

An artificial intelligence approach to classify archaeological wood charcoals: introduction of a case study with southern African Prehistoric charcoals

PUECH Elysandre^{1,2}, COLI Vanna Lisa¹, CORNELI Marco^{1,3}, LE ROUX¹ Quentin,
THERY-PARISOT Isabelle¹

¹Université Côte D'Azur, CNRS, CEPAM, UMR 7264, Nice, France

²Evolutionary Studies Institute, University of the Witwatersrand, Johannesburg, 2050,
South Africa

³INRIA, Maasai, CNRS, LJAD UMR 7351, Nice, France

elysandre.puech@cepam.cnrs.fr

Key words: Archaeological Wood Charcoal, Artificial Intelligence, 2D images, Deep Neural Networks, Southern Africa

Context and objectives

Wood anatomy of species-rich biomes

The study of archaeological wood charcoal remains allows documenting past vegetation, as well as fuelwood management of past societies (*e.g.* Asouti and Austin 2005). The structure of the wood, characteristic of the taxa, is preserved after carbonization and the methods of identification of charcoal fragments are based on descriptions of wood anatomy observed with microscope instruments. The taxonomic identification of charcoal is an inherent step in any anthracological analysis, but it can be challenging in some parts of the world with rich woody species, as tropical or subtropical areas. A combination of several anatomical criteria, or features, is necessary to describe a charcoal specimen (Wheeler *et al.* 1989). Regions with rich woody species are characterised by a high level of similar morphological characteristics that often limit wood identification at the family level, especially when archaeological charcoals are not well preserved, or very fragmented (*e.g.* Höhn and Neumann 2018).

Example of southern African wood charcoals

Charcoal remains are remarkably preserved and abundant in the archaeological site of Bushman Rock Shelter (BRS) located in the Limpopo basin in South Africa (Porraz *et al.*, 2015). Their analyses in the Later Stone Age layers allow, in particular, documenting human behaviour and adaptation regarding wood use as a fuel during the critical climatic period of the Late Pleistocene / early Holocene transition *ca.* 11 700 yrs. cal. BP (Fig. 1, Puech *et al.* 2021).

In the southern African interior plateau in central Limpopo, the richness of the native woody plants is estimated to include more than 950 ligneous species for the area (Thuiller *et al.* 2006, Schmidt *et al.* 2007). The diverse climatic regimes and landforms that characterize the region also contribute to the high *inter*- and *intra*-specific variability in wood anatomy, intrinsically linked to the physiological conditions of tree growth. Therefore, the accuracy of identification and reliability of reconstruction of past woody vegetation depend on the degree of knowledge of the anatomy of the taxa and the ecology of the current vegetation. However, there is currently no comprehensive atlas or identification key for the study area that takes into account both the wide diversity of woody species and their anatomical variability. Following standard protocol, new specimens of actual wood branches were collected to complete the existing comparison charcoal collection for the study area (Puech 2022). For each specimen sampled, systematic

anatomical descriptions and micrographs were processed with optical reflected light microscope as well as scanning electron microscope (SEM).



Fig. 1: Location, landscape and stratigraphic sequence of BRS site
(adapted from Porraz *et al.*, 2015, Puech *et al.*, 2021)

Exploring an artificial intelligence approach

Both measurement techniques and observation instruments continuously improve, allowing researchers to better characterise artefacts and bioarchaeological remains. In particular, new methods of artificial intelligence (AI), machine learning (ML) and computer graphics in general, recently contributed to raise the interpretative potential of historical and archaeological data. In this regard, the recent research project *Arch-AI-Story* (dir. I. Théry-Parisot, CEPAM) aims, among others, to automate the identification of diverse remains by deploying shape recognition and classification algorithms in the context of supervised learning.

In very general terms, AI can perform a wide class of tasks, such as classification or object detection, through the analysis of digital images (2D, 3D, SEM, Tomography, *etc.*) from different domains: rock art and wall painting, manuscripts, archaeological archives of biological (wood, seeds, bones) and geological (flint, ceramics) origin, among others. In each of these contexts, new heterogeneous, incomplete, discontinuous and even possibly biased datasets are useful for the AI community that exploits them for its own developments (see the PI@ntNet project for an example, Joly *et al.* 2016). Thus, the research activity emerging from the interdisciplinary project *Arch-AI-Story* represents a conceptual and practical advance for both archaeological, historical and textual sciences, on the first hand, and mathematical modelling and AI, on the other hand.

As part of the launch of the *Arch-AI-Story* project, datasets of charcoal micrographs and descriptions from the reference collection acquired around the BRS site were made available to test the potential of AI use for the taxonomic classification of archaeological charcoals in the area. If the purpose is not to replace the skill of the wood anatomists, it could be a useful tool to overcome the lack of reference collections and descriptions for rich woody area as well as to discriminate selected similar taxa with ecological interest. The framework of this first approach on southern African wood is structured as follows:

- Task 1: Mapping of data sets / Definition of corpus standardization methods / Creation of specific data banks;
- Task 2: Adaptation and development of machine learning/modelling/simulation algorithms;
- Task 3: Return on the corpus and reinterpretation of historical and archaeological datasets.

Here we present Task 1 and the progress on Task 2.

Material and methods

Dataset of 2D charcoal micrographs from southern African woods

The modern wood charcoal specimens currently include 83 specimens (20 families, 57 genus and 73 species) and are examined according to the three natural wood sections (*i.e.* transverse, radial, and tangential), for standard observation with:

- a reflected light microscope (100x, 200x, and 500x magnification) performed at the CEPAM, University Côte d'Azur (Fig. 2 left) ;
- a scanning electron microscope (SEM) allowing higher magnification and resolution (up to 5000x for some anatomical characters), performed on the CCMA EM Core Facility, University Côte d'Azur (Fig. 2 right).

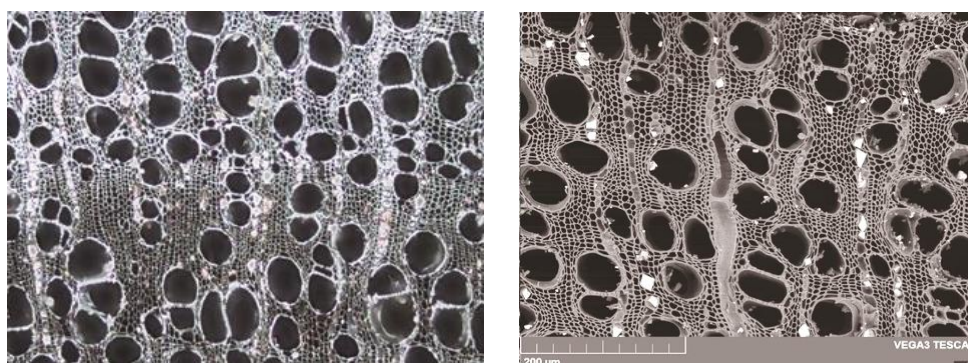


Fig. 2: Micrographs of modern wood charcoal from the reference collection (*Searsia rigida* var. *margaretae*) with an optical microscope (left) and a SEM (right)

All the images from the reference collection generated for the anthracological analysis of BRS have been uploaded to the online database *Anatom-IA*, representing 1567 micrographs acquired by reflected light and SEM. Some members of the GDR 3544 Sciences du bois have initiated this image bank of wood anatomy within the framework of the Working Group “Imaging”. It is a collaborative and interoperable tool dedicated to research and training in wood anatomy on the one hand and exploitable for anatomical identification based on AI and used for the *Arch-AI-Story* project (Fig. 3). To feed this image bank, it is necessary to fill in the metadata for each image, among which the geographical origin, taxonomy, anatomical section, magnification, microscope used, *etc.*, that is, more than 30 fields to fill per image.

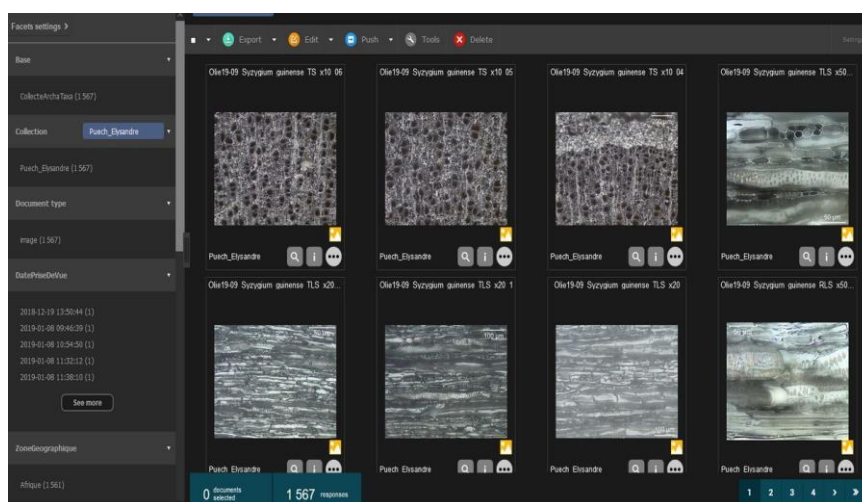


Fig. 3: Screen shot of the online image database *Anatom-IA*
(<https://mediasrv.cepam.cnrs.fr/prod/>)

Dataset of anatomical features from southern African woods

The need to compile in a single database the various wood anatomical descriptions extracted from miscellaneous sources (field collection, bibliography and wood library) for the study area naturally arose in order to allow rigorous taxonomic identification while facilitating the work of bibliographic research. These descriptions, which include more than a hundred anatomical characters per specimen, are regularly updated and adjusted according to the International Association of Wood Anatomists (IAWA) (Wheeler *et al.*, 1989). They currently include the descriptions of 83 specimens from fieldwork sampling, as well as from the *ca.* 1800 southern African wood specimen extracted from the InsideWood online database (InsideWood 2004-onwards; Wheeler 2011).

Supervised learning: setup

First, we only focus on the transverse sections (TS) of the charcoal images, as done in similar works (*e.g.* Maruyama *et al.*, 2018, Menon *et al.*, 2019). On the reference collection, our aim is to train a classifier (*i.e.* a mathematical function whose parameters are learned) that can associate a specimen to its family based on the TS images available for that specimen. Due to the incredible performances of convolutional neural networks (CNNs) in 2D images classification, we decided to test some popular architectures such as ResNet, VGG or INCEpTION as classifiers to reinforce our identification process. Since the number of specimens/observations (83) is not much higher than the number of classes/families (20), data augmentation is required. In addition, a consequent pre-processing step is required first. For instance, all the SEM images need to be cropped in order to remove the metadata bar and only keep the raw microscope image (Fig. 4). Finally, several views exist for the same specimen at different magnifications, for which the model might not converge to an optimal solution, such as with an overfit model. Therefore, we need to inform the classifier that two images refer to the same specimen.

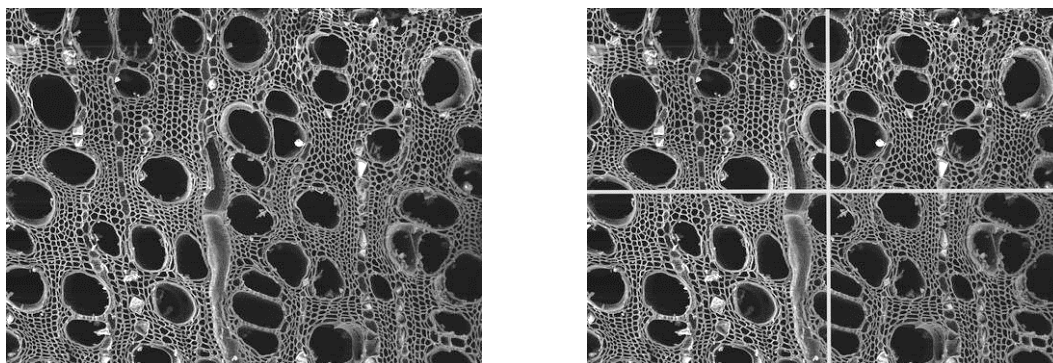


Fig. 4: Cropped version of the image in Fig. 2 (left) and data augmentation strategy by splitting (right).

The principle is to adapt the above-mentioned architectures in such a way to make them work in multi-view classification contexts. A simpler alternative is to consider a fixed magnification (*e.g.* 500x) for each specimen. In this way, data augmentation can be easily obtained by splitting an image into four sub-images for instance (Fig. 4). Although a higher number of sub-images could be extracted, a finer partitioning is not suitable as a global view of the anatomical features would be lost, such as the vessels density and distribution. However, due to imbalanced class number, some transfer learning techniques are nevertheless required by either fine-tuning the layers of the pre-trained CNN, or by using the CNN as a visual (anatomical) features extractor.

Expected results, comments and perspectives

Based on the results obtained by other authors on their modern collections of wood micrographs, we expect our models to yield high classification accuracies. At a latter point, longitudinal anatomical sections (radial and tangential) should be included with the development of the data augmentation process to strengthen the overall identification process. In this first project stage, we evidenced the importance of understanding how to pre-process (or not) both the optical and the SEM images to allow a classifier to deal with both dataset without introducing bias based on the acquisition method. Furthermore, it is worth mentioning that the taphonomy of archaeological charcoals (biotic and abiotic alterations before and after wood combustion) may limit the performance of the classifier in the first exploratory tests with archaeological charcoal images (see *e.g.* Rosa da Silva *et al.* 2022). This issue will be addressed at the later stage of the project, once the important taxa (family or tribe/genus) have been identified as those to focus on, while reflecting the data available to train the algorithm.

Acknowledgements

The authors greatly acknowledge the CCMA (Centre Commun de Microscopie Appliquée Université Côte d'Azur, Microscopy and Imaging platform Côte d'Azur, MICA) and in particular François Orange, who carried out the settings for the acquisition of the SEM images. Anne Lavalette and Anna Duplex (Alt R&D) as well as Antoine Pasqualini (CEPAM, Université Côte d'Azur) are hereby thanked for supervising the editing of the southern African charcoals images in the Anatom-IA database. Least but not last, the authors acknowledge all the past students who worked on the data set, including Elias Bughsn, Martiros Yeghiazaryana and Evariste Njomgue in the framework of their master project at the University Côte d'Azur.

References

- Asouti E., Austin P. (2005). Reconstructing woodland vegetation and its exploitation by past societies, based on the analysis and interpretation of archaeological wood charcoal macro-remains. *Environmental Archaeology*, 10(1), 1-18.
- Höhn A., Neumann K. (2018) Charcoal identification in a species-rich environment: The example of Dibamba, Cameroon. *IAWA Journal* 39, 87-147.
- IAWA Committee, Wheeler E.A., Baas P., Gasson P.E. (1989) IAWA list of microscopic features for hardwood identification: with an appendix on non-anatomical information. *IAWA Bull.* n.s. 10, 219–332.
- InsideWood (2004-onwards) Published on the Internet. <http://insidewood.lib.ncsu.edu/search> [10.10.2022].
- Joly A., Bonnet P., Goëau H., Barbe J., Selmi S., et al. (2016). A look inside the PI@ntNet experience. *Multimedia Systems*, Springer Verlag, 22 (6), pp.751-766.
- Maruyama T.M., Oliveira L.S., Britto Jr A.S., Nisgoski S. (2018). Automatic classification of native wood charcoal. *Ecological Informatics*, 46, 1-7.
- Menon L.T., Laurensi I.A., Penna M.C., Oliveira L.E., Britto A.S. (2019, June). Data augmentation and transfer learning applied to charcoal image classification. In 2019 International Conference on Systems, Signals and Image Processing (IWSSIP) (pp. 69-74). IEEE.
- Porraz G., Val A., Dayet L., De La Peña P., Douze K., Miller C.E., Murungi M.L., Tribolo C., Schmid V.C., Sievers C. (2015) Bushman Rock Shelter (Limpopo, South Africa): A perspective from the edge of the highveld. *The South African Archaeological Bulletin* 70, 166–179.

Puech E., Bamford M., Porraz G., Val A., Théry-Parisot I. (2021) Evaluating sampling methods in charcoal-rich layers and high diversity environment: A case study from the Later Stone Age of Bushman Rock Shelter, South Africa. *Quaternary International, Anthracology: Charcoal Science in Archaeology and Palaeoecology* 593–594, 36–49.

Puech E. (2021) Analyse des charbons de bois archéologiques de la séquence Later Stone Age du site préhistorique de Bushman Rock Shelter, Limpopo, South Africa : du terrain au travail de laboratoire. *Récit de terrain. Antipodes, Annales de la Fondation Martine Aublet*.

Rosa da Silva N., Deklerck V., Baetens, J.M. et al. (2022) Improved wood species identification based on multi-view imagery of the three anatomical planes. *Plant Methods* 18, 79

Schmidt E., Lotter M., McClelland W. (2007) *Trees and Shrubs of Mpumalanga and Kruger National Park*. Second edition. Jacana Media.

Thuiller W., Midgley G.F., Rougeti M., Cowling R.M. (2006) Predicting patterns of plant species richness in megadiverse South Africa. *Ecography* 29, 733–744.

Wheeler E.A. (2011) InsideWood - a web resource for hardwood anatomy. *IAWA Journal* 32 (2): 199-211.