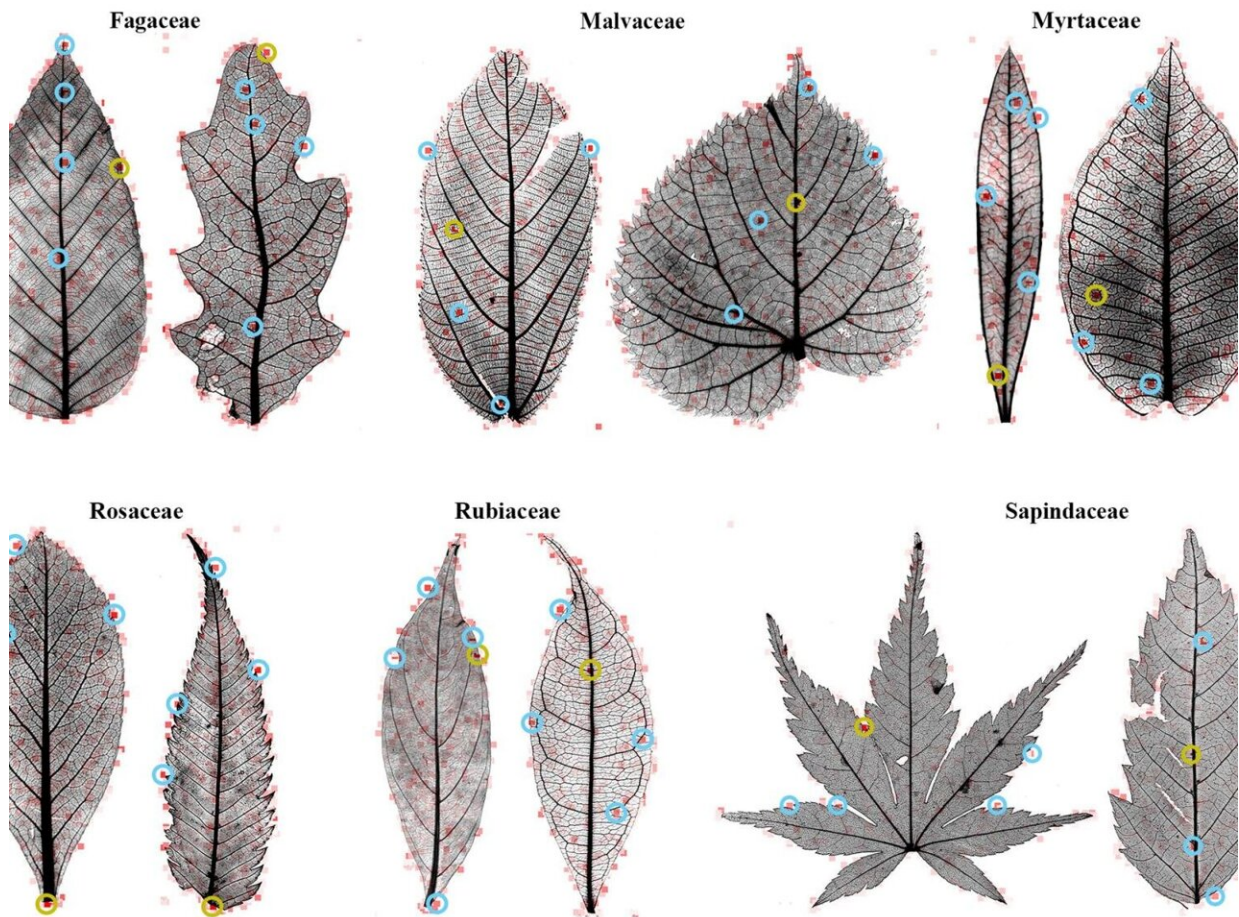# Decoding the leaf: Scientists search for features to ID modern, fossil leaves

May 13 2022, by Matthew Carroll

A machine learning program placed small red boxes on leaf images, indicating tiny sections the computer considered important in identifying the leaf by family group. Scientists placed the circles on the images to represent the hottest areas on the heat maps and compared them across families to find new ways to describe what the computer is seeing in traditional botanical terms. Credit: Penn State

Machine learning programs that can classify leaves and place them in biological families may unlock new clues about the evolution of plant life, but only if scientists understand what the computers are seeing. A team led by Penn State scientists combined a machine learning approach and traditional botanical language to find and describe new features for fossil identification.

"You have the computer saying, 'look over here, this is important,' but there has to be someone who can translate the results into human friendly terms," said Edward Spagnuolo, a recent Penn State graduate with a bachelor's degree in geobiology who led the research. "So that's really what we did. This is very much a first step in merging artificial intelligence to botany and paleobotany."

The team took heat maps produced by machine learning programs—leaf images covered with small red boxes that highlight areas the computer identified as important for identification—and developed a manual scoring system to analyze these regions areas across different plant families.

"We basically found that each family had a unique suite of features that were emphasized by the heat maps," Spagnuolo said. "And all these features provide new leads to identify fossil leaves. You can't take these out and directly identify fossils yet, but this is a first step. For some families, these are the only leads we have"

Leaves are the most common non-microscopic plant part found today and in the fossil record, but they are also the most difficult to identify. Variation in leaf shape and venation—the pattern of veins in the blade of a leaf—is too complex for botanical terminology to capture, the scientists said.

This is especially challenging for paleobotanists, who most often find

isolated fossil leaves without seeds, fruits or flowers that could help identify the plants. Further compounding the challenge, many of the individual fossils represent plants that are extinct.

"The [evolutionary history](#) and fossil record are very poorly understood for even some of the most important and diverse plant families alive today, and that's the impetus for this study," Spagnuolo said. "There are millions and millions of fossil leaves stored in museum collections worldwide that cannot be identified because we just don't have well-defined leaf structures to place them in proper groups."

Describing a single leaf could take hours for a trained researcher, but computer programs can learn to spot differences and sort leaves into taxonomic families quickly and accurately, the scientists said.

Peter Wilf, professor of geosciences at Penn State and Spagnuolo's adviser, and Thomas Serre, professor in computer science at Brown, led a prior machine learning study of more than 7,500 images of cleared leaves, which are specimens that have been chemically bleached, stained and mounted on slides to reveal venation patterns. The program placed the leaves into families with 72% accuracy and produced the heat maps that scientists can use to learn what the computer viewed as important for identification.

"This approach is different from most botanical and palaeobotanical leaf studies, which will look at large scale leaf features—the number of veins, how the leaf is shaped," Spagnuolo said. "These are really small crops of images. And moving forward we need a way to combine the larger scale botanical features we've used for centuries that also takes in these smaller scale features that have been missed because they are so hard to see without this help from the artificial intelligence algorithm."

Spagnuolo analyzed more than 3,000 of the heat maps featuring leaves

of 930 genera in 14 angiosperm, or flowering plant, families. He scored the top-five and top-one hot spot regions and used traditional botanical language to describe their locations on the leaves.

"We attempted to decode the machine-learning algorithm's family-level identification of cleared leaves through location-mapping the hottest hot spots," Spagnuolo said. "This is, to our knowledge, the first attempt to back-translate and interpret computer vision heat maps into botanical language."

They recently reported their findings in the *American Journal of Botany*.

Some families like Rosaceae—which include plants that produce apples, strawberries, plums, cherries, peaches and almonds—have distinctive features that botanists and paleobotanists can easily identify, like narrow teeth. The hot spots on these families seem to echo traditional observations, the scientists said.

Other families like Rubiaceae, or the coffee family, lack distinctive features and largely go unidentified in the fossil record. On these untoothed leaves, the computer pointed to the microcurvature of little-studied leaf margins.

"These new features can lead to additional studies to hopefully delineate new fossil-identifying characters," Spagnuolo said. "This could someday help to unlock the immense amount of evolutionary dark data that we just have not tapped into yet."

Wilf and Serre contributed to this work.

Provided by Pennsylvania State University

Citation: Decoding the leaf: Scientists search for features to ID modern, fossil leaves (2022, May 13) retrieved 12 May 2024 from https://phys.org/news/2022-05-decoding-leaf-scientists-features-id.html