

## Plant Biology

# Machine-learning boosts the conservation of endangered plant species

by **Tara Pelletier**<sup>1</sup> | Assistant Professor; **Anahí Espíndola**<sup>2</sup> | Assistant Professor

<sup>1</sup>: Radford University, Radford, Virginia, USA

<sup>2</sup>: University of Maryland, College Park, USA

This Break was edited by Max Caine, *Editor-in-chief* - TheScienceBreaker

### ABSTRACT

*To improve the protection of species, we developed a machine-learning method that lets us successfully identify plant species that are likely at risk. This method is a very helpful complement to other more expensive and time-consuming approaches, and can be used at both local and global scales.*



Image credits: Pixabay – CC0

Pandas, whales, elephants, and lions are extinct. Even though we know that's not true, that could be true in a world in which those species had never been studied and protected, and it is a real threat for many species that are unfortunately not as popular as these iconic mammals. To avoid that, many organizations study and evaluate whether species should be classified as in need of protection. One of those is the International Union for Conservation of Nature (IUCN), which has developed a detailed protocol to study species one-by-one and classify them based on trends in their population sizes, on their ranges, or on potential exposure to threats such as urbanization or deforestation. Indeed, the IUCN is and has been a central player in identifying species

the most in need of conservation, and has provided an internationally credible structure to identify them. The problem the IUCN runs into, however, is that there are too many species! There is not enough time and/or money to study them all.

To assist in this problem, we developed a method that uses the analytical strengths of something called machine-learning. Machine-learning is a family of methods that allows analyzing very large datasets, to find general trends, and use the observed trends to make predictions on new data. Even though this may sound very unfamiliar, it is these types of methods that Facebook, YouTube or Netflix use to 'suggest' new pages to visit, new movies to watch, and new

music to listen. They do that by analyzing the data about your past behaviors with their websites or apps, and suggesting ('predicting') other products you may like based on your interests. Unlike facebook or youtube, our goal in this study was not to sell new products but to use the large amount of publicly available data on species and use machine-learning methods to make predictions of which species are the most likely to be in need of conservation. Even though these methods handle large datasets, all species of the world were just too many species to study at once, so we focused on plant species, which are surprisingly understudied, but still abundant in public databases.

For all plant species, we decided to use different variables (the equivalent of 'behaviors' that facebook or youtube use) to predict the level of conservation need. These data related to characteristics of the plants' geographic ranges, their climatic preferences, and their morphology. After collecting this information for all species we identified about 165,000 species that had enough data to be analyzed. To start our analysis, we first identified species that had been classified as in need of conservation or not under the IUCN protocol (about 15,000 species), from those that had not (the remainder 150,000 species). With these two sets of data, we used the one with already classified species to let the method 'learn' how the range, climatic and morphological characteristics of those species related to their IUCN classification. Then, using the unclassified set of species, our method used those 'learned' relationships to predict whether or not a

species with known characteristics, but unknown conservation needs, was likely to be in need of conservation.

Our results show that about 10% of all the unstudied species are likely in need of conservation at a probability higher than 80%. Further, when we individually checked information on the species that were predicted with the highest probability, we found that a vast majority of them had already indications of being likely in danger (for example, have very small ranges, are hyper-endemic, occur in regions environmentally threatened). An interesting pattern we identified was that these species seemed to not be randomly distributed across the world and that many of these regions of high conservation needs corresponded to regions of high species diversity (like California, Southwest Australia, the Philippines, Central America, or Madagascar). Therefore, this method identified regions of conservation need and these regions make sense biologically.

We found that our method can be used to direct the evaluation of future species using the IUCN protocol, even though we do not recommend that it replaces those formal and in-depth evaluations. A more sensible use of our method is for assisting in the prioritization of which species to need be evaluated, using our predicted probabilities as a guide. By doing so, research resources can be allocated to species that are the most likely to be at risk, and thus need the most urgent action.