A Granularity-Based Intelligent Tutoring System for Zooarchaeology

Laia Subirats [1,2], **Leopoldo Pérez** [3,4], **Cristo Hernández** [5], **Santiago Fort** [1] and **Gomez-Monivas Sacha** [6,*]

1   Eurecat, Centre Tecnològic de Catalunya, C/Bilbao, 72, 08005 Barcelona, Spain; laia.subirats@eurecat.org or lsubirats@uoc.edu (L.S.); santi.fort@eurecat.org (S.F.)
2   Faculty of Computer Science, Multimedia and Telecommunications, Universitat Oberta de Catalunya, Rambla del Poblenou, 156, 08018 Barcelona, Spain
3   IPHES, Institut Català de Paleoecologia Humana i Evolució Social, Zona Educacional 4, Campus Sescelades URV (Edifici W3), 43007 Tarragona, Spain; ljperez@iphes.cat
4   Àrea de Prehistoria, Universitat Rovira i Virgili (URV), Avinguda de Catalunya, 35, 43002 Tarragona, Spain
5   Departamento de Geografía e Historia, Área de Prehistoria (Facultad de Humanidades), Universidad de La Laguna, Campus de Guajara, La Laguna, 38071 Tenerife, Spain; chergomw@gmail.com
6   Universidad Autónoma de Madrid, Ciudad Universitaria de Cantoblanco, 28049 Madrid, Spain
*   Correspondence: sacha.gomez@uam.es; Tel.: +34-914-97-32-07

Featured Application: **This study creates an intelligent tutoring system in archaeology for helping students in specialized tasks that require analysis of huge amounts of data. The method proposed here implies the application of classification algorithms that must be trained with a complete data set in order to give accurate results. We have tested our method by developing an intelligent tutoring system in the field of zooarchaeology.**

Abstract: This paper presents a tutoring system which uses three different granularities for helping students to classify animals from bone fragments in zooarchaeology. The 3406 bone remains, which have 64 attributes, were obtained from the excavation of the Middle Palaeolithic site of El Salt (Alicante, Spain). The coarse granularity performs a five-class prediction, the medium a twelve-class prediction, and the fine a fifteen-class prediction. In the coarse granularity, the results show that the first 10 most relevant attributes for classification are width, bone, thickness, length, bone fragment, anatomical group, long bone circumference, X, Y, and Z. Based on those results, a user-friendly interface of the tutor has been built in order to train archaeology students to classify new remains using the coarse granularity. A pilot has been performed in the 2019 excavation season in Abric del Pastor (Alicante, Spain), where the automatic tutoring system was used by students to classify 51 new remains. The pilot experience demonstrated the usefulness of the tutoring system both for students when facing their first classification activities and also for seniors since the tutoring system gives them valuable clues for helping in difficult classification problems.

Keywords: supervised learning; zooarchaeology; intelligent tutoring system

1. Introduction

The use of technology opens new frontiers in learning and improves data mining from different sources in order to improve students' learning processes [1]. One of the biggest challenges of including communication technologies in learning is the way in which interaction between teachers and students is simulated by automatic methods [2,3]. In this context, tutoring and the possibility of turning an automatic system into an effective instrument for counseling students are stirring a great amount of

interest in many areas [4,5]. The interest in this concept is demonstrated by the many initiatives related to intelligent tutoring systems (ITS). For example, Graesser et al. [6] incorporated natural language processing (NLP) to the tutor in order to enhance the student's learning process. In other studies, authors used gamification techniques to enhance the student's learning process [7,8] or integrate affect sensors [9]. Other more recent ITS guided students to choose their learning path according to their profile [10].

The use of ITS in archaeology is a very interesting topic because of the great amount of data that is available for analysis. In this discipline, data analysis has been performed by many statistical methods [11], including artificial intelligence (AI) in both supervised and unsupervised forms [12–14]. In [15], authors focused on computer vision and gave a visual analysis in archaeology. Random forest [16] has also been used as a supervised learning method to classify archaeological remains. There have also been initiatives [17] which use artificial neural networks in the archaeological domain claiming that they are suitable for the sparse data commonly found in this discipline. Augmented reality and serious games have also been used as tools for improving learning in related subjects such as history or archaeology [18,19]. Virtual reconstructions of ancient cities have also been developed in many different formats [20,21].

Unfortunately, there is a lack of references of ITS in more technical tasks that could help students in more specialized techniques, such as zooarchaeology, which is the topic included in this article. In archaeological sites, there are many different specialized tasks that must be performed by well-trained students [22]. In the case of zooarchaeology, the identification and characterization of animals or species from bone fragments is a crucial task that requires the analysis of many parameters obtained from the characteristics of the remains [16]. First, students require good training in the taxonomical and anatomical characteristics. Later, they must identify the animals and all the information related with their use by humans (trapping, butchering process, discard, technological use, etc.). This data analysis is a great effort and an ITS would be of great help. Moreover, implementing and defining an effective protocol for implementing an intelligent tutor in bone fragment characterization will be also useful in other similar disciplines that also need the analysis of huge amounts of data in archaeology such as the identification of constructive styles [23] or reconstruction of ceramics from small fragments [24].

In this article, we show the steps, for creating an ITS, that have been developed from the application of AI classification algorithms to a data set that includes bone fragments correctly identified in the archaeological site of El Salt. A comparison between different granularities and algorithms is also shown. By choosing the most effective combination, we have developed a user-friendly interface of the ITS with the objective of testing it in the 2019 season in the archaeological site. Finally, we show the results and conclusions of the pilot experience in the archaeological site during August 2019.

This article is structured as follows: Section 2 explains the materials and methods used in the study, Section 3 presents the results, Section 4 discusses the results, and finally Section 5 details the conclusions and future work.

2. Materials and Methods

This section is divided into materials, where the dataset of the study is explained, and methods, where the followed methodology to build the tutoring system and to evaluate it is explained.

2.1. Materials

The used dataset had 3406 instances of archaeological remains and 64 attributes (the last one is the predicted class) extracted from faunal assemblage of Stratigraphic Unit Xa of El Salt (Figures 1 and 2) [25–27]. The excavation of "El Salt" can be seen in Figure 1 and examples of the archaeological remains of "El Salt" can be seen in Figures 2 and 3.

The Salt is one of the fundamental sites of the Middle Palaeolithic in the western Mediterranean region due to the entity of its archaeological sequence, for its extraordinary state of conservation, which

affects even organic matter and for the integrated and multidisciplinary nature of the research process which takes place in it, with the confluence of different Hispanic, European, and North American Universities and Research Centers. This enclave has been located at the head of the Serpis River, in the locality of Alcoy (Alicante, Spain). It is presented as an outdoor site of about 300 m^2 located at the foot of a large wall that rises up to 38 m tall. The space inhabited at the foot of the wall was protected by a large roof, which at times of maximum development came to serve as protection for almost the entire surface. Highlights are the strategic condition of its location, in the middle of various biotopes of plain, mountains, river valley, and lake-palm environment, etc. and immersed in a territory of the mountains of Alicante, very rich in diversified resources.

Figure 1. Geographical location of "El Salt", site overview and current excavation surface.

Research at the El Salt site attempts to deepen the knowledge of its paleolithic record, from an integrative and multidisciplinary perspective. This record does not stop growing by virtue of the application of increasingly sophisticated excavation techniques and high resolution analytical procedures. To the traditional material record of this type of deposits, consisting of lithic remains, fauna, anthracological material, etc. Currently, a microscopic and even molecular registry is being added that is decisively contributing to enrich the information and improve its quality. These facts give us the opportunity of obtaining a complete and accurate data-set in zooarchaeology, among others, which is a key step for training classification algorithms as the ones needed for developing the ITS presented here.

Figure 2. Taphonomic damage and bone remains recovered examples: manganese (**A**), concreteness (**B**), root-marks (**C**), bone flake (**D**), burned bone (**E**), erosion (**F**), bones, teeth and deer antler (**G**), bones and wild goat teeth (**H**), horse tooth (**I**), phalanges and lynx maxillary (**J**), and rabbit bones (**K**).
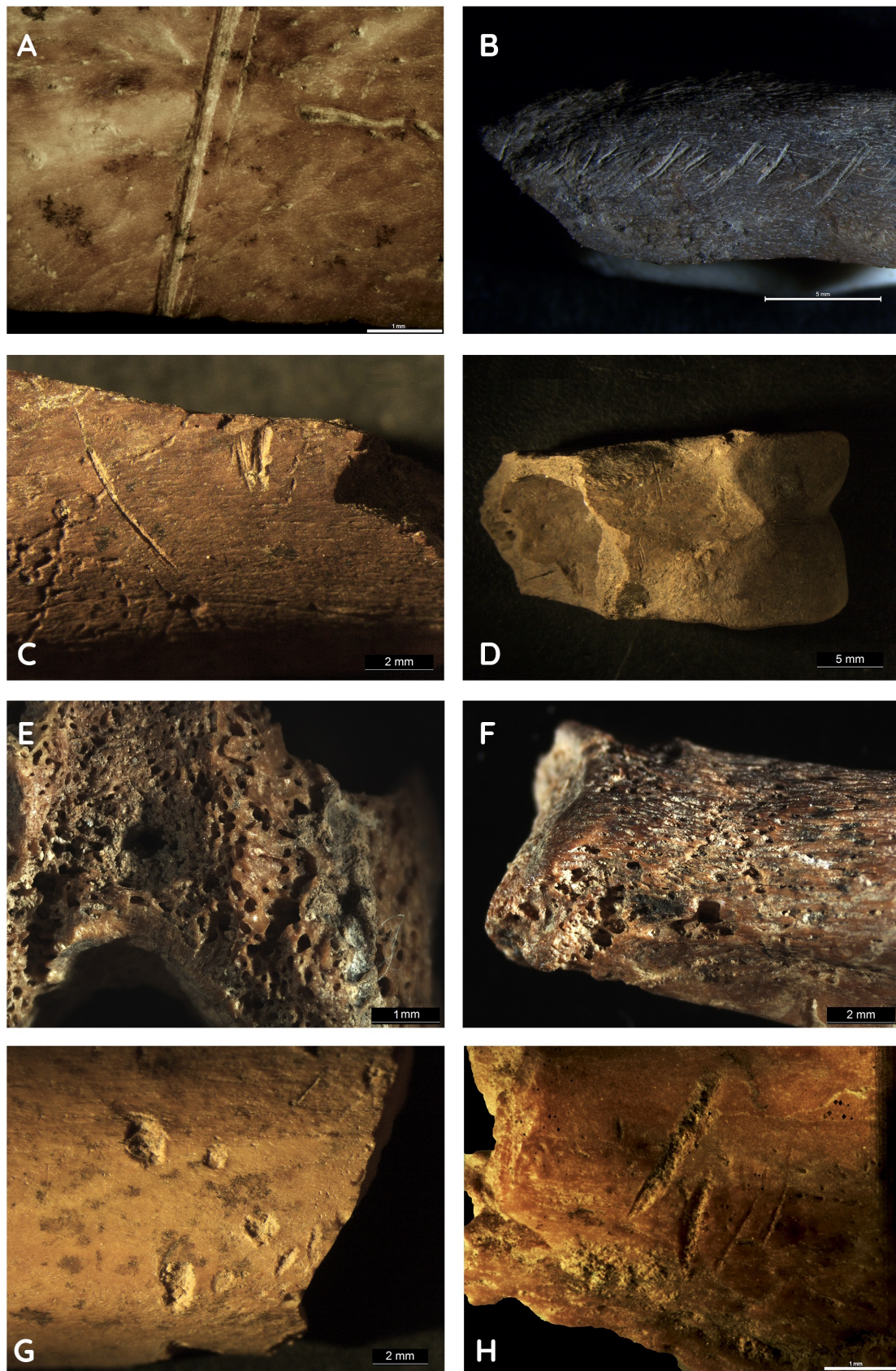
Figure 3. Bioestratinomic damage examples. There are anthropogenic marks: slicing and scraping marks (**A**–**D**), Predator non-anthropogenic marks: digestion (**E**,**F**), punctures (**G**), and scores (**H**).

In Figure 2, we can observe taphonomic damage and bone remains recovered examples. In Figure 3, we can observe bioestratinomic damage examples.

In Table 1, a descriptive analysis of the sample is shown. There are two columns in the table. In the first one, we have listed the most relevant parameters (according to the classification algorithm with best performance in the coarse, medium, and fine granularities, which will be explained in the following section) and the predicted class 'Family' (taxonomic rank between Order and Genus) both in coarse, medium, and fine granularities. Following the name of each parameter, we have included a description of the possible values that it can take. For instance, "Bone" can take the listed 93 unique values, while the anatomical group can take only 11 values. It is worth noting that, in the case of the prediction classes, the coarse granularity can take five values, the medium granularity 12, and the fine granularity 15. The second column of the table describes briefly the attribute. If the attribute is numeric, count (number of not null values), mean, standard deviation (std), minimum (min), first quartile (Q1), second quartile (Q2), third quartile (Q3), and maximum (max) are given. If the attribute is categorical, count (number of not null values), unique (number of categories), top (top category), and frequency (freq, which is the frequency of the top category) are given. In the categorical granularity families, the number of each category is given because it is relevant for the predicting classifier that will be explained in Section 2.2.

Table 1. Exploratory data analysis of the dataset.

Attribute	Description
Width	Count: 2036, Mean: 14.822829, Std: 8.973862, Min: 0.99, Q1: 8.457500, Q2: 13.43, Q3: 19.32, Max: 71.01
Bone: Ca, T, Cx, F, H, U, Fa1, Mtp, Mt2, Mc4, Mt4, Cr, Vl, Vcd, Mc2, Ta, Hem, I/1, Fa, Mt3, Vc, I1/, Mx, As, Pa, R, Es, Fa2, Fa3, Mt5, nan, Mc5, Ct, Vs, Vt, Mc3, V, Das, Dai, Da, Mt, M3/, M1/, Hy, Mc, P2/, P4/-M1/, M/1, Lt2, Art, Cc, PT, M2/, In, M/P, M/1-M/2, P4/, M3, M1/-M2/, P/2, I, M/X, Pl, P/M, R/, Co, L, Lt1, Lt3, P/4, P/3-P/4, Se, dp2/, P3/, M/x, P4/-P3/, M/3, PTPer8, PTPle, PTPL, P/3, M/, M/2, M, Fa1-V, Fa2-II, Fa2-IV, I/C, Fi, Asta, I/3, d/3, Mc2-3, I/2	Count: 3403, Unique: 93, Top: T, Freq: 374
Thickness	Count: 714, Mean: 4.887451, Std: 2.873843, Min: 0.460000, Q1: 3.072500, Q2: 4.595000, Q3: 6.4, Max: 16.52
Length	Count: 2492, Mean: 36.25859, Std: 28.28524, Min: 2.42000, Q1: 15.94500, Q2: 27.20000, Q3: 47.80000, Max: 284.96000
Bone fragment	Count: 3242, Mean: 189.635102, Std: 204.909877, Min: 1, Q1: 50, Q2: 111, Q3: 500, Max: 555
Anatomical group: Mp, Ma, E, C, A, nan, In, ES, Cr, PT, PTPL, PTPl	Count: 3399, Unique: 11, Top: Mp, Freq: 1275
Long bone circumference	Count: 1378, Mean: 2.526851, Std: 1.107021, Min: 1, Q1: 2, Q2: 2, Q3: 4, Max: 5
X	Count: 1.187000×10^3, Mean: 1.864528×10^3, Std: 6.292228×10^4, Min: 0, Q1: 8.519905, Q2: 2.900000×10^1, Q3: 6.775000×10^1, Max: 2.167891×10^6
Y	Count: 1186, Mean: 1272.859123, Std: 30223.002072, Min: 0, Q1: 11.809257, Q2: 24.500000, Q3: 63, Max: 803982

Table 1. *Cont.*

Attribute	Description
Z	Count: 1232, Mean: 403.047662, Std: 9767.169421, Min: -143.8, Q1: -0.156130, Q2: 166.2, Q3: 192.6, Max: 342938
Manganeso	Count: 1416, Mean: 1.592514, Std: 0.835553, Min: 1, Q1: 1, Q2: 1, Q3: 2, Max: 5
Concretion	Count: 1177, Mean: 1.806287, Std: 1.033386, Min: 1, Q1: 1, Q2: 1, Q3: 2, Max: 5
Coarse granularity family: Bovidae, Cervidae, Equidae, Leporidae, unknown	Bovidae: 420 Cervidae: 516, Equidae: 240, Leporidae: 2164, Unknown: 66, Total: 3406
Medium granularity family: Bovidae, Canidae, Cervidae, Corvidae, Equidae, Felidae, Leporidae, Phasianidae, Rhinocerotidae, Suidae, Testudinidae, unknown	Bovidae: 420 Canidae: 8, Cervidae: 516, Corvidae: 5, Equidae: 240, Felidae: 24, Leporidae: 2164, Phasianidae: 7, Rhinocerotidae: 2, Suidae: 7, Testudinidae: 9, Unknown: 4, Total: 3406
Fine granularity family: Anatidae, Bovidae, Bufonidae, Canidae, Cervidae, Corvidae, Equidae, Erinaceidae, Felidae, Leporidae, Phasianidae, Rhinocerotidae, Suidae, Testudinidae, Ursidae	Anatidae: 1 Bovidae: 420, Bufonidae: 1, Canidae: 8, Cervidae: 516, Corvidae: 5, Equidae: 240, Erinaceidae: 1, Felidae: 24, Leporidae: 2164, Phasianidae: 7, Rhinocerotidae: 2, Suidae: 7, Testudinidae: 9, Ursidae: 1, Total: 3406

2.2. Methods

We used 33% of the remains for testing and 67% for training. A pipeline with a standard scaler with a SMOTE (synthetic minority over-sampling technique) or ADASYN (adaptive synthetic) method was used, together with 10-fold cross-validation. The parameter that we want to maximize is f1-score_macro. We have chosen it because F1-score is the harmonic mean of precision and recall, and the macro option calculates metrics for each label, and finds their unweighted mean, and this does not take label imbalance into account. The used algorithms with their parameters have been:

- Random forest with parameters:

    - classifier__n_estimators (number of trees in the forest): [100, 500, 700]
    - classifier__max_features (number of features to consider when looking for the best split): ['auto', 'sqrt', 'log2']
    - classifier__class_weight (weights associated with classes; if not given, all classes are supposed to have weight one; the balanced mode uses the values to automatically adjust weights inversely proportional to class frequencies in the input data): ['balanced',None]

- Support vector machine (SVM) with parameters:

    - classifier__kernel (specifies the kernel type to be used in the algorithm): ['linear','rbf']
    - classifier__gamma (kernel coefficient for 'rbf', 'poly', and 'sigmoid'): [1×10^{-3},1×10^{-4}]
    - classifier__C (penalty parameter C of the error term): [1, 10, 100]

- Naive bayes.
- Neural networks with parameters:

    - classifier__solver (the solver for weight optimization. 'lbfgs' is an optimizer in the family of quasi-Newton methods. 'sgd' refers to stochastic gradient descent. 'adam' refers to a stochastic gradient-based optimizer): ['lbfgs', 'sgd', 'adam']
    - classifier__alpha (L2 penalty (regularization term) parameter): [1×10^{-4}, 1×10^{-5}]

- k-nearest neighbors (KNN) with parameters:

– classifier__n_neighbors (number of neighbors to use): [3, 5, 7, 9]

More information about the algorithms can be found in the literature [28]. These algorithms have been implemented using the Python library scikit-learn and the ITS has been implemented using Flask.

Regarding the criteria used to distinguish the different taphonomic alterations, zooarchaeological and taphonomic analysis are performed using already established standard methods [29,30]. Faunal remains are taxonomically and anatomically identified whenever possible, while non-identified specimens with insufficient information were classified into three bone categories (long, flat, or articular) and associated with a weight-size category based on bone density, circumference, and thickness of the cortical surface: large-sized >300 kg, medium-sized 100–300 kg, small-sized 5–100 kg, and very small-sized <5 kg [31–33]. On the other hand, all remains were analysed taphonomically using macroscopic and microscopic techniques to identify biostratinomic and diagenetic modifications. In the fracture analysis, we classified all of the fragments by fracture type (recent, old fresh, old dry, or indeterminate) following the criteria established by [34] and the morphotypes created by Real [35,36]. All dimensions were measured for each bone fragment. Also, bone surface modifications were observed and quantified to identify different damage caused by anthropogenic activity (percussion, butchering marks, or thermal alteration) or predator action (tooth marks, or digestion), as well as the diverse diagenetic processes that produce alterations on bone surfaces as erosion, sediment concreteness, roots marks, weathering, pigmentation, and trampling [37–44]. Damage caused after the prey's death and previously their deposit and sedimentation, the biostratinomic process, was classified and analysed individually in a specific database according to the type of damage, origin, agent, location, morphology, distribution, direction, intensity, quantity, and dimensions. Diagenetic processes were recorded based on their presence and the degree of alteration. In the specific case of butchering marks, different activities were established based on the ethnoarchaeological literature [37,45–47].

The questionnaire used to test the experience of the tutor in the pilot of El Salt 2019 season followed partially the methodology explained in [48]. The following areas were tested:

1. Constructive/active learning: The tutor stimulated us to understand underlying mechanisms/theories.
2. Self-directed learning: The tutor stimulated us to search for various resources by ourselves.
3. Contextual learning: The tutor stimulated us to apply knowledge to the discussed problem.
4. Global score: Overall performance of the tutor.
5. Open answer: Give some tips for improvement.

A rating between 1 (strongly disagree) and 5 (strongly agree) was given to questions 1–4.

3. Results

Results are divided into the experiments for the prediction of animals in the coarse, medium and fine granularities, and the ITS development and use. The performance of the different methods in the coarse granularity can be seen in Table 2, and the confusion matrices of the coarse, medium, and fine granularities using the random forest method can be seen in Figures 4–6. In the fine granularity (Figure 6), classes with less than three instances do not appear because a stratified split between training and test has been applied.

In the three granularities, the 10 more relevant attributes for the classification have been (by this order):

- Coarse granularity: width, bone, thickness, length, bone fragment, anatomical group, long bone circumference, X, Y, and Z.
- Medium granularity: bone, width, length, anatomic group, bone fragment, Y, X, Z, thickness and manganese.
- Fine granularity: width, bone, length, bone fragment, anatomic group, Z, Y, X, thickness and concretion.

The best configuration parameters for the random forest were:

- Coarse granularity: 'classifier__max_features': 'auto', 'classifier__class_weight': 'balanced', 'classifier__n_estimators': 500.
- Medium granularity: 'classifier__max_features': 'sqrt', 'classifier__class_weight': 'balanced', 'classifier__n_estimators': 700.
- Fine granularity: 'classifier__max_features': 'auto', 'classifier__class_weight': 'balanced', 'classifier__n_estimators': 700

Table 2. Performance of the different methods in the coarse granularity. Synthetic minority over-sampling technique (SMOTE), support vector machine (SVM), k-nearest neighbors (KNN), adaptive synthetic (ADASYN).

Method (Parameters)	Accuracy	Precision (Weighted)	Recall (Weighted)	F1-Score (Weighted)
Random forest, SMOTE (classifier class weight: balanced, classifier max features: auto, classifier n estimators: 500)	0.86	0.86	0.86	0.86
SVM, SMOTE (classifier C: 100, classifier gamma: 0.001, classifier kernel: rbf)	0.74	0.81	0.74	0.77
Naive Bayes, SMOTE	0.68	0.76	0.68	0.66
Neural Networks, SMOTE (classifier solver: lbfgs, classifier alpha: 1×10^{-5})	0.67	0.75	0.67	0.71
KNN, SMOTE (classifier n neighbors: 3)	0.75	0.79	0.75	0.77
Random Forest, ADASYN (classifier class weight: balanced, classifier max features: auto, classifier n estimators: 100)	0.86	0.85	0.86	0.86
SVM, ADASYN (classifier C: 100, classifier gamma: 0.001, classifier kernel: rbf)	0.72	0.80	0.72	0.75
Naive Bayes, ADASYN	0.68	0.78	0.68	0.65
Neural Networks, ADASYN (classifier solver: adam, classifier alpha: 1×10^{-5})	0.67	0.79	0.67	0.72
KNN, ADASYN (classifier n neighbors: 3)	0.74	0.79	0.74	0.76

In Figure 7, we show the interface of the ITS implemented for being used in the season. As we can see in the interface, there is a first region where the user must be identified. Right below this first section, the user should start filling the different characteristics and parameters that have been obtained from the bone fragment. The ITS also gives information about the different options of a parameter when applicable. This is very important since the ITS could be used by archaeologists from different sites, and we must be sure that they agree in the format of the data.

In Figure 8, we show the output screen of the ITS right after a user has finished the characterization of a fragment and decides to ask for a prediction. In this case, the prediction is shown at the beginning of the interface. As soon as the prediction is done, the ITS is ready for another prediction. For this reason, the fields below the prediction do not have any value. It is worth noting that all the information filled by the user has been stored in the ITS server and is available for users upon request. Figure 8 is an example of a table used by students in the archaeological site to compare their own predictions and the ITS answers. We have censored information about the fragment identification and the identity of the students.
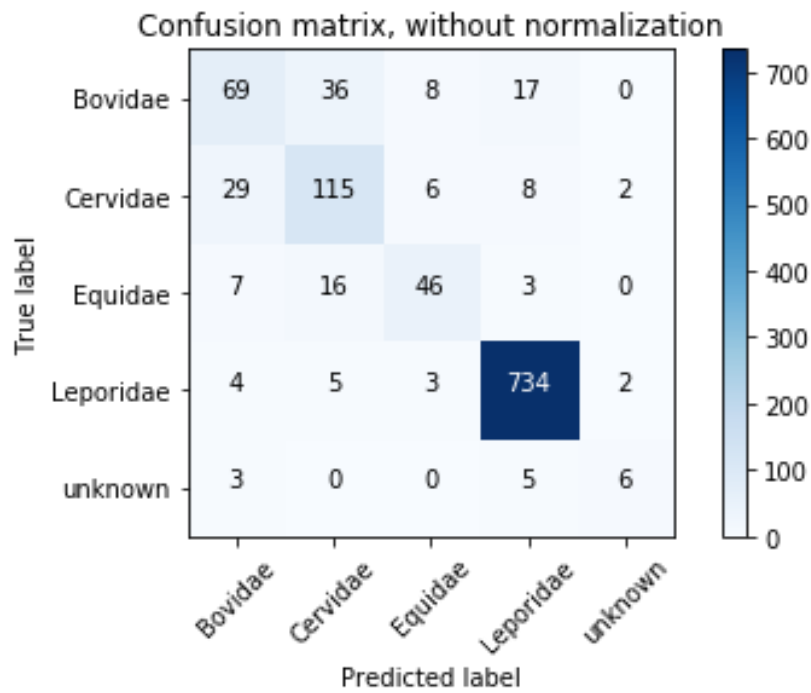
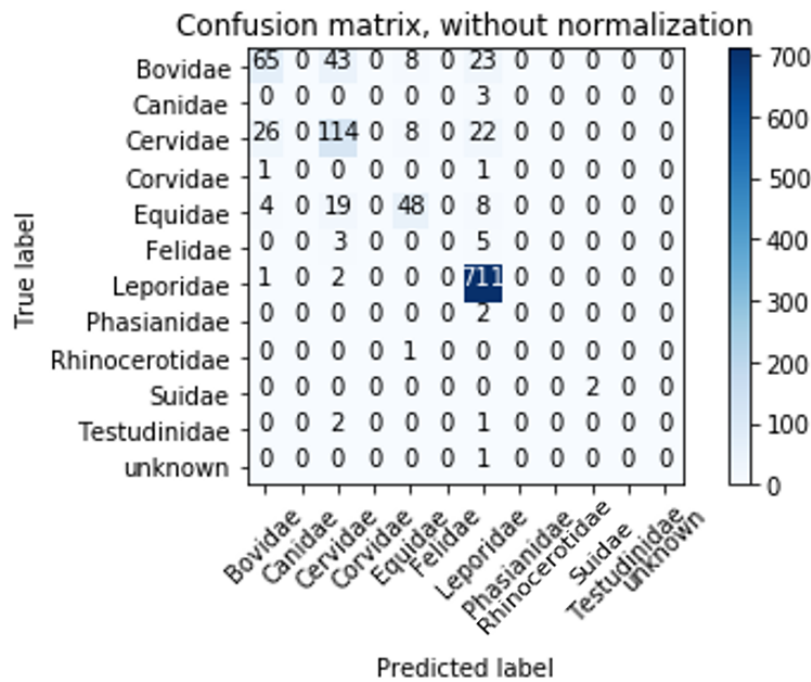**Figure 4.** Confusion matrix of the coarse granularity.

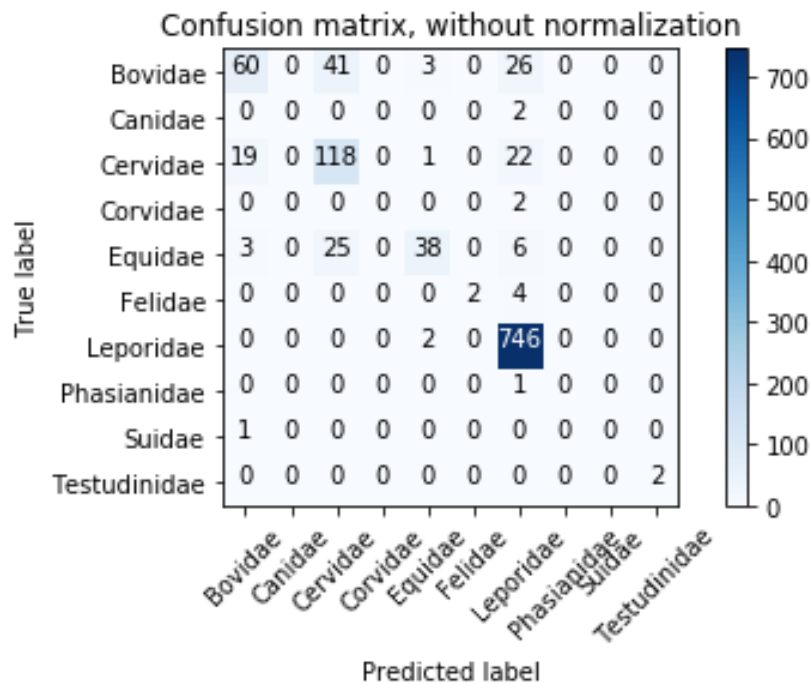**Figure 5.** Confusion matrix of the medium granularity.

Figure 6. Confusion matrix of the fine granularity.

This first version of the ITS included the coarse granularity and was used in August 2019 by 3 students to classify 51 remains that were characterized following the format of the same database but belonged to a different archaeological site called "Abric del Pastor". Abric del Pastor belongs also to Middle Palaeolithic and is near "El Salt". Two of the three students also answered the questionnaire. The results of the tutor, compared to the predictions of the students frequently agreed even when the remains used for training corresponded to a different site ("El Salt"). Students, due to their limited knowledge and the difficulties of analyzing very fragmented bone remains, usually described the remains based only on the size, which usually agree with the families predicted by the tutor. For example, a lot of bone remains predicted as "small size" are described by the tutor as Bovidae, which is a family included in the category of small animals because most bovids belong to the wild goat (*Capra pyrenaica*). This information can help students to give more accurate answers or verify their initial guess. On the other side, sometimes the tutor fails and not all the sizes and families are correctly classified. For example, some bones that are correctly described as "small size" by students are characterized as Leporidae by the tutor, which is a mistake. Other mistakes are related to the wrong characterization of birds as Leporidae.

Figure 7. Tutor interface to introduce the bone fragments.

Analyzing the results of the questionnaire, students appreciated the use of the tutor as a complementary tool in their activities. In the subjective open answer, they highlighted two concepts: (1) The need of improving the interface and (2) the need of increasing the accuracy of the predictions.
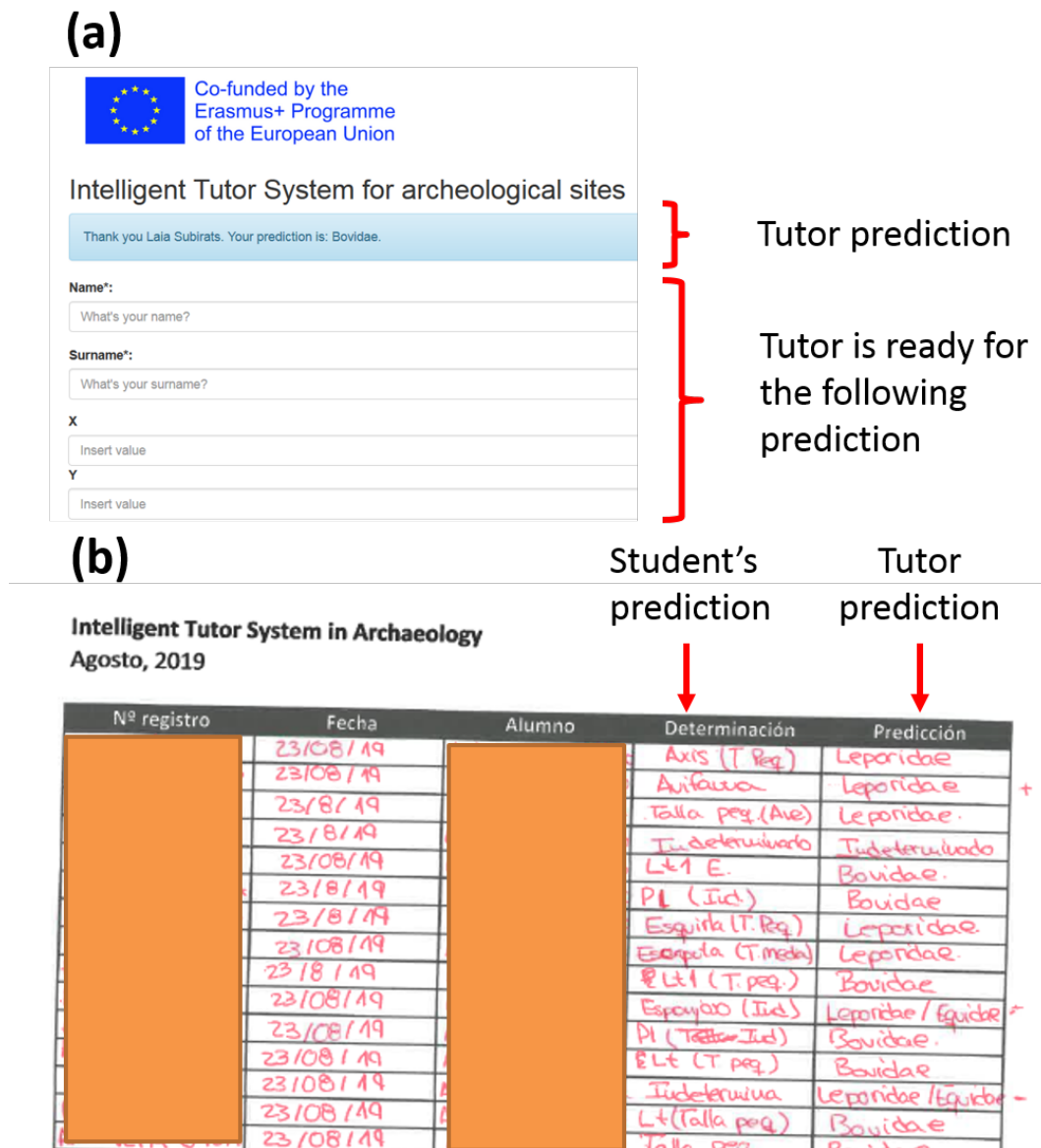
(a)

Co-funded by the Erasmus+ Programme of the European Union

Intelligent Tutor System for archeological sites

Thank you Laia Subirats. Your prediction is: Bovidae.

Name*:
What's your name?

Surname*:
What's your surname?

X
Insert value

Y
Insert value

Tutor prediction

Tutor is ready for the following prediction

(b)

Student's prediction   Tutor prediction

Figure 8. Tutor prediction (**a**) and student's/tutor prediction table (**b**).

4. Discussion

Random forests using SMOTE is the best machine learning method for this domain. In Table 3 the F1-score of the different animals is displayed. We can see that coarse and fine granularities have the best F1-scores (0.86 and 0.85 respectively). However, the fine granularity is more suitable because it has more granularity and it classifies better small animals.

Let us now analyze the feedback obtained from the archaeological students and teachers. From the pedagogical point of view, and analyzing their answers, we can see that the ITS moderately stimulates underlying mechanisms but, on the other hand, it strongly stimulates the search for resources by the users. There are a few different opinions about ITS's ability to stimulate the application of knowledge related to the discussed problem, but the overall performance of the ITS is quite high. Finally, the final tips for improvement were to improve the interface and to have more data in order to have a more accurate prediction.

Although predictions given by the ITS are usually accurate and useful, the presence of some mistakes must be analyzed. There are two ways of improving it. First, the excessive number of characterization as Leporidae can be related to the granularity used in this version. An increment

in the number of classes would increase the accuracy of the prediction by increasing the amount of information in the system. The second way of improving the accuracy of the tutor is taking into account the spatial distribution of the fragments in the sites. Since the X, Y, and Z variables are important in the characterization, using information from different archaeological sites in the training would help in the correct characterization of the tutor by avoiding an excessive importance of the bone distribution in a specific site.

Table 3. F1-score of the random forest in the different granularities.

Animal	Coarse	Medium	Fine
Bovidae	0.57	0.59	0.56
Cervidae	0.69	0.69	0.69
Equidae	0.68	0.63	0.66
Leporidae	0.97	0.94	0.96
Unknown	0.50	0	-
Canidae	-	0	0
Corvidae	-	0	0
Felidae	-	0.22	0.5
Phasianidae	-	0	0
Rhinocerotidae	-	0	0
Suidae	-	1	0
Testudinidae	-	1	1
Anatidae	-	-	0
Bufonidae	-	-	0
Erinaceidae	-	-	0
Ursidae	-	-	0
Total weighted	0.86	0.83	0.85

5. Conclusions

In this article we have presented a method to create intelligent tutoring systems in archaeology to help students in specialized tasks that require analysis of huge amounts of data. The method proposed here implies the application of classification algorithms that must be trained with a complete data set in order to give accurate results. We have tested our method by developing an intelligent tutoring system in the field of zooarchaeology. We have found that random forest is the method with the best performance in this domain, and the fine granularity approach is the more useful approach for archaeologists because the main doubts appear with small animals, which are not common. In the coarse granularity, random forest was the method which had better results between the tested ones, with an accuracy and f1-score of 0.86 for both. As questionnaires to classify the remains are quite long (63 attributes), the first 10 more relevant attributes for the classification have been shown, which are width, bone, thickness, length, bone fragment, anatomical group, long bone circumference, X, Y, and Z. In addition, a classification using a medium and fine granularity has also been performed obtaining an accuracy of 0.84 and 0.86 and a f1-score of 0.83 and 0.85 respectively.

Regarding the pilot with the tutor, it correctly classified the families in most of the cases. However, some mistakes were also found. Mistakes can be partially corrected by changing the granularity since some of the mistakes are related to the reduced amount of classes in the granularity implemented in the pilot experience. The satisfaction questionnaire showed that the tutor was helpful. Students suggested an improvement in the interface and the accuracy of the predictions.

In future work, more data from different archaeological sites can be included in the training phase in order to increase accuracy and avoid effects related to specific characteristics of a single excavation. Also, natural language processing (NLP) and gamification techniques could be added to the tutor in order to enhance the student's learning process.

Author Contributions: Conceptualization, L.S., L.P., C.H., S.F. and G.-M.S.; Data curation, L.P. and C.H.; Funding acquisition, S.F. and G.-M.S.; Investigation, L.S., L.P., C.H., S.F. and G.-M.S.; Methodology, L.S., L.P., C.H., S.F. and G.-M.S.; Project administration, S.F. and G.-M.S.; Software, L.S.; Supervision, S.F. and G.-M.S.; Writing—original draft, L.S., L.P., C.H., S.F. and G.-M.S.; Writing—review & editing, L.S., L.P., C.H., S.F. and G.-M.S.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ADASYN	Adaptive Synthetic
AI	Artificial intelligence
ITS	Intelligent Tutoring System
KNN	k-nearest neighbors
NLP	Natural language processing
SMOTE	Synthetic Minority Over-sampling Technique
SVM	Support vector machine

References

1. Maldonado-Mahauad, J.; Perez-Sanagustin, M.; Kizilcec, R.; Morales, N.; Munoz-Gama, J. Mining theory-based patterns from Big data: Identifying self regulated learning strategies in Massive Open Online Courses. *Comput. Hum. Behav.* **2018**, *80*, 179–196. [CrossRef]

2. González-Sacristán, C.; Molins-Ruano, P.; Díez, F.; Rodríguez, P.; Sacha, G.M. Computer-assisted Assessment with Item Classification for Programming Skills. In Proceedings of the First International Conference on Technological Ecosystem for Enhancing Multiculturality (TEEM '13), Salamanca, Spain, 14–15 November 2013; ACM: New York, NY, USA, 2013; pp. 111–117. [CrossRef]

3. Molins-Ruano, P.; González-Sacristán, C.; Díez, F.; Rodriguez, P.; Sacha, G.M. An Adaptive Model for Computer-Assisted Assessment in Programming Skills. *Int. J. Eng. Educ.* **2015**, *31*, 764.

4. Bravo, C.; van Joolingen, W.; de Jong, T. Using co-lab to build system dynamics models: Students' actions and on-line tutorial advice. *Comput. Educ.* **2009**, *53*, 243–251. [CrossRef]

5. Sabitha, A.; Mehrotra, D.; Bansal, A. An ensemble approach in converging contents of LMS and KMS. *Educ. Inf. Technol.* **2017**, *22*, 1673–1694. [CrossRef]

6. Graesser, A.C.; Chipman, P.; Haynes, B.C.; Olney, A. AutoTutor: An intelligent tutoring system with mixed-initiative dialogue. *IEEE Trans. Educ.* **2005**, *48*, 612–618. [CrossRef]

7. Baker, R.S.J.; Corbett, A.; Koedinger, K.; Evenson, S.; Roll, I.; Wagner, A.Z.; Raspat, M.N.; Baker, D.J.; Beck, J.E. Adapting to When Students Game an Intelligent Tutoring System. In *Intelligent Tutoring Systems*; Ikeda, M., Ashley, K.D., Chan, T.W., Eds.; ITS 2006; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2006; Volume 4053.

8. Long, Y.; Aleven, V. Educational Game and Intelligent Tutoring System: A Classroom Study and Comparative Design Analysis. *ACM Trans. Comput.-Hum. Interact.* **2017**, *24*, 20. [CrossRef]

9.   D'Mello, S.; Craig, S.; Gholson, B.; Franklin, S.; Picard, R.; Graesser, A. Integrating Affect Sensors in an Intelligent Tutoring System. In Proceedings of the Computer in the Affective Loop Workshop at 2005 International Conference Intelligent User Interfaces, San Diego, CA, USA, 10–13 January 2005; pp. 7–13.

10. Verdú, E.; Regueras, L.; Gal, E. Integration of an intelligent tutoring system in a course of computer network design. *Educ. Technol. Res. Dev.* **2017**, *65*, 653–677. [CrossRef]

11. Baxter, M. *Statistics in Archaeology*; Arnold Publ.: London, UK, 2003.

12. Baxter, M. A review of supervised and unsupervised pattern recognition in archaeometry. *Archaeometry* **2006**, *48*, 671–694. [CrossRef]

13. Barcelo, J.A. *Computational Intelligence in Archaeology*; Information Science Reference: Hershey, PA, USA, 2009.

14. Barcelo, J.A.; Bogdanovic, I. *Mathematics in Archaeology*; CRC Press: Boca Raton, FL, USA, 2015.

15. Barceló, J. Visual Analysis in Archaeology: An Artificial Intelligence Approach. In *Morphometrics for Nonmorphometricians*; Elewa, A.M.T., Ed.; Springer: Berlin, Germany, 2010; pp. 93–156.

16. Subirats, L.; Fort, S.; Hernández, C.; Pérez, L.; Vesisenaho, M.; Nousiainen, T.; Peltonen, M.; Miakush, I.; Sacha, G.M. Intelligent Tutoring System in Archaeology. In Proceedings of the EdMedia + Innovate Learning, Amsterdam, The Netherlands, 24 June 2019; pp. 745–750.

17. Bell, S.; Croson, C. Artificial neural networks as a tool for archaeological data analysis. *Archeometry* **1998**, *40*, 139–151. [CrossRef]

18. Ardito, C.; Buono, P.; Costabile, M.; Lanzilotti, R.; Pederson, T. Re-experiencing History in Archaeological Parks by Playing a Mobile Augmented Reality Game. In *On the Move to Meaningful Internet Systems 2007: OTM 2007 Workshops. OTM 2007. Lecture Notes in Computer Science*; Meersman, R., Tari, Z., Herrero, P., Eds.; Springer: Berlin/Heidelberg, Germany, 2007; Volume 4805.

19. Vidal-Matutano, P.; Pérez-Jordà, G.; Hernández, C.M.; Galván, B. Macrobotanical evidence (wood charcoal and seeds) from the Middle Palaeolithic site of El Salt, Eastern Iberia: Palaeoenvironmental data and plant resources catchment areas. *J. Archaeol. Sci. Rep.* **2018**, *19*, 454–464. [CrossRef]

20. Atienza-Fuente, J. The monumental thermal complex of the Roman city of Valeria: Proposal for virtual reconstruction of its cold rooms based on the archaeological findings. *Virtual Archaeol. Rev.* **2019**, *10*, 116–131. [CrossRef]

21. Ferrari, I.; Quarta, A. The Roman pier of San Cataldo: from archaeological data to 3D reconstruction. *Virtual Archaeol. Rev.* **2019**, *10*, 28–39. [CrossRef]

22. Molinero-Polo, M.A.; Pérez-Ruiz, S.; Acebo, A.; Sacha, G.M. Análisis del potencial motivador y docente del uso de las nuevas tecnologías en el aula. Interacción entre el aumento de la motivación y el aprendizaje; Artículo invitado en el monográfico "Cambios metodológicos: Nuevas formas en la enseñanza" de la revista OGE (Organización y Gestión Educativa) del fórum europeo de administradores de la educación. 2016. Volume 3, p. 26. Available online: https://dialnet.unirioja.es/servlet/articulo?codigo=5502970 (accessed on 29 September 2019).

23. Blanquez-Perez, J.; Roldan-Gomez, L. The Urban Reform of the Colonia Libertinorum Carteia during the Augustan Era. The Basilica Building. *Gerion-Rev. Hist. Antigua* **2017**, *35*, 443–468. [CrossRef]

24. Garcia-Vegas, G.; Castelo-Ruano, R.; Lopez-Perez, A. Virtual Reconstructions of Archaeological Heritage. The Convivial Space of the Roman Villa of El Saucedo, Talavera la Nueva, Toledo. *Arqueologia Iberoamericana* **2017**, *35*, 3–9.

25. Machado, J.; Pérez, L. Temporal frameworks to approach human behavior concealed in Middle Palaeolithic palimpsests: A high-resolution example from El Salt Stratigraphic Unit X (Alicante, Spain). *Quat. Int.* **2016**, 66–81. [CrossRef]

26. Pérez, L.J.; Sanchis, A.; Hernández, C.M.; Galván, B. Paleoecología de macromamíferos aplicada a los conjuntos zooarqueológicos de El Salt y el Abric del Pastor (Alcoy, Alicante). In *Interaccions entre felins i humans. III Jornades d'arqueozoologia*; Museu de Prehistòria de València: València, Spain, 2017; pp. 327–354.

27. Pérez, L. Estrategias de subsistencia y dinámicas de asentamiento en los Valles de Alcoy durante el Paleolítico medio. Análisis zooarqueológico, tafonómico y paleoecológico de la secuencia arqueológica de El Salt (Alcoy, Alicante). Ph.D. Thesis, University Rovira i Virgili, Tarragona, Spain, 2019.

28. Cady, F. *The Data Science Handbook*; Wiley: Hoboken, NJ, USA, 2017.

29. Lyman, R. *Vertebrate Taphonomy, Cambridge Manuals in Archaeology*; Cambridge University Press: Camdridge, UK, 1994.

30. Reitz, E.J.; Wing, E.S. *Zooarchaeology*; Cambridge University Press: Cambridge, UK, 2008.

31. Uerpmann, H.P. Animal bone finds and economic archaeology: A critical study of Osteoarchaeological method. *World Archaeol.* **1973**, *4*, 307–322. [CrossRef]

32. Bunn, H.T. Patterns of skeletal representation and hominid subsistence activities at Olduvai gorge, Tanzania, and Koobi fora, Kenya. *J. Hum. Evol.* **1986**, *15*, 673–690. [CrossRef]

33. Palomo, L.J.; Gisbert, J.; Blanco, J.C. *Atlas y Libro Rojo de los Mamíferos Terrestres de España*; Dirección General para la Biodiversidad-SECEM-SECEMU: Madrid, Spain, 2007.

34. Villa, P.; Mahieu, E. Breakage patterns of human long bones. *J. Hum. Evol.* **1991**, *21*, 27–48. [CrossRef]

35. Real, C. Aproximación metodológica y nuevos datos sobre los conjuntos arqueozoológicos del Magdaleniense superior de la Cova de les Cendres. *Archivo De Prehistoria Levantina* **2012**, *29*, 99–120.

36. Real, C. Estudio arqueozoológico y tafonómico del Magdaleniense de la Cova de les Cendres (Teulada-Moraira, Alicante). Doctoral Dissertation, Universitat de València, Valencia, Spain, 2016.

37. Binford, L.R. *Bones: Ancient Men and Modern Myths, Studies in Archaeology*; Academic Press: New York, NY, USA, 1981.

38. Shipman, P. *Life History of a Fossil. An Introduction to Taphonomy and Paleoecology*; Harvard University Press: Boston, MA, USA, 1981.

39. Shipman, P.; Rose, J. Early hominid hunting, butchering, and carcass-processing behaviors: Approaches to the fossil record. *J. Anthropol. Archaeol.* **1983**, *2*, 57–98. [CrossRef]

40. Blasco Sancho, M.F. *Tafonomía y Prehistoria. Métodos y procedimientos de investigación*; Departamento de Ciencias de la Antigüedad—Universidad de Zaragoza; Departamento de Cultura y Educación-Gobierno de Aragón: Zaragoza, Spain, 1992.

41. Yravedra, J. *Tafonomía aplicada a Zooarqueología*; UNED: Madrid, Spain, 2006.

42. Domínguez Rodrigo, M.; Yravedra, J. Why are cut mark frequencies in archaeofaunal assemblages so variable? A multivariate analysis. *J. Archaeol. Sci.* **2009**, *36*, 884–894. [CrossRef]

43. Denys, C.; Patou-Mathis, M. *Manuel De Taphonomie*; Archéologiques; Errance: Arles, France, 2014.

44. Fernandez-Jalvo, Y.; Andrews, P. *Atlas of Taphonomic Identifications. 1001+ Images of Fossil and Recent Mammal Bone Modification*; Springer: Chame, The Netherlands, 2016. [CrossRef]

45. Binford, L.R. *Nunamiut Ethnoarchaeology*; Studies in Archaeology Series; Academic Press: New York, NY, USA, 1978.

46. Shipman, P.; Rose, J.T. Cutmarks mimics on modern and fossil bovid bones. *Curr. Anthropol.* **1984**, *25*, 116–117. [CrossRef]

47. Capaldo, S.D.; Blumenschine, R.J. A quantitative diagnosis of notches made by hammerstone percussion and carnivore gnawing on bovid long bones. *Am. Antiquity* **1994**, *59*, 724–748. [CrossRef]

48. Dolmans, D.H.J.M.; Ginns, P. A short questionnaire to evaluate the effectiveness of tutors in PBL: Validity and reliability. *Med. Teach.* **2005**, *27*, 534–538. [CrossRef]