

Artificial intelligence identifies plant species for science

Deep-learning methods successfully classify thousands of herbarium samples.

Heidi Ledford

11 August 2017



Peter Macdiarmid/Getty Images

Digitizing plant specimens is opening up a whole new world for researchers looking to mine collections from around the world.

Computer algorithms trained on the images of thousands of preserved plants have learned to automatically identify species that have been pressed, dried and mounted on herbarium sheets, researchers report.

The work, published in *BMC Evolutionary Biology* on 11 August¹, is the first attempt to use deep learning — an artificial-intelligence technique that teaches neural networks using large, complex data sets — to tackle the difficult taxonomic task of identifying species in natural-history collections.

It's unlikely to be the last attempt, says palaeobotanist Peter Wilf of Pennsylvania State University in University Park. "This kind of work is the future; this is where we're going in natural history."

Natural-history museums around the world are racing to digitize their collections, depositing images of their specimens into open databases that researchers anywhere can rifle through. One data aggregator, the US National Science Foundation's iDigBio project, boasts more than 150 million images of plants and animals from collections around the country.

Related stories

- Natural-history collections face fight for survival
- Deep learning boosts Google Translate tool
- Plant collections left in the cold by cuts

[More related stories](#)

Related stories

- Natural-history collections face fight for survival
- Deep learning boosts Google Translate tool
- Plant collections left in the cold by cuts

[More related stories](#)

Complementary computing

There are roughly 3,000 herbaria in the world, hosting an estimated 350 million specimens — only a fraction of which has been digitized. But the swelling data sets, along with advances in computing techniques, enticed computer scientist Erick Mata-Montero of the Costa Rica Institute of Technology in Cartago and botanist Pierre Bonnet of the French Agricultural Research Centre for International Development in Montpellier, to see what they could make of the data.

Bonnet's team had already made progress automating plant identification through the PI@ntNet project. It has accumulated millions of images of fresh plants — typically taken in the field by people using its smartphone app to identify specimens.

Researchers trained similar algorithms on more than 260,000 scans of herbarium sheets, encompassing more than 1,000 species. The computer program eventually identified species with nearly 80% accuracy: the correct answer was within the algorithms' top 5 picks 90% of the time. That, says Wilf, probably out-performs a human taxonomist by quite a bit.

Such results often worry botanists, Bonnet says, many of whom already feel that their field is undervalued. “People feel this kind of technology could be something that will decrease the value of botanical expertise,” he says. “But this approach is only possible because it is based on the human expertise. It will never remove the human expertise.” People would also still need to verify the results, he adds.

Helping hand

This approach can help herbaria process new samples, simplifying an arduous task that sometimes requires hours of work. And similar efforts could help with other projects, such as a current crowdsourcing project that asks people to manually tick off which herbarium specimens feature a flower or a fruit. Researchers would certainly welcome an automated way of doing that, says botanist Gil Nelson of Florida State University in Tallahassee and a digitization specialist at iDigBio.

The algorithm could also aid smaller herbaria with their species identifications, Bonnet says. His team found that algorithms trained on large data sets from big herbaria improved the identification of plants from relatively data-poor regions of the world — a finding that could be particularly useful for areas that are rich in biodiversity but have smaller plant collections.

And this deep-learning approach will allow researchers to perform additional analyses. Herbaria samples contain a wealth of data: when and where the sample was collected, for example, and characteristics such as whether the plant was flowering or fruiting at collection time and how densely clustered the flowers were. Because some samples are centuries old, that data can paint a portrait of how plants have adapted to shifting climates — an area of growing interest in the face of concerns about climate change.

Moving ahead

Such efforts, including the identification study, are the next phase of digitization, Nelson says. “We’ve been trying to transition to methods that we can use to mine those images and to pull out useful data,” he says. “That’s our focus right now.”

The projects aren’t limited to herbaria. Nelson points to ongoing efforts to automate the identification of fly larvae, and Wilf is working with collaborators to carry out a similar analysis on plant fossils. Such fossils pose other problems, in part because they come in a variety of forms — fossilized fruits and flowers, petrified tree trunks or impressions of leaves in rock. Herbarium sheets, by contrast, are mercifully uniform: flat, dry and typically mounted on a standardized size of paper.

Still, Wilf has no doubt that the field will eventually work out these details. “It’s just going to get better,” he says. “Someday we’ll have students who won’t be able to remember when we didn’t have these sorts of tools.”

Nature doi:10.1038/nature.2017.22442

References

1. Carranza-Rojas, J., Goeau, H., Bonnet, P., Mata-Montero, E. & Joly, A. *BMC Evol. Biol.* <http://dx.doi.org/10.1186/s12862-017-1014-z> (2017).