

Study looks at potential of Deep Learning on herbarium species identification

August 14 2017, by Nancy Owano



Ardisia revoluta Kunth herbarium sheet sample taken from Arizona State University Herbarium. Credit: BMC Evolutionary Biology, [DOI: 10.1186/s12862-017-1014-z](https://doi.org/10.1186/s12862-017-1014-z)

"Let's leverage herbaria impact through Deep Learning," was the headline in the BMC Series blog of August 11.

To give you a sense of the numbers scientists are dealing with, "The number of institutions that maintain herbaria around the world is around 3,000; with more than 350,000,000 specimens stored under controlled [environments](#)."

Not only that, but, said *TechCrunch*, "It's suspected that hidden among them may be tens of thousands of [new](#) species—but the labor cost of manually going through all the samples to double-check them, modernize taxonomy and so on is prohibitive."

"Herbaria" refers to a collection of [plants](#) that have been mounted and classified systematically.

Small wonder, then, that, according to *Nature*, digitizing plant specimens is opening up a whole new world for researchers.

Heidi Ledford, *Nature*, said, "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."

The herbarium collections carry a heritage and knowledge of plants. Museums and institutions maintain the herbaria with their plant specimens collected over hundreds of [years](#).

Yet there are data relatively untapped. That is because identifying and classifying specimens have not been easy. The problems: (1) thousands of sheets have gone still unidentified at the species level (2) numerous sheets need reviewing and updating to reflect more recent taxonomic knowledge and (3) these tasks for botanists demand a lot of work in reasonable time.

In a BMC Series blog on Friday, José Carranza-Rojas, Erick Mata-Montero and Pierre Bonnet discussed their research, published today in *BMC Evolutionary Biology*, that uses deep learning computer vision techniques to automate specimen identification. The authors in their paper said that results showed the potential of Deep Learning on herbarium species identification, "particularly by training and testing across different datasets from different herbaria."

According to the paper's abstract, "Computer vision and machine learning approaches applied to herbarium [sheets](#) are promising but are still not well studied compared to automated species identification from leaf scans or pictures of plants in the field."

"Going deeper in the automated identification of Herbarium specimens," an Open Access article, is the title of their work, in *BMC Evolutionary Biology*.

Authors are Jose Carranza-Rojas, Herve Goeau, Pierre Bonnet, Erick Mata-Montero and Alexis Joly.

Heidi Ledford, *Nature*, said, "The work, published in BMC Evolutionary Biology on 11 August, 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."

Ledford in *Nature* described their research.

Computer algorithms trained on images of thousands preserved plants learned to automatically identify species pressed, dried and mounted on herbarium sheets, reported researchers. "Bonnet's team had already made progress automating plant identification through the Pl@ntNet project." (It has accumulated millions of images of fresh plants, typically

taken by people using its phone app to identify specimens.)

What's next?

The authors made their case for the use of Deep Learning in work with herbaria images.

"One of the problems with herbarium images is visual noise. Normally, specimens are placed on sheets without automated visual processing needs in mind. For instance, organs are juxtaposed and [elements](#) such as labels are also present in the image. Deep learning is a technology that has been proven to deal particularly well with visual noise and complex images. So now it is the right time to attempt to use it with herbaria images."

Quoted in the *Nature* article, palaeobotanist Peter Wilf of Pennsylvania State University said, "This kind of work is the future; this is where we're going in natural [history](#)."

Carranza-Rojas, Mata-Montero and Bonnet, who authored the blog, said that "There is a huge amount of work invested by field explorers, botanists, taxonomists, technicians, and data managers that has generated very useful data not only for the biological sciences but also for the computer science community. We hope that this work will open the door to stronger collaborations between these communities, particularly between Natural History Museums and Machine Learning / Computer Vision labs."

More information: Jose Carranza-Rojas et al. Going deeper in the automated identification of Herbarium specimens, *BMC Evolutionary Biology* (2017). [DOI: 10.1186/s12862-017-1014-z](https://doi.org/10.1186/s12862-017-1014-z)

© 2017 Tech Xplore

Citation: Study looks at potential of Deep Learning on herbarium species identification (2017, August 14) retrieved 18 April 2024 from <https://techxplore.com/news/2017-08-potential-deep-herbarium-species-identification.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.