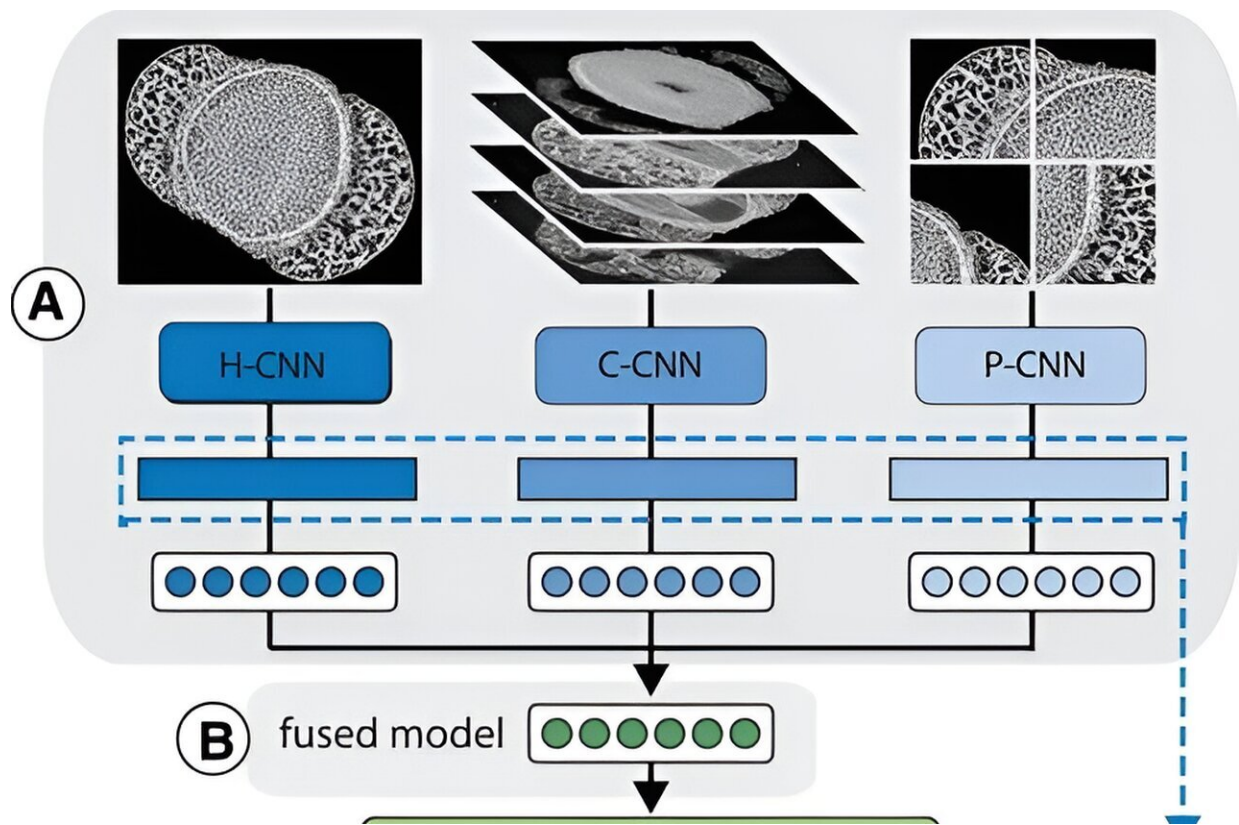


# Biologists use machine learning to classify fossils of extinct pollen

March 20 2024, by Shelby Lawson



lowchart illustrating the trained multimodal neural network pipeline. Credit: *PNAS Nexus* (2023). DOI: 10.1093/pnasnexus/pgad419

In the quest to decipher the evolutionary relationships of extinct organisms from fossils, researchers often face challenges in discerning

key features from weathered fossils, or with prioritizing characteristics of organisms for the most accurate placement within a phylogenetic tree. Enter neural networks, sophisticated algorithms that underlie today's image recognition technology.

While previous attempts to utilize neural networks in classifying extinct organisms within phylogenetic trees have struggled, a new study, [published](#) in *PNAS Nexus*, heralds a significant breakthrough. The model has been trained to recognize and rank organism features based on known phylogenetic information, and can accurately place new organisms, including those that are extinct, within the intricate branches of evolutionary trees.

The team includes Surangi Punyasena (CAIM), an associate professor of Plant Biology at the University of Illinois Urbana-Champaign, Shu Kong, an assistant professor of science and technology at the University of Macau, and Marc-Élie Adaimé, a graduate student in Punyasena's lab and first author on the study.

According to Adaimé, the reason neural networks have trouble accurately classifying extinct organisms as opposed to living ones is often a matter of how they are trained.

"Most paleontological AI studies typically focus on straightforward classification tasks, such as distinguishing between different fossil types," explained Adaimé.

"This approach works well within the scope of clearly defined categories, but less so with data that doesn't fit these categories. Think of a model that has only been trained to classify images of dogs or cats. If it were presented with an image of a snake, the model would try to categorize it as either a dog or cat because it's limited to what it was trained on.

"Similarly, there was no method previously that included phylogeny a priori into the model, so models did not learn to make sense of the features in an evolutionary or phylogenetic context. The goal of our research was to create a new modeling approach that would be trained on images in a phylogenetic context."

To accurately position organisms within a phylogenetic framework, [neural networks](#) must be trained not only to discern defining traits of various organism classes but also to recognize phylogenetic synapomorphies—derived features shared between organisms due to their common ancestry. This enables the network to determine the placement of organisms within a [phylogenetic tree](#).

The team chose to apply their model to the classification of pollen and spores—a ubiquitous and ancient entity found throughout the [fossil record](#), with earliest fossils dating back hundreds of millions of years.

The researchers first gathered optical super-resolution images of modern and fossil pollen that had been taken at the Carl R. Woese Institute for Genomic Biology Core Facility. They trained their model using micrograph images of 30 extant (living) Podocarpus species. During this process, the model identified features it deemed important for classifying the pollen into different classes.

Subsequently, these features were inputted into a secondary model, along with established phylogenetic data on the species, which then reweighted the features based on their phylogenetic significance. This approach enabled the model to generate a phylogenetically-informed distance function, applicable to new pollen images provided to the model.

To validate the model's efficacy, the researchers tested it on micrograph specimens of extinct pollen from Panama, Peru, and Columbia. While the exact phylogenetic relationships were not definitively known,

paleoecologists had previously placed the pollen within *Podocarpus* based on morphological traits and geographical distribution.

Impressively, the neural network model mirrored the placements made by the paleoecologists for nearly all specimens, underscoring its capacity to leverage morphological features learned during training to accurately position extinct species within a phylogenetic context.

Punyasena noted that her lab is collaborating with colleagues at the Smithsonian National Museum of Natural History and the Smithsonian Tropical Research Institute to expand this work and apply it to a broader set of fossil pollen data.

"International continental drilling projects are currently producing unimaginable amounts of fossil plant material," said Punyasena.

"Fully leveraging these new data sources means changing the way that we analyze and interpret [fossil pollen](#). As a community, we need to take advantage of advances in deep learning and computer vision. This work demonstrates that the amount of evolutionary information captured in pollen morphology had been previously underestimated. The history of a plant species is captured in its shape and form. Machine learning allows us to discover these novel phylogenetic traits."

The researchers plan to enhance their model's accuracy and adaptability by expanding the sample size of images used for training. Furthermore, they aim to ensure the model remains current by integrating emerging advancements in machine learning. Adaimé emphasizes the model's versatility beyond pollen classification, foreseeing its potential application in categorizing various fossil organisms.

"Machine learning models can make it easier to find features that are informative, because the way machine learning models think is obviously

very different from what the way humans think," said Adaimé.

"It's going to be able to find patterns that make sense but probably aren't intuitive to humans. And the benefit of this approach isn't just limited to pollen, we expect these models will be generalizable to classifying fossils of other organisms as well."

**More information:** Marc-Élie Adaimé et al, Deep learning approaches to the phylogenetic placement of extinct pollen morphotypes, *PNAS Nexus* (2023). [DOI: 10.1093/pnasnexus/pgad419](https://doi.org/10.1093/pnasnexus/pgad419)

Provided by University of Illinois at Urbana-Champaign

Citation: Biologists use machine learning to classify fossils of extinct pollen (2024, March 20) retrieved 27 April 2024 from <https://phys.org/news/2024-03-biologists-machine-fossils-extinct-pollen.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.