A machine learning based approach to the segmentation of micro CT data in archaeological and evolutionary sciences.

Authors:

Thomas O'Mahoney (1,2)*

Lidija Mcknight (3)

Tristan Lowe (3,4)

Maria Mednikova (5)

Jacob Dunn (1,6,7)

1. School of Life Sciences, Faculty of Science and Engineering, Anglia Ruskin University, Cambridge, UK
2. McDonald institute for Archaeological Research, University of Cambridge, Cambridge, UK
3. Interdisciplinary Centre for Ancient Life, University of Manchester, Manchester, UK
4. Henry Mosely X-Ray Imaging Facility, University of Manchester, Manchester, UK
5. Institute of Archaeology, Russian Academy of Sciences, Moscow, Russian Federation
6. Division of Biological Anthropology, Department of Archaeology, University of Cambridge, UK
7. Department of Cognitive Biology, University of Vienna

*correspondence to Thomas O'Mahoney: tomomahoney@gmail.com

KEY WORDS: MICROCT; IMAGE SEGMENTATION; BIOINFORMATICS; MACHINE LEARNING; ANTHROPOLOGY; PALEONTOLOGY

Abstract

Segmentation of high-resolution tomographic data is often an extremely time-consuming task and until recently, has usually relied upon researchers manually selecting materials of interest slice by slice. With the exponential rise in datasets being acquired, this is clearly not a sustainable workflow. In this paper, we apply the Trainable Weka Segmentation (a freely available plugin for the multiplatform program ImageJ) to typical datasets found in archaeological and evolutionary sciences. We demonstrate that Trainable Weka Segmentation can provide a fast and robust method for segmentation and is as effective as other leading-edge machine learning segmentation techniques.

Introduction

Three-dimensional imaging using micro CT scanning has rapidly become mainstream in the archaeological and evolutionary sciences. It enables the high-resolution and non-destructive analysis of internal structures of scientific interest. In archaeological sciences it has been used for a variety of purposes, from imaging pottery (Barron et al., 2017; Tuniz and Zanini, 2018) understanding soil compaction (McBride and Mercer, 2012) imaging early bone tools (Bello et al., 2013) and mummies, both human and animal (Charlier et al., 2014; Du Plessis et al., 2015; Romell et al., 2018). In evolutionary sciences, it is employed even more widely, from scanning hominin remains for morphological reconstruction (Gunz et al., 2012; Hershkovitz et al., 2018, 2015; Hublin et al., 2017) to diagnosing ancient pathologies (Anné et al., 2015; Odes et al., 2016; Randolph-Quinney et al., 2016). It is used extensively in vertebrate palaeontology (Abel et al., 2012; Chapelle et al., 2019; Hechenleitner et al., 2016; Laloy et al., 2013) and invertebrate palaeontology (Garwood and Dunlop, 2014; Wacey et al., 2017) and increasingly, palaeopalynology (Collinson et al., 2016).

In comparative anatomy, it is now part of the standard non-destructive analytical toolkit, alongside geometric morphometrics (GMM) and finite element analysis (FEA) (Borgard et al., 2019; Brassey et al., 2018, 2013; Brocklehurst et al., 2019; Cuff et al., 2015; Marshall et al., 2019; Polly et al., 2016). Unfortunately for researchers, if one wishes to quantify biological structures, the data does not simply appear from scanners ready to use. It requires processing through the segmentation of the structures of interest, followed (usually) by the generation of 3-dimensional models.

There are 5 main approaches to image segmentation.

- Global thresholding based upon greyscale values in scans.
- Watershed based segmentation
- Locally adaptive segmentation

- Manual segmentation of structures
- Label based segmentation, in conjunction with machine learning.

One can broadly classify greyscale segmentation and edge-based segmentation as passive approaches, as very little input is required from the user, and Region or label based segmentation as active, in that they require more explicit input from the user.

Greyscale thresholding

Greyscale thresholding is the oldest approach to the processing of tomographic data (Spoor et al., 1993) and has been refined to the use of the half width, full maximum height approach based upon stack histograms (Spoor et al., 1993). This however is only useful for materials which have a single range of X-ray absorption and several passes are therefore required for the segmentation of multiple tissue types.

Watershed based segmentation

Watershed based segmentation has enjoyed a lot of popularity for segmenting complex structures such as brain folds but is also of some utility when segmenting fossil structures. A recent innovation has been the application of Ray-casting and similar techniques to the processing of data (Dunmore et al., 2018; Scherf and Tilgner, 2009) which helps to ameliorate problems with fuzzy data and automates the processing of this. A problem is that it is only feasible to process a single material (although the others are also detected) and an aspect of 're-looping' the procedure is then required, which can create a bottleneck for scans where multiple materials are of equal interest (for example, mummies, where the skeleton, desiccated flesh and wrappings are all of equal scientific interest.

Locally adaptive segmentation

Locally adaptive segmentation is increasingly carried out using deep learning in an automated fashion. Algorithms use combinations of edge detection, texture similarity and image contrast to create rules for the classification of different materials. (Prasoon et al., 2013; Radford et al., 2015; Suzani et al., 2015). It has become increasingly popular with the availability of massive datasets from healthcare providers and several recent reviews cover this suite of techniques in-depth (Greenspan et al., 2016; Litjens et al., 2017; Shen et al., 2017; Suzuki, 2017). A criticism of unsupervised methods such as convoluted neural networks is that they can demand huge computing resources while still often yielding false positives including highlighting artefacts in data (e.g. ring artefacts in CT scans(Nguyen et al., 2015; Szegedy et al., 2013; Wang, 2016). Another set of related techniques include kmeans and c-means clustering algorithms. K-means is known as a 'hard' clustering algorithm, introduced independently by Forgy and MacQueen (Forgy, 1965; MacQueen, 1967). This method, and extensions of it, have been used widely in MRI processing (e.g. (Dimitriadou et al., 2004; Juang and Wu, 2010; Singh et al., 1996). An interesting note is that the original publication cautioned against using k-means clustering as a definitive algorithm, but as an aid to the user in interpreting clusters of data. Another popular clustering algorithm is that of fuzzy c-means (Bezdek, 1980, 1980, 1975; Pham and Prince, 1999) which is an example of 'soft' clustering methods, where probabilities of group allocation are given. Again, this is popular for the automated segmentation of MRI data (e.g. Dimitriadou et al., 2004) and computational speed can be further improved by an initial clustering using k-means partitioning (Dunmore et al., 2018).

Manual segmentation

Manual segmentation is usually carried out using a graphics tablet and the contours of each material of interest are manually traced by a researcher who is familiar with the characteristics of the material of interest. This technique is intrinsically reliant on the skill of the researcher carrying out the segmentation and is also extremely time consuming for large datasets.

Label based segmentation

Label based segmentation is commonly used to 'seed' areas of interest and a contour is propagated until a significant difference in the material absorption is observed (within user set parameters such as kernel size and diffuseness of boundaries). In more recent applications, these approaches have been combined with machine learning, such as in (Arganda-Carreras et al., 2017; Glocker et al., 2013). Most user guided approaches utilise a variant of the Random Forest Algorithm for training (Breiman, 2001; Tin Kam Ho, 1998). The Weka segmentation method is an example of supervised label based segmentation, augmented by machine learning and as such, is our preferred technique for the segmentation of complex anatomical and archaeological/palaeontological data which may suffer from artefacts in scanning and material inhomogeneity (defined here as differences in material x-ray absorption). It has previously been tested on ground truth images applicable to geological samples and found to perform at least as well as other leading algorithms (Berg et al., 2018). It provides an easy to interpret overlay on training datasets (see figure 1) and can be used to rapidly process multiple complex materials simultaneously.

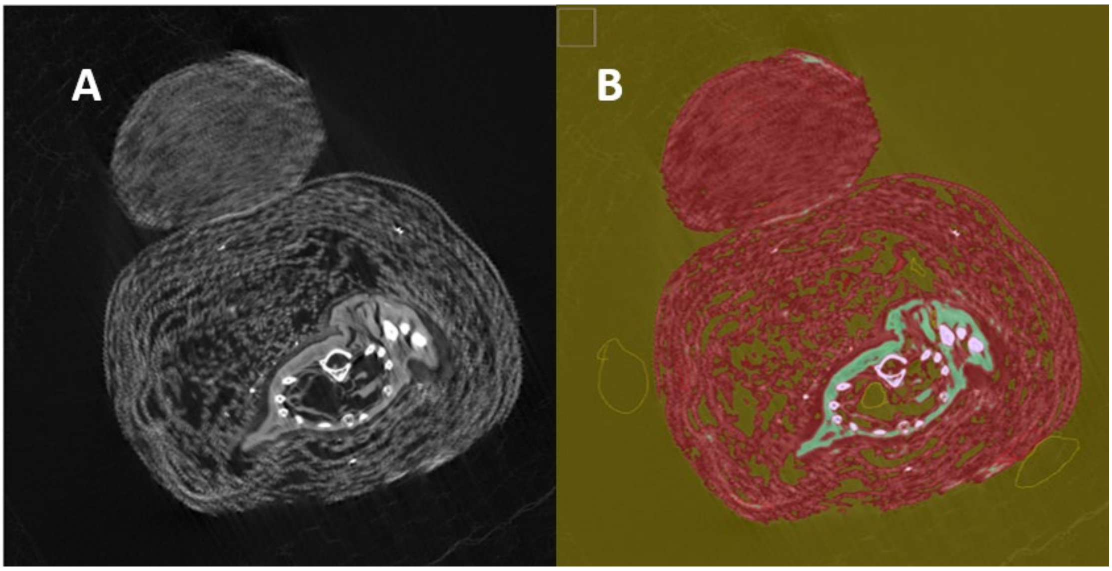

Figure 1 An example of Weka segmentation. A: Original slice B: Weka classification of rodent mummy showing previews of masking

In summary, a significant roadblock to more rapid and precise advances in micro CT imaging in archaeological and evolutionary sciences is that the structures of interest are often non-homogenous in nature and until recently, have required extensive manual processing of slices . Recent advances in machine learning, combined with user-friendly interfaces mean that an acceleration of data processing is seriously possible, especially when combined with the potential to process data through either clusters or multiple GPUs. In this article, we demonstrate for the first time the implementation of the Trainable Weka Segmentation (Arganda-Carreras et al., 2017), to typical micro CT data encountered in archaeological and evolutionary sciences. The Trainable Weka Segmentation is available through the FIJI fork of ImageJ (Schindelin et al., 2012). We demonstrate the efficacy of these algorithms as applied to six distinct examples: an entirely synthetic dataset; micro CT scans of a machine wire phantom; a defleshed mouse tibia; a lemur vocal tract; a juvenile Neanderthal humerus (Kostenki 2) and a small rodent animal mummy. These represent a range of the type of samples commonly encountered by researchers working in imaging in evolutionary sciences and each presents different segmentation challenges. To further demonstrate its efficacy, we compare this algorithm with other methods that have typically been used.

Materials and methods

Synthetic Dataset

A synthetic dataset of 12 images of white triangle outlines on a black background was made. The original was kept as the ground truth. To simulate partial volume averaging and scanner noise, the following filters were applied in ImageJ: Noise: Salt and Pepper; Gaussian Blur of radius 2.0 pixels; Shadow from south (base) of the image. The original data, the data with noise added, and, all segmentations are included in the supplementary material. A render is in figure 2 with an arbitrary voxel depth of 10.

MicroCT scans

A wire phantom object from (Dunmore et al., 2018) . This is a coil of randomly crunched stainless steel wire of thickness 4mm.

A wild type mouse proximal tibia and from (Ranzoni et al., 2018).

Rodent mummy Manchester Museum number 6033. This is thought to be a shrew, based upon size of the wrappings and earlier medical X-Rays (Adams, 2015).

Primate vocal tract-this is a scan of a wet preserved *Nycticebus pygmaeus* individual from the Duke Lemur Center, catalogue DLC_2901 and is more fully described in (Yapuncich et al., 2019)..

A partial proximal humerus from a juvenile Neanderthal from the site of Kiik-Koba. It has been described in detail by (Trinkaus et al., 2016) and has matrix and consolidant adhering to it which obscure some more detailed aspects of its morphology.

Full scan parameters are shown in Table 1 and volume renders of the tested datasets are shown in Figure 2.

Table 1. Scan parameters

Object	Reference number	Scanned with	KeV	µA	Voxel size/µm	Filter	Medium
Rodent mummy in wrappings	Manchester museum 6033	Nikon XTH225				Air	
Nycticebus pygmaeus vocal tract	Duke Lemur Center 2901	Nikon XTH225	155	110	35.47	None	Air
Kiik-Koba 2 juvenile neanderthal humerus	KunstKamera Museum	Custom MicroCT (microCT MPKT-01)	135	35	30	None	Air
Wild type mouse tibia	Number 6	Skyscan1172	49	200	5.06	Aluminium	70% Ethanol
10mm thickness machine wire artefact	n/a	SkyScan 1172	100	100	12.87	None	Air

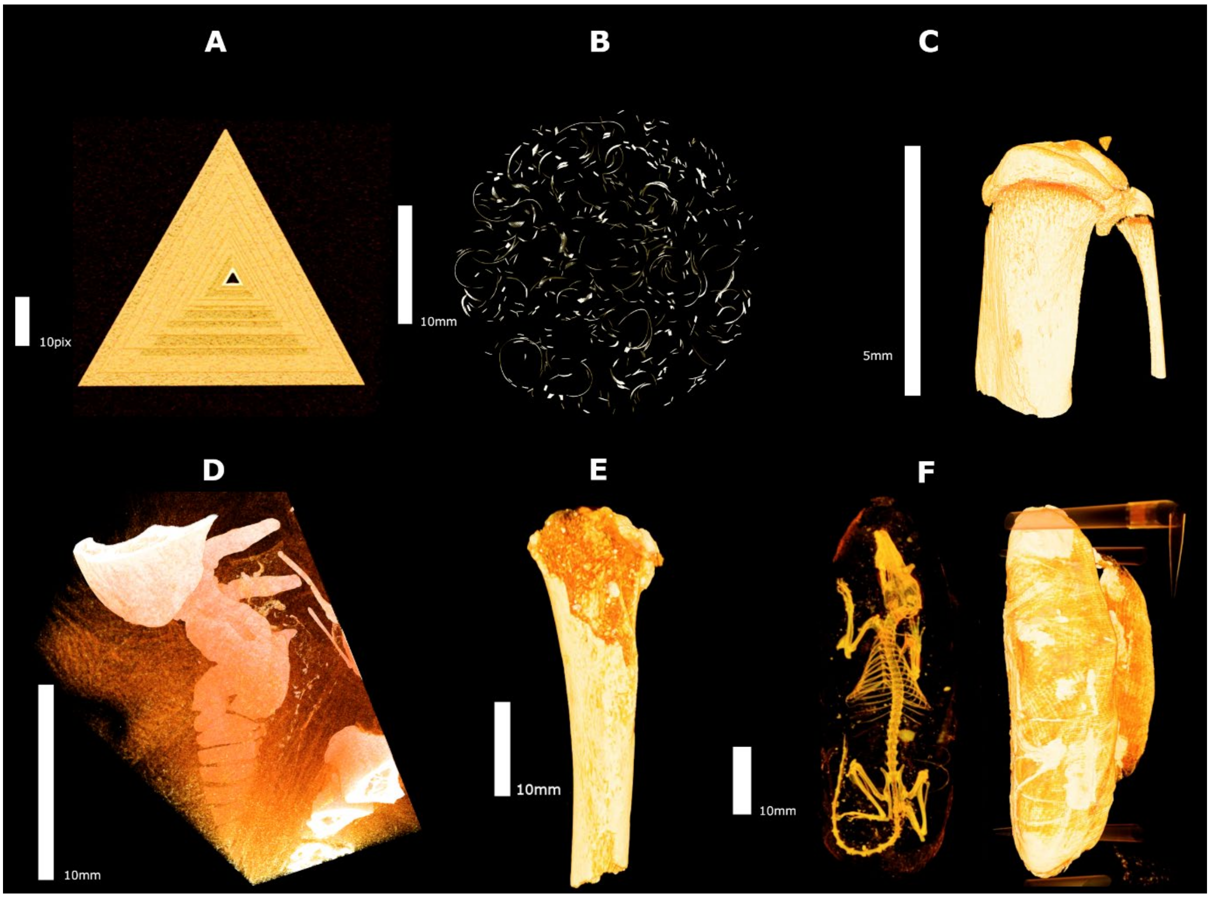

Figure 2. Volume renders of items analysed. A: Synthetic dataset; B: Machine Wire; C: Wild mouse type tibia section; D: Nycticebus pygmaeus vocal tract; E: Kiik Koba 2 partial humerus; E: Mummified rodent

Sample processing

All samples were subjected to segmentation using the Interactive Weka Segmentation editor plugin in ImageJ (Arganda-Carreras et al., 2017) with the following settings (adjusted after Somasundaram et al., 2018 who have applied this to medical images): Gaussian blur, Sobel Filter, Hessian Filter, Membrane projections (Thickness 1, patch size 10, difference of Gaussian filters, median filter (minimum sigma=1.0, maximum sigma=4.0); Kuwahara filter (Kuwahara et al., 1976). These filters help to counteract potential artefacts in the original scan slices and were found by (Somasundaram et al., 2018) to give the closest values to their 'gold standard' which was manual segmentation by a specialist.

Individual slices which contained all the materials of interest were trained using the Weka segmentation plugin. Briefly, areas containing each material of interest were selected using a graphics tablet and areas at the interface between materials were also selected (e.g. where a bone came into contact with air, near the edge of the bone was selected and added to the 'bone' label and a part near the edge of the air was selected and added to the 'air' label). This helped the algorithm to effectively select the correct labels at interfaces between materials. To propagate this label selection across the whole image, the Random Forest Algorithm was used, with 200 hundred trees. Although the use of fewer trees is more computationally efficient, the tradeoff between efficiency and efficacy starts to plateau after ~250 trees (Probst and Boulesteix, 2018). All images were then segmented using the appropriate training dataset.

All stacks were processed on of two machines with 32GB RAM, PCIeM2 SSD and either a 6 core i7 at 3.6GHz (4.2GHz at boost) or an 8 core AMD 2700 at 3.2 GHz (4Ghz at boost). Due to the way the Java virtual machine is configured, graphic card parameters are not currently relevant for this workflow.

All datasets were also in ImageJ segmented with the following competing algorithms:

- Greyscale thresholding (using the half-maximum-height algorithm)
- K-means segmentation
- C-means segmentation
- Localised fuzzy c-means segmentation, with pre-selection through k-means clustering

All variants of k-means and c-means segmentation used the ImageJ plugin available from https://github.com/arranger1044/SFCM. Spatial fuzzy c-means used the settings recommended by Dunmore et al. (2018).

For visual purposes, 3D renderings of each of the complete models were created using Avizo with no smoothing applied and the distances between the Weka segmentation and meshes generated with competing algorithms was visualised using the package Rvcg in R (Schlager, 2017).

Statistical comparisons

The effects of the varying segmentation algorithms on real world results is the most important consideration as it is sensible to anticipate that improvements will be made to the accuracy of these. In the case of the wire phantom and the tibia ROI it was also possible to compare average wire/trabecular thickness and thickness distribution of the samples (with the wire also having a ground truth thickness of 4mm). One further real-world test was a comparison of the ellipsoidness (after Salmon et al., 2015) and degree of anisotropy in the trabecular ROI, to demonstrate what effect the segmentation would have on biomechanical analyses. All thickness and anisotropy calculations were calculated with BoneJ 2 (Doube et al., 2010). We also assessed the degree of bone volume in the trabecular ROI, as many publications use this as a proxy for levels of bone formation in response to weight bearing or mechanical stimulation (e.g. Acquaah et al., 2015; Farooq et al., 2017; Li et al., 2016; Milovanovic et al., 2017; Turner, 2002).

Results

Segmentation times

Segmentation times for Weka segmentation are listed for representative individual slices from each stack in table 2.

Table 2. Segmentation times for each dataset

Scan	Training time/ms	Classification time/ms	Bit depth	Image x length	Image y length	No. materials
Synthetic dataset	15	0.5	8	814	814	2
Machine wire phantom	18	6.8	32	3240	3240	2
Mouse tibia	10	0.59	8	960	960	2
Lemur larynx	50	4	16	1329	1271	3
Kiik-Koba	49	0.3	8	480	576	3
Rodent mummy	50	1.5	8	1117	1141	3

Synthesised dataset

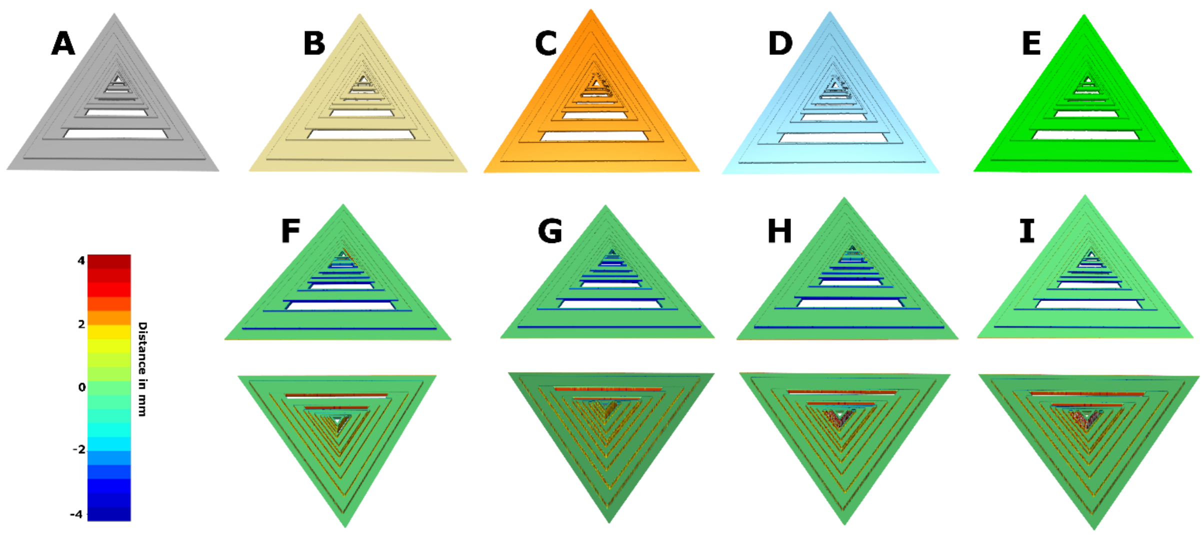

Figure 3. Synthesised data results. A: Original non-modified data; B: Weka Segmentation; C: local c-means; D: k-means; E: Watershed. F-I comparisons of above meshes with original data.

The majority of the data segmented relatively easily, but both k-means and local c-means struggled with the smaller triangles, where noise was closer to the dimensions of the object of interest.

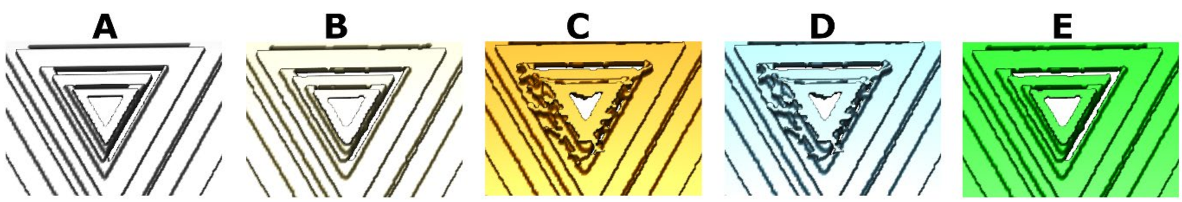

Figure 4. Closeups of the segmentations. Order as in Figure 3.

Wire phantom

The Weka segmentation performed as well as the local c-means segmentation and improved some aspects of fine detail retrieval (figure 5).

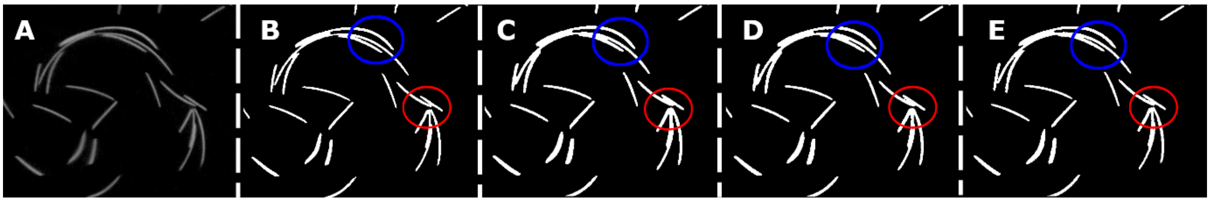

Figure 5. A: Original wire phantom scan; B: Weka segmentation. C: Local c-means segmentation; D: K-means segmentation; E: Watershed segmentation. Modified after Dunmore et al., (2018) Red circles indicate the areas highlighted by Dunmore et al., (2018) and blue are areas where Weka segmentation retrieves more fine detail.

Wild type mouse tibia

The Weka segmentation performed better than the other types of segmentation, with improved quality on fine features.

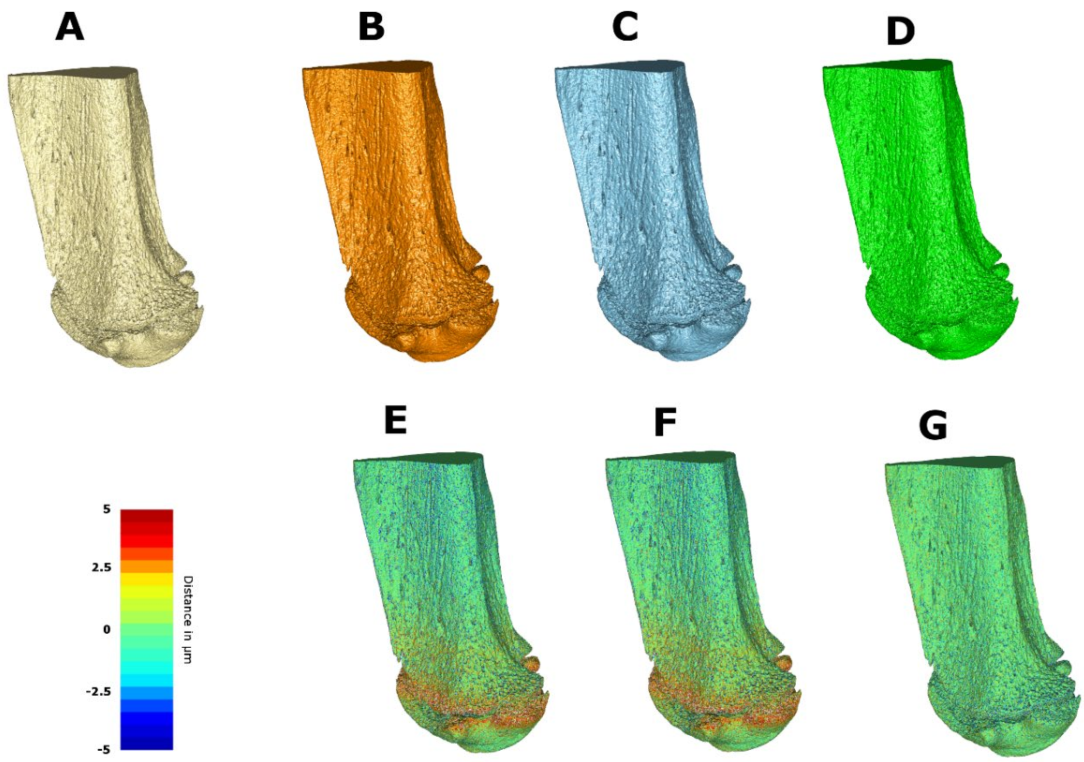

Figure 6. Comparisons of segmentations of the tibia. A: Weka; B: Local C-means; C: K-means; D: Watershed. E-G: heatmap comparisons of the above specimens with Weka segmentation. Blue to red scale, Blue indicates values which are concave compared with Weka; Red indicates areas that are concave compared with Weka.

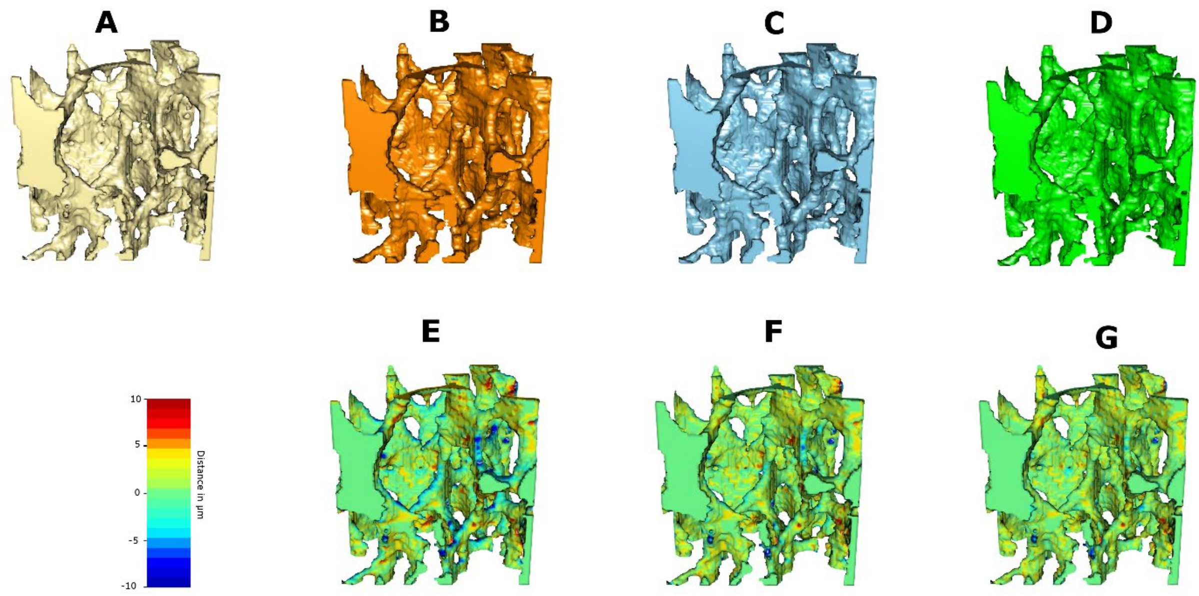

Figure 7. Comparisons of segmentations of ROI. A: Weka; B: Local C-means; C: K-means; D: Watershed. E-G: heatmap comparisons of the above specimens with Weka segmentation. Blue to red scale, Blue indicates values which are concave compared with Weka; Red indicates areas that are concave compared with Weka.

Violin plots indicate that alternative segmentation methods have subtly different distributions in terms of distance from the Weka segmentation. All suffer from arbitrary spiking in the data.

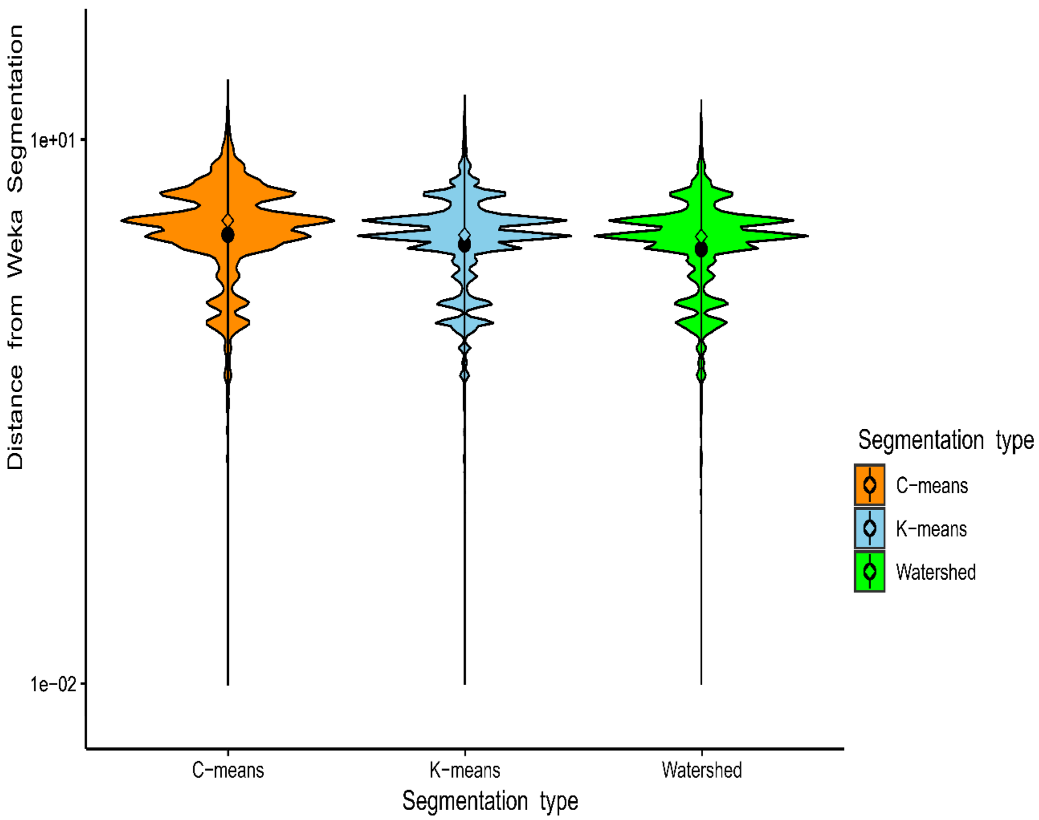

Figure 8. Violin plots of mesh distances. Cut off at 1e^{-2} to illustrate the main trends in the data

It is also apparent that differing segmentation techniques have a marked effect on the degree of anisotropy detected in trabecular bone, with Weka tending towards more anisotropic structures. This may be because of the lack of spiking in the resulting segmentation when compared with the other methods here. The Ellipsoid Factor (a replacement for the Structure model index (Doube, 2015; Salmon et al., 2015) also varies considerably, with a difference of almost 4% between Weka and watershed segmentation. It is noticeable also that Weka segmentation classifies a relatively low percentage of bone and also trabecular thickness.

Table 3 Comparison of the morphometric measures of the tibia ROI

Segmentation technique	Degree of anisotropy	% of foreground volume filled with ellipsoids	% of ROI classified as bone	Tb thickness/µm	Tb. Th S.D. /µm
Weka	0.44	14.32	12.7	35.1	12.7
Local c-means binarized	0.38	14.39	15.1	39.2	12.6
K-means	0.39	16.52	15.2	40.9	13.2
Conventional watershed	0.41	18.28	15.3	41.0	13.2

Lemur scan

The Weka segmentation was able to account for the ring artefacts in the scan and successfully segmented the materials of interest. It was also more successful at segmenting the finer structures in the larynx (see Figure 8). It also generated much cleaner data than all other segmentations.

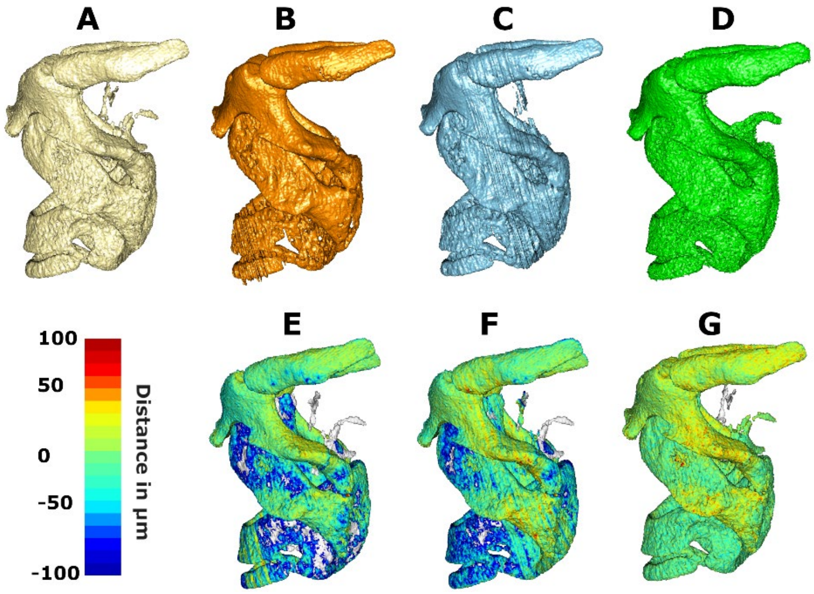

Figure 9. Comparison of Nycticebus pygmaeus scan segmentations. A: Weka; B: Local C-means; C: K-means; D: Watershed. E-G: heatmap comparisons of the above specimens with Weka segmentation. Blue to red scale, Blue indicates values which are concave compared with Weka; Red indicates areas that are concave compared with Weka.

Kiik-Koba Neanderthal humerus

The Weka segmentation was able to track trabecular structure successfully, without eroding the material. It also was able to take into account the slight 'halo' effect on the bone/air interface, which conventional segmentation used to create an external border of the matrix material. The c-means and k-means segmentation both created this 'halo' like border (Figure 11).

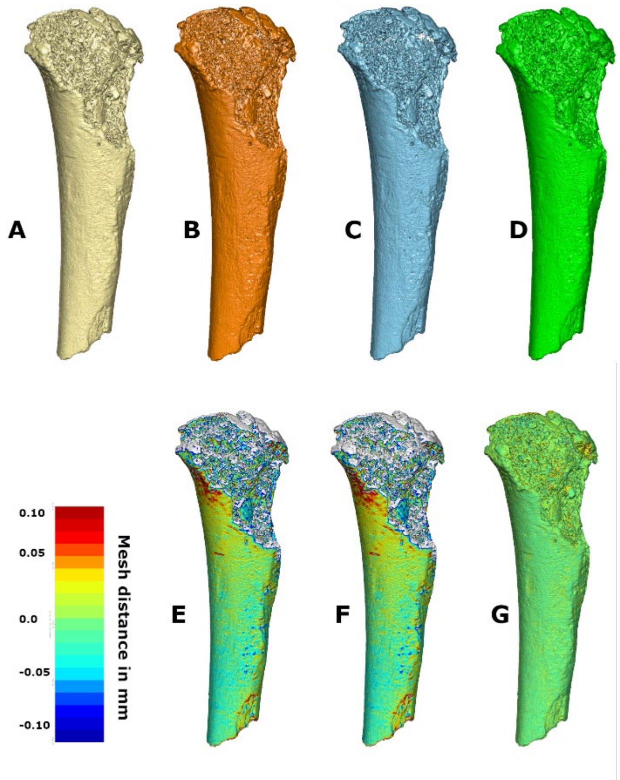

Figure 10. Kiik-Koba 2 reconstructions.

A: Weka; B: Local C-means; C: K-means; D: Watershed. E-G: heatmap comparisons of the above specimens with Weka segmentation. Blue to red scale, Blue indicates values which are concave compared with Weka; Red indicates areas that are concave compared with Weka.

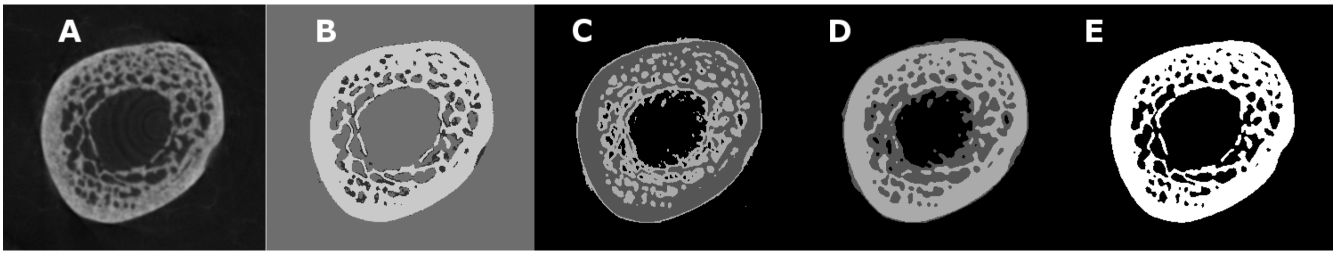

Figure 11. Orthoslice from Kiik-Koba microCT scan. A: Original scan; B: Weka; C: Local C-means; D: K-means; W: Watershed

Rodent Mummy

The Weka segmentation was able to detect the majority of the skeletal features and also discriminate the mummified tissues from the outer wrappings. There were a few artefacts around the front paws and mandible which would require some manual correction to fully delineate the structures. The smoothing steps introduced in the segmentation were able to remove many of the scanning artefacts which made borders of materials harder to resolve with conventional methods. K-means segmentation was able to discriminate materials relatively well also, although it did misclassify several slices. Localised C-means was unsuccessful in several slices, missing bone material entirely, where other segmentation techniques succeeded.

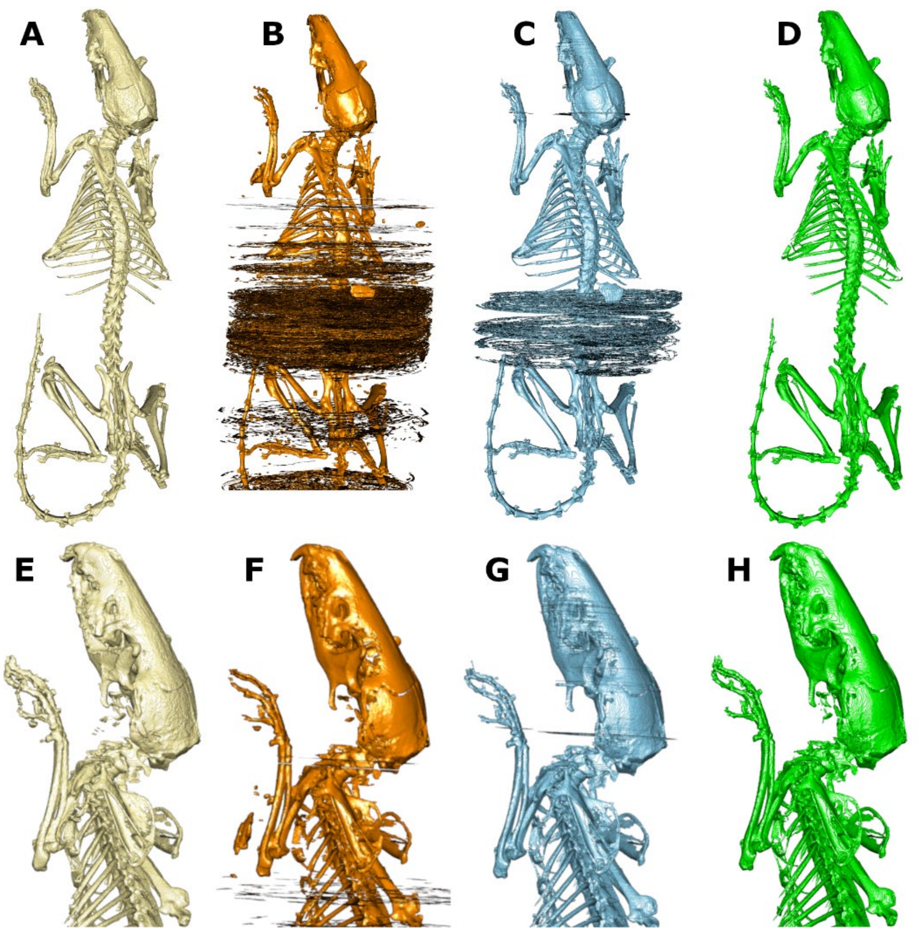

Figure 12. Rodent mummy 3D model reconstructions. Order as before. Second row are closeups showing extra detail

Discussion

The Weka segmentation algorithm generated improved results when compared with standard tools available in Avizo and was especially good at tracking fine structures throughout the samples. It is interesting that all algorithms apart from Weka seem to struggle to differentiate between materials as contrast is low and details approach the resolution of the scans (e.g. they struggle with details that are around 10 microns in size, when the scan is 5 microns in resolution). Although it did not perfectly segment out the bone in the rodent mummy sample, it is a noted improvement on current methods, and is extremely easy to implement. It compared favourably with the MCIA data and has the added advantage that ImageJ is widely available, easy to learn and platform independent. It also has the advantage over the MCIA algorithm in that it is able to compensate for the material interface artefacts in scans (Dunmore et al., 2018). A suggestion for further processing of data from especially noisy scan data is that the user applies the 'despeckle' filter in imageJ, either on the original data, or the resulting segmentation.

Our results with the trabecular ROI add to those of Verdelis et al., (2011), who caution that inter system microCT results at high resolutions may not necessarily be entirely comparable. We would add an extra caution to this and would suggest that segmentation choice should be explicitly referred to as this can also have a large effect on results. It also appears to contradict the results of Christiansen, (2016) who found that different watershed methods did not really affect results at high resolutions. This is definitely an area which warrants more investigation.

The Weka algorithms can be applied to a wide range of image types, as it was originally developed for microscopy (Arganda-Carreras et al., 2017). We have tested this on image datasets from 8 to 32-bit depth. Given workflow constraints, most images used in everyday analysis will be either 8 or 16-bit. DICOM data requires the conversion to TIFF or other standard image formats to processing with Weka. ImageJ/Weka segmentation is multi-platform and has a user-friendly GUI. This make it an ideal toolbox to teach researchers (who may be unfamiliar with the subtleties of image processing) a fast and free way to process their CT data. Key parameters t observe are to use a fast CPU with multiple cores, which will enable users to fully leverage multi-threading; as well as the use of fast hard drives (preferably Solid State Drives) if working on a desktop. Training the segmentation using fine structures will also improve delineation of edge features. Finally, the use of a graphics tablet is also recommended.

A major disadvantage currently is that the Weka algorithms are extremely CPU intensive but in ImageJ, do not utilise the GPU. K-means and fuzzy c-means algorithms are also extremely CPU intensive, regardless of if they are written in ImageJ or Matlab. Interestingly, the MCIA algorithms are very RAM intensive (Dunmore, pers. comm.). Implementation of the Weka algorithm, either through virtualised clusters (e.g. FIJI archipelago,) or through GPU optimisation (either through CLIJ (Haase et al., 2019) or Matlab bridging) may work to ameliorate bottlenecks in processing speed somewhat. The yield in minimising user time in segmentation (as once trained, segmentation can process independently of the user) does however make the current implementation an ideal approach for the first pass segmentation of structures in archaeological and evolutionary studies.

Localised C-means segmentation in both ImageJ and Matlab also tend to flip the order of labels in some images, which then necessitates further steps of interleaving different stacks to obtain one segmentation. This is probably fairly straightforward to address by a forcing of order of labels in the algorithm but is beyond the scope of this paper.

Conclusions

For the first time, we have presented the implementation of the Weka Machine learning library to archaeological and palaeontological material. It yields results that are the equal of leading edge-based methods and superior to conventional segmentations produced by commercial packages. The implementation of Weka segmentation is fast, with no software cost to the end user and it enables an easy introduction to both image segmentation and machine learning for the inexperienced user. Future work will seek to apply this algorithm to larger and more varied samples, as well as exploring the possibility of increasing the speed of computation, either through GPU based acceleration or use of virtual clusters.

Acknowledgements

We would like to thank the curators of the collections from which scan material was used for this paper, including Campell Price and Sam Sportun (Manchester Museum-acces to mummifieid samples); Dr.V.Khartanovich from Kunstkamera, Russian Academy of Sciences (access to Kiik Koba 2); The Duke Lemur Center provided access to the *Otolemur garnetti* data , originally appearing in Yapuncich et al. (2019), the collection of which was funded by NSF BCS 1540421 to Gabriel S. Yapuncich and Doug M. Boyer. The files were downloaded from www.MorphoSource.org, Duke University. The animal mummy was scanned at the Henry Moseley X-ray Imaging Facility, University of Manchester. We are also grateful to Christopher Dunmore for discussion of the MCIA algorithms and demands that they place on computer hardware. William Harcourt-Smith made the very helpful suggestion of including the anisotropy analyses in our comparisons and Charlotte Brassey kindly gave very helpful feedback on an earlier draft of the paper.

Funding

Thomas O'Mahoney and Jake Dunn were both funded by the Royal Society: (RSG\R1\180340); Rhinology and Laryngology Research Fund and Anglia Ruskin University. Lidija McKnight was funded by the Leverhulme Trust (RPG-2013-143) and the Arts and Humanities Research Council (AH/P005047/1). We also acknowledge the Engineering and Physical Science Research Council (EPSRC) for funding the Henry Moseley X-ray Imaging Facility which has been made available through the Royce Institute for Advanced Materials through grants (EP/F007906/1, EP/F001452/1,EP/I02249X, EP/M010619/1, EP/F028431/1, EP/M022498/1 and EP/R00661X/1).

References

Abel, R., Laurini, C., Richter, M., 2012. A palaeobiologist's guide to "virtual" micro-CT preparation. Palaeontol. Electron. https://doi.org/10.26879/284

Acquaah, F., Brown, R., A, K., Ahmed, F., Jeffery, N., Abel, R.L., 2015. Early Trabecular Development in Human Vertebrae: Overproduction, Constructive Regression, and Refinement. Front. Endocrinol. 6. https://doi.org/10.3389/fendo.2015.00067

Anné, J., Garwood, R.J., Lowe, T., Withers, P.J., Manning, P.L., 2015. Interpreting pathologies in extant and extinct archosaurs using micro-CT. PeerJ 3, e1130. https://doi.org/10.7717/peerj.1130

Arganda-Carreras, I., Kaynig, V., Rueden, C., Eliceiri, K.W., Schindelin, J., Cardona, A., Sebastian Seung, H., 2017. Trainable Weka Segmentation: a machine learning tool for microscopy pixel classification. Bioinforma. Oxf. Engl. 33, 2424–2426. https://doi.org/10.1093/bioinformatics/btx180

Barron, A., Turner, M., Beeching, L., Bellwood, P., Piper, P., Grono, E., Jones, R., Oxenham, M., Kien, N.K.T., Senden, T., Denham, T., 2017. MicroCT reveals domesticated rice (Oryza sativa) within pottery sherds from early Neolithic sites (4150–3265 cal BP) in Southeast Asia. Sci. Rep. 7, 7410. https://doi.org/10.1038/s41598-017-04338-9

Bello, S.M., De Groote, I., Delbarre, G., 2013. Application of 3-dimensional microscopy and micro-CT scanning to the analysis of Magdalenian portable art on bone and antler. J. Archaeol. Sci. 40, 2464–2476. https://doi.org/10.1016/j.jas.2012.12.016

Berg, S., Saxena, N., Shaik, M., Pradhan, C., 2018. Generation of ground truth images to validate micro-CT image-processing pipelines. Lead. Edge 37, 412–420. https://doi.org/10.1190/tle37060412.1

Bezdek, J.C., 1980. A convergence theorem for the fuzzy ISODATA clustering algorithms. IEEE Trans. Pattern Anal. Mach. Intell. 1–8.

Bezdek, J.C., 1975. Mathematical models for systematics and taxonomy, in: Estabrook, G. (Ed.), Proceedings of the 8th International Conference on Numerical Taxonomy. Freeman company, San Fransicso, pp. 143–166.

Borgard, H.L., Baab, K., Pasch, B., Riede, T., 2019. The Shape of Sound: a Geometric Morphometrics Approach to Laryngeal Functional Morphology. J. Mamm. Evol. https://doi.org/10.1007/s10914-019-09466-9

Brassey, C.A., Gardiner, J.D., Kitchener, A.C., 2018. Testing hypotheses for the function of the carnivoran baculum using finite-element analysis. Proc. R. Soc. B Biol. Sci. 285, 20181473. https://doi.org/10.1098/rspb.2018.1473

Brassey, C.A., Margetts, L., Kitchener, A.C., Withers, P.J., Manning, P.L., Sellers, W.I., 2013. Finite element modelling versus classic beam theory: comparing methods for stress estimation in a morphologically diverse sample of vertebrate long bones. J. R. Soc. Interface 10, 20120823. https://doi.org/10.1098/rsif.2012.0823

Breiman, L., 2001. Random Forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324

Brocklehurst, R., Porro, L., Herrel, A., Adriaens, D., Rayfield, E., 2019. A digital dissection of two teleost fishes: comparative functional anatomy of the cranial musculoskeletal system in pike (*Esox lucius*) and eel (*Anguilla anguilla*). J. Anat. joa.13007. https://doi.org/10.1111/joa.13007

Chapelle, K.E.J., Barrett, P.M., Botha, J., Choiniere, J.N., 2019. Ngwevu intloko: a new early sauropodomorph dinosaur from the Lower Jurassic Elliot Formation of South Africa and comments on cranial ontogeny in Massospondylus carinatus. PeerJ 7, e7240. https://doi.org/10.7717/peerj.7240

Charlier, P., Wils, P., Froment, A., Huynh-Charlier, I., 2014. Arterial calcifications from mummified materials: use of micro-CT-scan for histological differential diagnosis. Forensic Sci. Med. Pathol. 10, 461–465. https://doi.org/10.1007/s12024-014-9544-9

Christiansen, B.A., 2016. Effect of micro-computed tomography voxel size and segmentation method on trabecular bone microstructure measures in mice. Bone Rep. 5, 136–140. https://doi.org/10.1016/j.bonr.2016.05.006

Collinson, M.E., Adams, N.F., Manchester, S.R., Stull, G.W., Herrera, F., Smith, S.Y., Andrew, M.J., Kenrick, P., Sykes, D., 2016. X-ray micro-computed tomography (micro-CT) of pyrite-permineralized fruits and seeds from the London Clay Formation (Ypresian) conserved in silicone oil: a critical evaluation. Botany 94, 697–711. https://doi.org/10.1139/cjb-2016-0078

Cuff, A.R., Bright, J.A., Rayfield, E.J., 2015. Validation experiments on finite element models of an ostrich (Struthio camelus) cranium. PeerJ 3, e1294. https://doi.org/10.7717/peerj.1294

Dimitriadou, E., Barth, M., Windischberger, C., Hornik, K., Moser, E., 2004. A quantitative comparison of functional MRI cluster analysis. Artif. Intell. Med. 31, 57–71. https://doi.org/10.1016/j.artmed.2004.01.010

Doube, M., 2015. The Ellipsoid Factor for Quantification of Rods, Plates, and Intermediate Forms in 3D Geometries. Front. Endocrinol. 6. https://doi.org/10.3389/fendo.2015.00015

Doube, M., Kłosowski, M.M., Arganda-Carreras, I., Cordelières, F.P., Dougherty, R.P., Jackson, J.S., Schmid, B., Hutchinson, J.R., Shefelbine, S.J., 2010. BoneJ: Free and extensible bone image analysis in ImageJ. Bone 47, 1076–1079. https://doi.org/10.1016/j.bone.2010.08.023

Du Plessis, A., Slabbert, R., Swanepoel, L.C., Els, J., Booysen, G.J., Ikram, S., Cornelius, I., 2015. Three-dimensional model of an ancient Egyptian falcon mummy skeleton. Rapid Prototyp. J. 21, 368–372. https://doi.org/10.1108/RPJ-09-2013-0089

Dunmore, C.J., Wollny, G., Skinner, M.M., 2018. MIA-Clustering: a novel method for segmentation of paleontological material. PeerJ 6, e4374. https://doi.org/10.7717/peerj.4374

Farooq, S., Leussink, S., Sparrow, L.M., Marchini, M., Britz, H.M., Manske, S.L., Rolian, C., 2017. Cortical and trabecular morphology is altered in the limb bones of mice artificially selected for faster skeletal growth. Sci. Rep. 7, 10527. https://doi.org/10.1038/s41598-017-10317-x

Forgy, E.W., 1965. Cluster analysis of multivariate data: efficiency versus interpretability of classifications. biometrics 21, 768–769.

Garwood, R.J., Dunlop, J., 2014. Three-dimensional reconstruction and the phylogeny of extinct chelicerate orders. PeerJ 2, e641. https://doi.org/10.7717/peerj.641

Glocker, B., Zikic, D., Konukoglu, E., Haynor, D.R., Criminisi, A., 2013. Vertebrae Localization in Pathological Spine CT via Dense Classification from Sparse Annotations, in: Salinesi, C., Norrie, M.C., Pastor, Ó. (Eds.), Advanced Information Systems Engineering. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 262–270. https://doi.org/10.1007/978-3-642-40763-5_33

Greenspan, H., van Ginneken, B., Summers, R.M., 2016. Guest Editorial Deep Learning in Medical Imaging: Overview and Future Promise of an Exciting New Technique. IEEE Trans. Med. Imaging 35, 1153–1159. https://doi.org/10.1109/TMI.2016.2553401

Gunz, P., Neubauer, S., Golovanova, L., Doronichev, V., Maureille, B., Hublin, J.-J., 2012. A uniquely modern human pattern of endocranial development. Insights from a new cranial reconstruction of the Neandertal newborn from Mezmaiskaya. J. Hum. Evol. 62, 300–313. https://doi.org/10.1016/j.jhevol.2011.11.013

Haase, R., Royer, L.A., Steinbach, P., Schmidt, D., Dibrov, A., Schmidt, U., Weigert, M., Maghelli, N., Tomancak, P., Jug, F., Myers, E.W., 2019. CLIJ: Enabling GPU-accelerated image processing in Fiji (preprint). Bioinformatics. https://doi.org/10.1101/660704

Hechenleitner, E.M., Grellet-Tinner, G., Foley, M., Fiorelli, L.E., Thompson, M.B., 2016. Micro-CT scan reveals an unexpected high-volume and interconnected pore network in a Cretaceous Sanagasta dinosaur eggshell. J. R. Soc. Interface 13, 20160008. https://doi.org/10.1098/rsif.2016.0008

Hershkovitz, I., Marder, O., Ayalon, A., Bar-Matthews, M., Yasur, G., Boaretto, E., Caracuta, V., Alex, B., Frumkin, A., Goder-Goldberger, M., Gunz, P., Holloway, R.L., Latimer, B., Lavi, R., Matthews, A., Slon, V., Mayer, D.B.-Y., Berna, F., Bar-Oz, G., Yeshurun, R., May, H., Hans, M.G., Weber, G.W., Barzilai, O., 2015. Levantine cranium from Manot Cave (Israel) foreshadows the first European modern humans. Nature 520, 216–219. https://doi.org/10.1038/nature14134

Hershkovitz, I., Weber, G.W., Quam, R., Duval, M., Grün, R., Kinsley, L., Ayalon, A., Bar-Matthews, M., Valladas, H., Mercier, N., Arsuaga, J.L., Martinón-Torres, M., Bermúdez de Castro, J.M., Fornai, C., Martín-Francés, L., Sarig, R., May, H., Krenn, V.A., Slon, V., Rodríguez, L., García, R., Lorenzo, C., Carretero, J.M., Frumkin, A., Shahack-Gross, R., Bar-Yosef Mayer, D.E., Cui, Y., Wu, X., Peled, N., Groman-Yaroslavski, I., Weissbrod, L., Yeshurun, R., Tsatskin, A., Zaidner, Y., Weinstein-Evron, M., 2018. The earliest modern humans outside Africa. Science 359, 456–459. https://doi.org/10.1126/science.aap8369

Hublin, J.-J., Ben-Ncer, A., Bailey, S.E., Freidline, S.E., Neubauer, S., Skinner, M.M., Bergmann, I., Le Cabec, A., Benazzi, S., Harvati, K., 2017. New fossils from Jebel Irhoud, Morocco and the pan-African origin of Homo sapiens. Nature 546, 289.

Juang, L.-H., Wu, M.-N., 2010. MRI brain lesion image detection based on color-converted K-means clustering segmentation. Measurement 43, 941–949. https://doi.org/10.1016/j.measurement.2010.03.013

Kuwahara, M., Hachimura, K., Eiho, S., Kinoshita, M., 1976. Processing of RI-Angiocardiographic Images, in: Preston, K., Onoe, M. (Eds.), Digital Processing of Biomedical Images. Springer US, Boston, MA, pp. 187–202. https://doi.org/10.1007/978-1-4684-0769-3_13

Laloy, F., Rage, J.-C., Evans, S.E., Boistel, R., Lenoir, N., Laurin, M., 2013. A Re-Interpretation of the Eocene Anuran Thaumastosaurus Based on MicroCT Examination of a ‘Mummified’ Specimen. PLoS ONE 8, e74874. https://doi.org/10.1371/journal.pone.0074874

Li, B., Sankaran, J.S., Judex, S., 2016. Trabecular and Cortical Bone of Growing C3H Mice Is Highly Responsive to the Removal of Weightbearing. PLOS ONE 11, e0156222. https://doi.org/10.1371/journal.pone.0156222

Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A.W.M., van Ginneken, B., Sánchez, C.I., 2017. A survey on deep learning in medical image analysis. Med. Image Anal. 42, 60–88. https://doi.org/10.1016/j.media.2017.07.005

MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations, in: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. Oakland, CA, USA, pp. 281–297.

Marshall, A.F., Bardua, C., Gower, D.J., Wilkinson, M., Sherratt, E., Goswami, A., 2019. High-density three-dimensional morphometric analyses support conserved static (intraspecific) modularity in caecilian (Amphibia: Gymnophiona) crania. Biol. J. Linn. Soc. 126, 721–742. https://doi.org/10.1093/biolinnean/blz001

McBride, R.A., Mercer, G.D., 2012. Assessing Damage to Archaeological Artefacts in Compacted Soil Using Microcomputed Tomography Scanning: CT Scans of Damaged Artefacts in Soil. Archaeol. Prospect. 19, 7–19. https://doi.org/10.1002/arp.426

Milovanovic, P., Djonic, D., Hahn, M., Amling, M., Busse, B., Djuric, M., 2017. Region-dependent patterns of trabecular bone growth in the human proximal femur: A study of 3D bone microarchitecture from early postnatal to late childhood period: MILOVANOVIC ET AL . Am. J. Phys. Anthropol. 164, 281–291. https://doi.org/10.1002/ajpa.23268

Nguyen, A., Yosinski, J., Clune, J., 2015. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Boston, MA, USA, pp. 427–436. https://doi.org/10.1109/CVPR.2015.7298640

Odes, E.J., Randolph-Quinney, P.S., Steyn, M., Throckmorton, Z., Smilg, J.S., Zipfel, B., Augustine, T.N., Beer, F. de, Hoffman, J.W., Franklin, R.D., Berger, L.R., 2016. Earliest hominin cancer: 1.7-million-year-old osteosarcoma from Swartkrans Cave, South Africa. South Afr. J. Sci. 112, 5–5. https://doi.org/10.17159/sajs.2016/20150471

Pham, D.L., Prince, J.L., 1999. An adaptive fuzzy C-means algorithm for image segmentation in the presence of intensity inhomogeneities. Pattern Recognit. Lett. 20, 57–68.

Polly, P.D., Stayton, C.T., Dumont, E.R., Pierce, S.E., Rayfield, E.J., Angielczyk, K.D., 2016. Combining geometric morphometrics and finite element analysis with evolutionary modeling: towards a synthesis. J. Vertebr. Paleontol. 36, e1111225. https://doi.org/10.1080/02724634.2016.1111225

Prasoon, A., Petersen, K., Igel, C., Lauze, F., Dam, E., Nielsen, M., 2013. Deep Feature Learning for Knee Cartilage Segmentation Using a Triplanar Convolutional Neural Network, in: Salinesi, C., Norrie, M.C., Pastor, Ó. (Eds.), Advanced Information Systems Engineering. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 246–253. https://doi.org/10.1007/978-3-642-40763-5_31

Probst, P., Boulesteix, A.-L., 2018. To Tune or Not to Tune the Number of Trees in Random Forest. J. Mach. Learn. Res. 18, 1–18.

Radford, A., Metz, L., Chintala, S., 2015. Unsupervised representation learning with deep convolutional generative adversarial networks. ArXiv Prepr. ArXiv151106434.

Randolph-Quinney, P.S., Williams, S.A., Steyn, M., Meyer, M.R., Smilg, J.S., Churchill, S.E., Odes, E.J., Augustine, T., Tafforeau, P., Berger, L.R., 2016. Osteogenic tumour in Australopithecus sediba: Earliest hominin evidence for neoplastic disease. South Afr. J. Sci. 112, 1–7. https://doi.org/10.17159/sajs.2016/20150470

Ranzoni, A.M., Corcelli, M., Arnett, T.R., Guillot, P.V., 2018. Micro-computed tomography reconstructions of tibiae of stem cell transplanted osteogenesis imperfecta mice. Sci. Data 5, 180100. https://doi.org/10.1038/sdata.2018.100

Romell, J., Vågberg, W., Romell, M., Häggman, S., Ikram, S., Hertz, H.M., 2018. Soft-Tissue Imaging in a Human Mummy: Propagation-based Phase-Contrast CT. Radiology 289, 670–676. https://doi.org/10.1148/radiol.2018180945

Salmon, P.L., Ohlsson, C., Shefelbine, S.J., Doube, M., 2015. Structure Model Index Does Not Measure Rods and Plates in Trabecular Bone. Front. Endocrinol. 6. https://doi.org/10.3389/fendo.2015.00162

Scherf, H., Tilgner, R., 2009. A new high-resolution computed tomography (CT) segmentation method for trabecular bone architectural analysis. Am. J. Phys. Anthropol. 140, 39–51. https://doi.org/10.1002/ajpa.21033

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D.J., Hartenstein, V., Eliceiri, K., Tomancak, P., Cardona, A., 2012. Fiji: an open-source platform for biological-image analysis. Nat. Methods 9, 676–682. https://doi.org/10.1038/nmeth.2019

Schlager, S., 2017. Morpho and Rvcg - Shape Analysis in R, in: Zheng, G., Li, S., Székely, G. (Eds.), Statistical Shape and Deformation Analysis: Methods, Implementation and Applications, Computer Vision and Pattern Recognition Series. Academic Press, London, pp. 217–256.

Shen, D., Wu, G., Suk, H.-I., 2017. Deep Learning in Medical Image Analysis. Annu. Rev. Biomed. Eng. 19, 221–248. https://doi.org/10.1146/annurev-bioeng-071516-044442

Singh, M., Patel, P., Khosla, D., Kim, T., 1996. Segmentation of functional MRI by K-means clustering. IEEE Trans. Nucl. Sci. 43, 2030–2036. https://doi.org/10.1109/23.507264

Somasundaram, E., Deaton, J., Kaufman, R., Brady, S., 2018. Fully Automated Tissue Classifier for Contrast-enhanced CT Scans of Adult and Pediatric Patients. Phys. Med. Biol. 63, 135009. https://doi.org/10.1088/1361-6560/aac944

Spoor, C.F., Zonneveld, F.W., Macho, G.A., 1993. Linear measurements of cortical bone and dental enamel by computed tomography: Applications and problems. Am. J. Phys. Anthropol. 91, 469–484. https://doi.org/10.1002/ajpa.1330910405

Suzani, A., Seitel, A., Liu, Y., Fels, S., Rohling, R.N., Abolmaesumi, P., 2015. Fast Automatic Vertebrae Detection and Localization in Pathological CT Scans - A Deep Learning Approach, in: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Springer International Publishing, Cham, pp. 678–686. https://doi.org/10.1007/978-3-319-24574-4_81

Suzuki, K., 2017. Overview of deep learning in medical imaging. Radiol. Phys. Technol. 10, 257–273. https://doi.org/10.1007/s12194-017-0406-5

Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., Fergus, R., 2013. Intriguing properties of neural networks. ArXiv13126199 Cs.

Tin Kam Ho, 1998. The random subspace method for constructing decision forests. IEEE Trans. Pattern Anal. Mach. Intell. 20, 832–844. https://doi.org/10.1109/34.709601

Trinkaus, E., Mednikova, M.B., Cowgill, L.W., 2016. The Appendicular Remains of the Kiik-Koba 2 Neandertal Infant. PaleoAnthropology 185, 210.

Tuniz, C., Zanini, F., 2018. Microcomputerized Tomography (MicroCT) in Archaeology, in: Encyclopedia of Global Archaeology. Springer International Publishing, Cham, pp. 1–7. https://doi.org/10.1007/978-3-319-51726-1_675-2

Turner, C.H., 2002. Biomechanics of Bone: Determinants of Skeletal Fragility and Bone Quality. Osteoporos. Int. 13, 97–104. https://doi.org/10.1007/s001980200000

Verdelis, K., Lukashova, L., Atti, E., Mayer-Kuckuk, P., Peterson, M.G.E., Tetradis, S., Boskey, A.L., van der Meulen, M.C.H., 2011. MicroCT morphometry analysis of mouse cancellous bone: intra- and inter-system reproducibility. Bone 49, 580–587. https://doi.org/10.1016/j.bone.2011.05.013

Wacey, D., Battison, L., Garwood, R.J., Hickman-Lewis, K., Brasier, M.D., 2017. Advanced analytical techniques for studying the morphology and chemistry of Proterozoic microfossils. Geol. Soc. Lond. Spec. Publ. 448, 81–104. https://doi.org/10.1144/SP448.4

Wang, G., 2016. A Perspective on Deep Imaging. IEEE Access 4, 8914–8924. https://doi.org/10.1109/ACCESS.2016.2624938

Yapuncich, G.S., Kemp, A.D., Griffith, D.M., Gladman, J.T., Ehmke, E., Boyer, D.M., 2019. A digital collection of rare and endangered lemurs and other primates from the Duke Lemur Center (preprint). Zoology. https://doi.org/10.1101/688739