A Novel Machine Learning Framework for Predicting Scientific Names to Quantitatively Document and Analyze Bioindicators, Invasive, and Keystone Species for Assessing the Impacts of Urbanization and Climate Change on Indigenous Ecosystems

Li, Lin Wang (School: The Bronx High School of Science)

Human activities have led to an alarming shift in the global climate and ecosystems. The scientific community had made various attempts to analyze and counteract climate change and environmental catastrophes. Current efforts include identifying patterns and problems of bioindicators, invasive, and keystone species in habitats around the globe. However, due to budget priorities and constraints, many endeavors to analyze these species have been hindered. Thus, I developed a novel machine learning framework that accelerates the process of documenting and analyzing these species. Here, I focused mainly on small and unbalanced datasets because they mimic the real-world situation when target species are scarce. The datasets are queried from iNaturalist, an open source dataset that is supported by volunteers worldwide. I tested the efficacy of various machine learning models on the classification and analysis of species. Then, I evaluated the capabilities of these various models and integrated them into a novel framework that achieves an overall accuracy of above 90% with unbalanced and small datasets. This framework overcomes the limitations of conventional endeavors in species documentation, which have involved scarce professional resources and insufficient budgets. Moreover, by using data uploaded by the general public, I hope to foster more people to engage with the scientific community on fixing environmental issues.

Awards Won: