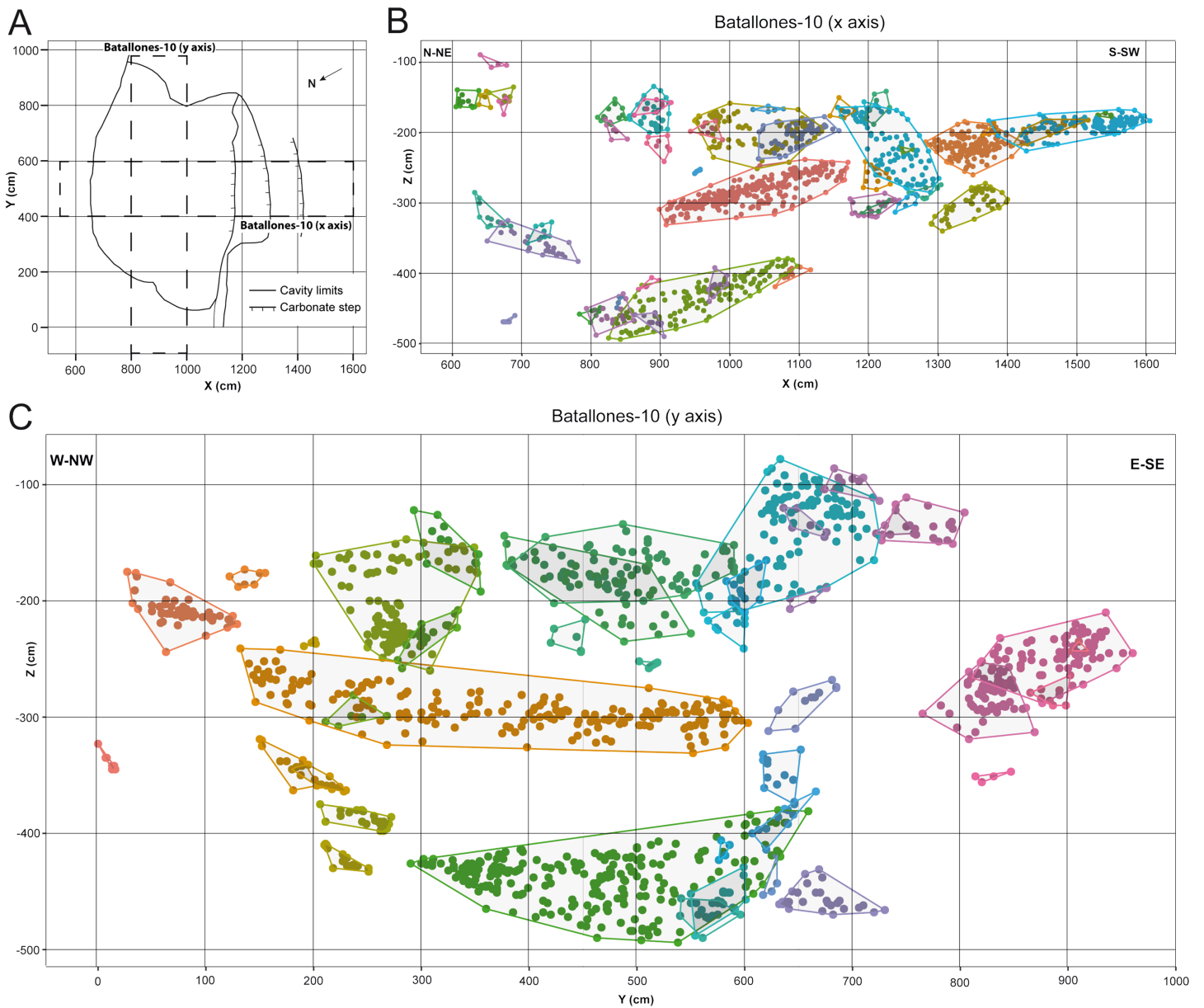
Figure 2 DBSCAN clustering for the studied slices at Batallones-3, each point representing a fossil remain. Z, depth. (A) Batallones-3 grid and slice orientation. (B) Batallones-3 left slice clusters (MinPts = 3, ε = 11.5). (C) Batallones-3 right slice clusters (MinPts = 5, ε = 15).
Full-size 🖼 DOI: 10.7717/peerj.8767/fig-2

geomorphology at Batallones-10, visible in the *x* axis slice (Fig. 4A) or to a fallen large carbonate block in Batallones-10, as seen in the *y* axis slice (Fig. 4B). This highlights the importance of the geologist-in-the-loop hybrid intelligence system for clarification and unification of clusters.

In other cases, numerous examples can be seen where depositional processes have placed fossiliferous levels closer together such as the sedimentary onlap between levels II and III in Batallones-10 *x* axis slice (Fig. 4A) and Batallones-3 right slice (Fig. 5B). Nevertheless, considering the algorithms' effectiveness processing 3D data, DBSCAN is still able to detect density patterns across the slice's *x*, *y* and *z* axes, thus separating the different levels.

Supervised learning

Once processed and cleaned by both DBSCAN and the EitL specialists, databases were channeled into supervised algorithms for training. In all cases, each of the databases

Figure 3 DBSCAN clustering for the studied slices at Batallones-10, each point representing a fossil remain. Z, depth. (A) Batallones-10 grid and slice orientation. (B) Batallones-10 *x* axis slice clusters (MinPts = 5, ε = 27.5). (C) Batallones-10 *y* axis slice clusters (MinPts = 5, ε = 25).
Full-size ⤓ DOI: 10.7717/peerj.8767/fig-3

generated by unsupervised learning algorithms proved to be highly efficient for the training of their supervised counterparts. Both SVM and RF obtained exceptional results with over 90% accuracy when differentiating between groups and small MSE values on all accounts (Table 1). While SVM proved to have greater balanced accuracy on training sets, RF in general obtained optimal MSE values in testing, especially in the case of Batallones-10. Interestingly Batallones-3 (both left and right slices) seems to create the most amount of confusion, nevertheless on all accounts κ values remain above the acceptable threshold.

Martín-Perea et al. (2020), *PeerJ*, DOI 10.7717/peerj.8767

10/25
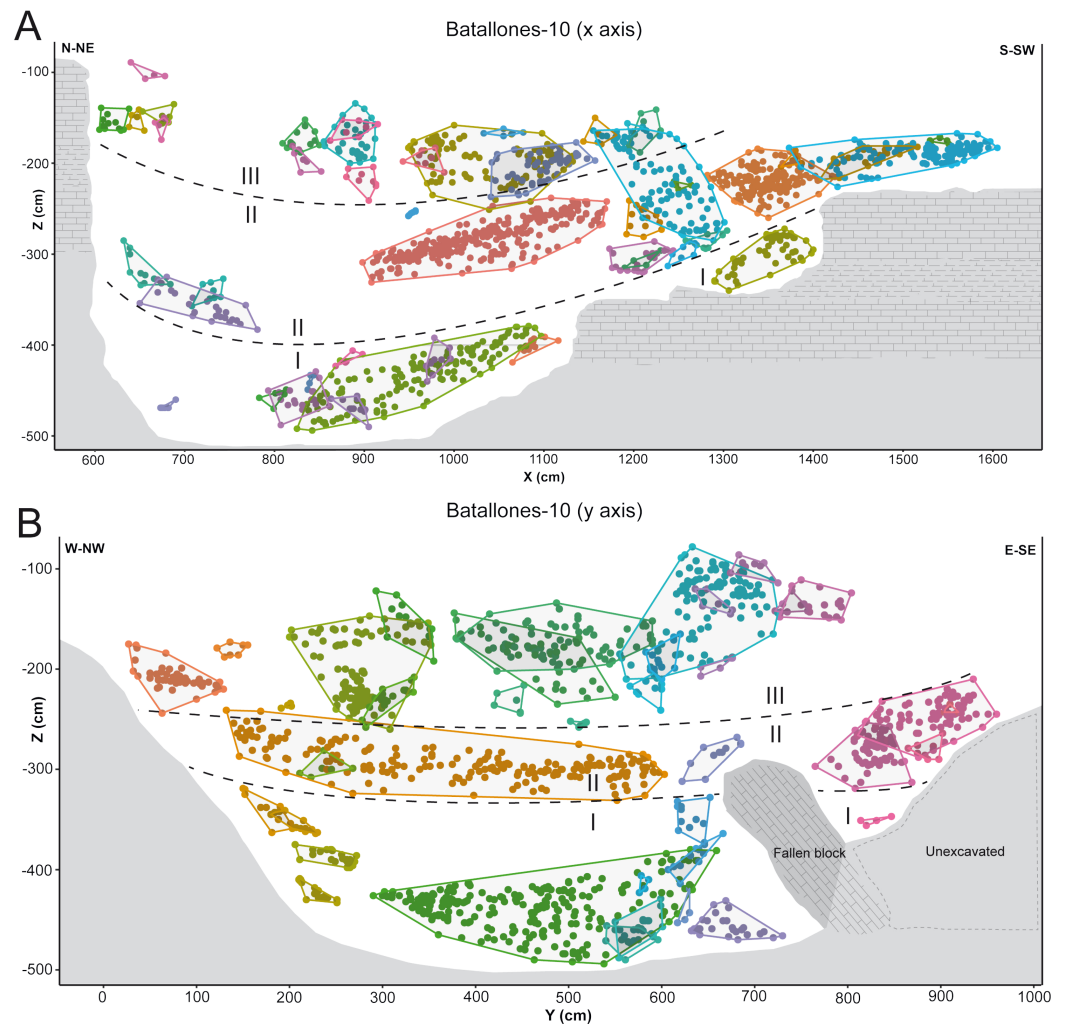
Figure 4 DBSCAN clustering and geological interpretation. In gray, site limits. (A) Batallones-10 *x* axis slice clusters and interpretation. (B) Batallones-10 *y* axis slice clusters and interpretation.

Full-size 🖼 DOI: 10.7717/peerj.8767/fig-4

Using AIAs to clean the final profiles and de-noise the datasets, both SVM and RF proved to be confident classifiers when making predictions. In most cases, over half the points denoted as noise by the previous phases were successfully classified and assigned into a palaeontostratigraphic level (Table 2). For Batallones-10 *y* axis slice and Batallones-3 left slice, RF outperformed SVM with greater MSE values and success in classifying indeterminable points. Nevertheless, in either case, RF tended to have greater decision making capabilities on some individual points than SVM, even though SVM obtained greater overall MSE in others.

Upon further evaluation of model performance, SVM maximized margin decision boundaries (dotted line in Fig. 6) are seen to expand or contract closer to the identified sedimentary onlap areas (in both Batallones-3 and Batallones-10) while an increase in MSE is observed for these particular areas as well (Table 2).
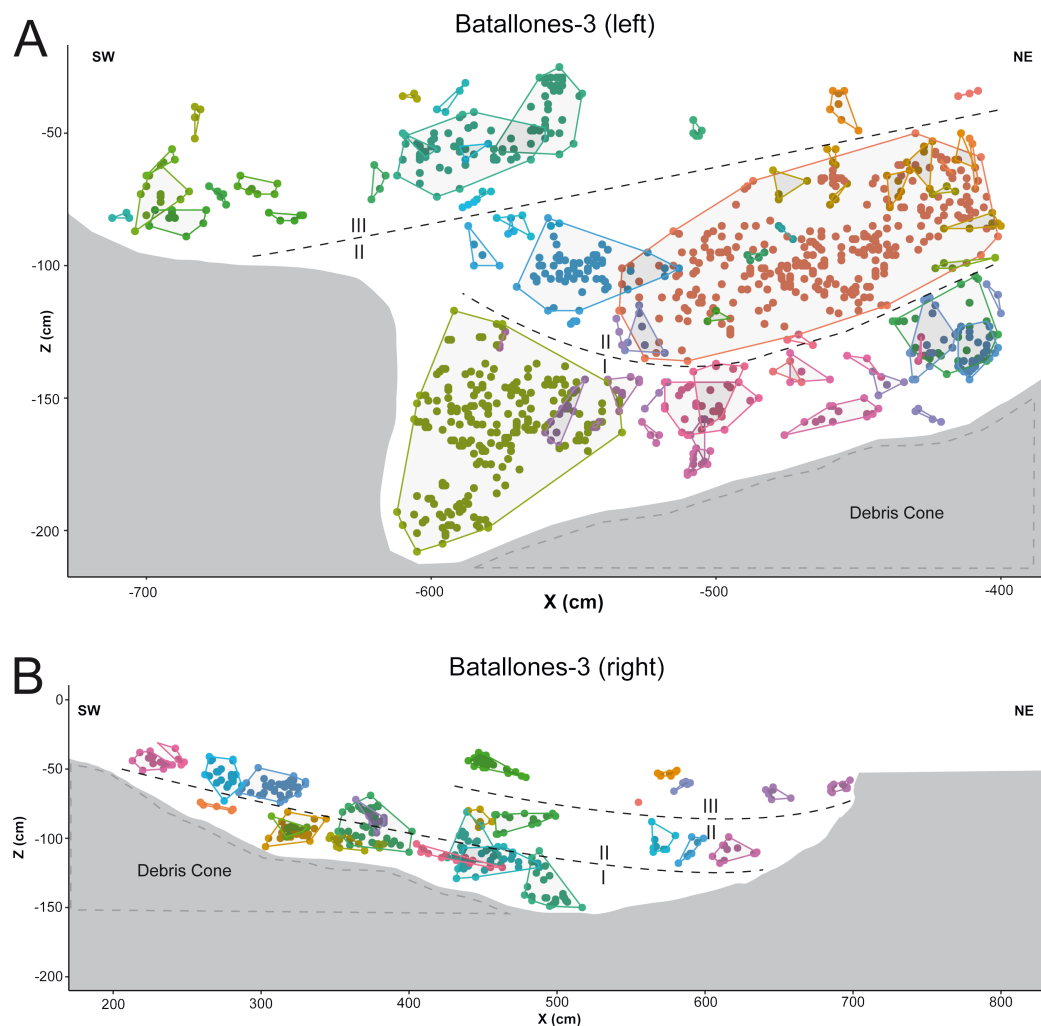
Figure 5 **DBSCAN clustering and geological interpretation. In gray, site limits.** (A) Batallones-3 left slice clusters and interpretation. (B) Batallones-3 right slice clusters and interpretation.

Full-size 🖼 DOI: 10.7717/peerj.8767/fig-5

Fine-tuned models

Once the data had been processed by each of the phases of the workflow, a final fine-tuned model of the fossiliferous levels was created.

The fine-tuned model produced for Batallones-3 shows three discrete fossiliferous levels on either side (left and right) of the debris cone (Fig. 7). The left profile (Fig. 7A) is laterally less extensive (~3 m) than the right profile (Fig. 7B), which extends over 5 m. All identified levels in both profiles dip from the debris cone outwards, with dip angles decreasing towards the outermost parts of the cavity and towards the top of the infilling, with Level 3 dipping only slightly in Batallones-3 left profile (Fig. 7A) and nearly horizontal in the outer part of the Batallones-3 right profile (Fig. 7B).

Batallones-10 fine-tuned model shows three different discrete fossiliferous levels (Fig. 8). A first level (Level I), dips northward (Fig. 8A) and towards the East and West dips

Table 1 Evaluation metrics of each ML algorithm on test sets, as derived from confusion matrices in File S1. Confidence Intervals (CI) are established for the accuracy metric.

	Random forest				Support vector machine			
	Batallones-10		Batallones-3		Batallones-10		Batallones-3	
	x axis	y axis	Left	Right	x axis	y axis	Left	Right
Kappa	0.99	0.99	0.97	0.97	1.00	1.00	0.99	1.00
Lower CI	0.98	0.98	0.93	0.93	0.99	0.99	0.96	0.96
Accuracy	0.99	0.99	0.98	0.98	1.00	1.00	1.00	1.00
Upper CI	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Sensitivity	0.99	1.00	0.98	0.98	1.00	1.00	1.00	1.00
Specificity	1.00	1.00	0.99	0.99	1.00	1.00	1.00	1.00
MSE	7.73E−06	3.39E−05	1.15E−03	1.15E−03	7.76E−05	6.84E−04	6.07E−05	9.37E−04

Note:
MSE, Mean squared error.

Table 2 Results obtained from ML algorithm for the classification of points considered indeterminable after both MT and EitL intervention.

		Batallones-10 x axis	Batallones-10 y axis	Batallones-3 left	Batallones-3 right
SVM	MSE	3.45E−03	1.21E−02	NA	7.11E−03
	Determinable	113	15	0	45
	Indeterminable	28	10	6	11
RF	MSE	4.01E−03	6.17E−05	3.18E−02	8.88E−03
	Determinable	106	23	4	39
	Indeterminable	35	2	2	17

Note:
MSE values represent the confidence of classifications.

towards the central part of the cavity, adapting to the cave limits (Fig. 8B). Level II conformably overlies Level I in the southern section, dipping towards the North and in the northern, eastern and western cavity limits dips towards the central part of the cavity (Figs. 8A and 8B). Finally, Level III is practically horizontal, with gentle folding in the outermost limits of the cavity (Figs. 8A and 8B).

DISCUSSION

The methodology proposed in this study, based on artificially intelligent systems, has been deemed effective in quantitatively identifying discrete fossiliferous levels in paleontological and archaeological sites based on the study of the spatial distribution of fossils. This new, more quantitative approach could be considered a valuable substitute for previous qualitative techniques that have been applied in the past.

Evaluating variable importance through testing for mean decrease in node impurity for RF algorithms (*Louppe et al., 2013*) explains how a boost in the weight of the third dimension of each slice helps support the differentiation between levels (Fig. 9). Batallones-3, for example, shows a shift in the value of the x axis between the left-hand and
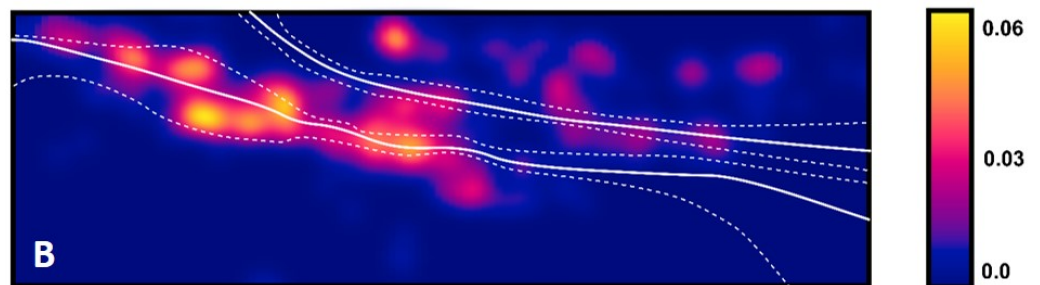
Batallones-10
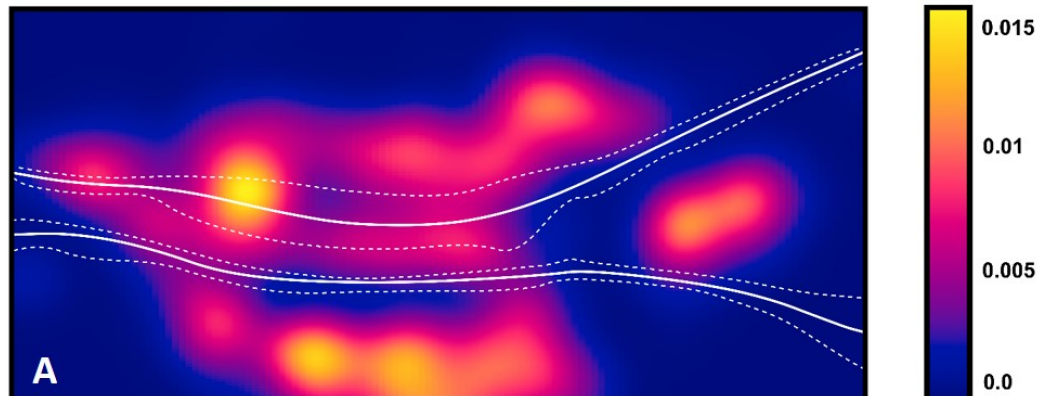
Batallones-3

Figure 6 Examples of SVM maximized margins (dotted) and decision boundaries (solid line), plotted against calculations of material densities. Density values are reported as the number of points per square unit (cm), per quadrat. (A) Batallones-10 *y* axis slice. (B) Batallones-3 right slice.
Full-size 🖼 DOI: 10.7717/peerj.8767/fig-6

right-hand sides of the debris cone, having more weight in the final decision making capabilities of RF than the y dimension. Data of this nature thus supports the need for 3D over 2D analysis in order to ensure an efficient detection of patterns that 2D data may not be able to reveal.

Fine-tuned models produced for Batallones-3, in the lower part of a hypothetical hourglass-shaped cavity, clearly show three discrete levels on either side of the debris cone (Fig. 7). Cave asymmetry, as explained in *Calvo et al. (2013)*, is clearly visible in these profiles, with greater cave development towards the Northeast, in Batallones-3 right profile (Fig. 7B). Batallones-10 fine-tuned models, in the upper part of the hourglass-shaped cavities, show another three discrete fossiliferous levels (Fig. 8). Similar geological structures were observed by *Calvo et al. (2013)* in Batallones-9, located 50 m to the north of Batallones-10.

These newly discovered levels in both the lower levels (Batallones-3) and upper levels (Batallones-10) of the hourglass-shaped cavities of Batallones Butte site support the preliminary inferences made by *Martín Escorza & Morales (2005)* about the possibility of discrete fossiliferous levels within Batallones-1. Careful revision should therefore be carried out considering previous paleoecological, palaeoenvironmental and taphonomical
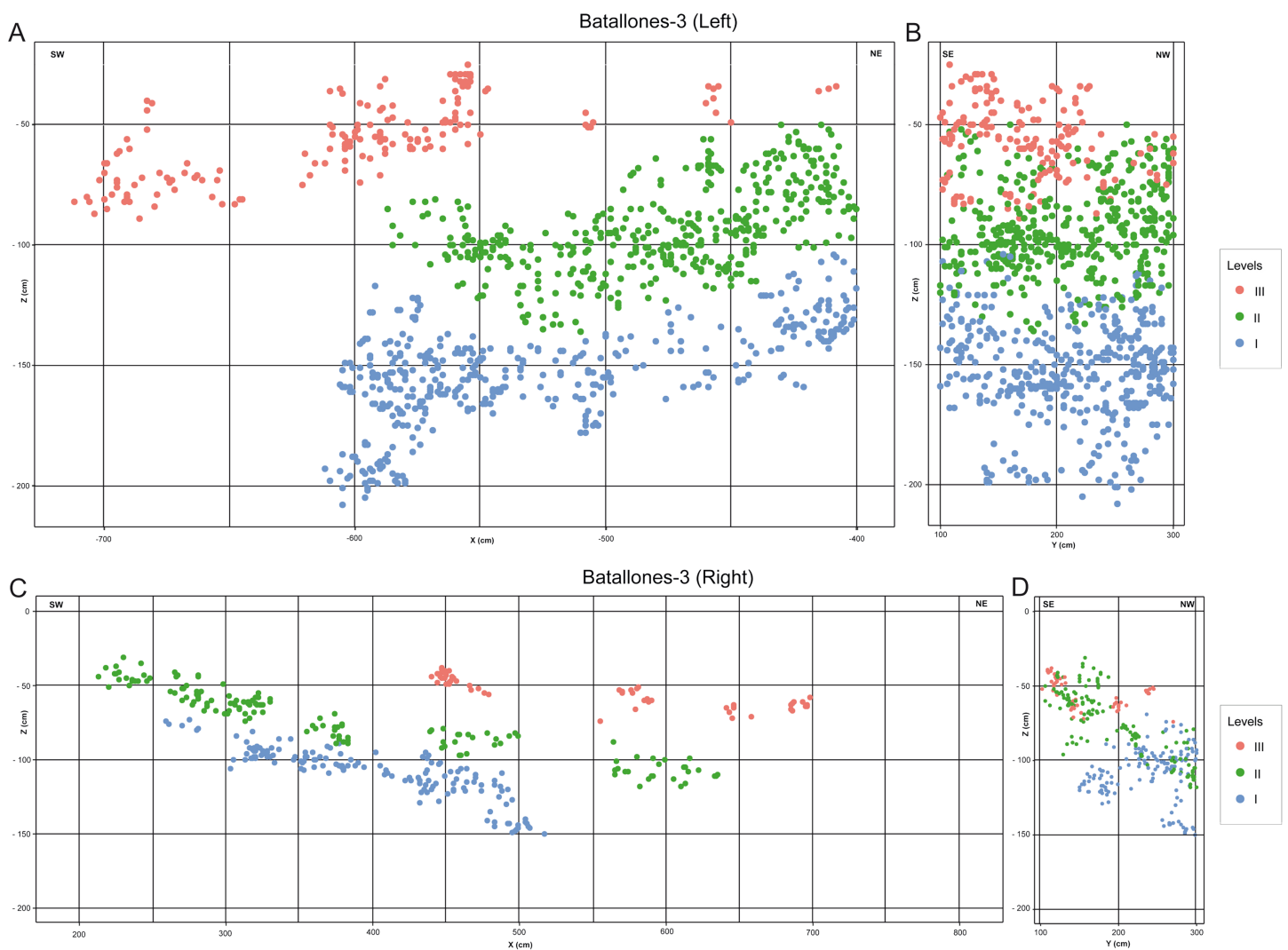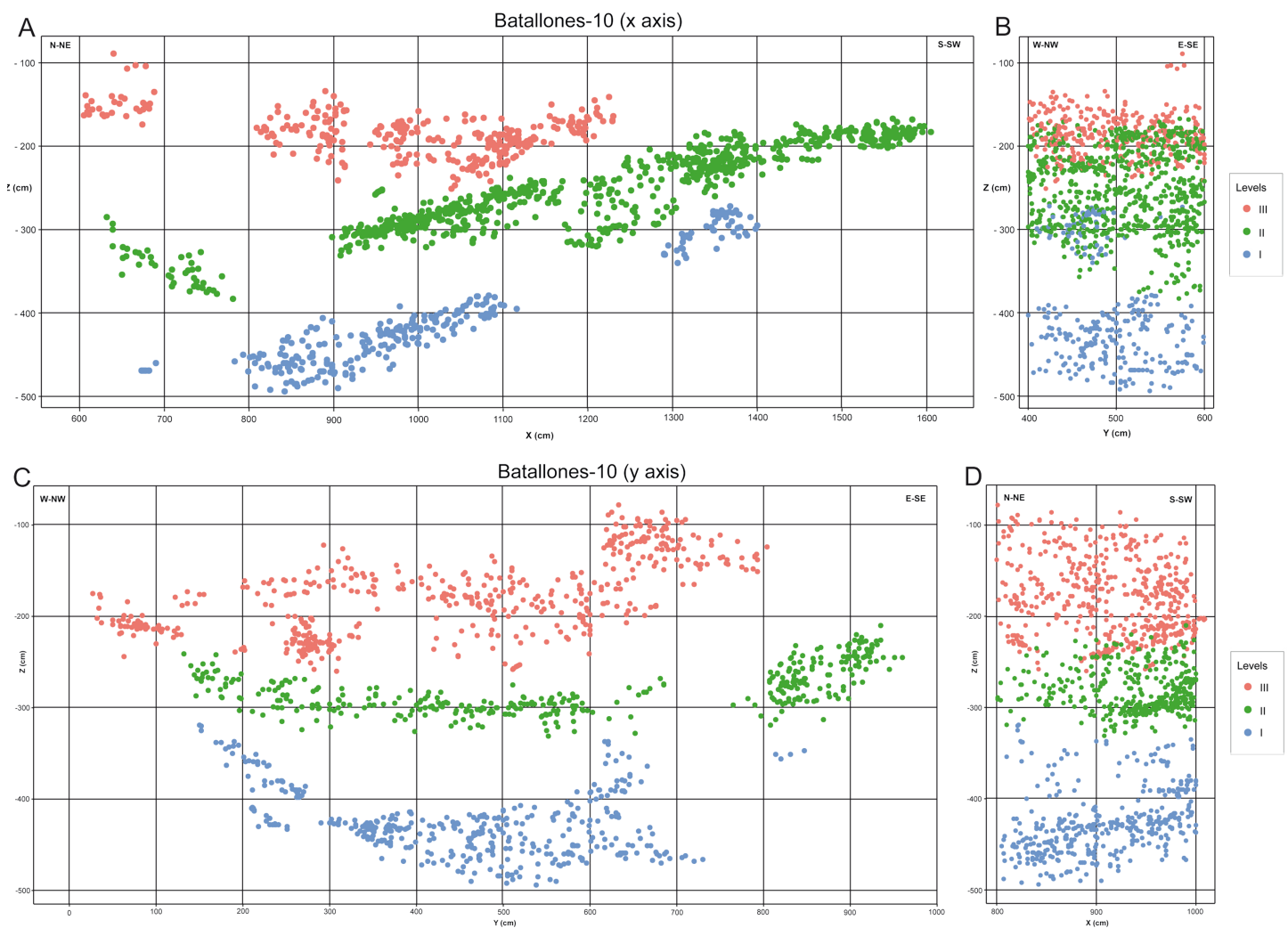
Figure 7 Fine-tuned fossiliferous level models. (A) Batallones-3 left slice. (B) Batallones-3 left slice associated 2-m-wide perpendicular slice. (C) Batallones-3 right slice. (D) Batallones-3 right slice associated 2-m-wide perpendicular slice. Full-size ⧉ DOI: 10.7717/peerj.8767/fig-7

analyses. As of this point, future studies at Batallones-3 and Batallones-10 should consider each of these individual discrete levels, instead of for the whole upper or lower part of the cavity.

As previously explained, Batallones-3 deposits and fossils are located in the lower part of the cavity and the fossil site is considered exhausted as the encasing rock was already found, for this reason, the three identified levels can be considered definitive. In the case of Batallones-10, the encasing rock has not yet been found and therefore fossil material recovered in future excavations may reveal additional discrete and separable fossiliferous layers. Thus, the increase in the number of levels at Batallones-10 cannot be discarded.

Future studies should analyze the embedded faunal assemblages separately for each of the described levels rather than the entire upper or lower part of the cavity, paying special attention to micromammal remains in order to establish chronological intra-site

Figure 8 Fine-tuned fossiliferous level models. (A) Batallones-10 *x* axis slice. (B) Batallones-10 *x* axis slice associated 2-m-wide perpendicular slice. (C) Batallones-10 *y* axis slice. (D) Batallones-10 *y* axis slice associated 2-m-wide perpendicular slice. Full-size ▣ DOI: 10.7717/peerj.8767/fig-8

relationships, such as those inter-site relationships described by *López-Antoñanzas et al. (2010)*. These detailed micromammal studies would greatly refine the geochronological framework of the Batallones Butte sites.

The overall workflow of this hybrid intelligence system has proven effective in detecting the presence of 3 levels in both Batallones-10 and Batallones-3. While in some slices the detection of these levels is clearer than in others, the use of two representative, thick and potent slices for each site helps empirically support their identification. Through the implementation of unsupervised algorithms for the initial detection of patterns, expert (geologist)-in-the-loop interactions for sense making and the final tuning of profiles using supervised algorithms, the overall system can be used to find areas where identification of fossiliferous levels is clearer. This can then be extrapolated for the supervised classification of the rest of the site.
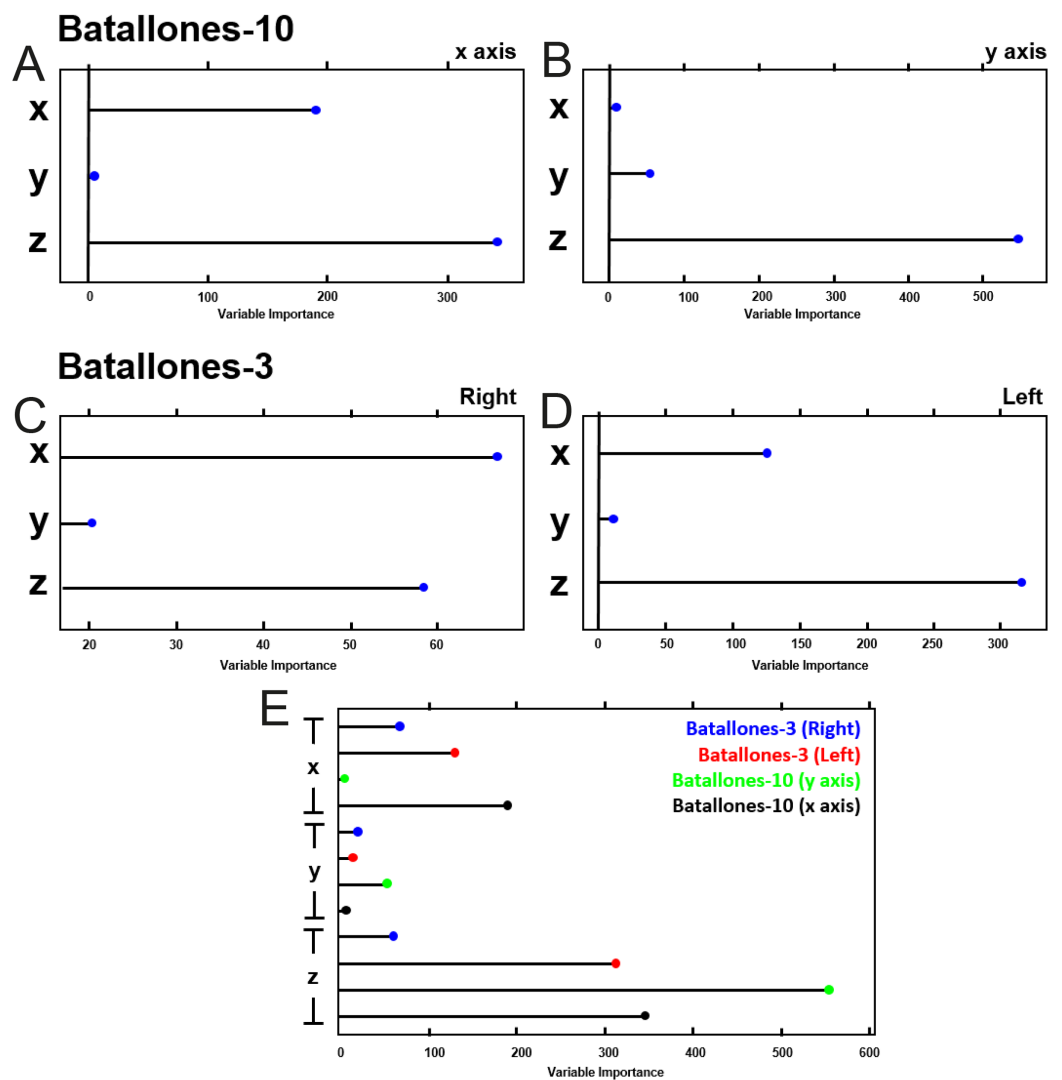
Figure 9 Variable importance plots for the RF model. Each panel represents the weight each variable has on the decision making capabilities of the RF model, indicating the importance in some cases of the use of 3D data for stratigraphic model definition. (A) Batallones-10 *x* slice. (B) Batallones-10 *y* slice. (C) Batallones-3 right slice. (D) Batallones-3 left slice. (E) Comparison of all slices.

Full-size ☑ DOI: 10.7717/peerj.8767/fig-9

Moreover, while some isolated points suspiciously appear to be classified into one layer and can be arguably considered out of place, future efforts should try to combine more data for the fine tuning of teaching and training processes. These could include taxonomic or taphonomic variables (*Brain, 1981*; *Domínguez-Rodrigo et al., 2018*), data regarding object density (*Kreutzer, 1992*; *Lam et al., 1998*, *2003*; *Lam, Chen & Pearson, 1999*), weight and size (*Bunn et al., 1980*; *Brain, 1981*; *Bunn, 1987*; *Bunn & Pickering, 2010*; *Domínguez-Rodrigo et al., 2018*), as well as orientation and plunge patterns (*Woodcock, 1977*; *Woodcock & Naylor, 1983*; *Lenoble & Bertran, 2004*; *Domínguez-Rodrigo & García-Pérez, 2013*).

While models and systems without humans-in-the-loop would be an optimal solution for the future (*Holzinger, 2016*), numerous geological factors that need extensive investigation for modeling are required before this can be achieved. Likewise, a number of underlying depositional and post-depositional processes may be generating confusion that machines are unable to understand or process for the time being. Such is the case for the confusion generated amongst algorithms caused by sedimentary onlaps, causing fossiliferous levels to lie closer to each other, at both Batallones-3 and Batallones-10. This may, however, be solved by more complex AIAs and the use of Deep MT systems, such as clustering AIAs using auto-encoders (*Xie, Girshick & Farhadi, 2016*; *Guo et al., 2017*; *Mrabah et al., 2019*; *Yang et al., 2019*), or those used for reinforcement learning tasks (*Lake et al., 2014*; *Mnih et al., 2015*; *Holzinger, 2016*; *Simard et al., 2017*). Efforts should therefore be made to investigate the effects of these numerous geological components on pattern detection algorithms.

CONCLUSIONS

This study presents a novel use of Artificially Intelligent Systems for the quantitative identification of discrete fossiliferous levels in paleontological and archaeological sites based solely on the study of the spatial three-dimensional distribution of fossil remains. These results have been able to reveal new discoveries in the Batallones Butte site inner structure, including three discrete levels at both Batallones-3 and Batallones-10.

The two lowermost levels of Batallones-3, Level I and Level II, dip outwards from the debris cone towards the outer limits of the cave, becoming progressively more horizontal, whereas the uppermost level, Level III is sub-horizontal. On the other hand, Batallones-10's lowermost Level I dips northward into the cavity and adapts laterally to the cave limits whereas Levels II and III also show northern dips but are more influenced by cave limits and dip towards the center of the cavity in the outermost areas.

Through these discoveries, AIAs have been proved to be a highly efficient and objective means of detecting spatial patterns in paleontological sites. Unsupervised and unsupervised techniques can greatly reduce (but not eliminate) the amount of subjectivity in discrete fossiliferous level identification. The possibilities provided by the combination of these algorithms and expert-in-the-loop systems are multiple, however it is important to point out that in order for these methods to be widely used, a need for further experimentation and investigation is essential. This should include an increase in the data used to feed these AIAs. For example, the incorporation of taphonomic features of the remains in the analyzed fossil may provide an even finer tuning of these models.

ACKNOWLEDGEMENTS

ADDITIONAL INFORMATION AND DECLARATIONS

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- David M. Martín-Perea conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Lloyd A. Courtenay conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- M. Soledad Domingo analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Jorge Morales analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:
Confusion matrices are in File S1. Raw coordinate data and the revised R script used are also available as Supplemental Files.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.8767#supplemental-information.

REFERENCES

Anemone RL, Emerson CW, Conroy GC. 2011. Finding fossils in new ways: an artificial neural network approach to predicting the location of productive fossil localities. *Evolutionary Anthropology: Issues, News, and Reviews* **20(5)**:169–180 DOI 10.1002/evan.20324.

Anemone RL, Conroy CG. 2018. *New geospatial approaches to the anthropological sciences.* Alburquerque: University of New Mexico Press/SAR Press.

Antón M, Morales J. 2000. Inferencias paleoecológicas de la asociación de carnívoros del yacimiento de Cerro Batallones, Paleontología y Etnografía. In: Morales J, ed. *Patrimonio Paleontológico de la Comunidad de Madrid.* Madrid: Arqueología, 190–201.

Arriaza MC, Domínguez-Rodrigo M. 2016. When felids and hominins ruled at olduvai gorge: a machine learning analysis of the skeletal profiles of the non-anthropogenic bed I sites. *Quaternary Science Reviews* **139**:43–52 DOI 10.1016/j.quascirev.2016.03.005.

Benhabiles H, Tabia H. 2016. Convolutional neural network for pottery retrieval. *Journal of Electronic Imaging* **26(1)**:011005 DOI 10.1117/1.JEI.26.1.011005.

Bergstra J, Bengio Y. 2012. Random search for hyper-parameter optimization. *Journal of Machine Learning Research* **13**:281–305.

Bewes J, Low A, Morphett A, Pate FD, Henneberg M. 2019. Artificial intelligence for sex determination of skeletal remains: application of a deep learning artificial neural network to human skulls. *Journal of Forensic and Legal Medicine* **62**:40–43 DOI 10.1016/j.jflm.2019.01.004.

Bishop C. 2006. *Pattern recognition and machine learning.* Singapore: Springer.

Block S, Saltré F, Rodríguez-Rey M, Fordham DA, Unkel I, Bradshaw CJA. 2016. Where to dig for fossils: combining climate-envelope, taphonomy and discovery models. *PLOS ONE* **11(3)**:e0151090 DOI 10.1371/journal.pone.0151090.

Breiman L, Friedman JH, Stone CJ, Olshen RA. 1984. *Classification and regression trees.* London: Chapman and Hall.

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

Brenning A. 2012. Spatial cross-validation and bootstrap for the assessment of prediction rules in remote sensing: The R package sperrorest. In: *2012 IEEE International Geoscience and Remote Sensing Symposium, Munich, Germany.* 5372–5375.

Brain CK. 1981. *The hunters or the hunted? An introduction to African cave taphonomy.* Chicago: The University of Chicago Press.

Bunn HT, Harris JWK, Isaac G, Kaufulu Z, Kroll E, Schick K, Toth N, Behrensmeyer A. 1980. FxJj50: an early pleistocene site in Northern Kenya. *World Archaeology* **12(2)**:109–136 DOI 10.1080/00438243.1980.9979787.

Bunn HT. 1987. Patterns of skeletal representation and hominid subsistence activities at Olduvai Gorge, Tanzania, and Koobi Fora, Kenya. *Journal of Human Evolution* **15(8)**:673–690 DOI 10.1016/S0047-2484(86)80004-5.

Bunn HT, Pickering TR. 2010. Methodological recommendations for ungulate mortality analyses in paleoanthropology. *Quaternary Research* **74(3)**:388–394 DOI 10.1016/j.yqres.2010.07.013.

Byeon W, Domínguez-Rodrigo M, Arampatzis G, Baquedano E, Yravedra J, Maté-González MA, Koumoutsakos P. 2019. Automated identification and deep classification of cut marks on bones and its palaeonthropological implications. *Journal of Computational Science* **32**:36–43 DOI 10.1016/j.jocs.2019.02.005.

Calvo JP, Pozo M, Silva PG, Morales J. 2013. Pattern of sedimentary infilling of fossil mammal traps formed in pseudokarst at Cerro de los Batallones, Madrid Basin, central Spain. *Sedimentology* **60(7)**:1681–1708 DOI 10.1111/sed.12048.

Canals A, Vallverdú J, Carbonell E. 2003. New archaeo-stratigraphic data for the TD6 level in relation to *Homo antecesor* (Lower Pleistocene) at the site of Atapuerca, north-central Spain. *Geoarchaeology* **18(5)**:481–504 DOI 10.1002/gea.10071.

Caspari G, Crespo P. 2019. Convolutional neural networks for archaeological site detection–finding "princely" tombs. *Journal of Archaeological Science* **110**:104998 DOI 10.1016/j.jas.2019.104998.

Conroy GC, Emerson CW, Anemone RL, Townsend KEB. 2012. Let your fingers do the walking: a simple spectral signature model for "remote" fossil prospecting. *Journal of Human Evolution* **63(1)**:79–84 DOI 10.1016/j.jhevol.2012.04.002.

Cortes C, Vapnik V. 1995. Support-vector networks. *Machine Learning* **20**:273–297.

Courtenay LA, Yravedra J, Huguet R, Aramendi J, Maté-González MA, González-Aguilera D, Arriaza MC. 2019. Combining machine learning algorithms and geometric morphometrics: a study of carnivore tooth marks. *Palaeogeography, Palaeoclimatology, Palaeoecology* **522**:28–39 DOI 10.1016/j.palaeo.2019.03.007.

Dellermann D, Calma A, Lipusch N, Weber T, Weigel S, Ebel P. 2019. The future of human-AI collaboration: a taxonomy of design knowledge for hybrid intelligence systems. In: *Hawaii International Conference on System Sciences, Maui, Hawaii, USA*. ScholarSpace/AIS Electronic Library (AISeL), 1–10.

Derech N, Tal A, Shimshoni H. 2018. Solving archaeological puzzles. *Available at http://arxiv.org/abs/1812.10553*.

Domingo MS, Alberdi MT, Azanza B. 2007. A new quantitative biochronological ordination for the upper neogene mammalian localities of Spain. *Palaeogeography, Palaeoclimatology, Palaeoecology* **255(3–4)**:361–376 DOI 10.1016/j.palaeo.2007.08.004.

Domingo MS, Domingo L, Sánchez IR, Alberdi MT, Azanza B, Morales J. 2011. New insights on the taphonomy of the exceptional mammalian fossil sites of Cerro de los Batallones (Late Miocene, Spain) based on rare earth element geochemistry. *Palaios* **26(1)**:55–65 DOI 10.2110/palo.2010.p10-047r.

Domingo MS, Alberdi MT, Azanza B, Morales J. 2012. Mortality patterns and skeletal physical condition of the carnivorans from the Miocene assemblage of Batallones-1 (Madrid Basin, Spain). *Neues Jahrbuch für Geologie und Paläontologie–Abhandlungen* **265(2)**:131–145 DOI 10.1127/0077-7749/2012/0251.

Domingo MS, Alberdi MT, Azanza B, Silva PG, Morales J. 2013a. Origin of an assemblage massively dominated by carnivorans from the Miocene of Spain. *PLOS ONE* **8(5)**:e63046 DOI 10.1371/journal.pone.0063046.

Domingo MS, Domingo L, Badgley C, Sanisidro O, Morales J. 2013b. Resource partitioning among top predators in a Miocene food web. *Proceedings of the Royal Society B: Biological Sciences* **280(1750)**:20122138 DOI 10.1098/rspb.2012.2138.

Domingo MS, Domingo L, Abella J, Valenciano A, Badgley C, Morales J. 2016. Feeding ecology and habitat preferences of top predators from two Miocene carnivore-rich assemblages. *Paleobiology* **42(3)**:489–507 DOI 10.1017/pab.2015.50.

Domínguez-Rodrigo M. 2019. Successful classification of experimental bone surface modifications (BSM) through machine learning: a solution to the controversial use of BSM in palaeoanthropology. *Archaeological and Anthropological Sciences* **11(6)**:2711–2725 DOI 10.1007/s12520-018-0684-9.

Domínguez-Rodrigo M, Cobo-Sánchez L, Yravedra J, Uribelarrea D, Arraiza C, Organista E, Baquedano E. 2018. Fluvial spatial taphonomy: a new method for the study of post-depositional

processes. *Archaeological and Anthropological Sciences* **10(7)**:1769–1789
DOI 10.1007/s12520-017-0497-2.

Domínguez-Rodrigo M, García-Pérez A. 2013. Testing the accuracy of different A-axis types or
measuring the orientation of bones in the archaeological and paleontological record. *PLOS ONE*
8(7):e68955 DOI 10.1371/journal.pone.0068955.

Eberth DA, Rogers RR, Fiorillo AR. 2007. A practical approach to the study of bonebeds.
In: Rogers RR, Eberth DA, Fiorillo AR, eds. *Bonebeds: Genesis, Analysis and Paleobiological
Significance*. Chicago: The University of Chicago Press, 265–331.

Egeland CP, Domínguez-Rodrigo M, Pickering TR, Menter CG, Heaton JL. 2018. Hominin
skeletal part abundances and claims of deliberate disposal of corpses in the middle pleistocene.
Proceedings of the National Academy of Sciences **115(18)**:4601–4606
DOI 10.1073/pnas.1718678115.

Emerson CW, Anemone RL. 2012. An artificial neural network-based approach to identifying
mammalian fossil localities in the Great Divide Basin, Wyoming. *Remote Sensing Letters*
3(5):453–460 DOI 10.1080/01431161.2011.621463.

Emerson CW, Bommersbach B, Nachman B, Anemone RL. 2015. An object-oriented approach
to extracting productive fossil localities from remotely sensed imagery. *Remote Sensing*
7(12):16555–16570 DOI 10.3390/rs71215848.

Ester M, Kriegel HP, Sander J, Xu X. 1996. A density-based algorithm for discovering clusters in
large spatial databases with noise. In: *Proceedings of the 2nd International Conference on
Knowledge Discovery and Data Mining, München, Germany*.

Gómez Cano AR, Hernández Fernández M, Álvarez-Sierra MA. 2011. Biogeographic
provincialism in rodent faunas from the Iberoccitanian Region (southwestern Europe) generates
severe diachrony within the mammalian neogene (MN) biochronologic scale during the late
Miocene. *Palaeogeography, Palaeoclimatology, Palaeoecology* **307(1–4)**:193–204
DOI 10.1016/j.palaeo.2011.05.014.

**Gravina B, Bachellerie F, Caux S, Discamps E, Faivre JP, Galland A, Michel A, Teyssandier N,
Bordes JG. 2018.** No reliable evidence for a Neanderthal-Châtelperronian association at La
Roche-à-Pierrot, Saint-Césaire. *Scientific Reports* **8(1)**:15134 DOI 10.1038/s41598-018-33084-9.

Guo X, Liu X, Zhu E, Yin J. 2017. Deep clustering with convolutional autoencoders. In: Liu D,
Xie S, Li Y, Zhao D, El-Alfy ES, eds. *Neural Information Processing, ICONIP 2017. Lecture Notes
in Computer Science*. Vol. 10635. Cham: Springer, 373–382.

Hengl T, Nussbaum M, Wright MN, Heuvelink GBM, Gräler B. 2018. Random forest as a
generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ*
6:e5518 DOI 10.7717/peerj.5518.

Hermoza R, Sipiarn I. 2018. 3D Reconstruction of incomplete archaeological objects using a
generative adversarial network. *Available at http://arxiv.org/abs/1711.06363v2*.

Holzinger A. 2016. Interactive machine learning for health informatics: when do we need the
human in the loop? *Brain Informatics* **3(2)**:119–131 DOI 10.1007/s40708-016-0042-6.

Kamar E. 2016. Directions in hybrid intelligence: complementing AI systems with human
intelligence. In: *IJCAI International Joint Conference on Artificial Intelligence, California, USA*.
4070–4073.

Kreutzer LA. 1992. Bison and deer bone mineral densities: comparisons and implications for the
interpretation of archaeological faunas. *Journal of Archaeological Science* **19(3)**:271–294
DOI 10.1016/0305-4403(92)90017-W.

Kuhn M. 2008. Building predictive models in R using the caret package. *Journal of Statistical
Software* **28(5)**:1–26 DOI 10.18637/jss.v028.i05.

Kuhn M, Johnson K. 2013. *Applied predictive modelling.* Heidelberg: Springer.

Lake BM, Ullman TD, Tenenbaum JB, Gershman SJ. 2014. Building machines that learn and think like people. *Behavioral and Brain Sciences* **40**:e253 DOI 10.1017/S0140525X16001837.

Lam YM, Chen X, Marean CW, Frey CJ. 1998. Bone density and long bone representation in archaeological faunas: comparing results from CT and photon densitometry. *Journal of Archaeological Science* **25(6)**:559–570 DOI 10.1006/jasc.1997.0256.

Lam YM, Chen X, Pearson OM. 1999. Intertaxonomic variability in patterns of bone density and the differential representation of bovid, cervid, and equid elements in the archaeological record. *American Antiquity* **64(2)**:343–362 DOI 10.2307/2694283.

Lam YM, Pearson OM, Marean CW, Chen X. 2003. Bone density studies in zooarchaeology. *Journal of Archaeological Science* **30(12)**:1701–1708 DOI 10.1016/S0305-4403(03)00065-7.

Lantz B. 2013. *Machine learning with R.* Birmingham: Packt Publishing.

Lenoble A, Bertran P. 2004. Fabric of palaeolithic levels: methods and implications for site formation processes. *Journal of Archaeological Science* **31(4)**:457–469 DOI 10.1016/j.jas.2003.09.013.

López-Antoñanzas R, Peláez-Campomanes P, Álvarez-Sierra MA, García-Paredes I. 2010. New species of *Hispanomys* (Rodentina, Cricetodontinae) from the upper Miocene of Batallones (Madrid, Spain). *Zoological Journal of the Linnean Society* **160(4)**:725–747 DOI 10.1111/j.1096-3642.2010.00618.x.

Louppe G, Wehenkel L, Sutera A, Geurts P. 2013. Understanding variable importance in forest of randomized trees. In: *Proceedings of the 26th International Conference on Neural Information Processing Systems.* Vol. 1. 431–439.

Lovelace R, Nowosad J, Muenchow J. 2019. *Geocomputation with R.* Boca Raton: CRC Press.

Martín Escorza C, Morales J. 2005. Estructura interna del yacimiento mioceno de Batallones (Madrid, España). *XVI Reunión Bienal de la Real Sociedad Española de Historia Natural* **6**:109–111.

Martín-Perea DM, Fesharaki O, Rey Samper JJ, Arroyo X, Uribelarrea D, Cobo-Sánchez L, Baquedano E, Mabulla A, Domínguez-Rodrigo M. 2019. Mineral assemblages and low energy sedimentary processes in the FLK-Zinj, DS, PTK and AMK complex palaeolandscape (Olduvai Gorge, Tanzania). *Quaternary International* **526**:15–25 DOI 10.1016/j.quaint.2019.09.018.

Miller H. 2004. Tobler's first law and spatial analysis. *Annals of the Association of American Geographers* **94(2)**:284–289.

Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, Graves A, Riedmiller M, Fidjeland AK, Ostrovski G, Petersen S, Beattie C, Sadik A, Antonoglou I, King H, Kumaran D, Wierstra D, Legg S, Hassabis D. 2015. Human-level control through deep reinforcement learning. *Nature Letters* **518(7540)**:529–533 DOI 10.1038/nature14236.

Moclán A, Domínguez-Rodrigo M, Yravedra J. 2019. Classifying bone breakage patterns: an experimental analysis of fracture planes to discern between Hominin and carnivore activity using machine learning (ML) algorithms. *Anthropological and Archaeological Sciences* **11(9)**:4663–4680 DOI 10.1007/s12520-019-00815-6.

Mondal M, Bertranpetit J, Lao O. 2019. Approximate Bayesian computation with deep learning supports a third archaic introgression in Asia and Oceania. *Nature Communications* **10(1)**:652 DOI 10.1038/s41467-018-08089-7.

Morales J, Capitán J, Calvo JP, Sesé C. 1992. Nuevo yacimiento de vertebrados del Mioceno Superior al Sur de Madrid (Cerro de los Batallones, Torrejón de Velasco). *Geogaceta* **12**:77–80.

Morales J, Alcalá L, Álvarez-Sierra MA, Antón M, Azanza B, Calvo JP, Carrasco P, Fraile S, García-Paredes I, Gómez E, Hernández Fernández M, Merino L, Van der Meulen A, Martín-Escorza C, Montoya P, Nieto M, Peigné S, Pérez B, Peláez-Campomanes P, Pozo M, Quiralte V, Salesa MJ, Sánchez IM, Sánchez-Marco A, Silva PG, Soria MD, Turner A. 2004. Paleontología del Sistema de yacimientos de mamíferos miocenos del cerros de los Batallones cuenca de Madrid. *Geogaceta* **35**:139–142.

Morales J, Pozo M, Silva PG, Domingo MS, López-Antoñanzas R, Álvarez-Sierra MA, Antón M, Martín-Escorza C, Quiralte V, Salesa MJ, Sánchez IM, Azanza B, Calvo JP, Carrasco P, García-Paredes I, Knoll F, Merino L, Van der MA, Montoya P, Peigné S, Peláez-Campomanes P, Sánchez-Marco A, Turner A, Abella J, Alcalde GM, Andrés M, Demiguel D, Cantalapiedra JL, Fraile S, García YB, Gómez CAR, López GP, Oliver PA, Siliceo G. 2008. El sistema de yacimientos de mamíferos miocenos del Cerro de los Batallones, Cuenca de Madrid: estado actual y perspectivas. *Publicaciones del Seminario de Paleontología de Zaragoza* **8**:41–117.

Mrabah N, Khan NM, Ksantini R, Lachiri Z. 2019. Deep clustering with a dynamic autoencoder. *Available at http://arxiv.org/abs/1901.07752v4*.

Patel AA. 2019. *Hands-on unsupervised learning with python: how to build applied machine learning solutions from unlabeled data*. Tokyo: O'Reilly.

Pohjankukka J, Pahikkala T, Nevalainen P, Heikkonen J. 2017. Estimating the prediction performance of spatial models via spatial k-fold cross validation. *International Journal of Geographical Information Science* **31(10)**:2001–2019.

Pozo M, Calvo JP, Silva PG, Morales J, Peláez-Campomanes P, Nieto M. 2004. Geología del sistema de yacimientos de mamíferos miocenos del Cerro de los Batallones. *Cuenca de Madrid. Geogaceta* **35**:143–146.

Püschel TA, Marcé-Nogué J, Gladman JT, Bobe R, Sellers WI. 2018. Inferring locomotor behaviours in Miocene New World monkeys using finite element analysis, geometric morphometrics and machine-learning classification techniques applied to talar morphology. *Journal of the Royal Society Interface* **15(146)**:20180520 DOI 10.1098/rsif.2018.0520.

Sander J, Ester M, Kriegel HP, Xu X. 1998. Density-based clustering in spatial databases: the algorithm GDBSCAN and its applications. *Data Mining and Knowledge Discovery* **2(2)**:169–194 DOI 10.1023/A:1009745219419.

R Development Core Team. 2018. *A language and environment for statistical computing*. Vienna: The R Foundation for Statistical Computing. *Available at http://www.R-project.org/*.

Sañudo P, Blasco R, Fernández Peris J. 2016. Site formation dynamics and human occupations at Bolomor Cave (Valencia, Spain): an archaeolstratigraphic analysis of levels I to XII (100–200 ka). *Quaternary International* **417**:94–104 DOI 10.1016/j.quaint.2015.09.044.

Satopa V, Albrecht J, Irwin D, Raghavan B. 2017. Finding a "Kneedle in a Haystack: detecting knee points in system behavior, Python PYPI. *Available at https://github.com/arkevi/kneed.git*.

Schubert E, Sander J, Ester M, Kriegel HP, Xu X. 2017. DBSCAN revisited: why and how you should (still) use DBSCAN. *ACM Transactions on Database Systems* **42(3)**:1–21 DOI 10.1145/3068335.

Simard PY, Saleema A, Chickering DM, Pelton AE, Ghorashi S, Meek C, Ramos G, Suh J, Verwey J, Wang M, Wernsing J. 2017. Machine teaching: a new paradigm for building machine learning systems. *Available at http://arxiv.org/abs/1707.06742v3*.

Thorndike RL. 1953. Who belongs in the family? *Psychometrika* **48(4)**:1267–1276.

Uribelarrea D, Domínguez-Rodrigo M, Pérez-González A, Vegas Salamanca J, Baquedano E, Mabulla A, Musiba C, Barboni D, Cobo-Sánchez L. 2014. Geo-archaeological and

geometrically corrected reconstruction of the 1.84 Ma FLK Zinj paleolandscape at Olduvai Gorge. *Tanzania Quaternary International* **322**:7–31.

Verschoof-van der Vaart WB, Lambers K. 2019. Learning to Look at LiDAR: the use of R-CNN in the automated detection of archaeological objects in LiDAR data from the Netherlands. *Journal of Computer Applications in Archaeology* **2(1)**:31–40 DOI 10.5334/jcaa.32.

Wills S, Choiniere JN, Barrett PM. 2018. Predictive modelling of fossil-bearing locality distributions in the Elliot formation (Upper Triassic–Lower Jurassic), South Africa, using a combined multivariate and spatial statistical analysis of present-day environmental data. *Palaeogeography, Palaeoclimatology, Palaeoecology* **489**:186–197 DOI 10.1016/j.palaeo.2017.10.009.

Woodcock NH. 1977. Specification of fabric shapes using the eigenvalue method. *Geological Society of America Bulletin* **88**:1231–1236 DOI 10.1130/0016-7606(1977)88<1231:SOFSUA>2.0.CO;2.

Woodcock NH, Naylor MA. 1983. Randomness testing in three dimensional orientation data. *Journal of Structural Geology* **5(5)**:539–548 DOI 10.1016/0191-8141(83)90058-5.

Xie J, Girshick R, Farhadi A. 2016. Unsupervised deep embedding for clustering analysis. *Available at* http://arxiv.org/abs/1511.06335.

Yang X, Deng C, Zheng F, Yan J, Liu W. 2019. Deep spectral clustering using dual autoencoder network. *Available at* http://arxiv.org/abs/1904.13113.