Deep learning *versus* geometric morphometrics for archaeobotanical domestication study and subspecific identification

Vincent Bonhomme[1,2,@], Laurent Bouby[1], Julien Claude[1], Camille Dham[1], Muriel Gros-Balthazard[4], Sarah Ivorra[1], Angèle Jeanty[1], Clémence Pagnoux[5], Thierry Pastor[1], Jean-Frédéric Terral[1] and Allowen Evin[1]

[1] ISEM, Univ Montpellier, CNRS, EPHE, IRD, Montpellier, France.
[2] Athéna, Lacamp, Roquedur, France.
[3] Department of Biology, Faculty of Science, University of Chulalongkorn, Bangkok, Thailand.
[4] DIADE, Univ Montpellier, IRD, CIRAD, Montpellier, France
[5] AASPE, CNRS, MNHN, Paris, France.

@ Corresponding author: bonhomme.vincent@gmail.com

Abstract

Taxonomical identification of archaeological fruit and seed is of prime importance for any archaeobotanical studies. We compared the relative performance of deep learning and geometric morphometrics at identifying pairs of plant taxa. We used their seeds and fruit stones that are the most abundant recovered organs in archaeobotanical assemblages, and whose morphological identification, chiefly between wild and domesticated types, allow to document their domestication and biogeographical history. We used existing modern datasets of four plant taxa (date palm, barley, olive and grapevine) corresponding to photographs of two orthogonal views of their seeds that were analysed separately to offer a larger spectrum of shape diversity. On these eight datasets, we compared the performance of a deep learning approach, here convolutional neural networks (CNN), to that of a geometric morphometric approach, here outline analyses using elliptical Fourier transforms (EFT). Sample sizes were at minimum eight hundred seeds in each class, which is quite small when training deep learning models but of typical magnitude for archaeobotanical studies. Our objectives were twofold: i) to test whether deep learning can beat geometric morphometrics in taxonomic identification and if so, ii) to test which minimal sample size is required. We ran simulations on the full datasets and also on subsets, starting from 50 images in each binary class. For CNN networks, we deliberately used a candid approach relying on pre-parameterised VGG16 network. For EFT, we used a state-of-the art morphometrical pipeline. The main difference rests in the data used by each model: CNN used bare photographs where EFT used (x, y) outline coordinates. This "pre-distilled" geometrical description of seed outlines is often the most time-consuming part of morphometric studies. Results show that CNN beats EFT in most cases, even for very small datasets. We finally discuss the potential of CNN for archaeobotany, why outline analyses and morphometrics have not yet said their last word by providing quantitative descriptions, and how bioarchaeological studies could embrace both approaches, used in a complementary way, to better assess and understand the past history of species.

Keywords: convolutional neural networks, elliptical Fourier transforms, archaeobotany, domestication studies.

Introduction

From Aristotle to Darwin, the form of organisms has long inspired our understanding of the living world. In some disciplines such as archaeobotany, the shape of plant remains is, most often, the only available datum. Both qualitative and quantitative morphological criteria first allowed to identify plant remains, particularly seeds and fruit stones, often at the species level (Jacomet, 2008; Zohary *et al.*, 2012). Then, purely quantitative tools, and chiefly geometric morphometrics, allowed for finer-grained, statistically assessed identifications, to further explore the morphological size and shape variation.

Geometric modern morphometrics (further abbreviated GMM), is the statistical description of shape and its covariation (Kendall, 1989). It uses generic mathematical transformations to convert shape into quantitative variables. Most GMM studies either uses configuration of landmark coordinates, the entire geometry of curves (closed or not), or more recently of entire surfaces. Curves analyses are often favoured in archaeobotany due to the absence of clear landmarks to quantify the object geometric and elliptical Fourier transforms (further abbreviated EFT) is the most popular approach for domesticated plants since the organs of interests come as outlines (closed curves) and only present few landmarks, if any.

By comparing archaeological material to modern collections of reference, GMM and EFT in particular for plants, allowed fine-grained inferences, in particular to document the emergence of new morphological types, reconstruct the dynamics of their diffusion in both time and space, and overall gain insights into the intertwined histories of human societies and domesticated plants (Terral *et al.*, 2004, 2010, 2012; Burger *et al.*, 2011; Bouby *et al.*, 2013; Ros *et al.*, 2014; Pagnoux *et al.*, 2015; Bonhomme *et al.*, 2017a, 2021b; Wallace *et al.*, 2018; Bourgeon *et al.*, 2018; Tarongi *et al.*, 2021; Jesus *et al.*, 2021; Evin *et al.*, 2022; Roushannafas *et al.*, 2022).

On the other hand, deep learning quickly became a game-changer from academia to industry, through its versatility and cutting-edge achievements. Computer vision has largely benefited the synergy between the massive democratization of computational power and the arrival of software frameworks on top of solid mathematical foundations. Convolutional neural networks (further abbreviated CNN), in particular, have been at heart of very diverse supervised classification tasks, from autonomous vehicles to plant identification (Alzubaidi *et al.*, 2021). However deep learning still remain rare in archaeological studies (Soroush *et al.*, 2020; Garcia-Molsosa *et al.*, 2021) and in morphometrics (Miele *et al.*, 2020; Le *et al.*, 2020).

Date palm (*Phoenix dactylifiera* L.), olive (*Olea europaea* L.), grapevine (*Vitis vinifera* L.) and barley (*Hordeum vulgare* L.) archaeological seeds and fruits have been intensively studied in archaeobotany using geometric morphometrics. They are four important taxa of human subsistence in the Mediterranean basin since millenia. The presence of the wild progenitors of the domestic forms in vast geographic ranges make the identification of the wild or domestic status of the archaeobotanical

remains of date palm, olive and grapevine particularly difficult. In addition, the presence of multiple types for barley in the region, exploited for diverse use and with different agricultural practices require intra-specific identification.

The morphological distinction between wild and domestic types using GMM is now very accurate for olive (Terral *et al.*, 2004, 2021) and grapevine (Terral *et al.*, 2010; Bonhomme *et al.*, 2022). On the other hand, distinguishing between wild and domestic date palm seeds (Terral *et al.*, 2012; Gros-Balthazard *et al.*, 2017), as well as between two- and six-row barley grains remains challenging (Ros *et al.*, 2014; Bonhomme *et al.*, 2017b; Wallace *et al.*, 2018; Jeanty *et al.*, 2023).

In that context, this paper aims to test the potential of deep learning for archaeobotanical identification and ask the following questions: can CNN beat baselines obtained with GMM and if so, how much data are required to train the models? Here, we used four plant models presenting binary challenges below the species level, e.g. to distinguish between wild and domesticated types of date palm, olive and grapevine, and between two- and six-row barley, at core in archaeobotanical studies.

CNN models correctly trained on large datasets are expected to beat EFT approaches, providing taxonomical differences are reflected in some morphological contrasts, at least because EFT are limited to the geometrical differences of outlines, while CNN can capture any morphological discriminant feature beyond shape, texture for example. That being said, several conditions of our models make such expectation far from granted here:

i) *Low inter-class differences*: differences tested here, chiefly shape differences, ranged from subtle at best to extremely challenging including for experts; the group labelling was certain only because the identification was obtained through molecular markers (for the date palm) or on entire plants cultivated in biological conservation centres (other models).

ii) *Small sample sizes*: the available datasets were particularly small compared to those usually deployed in deep learning tasks. The datasets used here were obtained through 2D images acquired following rigorous (and time-consuming protocols), which limit the number of biological objects that can be analysed in the context of archaeobotanical studies.

iii) *Challenging baselines*: existing baselines obtained through GMM are already good to very good.

iv) *Accessible models*: our intention was to develop CNN-based pipelines, reasonably easy to run for non-expert users and using regular desktop computers facilitating similar analyses on other models.

We first present the models used and compare their performance to geometric morphometrics. Finally, we discuss the pros and cons of deep learning versus geometric morphometrics and propose an agenda of future researches.

Material and methods

Statistical environment

All analyses were run using R 4.1.3 (R Development Core Team, 2023). We used a spare and quite antique MacBook Pro 2013 model with a 2,6 GHz Intel Core i5 CPU and 16 Go 1600 MHz DDR3 RAM. Data manipulation and visualization was done using *tidyverse* (Wickham *et al.*, 2019). Image manipulation was done using *magick* (Ooms, 2021). All morphometric analyses were performed using *Momocs* (Bonhomme *et al.*, 2014, 2020a) and linear discriminant analyses using *MASS* (Venables & Ripley, 2002). Deep learning was done using *keras* (Allaire & Chollet, 2020), the R interface to the eponym Python architecture, which here ran on CPU alone.

Datasets used

Among the model species studied by our team, we retained those for which we have enough material, secure identification and associated publication record: grapevine pips (Pagnoux *et al.*, 2015; Bonhomme *et al.*, 2020b), barley grains (Jeanty *et al.*, 2023), olive stones (Bourgeon *et al.*, 2018; Terral *et al.*, 2021) and dates seeds (Terral *et al.*, 2012; Gros-Balthazard *et al.*, 2017) (Table 1). All models corresponded to a binary classification task with 2- versus 6-row types of barley (*Hordeum vulgare*), and wild versus domestic for the three other taxa.

All seeds/stones/fruits were photographed in dorsal and lateral views using a stereomicroscope coupled with a digital camera. It is worth noting that GMM identification is usually obtained by combining the information brought by the two orthogonal views but we chose here to *not* combine these views to increase the number of "independent" datasets and have a larger spectrum of shapes (Figure 1).

Convolutional neural networks

Our CNN models used the VGG16 architecture (Simonyan & Zisserman, 2014) with the weights trained on the ImageNet reference dataset (Deng *et al.*, 2009), as available in *keras*. The convolutional base, with feature hierarchies learnt on ImageNet, was frozen. Given we did not want to predict ImageNet classes, the last three fully connected layers were replaced with two fully connected dense layers and only these last layers were fine-tuned. The first has 32 units and a rectified linear unit (ReLU). Because all models were binary classification, the last layer has two units and a sigmoid activation (Figure 1).

The loss was calculated using binary cross-entropy for binary classification tasks. We used two callbacks to control the training step. The first controlled the learning rate, initially fixed to 10^{-2} with a decay factor of 10, a patience of 10 epochs and a minimal value of 10^{-7}. The second stopped the training with a patience of 20 epochs on the accuracy. In other words, for each model, the training was stopped when the last twenty epochs did not show a decrease in loss.

For each dataset, the number of images was balanced between classes, using a random sampling without replacement among available images (Table 1). For each sample size tested, 60% of the total number of images was used for the training set, 20% for the validation set and the last 20% for evaluating the model. The training set was used to adjust weights while the validation set was used to evaluate model performance back-propagate results to the unfrozen layers at the end of each epoch. The evaluating set was used only once and after the training step, to report the model performance on images never seen before by the model.

The rgb images were reduced to a resolution of 120×90 pixels, while maintaining the aspect ratio (Figure 1). The first layer of the VGG16 convolutional based was adapted accordingly.

Geometric morphometrics baseline

We used outline analysis using elliptical Fourier transforms (EFT) (Kuhl & Giardina, 1982; Claude, 2008; Bonhomme *et al.*, 2014). We first converted full-sized images into silhouette masks on which 360 outline coordinates were sampled, equally spaced along the curvilinear abscissa. We then normalized outlines for their size, position, rotation and first point and obtained enough harmonics to gather 95% of the total harmonic power (6 for all datasets). Then, a linear discriminant analysis was trained on the same dataset as for CNN yet combining the training and validation sets (Figure 1). The general methodology is detailed elsewhere (Bonhomme *et al.*, 2014, 2020c).

Model comparisons

Five replicates were used for each of the eight datasets and each was tested with increasing sample sizes (Table 1 and Table 2). Given one of the 280 runs, the very same sets of images (or masks) was submitted to both CNN and EFT. For CNN, the images were partitioned into 60% training, 20% validation then evaluated on the last 20%; for EFT the two training and validation sets were combined for training then evaluated on the same 20% as for CNN (Figure 1). This cross-validation scheme allowed direct comparisons between the respective performances of each model. Performance was measured with accuracy, that is the proportion of correctly identified individuals.

Results

In most cases, CNN beat EFT (213 cases over 280, that is 76% - Figure 2, Figure 3, Table 2). This is particularly true for larger training sets.

Among the eight datasets (further referred using their vernacular name) two groups can be made: olive and grapevine in one hand, barley and date palm in the other, no matter the view considered. For grapevine and olive, where EFT "already" provided good accuracies, CNN perform even better, particularly for the large sample sizes. For grapevine with 700 images, average CNN accuracies range from 87 to 97% for ventral view and from 88 to 94% for lateral view. For olive with 500 images,

performances range from 91 to 99% for ventral view, and from 92 to 99% for lateral view. For barley and date palm, the results seem more mitigated at first glance (Figure 2), yet, on average (Table 2) the CNN also achieve better accuracies when the datasets are large enough. For sample sizes above 150 individuals, CNN are better in most cases for barley and consistently for date palm. These two groups of results are reflected in the mean differences between models for the largest sample size: olive and grapevine gained ~10% accuracy where date palm and barley gained less than 5%.

Finally, to give an idea of computational time, a single iteration of the 280 models pairs took ~17 hours to complete, with less than 1% dedicated to the EFT. In the other hand, post-treatment time for preparing pictures is virtually zero for CNN and about 1 min per picture for EFT, that is about a full-time week for each dataset here.

Discussion

Our results show that even a candid CNN approach could outperform state-of-the-art EFT to identify plant seeds and fruits below the species level. Even if the performance boost is not dramatical for all four studied taxa, this was a quite surprising result since the CNN beat almost consistently our EFT baselines even when the sample sizes were small.

Regarding the four pairs of taxa studied, identifying wild and domestic types for olive and grapevine is relatively easy using the seed shape but distinguishing between the wild and domestic date palm, and the two- and six-row barley is challenging, not to say troublesome.

Here, when geometrical differences between studied pairs are quite obvious macroscopically (olive and grapevine), the CNN clearly beat GMM identification and is close to perfect when the sample sizes of the training sets exceeded 500 hundred seeds. For instance, over the 5 replicates, a single olive stone in lateral view was wrongly identified among the 700 evaluated images (5*700*20%). Accuracies around 95% are now common for certain taxa (e.g. olive and grapevine), particularly when combining several views (Bonhomme *et al.*, 2020c), but here CNN only have raw 120×90 images as inputs with the size of the seeds even smaller than that in the image (Figure 1).

Given the number of parameters to fine-tune in the CNN architecture used, the most surprising result is that CNN also beat EFT baseline even when trained with only ~100 images in each class, at least for these two "easy" models (here grapevine and olive). Given how costly and time-consuming is the constitution of a reference collection, this means that CNN can be tested early and possibly cut off these costs. Also, that methods applied here could be easily tested in many other archeological models whether they are plant, animal organs or non-biological artefacts, imprints, etc.

One important result here is that CNN can still improve their classification score when improving the learning sample size well after the classification score of GMM can no more be improved (because of method limitation or because the number of available variables is limited). Our results seem also to indicate that GMM and linear discriminant analysis allow to fast reach their maximal accuracy

but rapidly reach a plateau (corresponding to 50 to 150 analyzed individuals). With larger sample size, they are clearly performing less well than CNN.

Deep learning approaches are now quite common for animal and plant species identification, particularly for citizen science projects (Willi *et al.*, 2019; Picek *et al.*, 2022), but remain so far very new when it comes to archaeological material (but see Miele et al., 2020) or morphometrics (but see Le et al., 2020). To the best of our knowledge, this is the first time CNN are used for such sub-specific identification task in plants, a fortiori on four different model taxa. The results shown here appeal to further studies to test how they could be extended to other archaeological material, other plant or animal taxa and at the species level. Here we show that, at least in some cases, the diversity at even lower taxonomic levels can be explored. This would be of prime interest to develop tools that can be used not only by archaeobotanists but also by any people interested in identifying variety (*e.g.* for conservation purposes). In palynology, another field that may be developed in an archaeological context, deep learning using CNN has already proven to be helpful in the fastidious task of pollen identification and counting (Sevillano *et al.*, 2020).

More generally, should we expect rivalry or synergy emerge between deep learning and geometric morphometrics? For rough identification, CNN will likely become a new tool for future archaeobotanical studies, and possibly the next best one. Here, we insist on the fact that our CNN architecture was deliberately kept simple for both practical and conservative reasons: we had many models to run that needed to be generic and the point was to test if a candid CNN approach could beat state-of-the-art EFT. There is definitely room for improvement by using better models, fine-tuning them, larger datasets, larger images and by combining views or even using 3D models of the objects. That being said, with the best will in the world a model cannot see what simply does not exist. In some cases, a single ratio of lengths can achieve nearly perfect identification when morphological differences are trivial. This is the case for example for grapevine (Bonhomme *et al.*, 2022). On the other hand, meaningful differences for human use may just not be reflected on the studied organs. Somewhere between these two extremes are a wide range of real differences that can only be detected by statistical means (Bonhomme *et al.*, 2021a,b). This is where methodological refinement makes the most sense and a natural playground for deep learning approaches.

Geometric morphometrics - and EFT in particular for plant organs - have the advantage of traducing the shape into coefficients that can be directly treated as quantitative variables. Also, the inverse transformations are mathematically defined, so that one can go “back” to shape from coefficients, which allows rich insights into the morphological space of taxa of interest and the comparison between the relative occupancies between taxonomic, diachronic or synchronic assemblages. For that matter, the best equivalent CNN have to offer are activation maps where one can visualize for each image, the regions that triggered the final vote. Even though the reputation of being black boxes is largely erroneous, CNN are and will likely remain less handy to that respect. More generally, “to predict is not to explain” (Thom, 2010), and in our opinion, CNN and EFT should be seen

as complementary approaches rather than competitors. Future studies will explore this assumption but CNN may soon become the go-to tool when identification is of prime interest. Paradoxically, CNN are more computationally intensive than GMM models but may reveal easier to deploy as app and use for a broad audience, because they train and predict raw images where GMM approaches require meticulously prepared images.

Finally, if deep learning was here restricted to identification using convolutional neural networks, it has much more to offer to archaeology and morphometrics: its versatility extends to regression problems (e.g. Reese, 2021), segmentation (i.e. automating and/or improving pre-morphometrics treatment (e.g. Lee et al., 2017), adversarial reconstruction for broken or missing parts (e.g. Hermoza & Sipiran, 2018), pose and parallax correction for data acquisition (e.g. Zhang et al., 2021).

Acknowledgements

This work has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant Agreement 852573), the French National Agency (ANR-16-CE27-0013 and ANR-06-BLAN-0212-02-PHOENIX) the International Research Program EVOLEA (France - Morocco) (CNRS-INEE) and the Défi Clé "Sciences du passé" - Occitanie region / University of Toulouse, France (PATRIMOLEA programme). We are grateful to the INRA Vassal-Montpellier grapevine collection (Marseillan-Plage, France) that provided all the pips from cultivated varieties and the OSU-OREME (https://oreme.org/) that helped to the constitution of the wild grape pip collection. We thank the Melgueil experimental domain (INRA, Mauguio, France) and the Tessaout worldwide collection of the olive tree (INRA Marrakech, Morocco) that provide the olives stones from cultivated varieties. We also thank the Biological Resource Center-INRAE Clermont Ferrand for providing all the barley grains studied.

Tables

Table 1: Material used. Each of the four taxa provided two views treated separately.

dataset	taxa	classification problem	levels	sample sizes
hordeum	*Hordeum vulgare*	2- vs 6-row	2-row	981
			6-row	473
olea	*Olea europaea*	wild vs domesticated	domesticated	1589
			wild	630
phoenix	*Phoenix dactylifera*	wild vs domesticated	domesticated	776
			wild	662
vitis	*Vitis vinifera*	wild vs domesticated	domesticated	1769
			wild	872

Table 2: Mean performance (CNN – EFT) differences, expressed as percentages, for each of the 5 replicates. Sample sizes are expressed as the total number of images used per class (training/validation + evaluation). Cells with '-' could not be calculated due to sample size limitations.

		sample sizes								
dataset	**view**	**50**	**100**	**150**	**200**	**300**	**400**	**500**	**600**	**700**
hordeum	VD	-2	-3.2	2.3	4.6	-0.2	-	-	-	-
hordeum	VL	-4	2.5	-1.6	2.5	2	-	-	-	-
olea	VD	7	4.8	5.6	6.5	8.6	9	9	-	-
olea	VL	1.3	1.7	5.9	6.6	7.1	8.6	7.4	-	-
phoenix	VD	-0.9	-0.6	1.3	0.3	1.8	2.8	3.2	-	-
phoenix	VL	-8.3	-2	1.5	1.3	0.1	1.1	3.5	-	-
vitis	VD	-0.8	0.8	9.2	5.1	7.5	8.6	7.6	8.9	10.1
vitis	VL	-4.1	1	1.3	3.3	2	4	4.3	3.7	6.5

Figures

Figure 1: Schematic description of the GMM and CNN models used. For each taxon, an archetypical individual seed is presented.

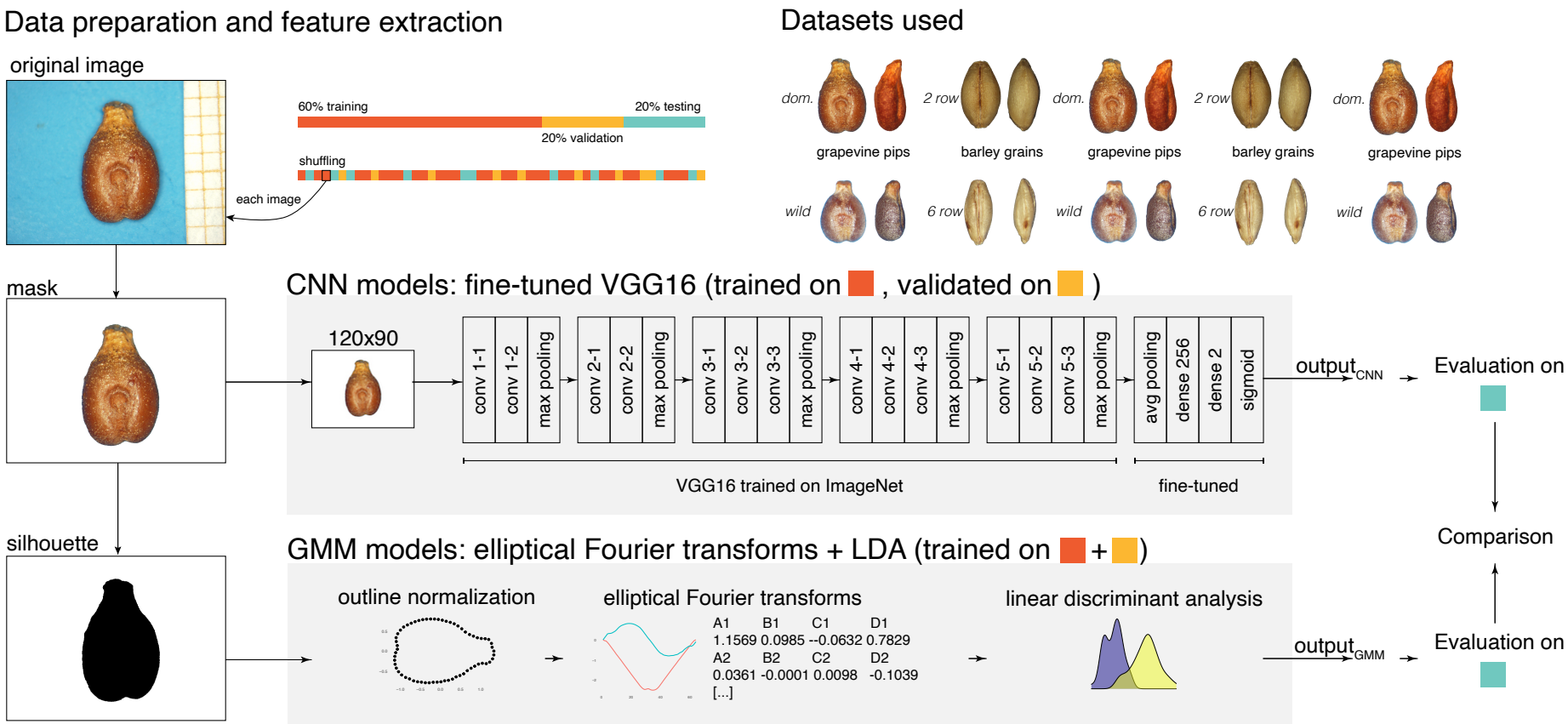


Figure 2: Model performances for each dataset, training sample sizes and replicates for deep learning (cnn) and geometric morphometrics (eft). Sample sizes are expressed as the total number of images used per class (training/validation + evaluation). The models are run for the ventral (VD) and lateral (VL) views. Two-row vs six-row barley (hordeum) are compared, as well as the wild and domestic forms of olive (olea), date palm (phoenix) and grapevine (vitis).

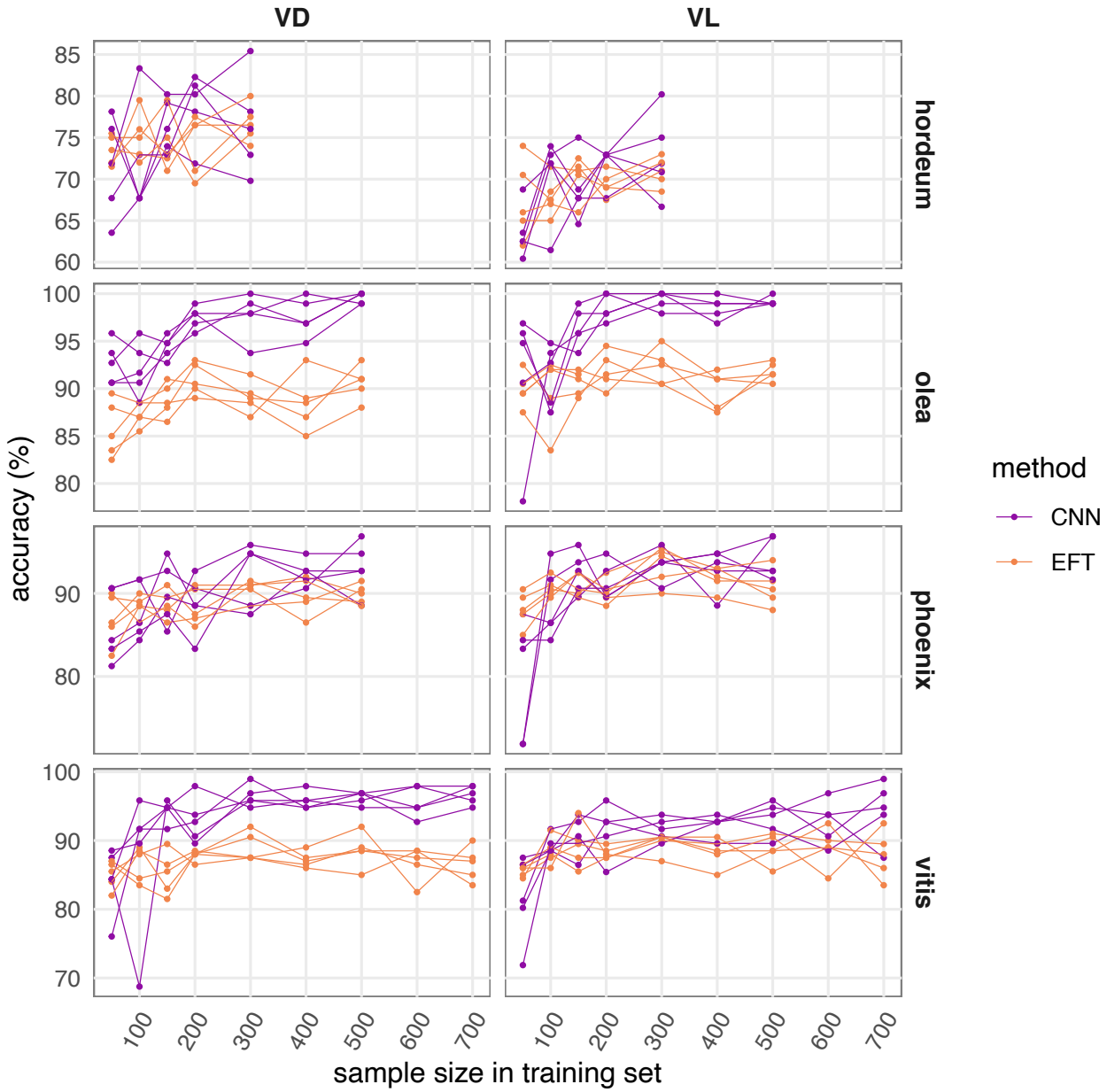
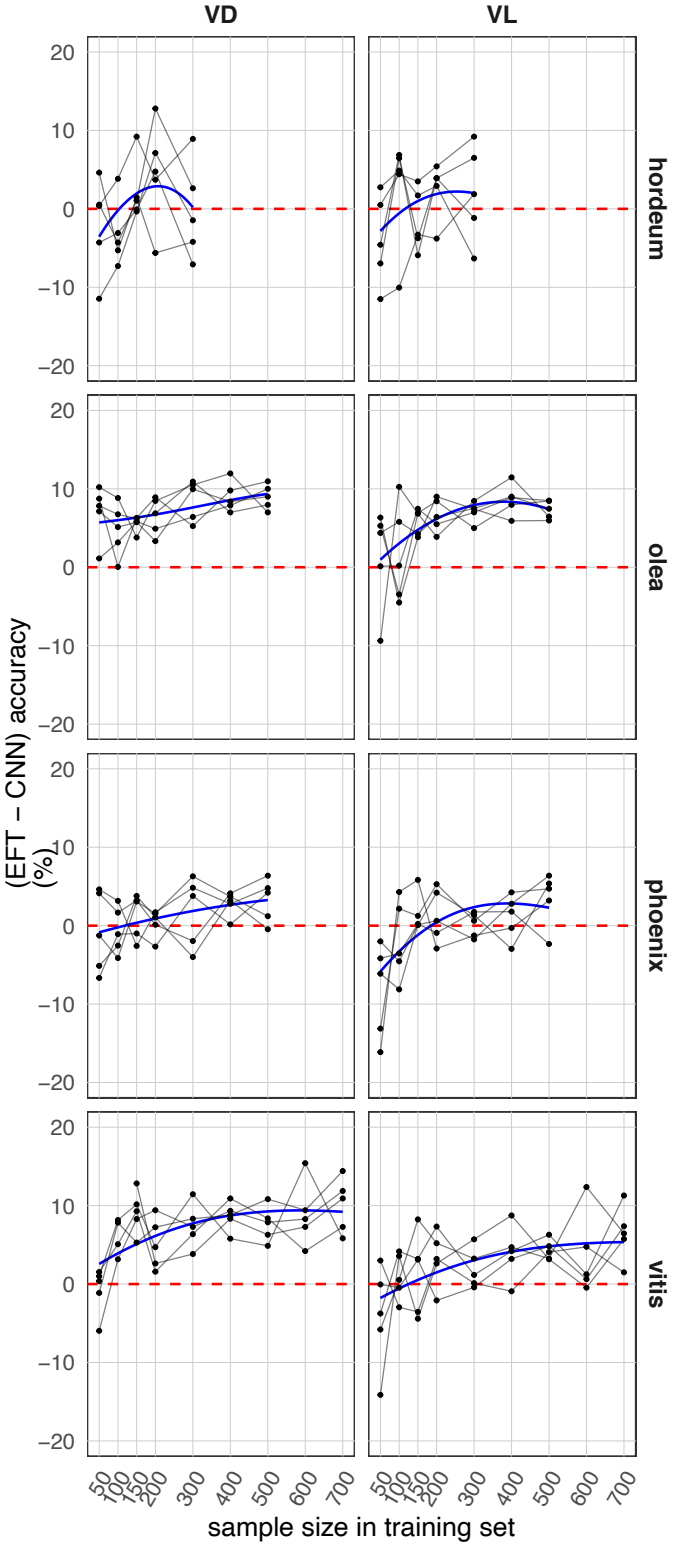


Figure 3: Model performances for each dataset, training sample sizes and replicates, presented as absolute percentage differences. Sample sizes are expressed as the total number of images used per class (training/validation + evaluation). The models are run for the ventral (VD) and lateral (VL) views. Two-row vs six-row barley (hordeum) are compared, as well as the wild and domestic forms of olive (olea), date palm (phoenix) and grapevine (vitis).



References

Allaire J, Chollet F. 2020. keras: R Interface to 'Keras'.

Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, Santamaría J, Fadhel MA, Al-Amidie M, Farhan L. 2021. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data* **8**: 53.

Bonhomme V, Forster E, Wallace M, Stillman E, Charles M, Jones G. 2017a. Identification of inter- and intra-species variation in cereal grains through geometric morphometric analysis, and its resilience under experimental charring. *Journal of Archaeological Science* **86**: 60–67.

Bonhomme V, Forster E, Wallace M, Stillman E, Charles M, Jones G. 2017b. Identification of inter- and intra-species variation in cereal grains through geometric morphometric analysis, and its resilience under experimental charring. *Journal of Archaeological Science* **86**: 60–67.

Bonhomme V, Ivorra S, Lacombe T, Evin A, Figueiral I, Maghradze D, Marchal C, Pagnoux C, Pastor T, Pomarèdes H, *et al.* 2021a. Pip shape echoes grapevine domestication history. *Scientific Reports* **11**: 21381.

Bonhomme V, Pagnoux C, Bouby L, Evin A, Ivorra S, Allen S, Valamoti S-M. 2022. Early viticulture in Neolithic and Bronze Age Greece: looking for the best traditional morphometric method to distinguish wild and domestic grape pips. In: Valamoti S-M, Dimoula A, Ntinou M, eds. Cooking with Plants in Ancient Europe and Beyond: interdisciplinary approaches to the Archaeology of Plant Foods (in press.). Sidestone Press.

Bonhomme V, Picq S, Claude J. 2020a. Momocs package (github.com/MomX/Momocs).

Bonhomme V, Picq S, Gaucherel C, Claude J. 2014. Momocs : Outline Analysis Using R. *Journal of Statistical Software* **56**.

Bonhomme V, Picq S, Ivorra S, Evin A, Pastor T, Bacilieri R, Lacombe T, Figueiral I, Terral J-F, Bouby L. 2020b. Eco-evo-devo implications and archaeobiological perspectives of trait covariance in fruits of wild and domesticated grapevines (M Singh, Ed.). *PLOS ONE* **15**: e0239863.

Bonhomme V, Picq S, Ivorra S, Evin A, Pastor T, Bacilieri R, Lacombe T, Figueiral I, Terral J-F, Bouby L. 2020c. Eco-evo-devo implications and archaeobiological perspectives of trait covariance in fruits of wild and domesticated grapevines (M Singh, Ed.). *PLOS ONE* **15**: e0239863.

Bonhomme V, Terral J-F, Zech-Matterne V, Ivorra S, Lacombe T, Deborde G, Kuchler P, Limier B, Pastor T, Rollet P, *et al.* 2021b. Seed morphology uncovers 1500 years of vine agrobiodiversity before the advent of the Champagne wine. *Scientific Reports* **11**: 2305.

Bouby L, Figueiral I, Bouchette A, Rovira N, Ivorra S, Lacombe T, Pastor T, Picq S, Marinval P, Terral J-FF. 2013. Bioarchaeological Insights into the Process of Domestication of Grapevine (Vitis vinifera L.) during Roman Times in Southern France (P Degryse, Ed.). *PLoS ONE* **8**: e63195.

Bourgeon O, Pagnoux C, Mauné S, Vargas EG, Ivorra S, Bonhomme V, Ater M, Moukhli A, Terral J-F. 2018. Olive tree varieties cultivated for the great Baetican oil trade between the 1st and the 4th centuries ad: morphometric analysis of olive stones from Las Delicias (Ecija, Province of Seville, Spain). *Vegetation History and Archaeobotany* **27**.

Burger P, Terral J-F, Ruas M-P, Ivorra S, Picq S. 2011. Assessing past agrobiodiversity of *Prunus avium* L. (Rosaceae): a morphometric approach focussed on the stones from the archaeological site Hôtel-Dieu (16th century, Tours, France). *Vegetation History and Archaeobotany* **20**: 447–458.

Claude J. 2008. *Morphometrics with R*. New York, NY: Springer New York.

Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. 2009. ImageNet: A Large-Scale Hierarchical Image Database. In: CVPR09.

Evin A, Bouby L, Bonhomme V, Jeanty A, Jeanjean M, Terral J-F. 2022. Archaeophenomics of ancient domestic plants and animals using geometric morphometrics : a review. *Peer Community Journal* **2**: e27.

Garcia-Molsosa A, Orengo HA, Lawrence D, Philip G, Hopper K, Petrie CA. 2021. Potential of deep learning segmentation for the extraction of archaeological features from historical map series. *Archaeological Prospection* **28**: 187–199.

Gros-Balthazard M, Galimberti M, Kousathanas A, Newton C, Ivorra S, Paradis L, Vigouroux Y, Carter R, Tengberg M, Battesti V, *et al.* 2017. The Discovery of Wild Date Palms in Oman Reveals

a Complex Domestication History Involving Centers in the Middle East and Africa. *Current Biology* **27**: 2211-2218.e8.
Hermoza R, Sipiran I. **2018**. 3D Reconstruction of Incomplete Archaeological Objects Using a Generative Adversarial Network. In: Proceedings of Computer Graphics International 2018 on - CGI 2018. New York, New York, USA: ACM Press, 5–11.
Jacomet S. **2008**. *Identification of cereal remains from archaeological sites*. https://ipna.unibas.ch/archbot/pdf/Cereal_Id_Manual_engl.pdf: IPAS, Basel University.
Jeanty A, Bouby L, Bonhomme V, Balfourier F, Debiton C, Dham C, Ivorra S, Ros J, Evin A. **2023**. Barley systematics and taxonomy foreseen by seed morphometric variation. *PLoS ONE* **18**: 1–17.
Jesus A, Bonhomme V, Evin A, Ivorra S, Soteras R, Salavert A, Antolín F, Bouby L. **2021**. A morphometric approach to track opium poppy domestication. *Scientific Reports* **11**: 9778.
Kendall D. **1989**. A survey of the statistical theory of shape. *Statistical Science* **4**: 81–120.
Kuhl FFP, Giardina CCR. **1982**. Elliptic Fourier features of a closed contour. *Computer graphics and image processing* **18**: 236–258.
Le V-L, Beurton-Aimar M, Zemmari A, Marie A, Parisey N. **2020**. Automated landmarking for insects morphometric analysis using deep neural networks. *Ecological Informatics* **60**: 101175.
Lee H, Troschel FM, Tajmir S, Fuchs G, Mario J, Fintelmann FJ, Do S. **2017**. Pixel-Level Deep Segmentation: Artificial Intelligence Quantifies Muscle on Computed Tomography for Body Morphometric Analysis. *Journal of Digital Imaging* **30**: 487–498.
Miele V, Dussert G, Cucchi T, Renaud S. **2020**. Deep learning for species identification of modern and fossil rodent molars. *bioRxiv*.
Ooms J. **2021**. magick: Advanced Graphics and Image-Processing in R.
Pagnoux C, Bouby L, Ivorra S, Petit C, Valamoti S-M, Pastor T, Picq S, Terral J-F. **2015**. Inferring the agrobiodiversity of Vitis vinifera L. (grapevine) in ancient Greece by comparative shape analysis of archaeological and modern seeds. *Vegetation History and Archaeobotany* **24**: 75–84.
Picek L, Šulc M, Patel Y, Matas J. **2022**. Plant recognition by AI: Deep neural nets, transformers, and kNN in deep embeddings. *Frontiers in Plant Science* **13**.
R Development Core Team. **2023**. R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing, Vienna, Austria.*
Reese KM. **2021**. Deep learning artificial neural networks for non-destructive archaeological site dating. *Journal of Archaeological Science* **132**: 105413.
Ros J, Evin A, Bouby L, Ruas M-P. **2014**. Geometric morphometric analysis of grain shape and the identification of two-rowed barley (*Hordeum vulgare* subsp. *distichum* L.) in southern France. *Journal of Archaeological Science* **41**: 568–575.
Roushannafas T, Bogaard A, Charles M. **2022**. Geometric morphometrics sheds new light on the identification and domestication status of 'new glume wheat' at Neolithic Çatalhöyük. *Journal of Archaeological Science* **142**: 105599.
Sevillano V, Holt K, Aznarte JL. **2020**. Precise automatic classification of 46 different pollen types with convolutional neural networks (M Raza, Ed.). *PLOS ONE* **15**: e0229751.
Simonyan K, Zisserman A. **2014**. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
Soroush M, Mehrtash A, Khazraee E, Ur JA. **2020**. Deep Learning in Archaeological Remote Sensing: Automated Qanat Detection in the Kurdistan Region of Iraq. *Remote Sensing* **12**: 500.
Tarongi M, Bonhomme V, Evin A, Ivorra S, López D, Alonso N, Bouby L. **2021**. A new way of seeing pulses: preliminary results of geometric morphometric analyses of Iron Age seeds from the site of La Font de la Canya (Barcelona, Spain). *Vegetation History and Archaeobotany* **30**: 77–87.
Terral J-F, Alonso N, Chatti N, Capdevila RB i, Fabre L, Fiorentino G, Marinval P, Jordá GP, Pradat B, Rovira N, *et al.* 2004. Historical biogeography of olive domestication (Olea europaea L.) as revealed by geometrical morphometry applied to biological and archaeological material. *Journal of Biogeography* **31**: 63–77.
Terral J-F, Bonhomme V, Pagnoux C, Ivorra S, Newton C, Paradis L, Ater M, Kassout J, Limier B, Bouby L, *et al.* 2021. The Shape Diversity of Olive Stones Resulting from Domestication and Diversification Unveils Traits of the Oldest Known 6500-Years-Old Table Olives from Hishuley Carmel Site (Israel). *Agronomy* **11**: 2187.
Terral J-F, Newton C, Ivorra S, Gros-Balthazard M, de Morais CT, Picq S, Tengberg M, Pintaud

J-C. **2012**. Insights into the historical biogeography of the date palm (*Phoenix dactylifera* L.) using geometric morphometry of modern and ancient seeds. *Journal of Biogeography* **39**: 929–941.
Terral J-F, Tabard E, Bouby L, Ivorra S, Pastor T, Figueiral I, Picq S, Chevance J-BB, Jung C, Fabre L, *et al.* 2010. Evolution and history of grapevine (Vitis vinifera) under domestication: new morphometric perspectives to understand seed domestication syndrome and reveal origins of ancient European cultivars. *Annals of Botany* **105**: 443–455.
Thom R. **2010**. *To Predict is not To Explain: : Conversations on Mathematics, Science, Catastrophe Theory, Semiophysics, Morphogenesis and Natural Philosophy*. Thombooks Press.
Venables WN, Ripley BD. **2002**. *Modern Applied Statistics with S*. New York: Springer.
Wallace M, Bonhomme V, Russell J, Stillman E, George TS, Ramsay L, Wishart J, Timpany S, Bull H, Booth A, *et al.* 2018. Searching for the Origins of Bere Barley: A Geometric Morphometric Approach to Cereal Landrace Recognition in Archaeology. *Journal of Archaeological Method and Theory*.
Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, Grolemund G, Hayes A, Henry L, Hester J, *et al.* 2019. Welcome to the Tidyverse. *Journal of Open Source Software* **4**: 1686.
Willi M, Pitman RT, Cardoso AW, Locke C, Swanson A, Boyer A, Veldthuis M, Fortson L. **2019**. Identifying animal species in camera trap images using deep learning and citizen science. *Methods in Ecology and Evolution* **10**: 80–91.
Zhang S, Lu S, He R, Bao Z. **2021**. Stereo Visual Odometry Pose Correction through Unsupervised Deep Learning. *Sensors* **21**: 4735.
Zohary D, Hopf M., Weiss E. **2012**. *Domestication of Plants in the Old World: The origin and spread of domesticated plants in Southwest Asia, Europe, and the Mediterranean Basin*. OUP Oxford.