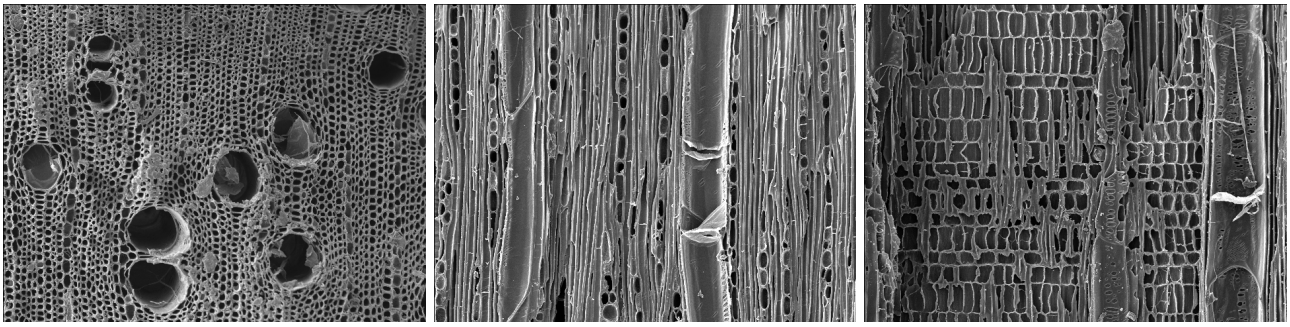
Supervised classification of wood and charcoal from microscopic images

M1/M2 internship proposal + possible PhD thesis

Supervisors: Marco Corneli (CEPAM, INRIA, UCA), Diane Lingrand (I3S, UCA), Frédéric Precioso (I3S, UCA), Isabelle Théry (CEPAM, UCA),
Laboratories: I3S, INRIA-MAASAI and CEPAM, UniCA.

Scientific project

Within the framework of the recently funded ANR project **AI-WOOD**, researchers at CEPAM, I3S and INRIA Université Côte d'Azur are collaborating to the development of new machine/deep learning approaches aiming at performing the taxonomical identification (i.e. classification at the species, genus or family level) of wood and charcoal from microscopic 2D images (Figure 1). The project has a main interest from an archaeological point of view, the main idea being to train a classifier on a modern collection (about 6000 images for 120 species) and then use it to identify ancient charcoals. The anthracologists (i.e. the archaeologists specialized in the identification and analysis of ancient charcoal) actually perform this identification relying on comparative anatomy and based on anatomical features settled by the IAWA[1] that they build manually through microscopic observation. Apart from being long and tedious, this identification routine is not entirely satisfying, (also) due to the anatomical proximity of some essences. Hence, the aim of this project is to explore the potential of machine/deep learning to directly identify the taxon of a specimen from the microscopic image and possibly to boost the identification routine. Although some attempts in this direction have been made in the literature (Rosa da Silva et al., 2022; Silva et al., 2022) there is still considerable room for improvement.



(a) Cross section (b) Tangential section (c) Radial section

Three anatomical sections from the microscopic observation of charcoal (specimen from the Anacardiaceae family).

In this internship we plan to explore several avenues: 1) conceive multiview classifiers (Wang et al., 2020) since each observation/specimen has several "faces", corresponding to different acquisition zooms and three sections (cross, longitudinal and radial); 2) exploit the IAWA features (available for the training dataset) in order to merge learning approaches and symbolic AI techniques (see e.g. Diligenti et al., 2017; Ciravegna et al., 2023). This might be helpful in order to reduce the number of observations needed to train the model; 3) develop ad-hoc solution (i.e. dedicated convolutional layers) accounting for the specific nature of the data.

Candidate profile, timeline and stipend

Master student (M2 level) in signal/image processing, applied mathematics, data science and artificial intelligence with a background in the following disciplines: learning, classification, image processing and use of libraries for deep learning (PyTorch/Tensorflow...). A general interest in botany and/or archaeology would be a plus. The starting date of the project is April 2023. The duration of the project is 4/5 months. Stage internship gratification ($\approx$ 600 €/month).A PhD could be envisaged, with funding provided for in the ANR project.

Application procedure

Interested candidates should send an e-mail with their CV, their transcripts and two possible referee contacts to marco.corneli@univ-cotedazur and diane.lingrand@univ-cotedazur.fr. Selected candidates will be invited for an interview organised through video-conferencing services.

References

Ciravegna, G., Precioso, F., Betti, A., Mottin, K., and Gori, M. (2023). Knowledge-driven active learning. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 38–54. Springer.

[1] International Association of Wood Anatomy

Diligenti, M., Gori, M., and Sacca, C. (2017). Semantic-based regularization for learning and inference. *Artificial Intelligence*, 244:143–165.

Rosa da Silva, N., Deklerck, V., Baetens, J. M., Van den Bulcke, J., De Ridder, M., Rousseau, M., Bruno, O. M., Beeckman, H., Van Acker, J., De Baets, B., et al. (2022). Improved wood species identification based on multi-view imagery of the three anatomical planes. *Plant Methods*, 18(1):1–17.

Silva, J. L., Bordalo, R., Pissarra, J., and de Palacios, P. (2022). Computer vision-based wood identification: A review. *Forests*, 13(12):2041.

Wang, W., Tran, D., and Feiszli, M. (2020). What makes training multi-modal classification networks hard? In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12695–12705.