LITERATURE SURVEY

DIGITAL NATURALIST AI ENABLED TOOL FOR BIODIVERSITY RESEARCHERS

SUMMARY :

The increasing availability of digital images, coupled with sophisticated artificial intelligence (AI) techniques for image classification, presents an exciting opportunity for biodiversity researchers to create new datasets of species observations. We investigated whether an AI plant species classifier could extract previously unexploited biodiversity data from social media photos (Flickr). We found over 60,000 geolocated image stagged with the keyword ‘‘flower’’ across an urban and rural location in the UK and classified these using AI, reviewing these identifications and assessing the representativeness of images. Images were predominantly biodiversity focused, showing single species. Non-native garden plants dominated, particularly in the urban setting. The AI classifier performed best when photos were focused on single native species in wild situations but also performed well at higher taxonomic levels (genus and family), even when images substantially deviated from this. We present a checklist of questions that should be considered when undertaking a similar analysis .

INTRODUCTION:

The ever-growing number of digital sensors in the environment has led to an increase in the amount of digital data being generated. This includes data from satellites, weather stations, data from “internet of things” devices, and data collected by members of the public via smartphone applications, to name but a few. These new sources of data have contributed to the era of “Big Data” characterized by large volumes of data, of numerous types and quality, being generated at an increasing speed.1 This presents challenges and opportunities across a number of domains, including water management,2 camera trapping,3 and acoustic4 analysis. To process these data into useful information there are many tools available, including classical statistical analyses5 and classification by citizen scientists.6 However, at some point traditional approaches may become inefficient or even impossible given the volume, diversity, and heterogeneity of these data. Storage, exploration, curation, and revision of data may have to be re-thought to allow for their quick and efficient transformation, annotation, or analysis. This is particularly difficult for multimedia data which are typically much more complex than other data types. For example, biodiversity and environmental records in the form of audio, video, or image files are typically larger and more complex than text or numeric data. Large-scale analysis of multimedia data has only been possible in recent years since the development of large computational facilities, both academic and commercial. Regardless, the analysis of multimedia data is often further complicated because of their nonstandardized methods of acquisition, with highly diverse devices, sensors, formats, scales, environmental contexts, and taxonomic scope. Building efficient, scalable, and robust approaches to solve these problems is a difficult scientific challenge at the forefront of data science and machine learning specifically .

DISCUSSION :

In this work, we implemented a collaborative biodiversity project on the subject of biology for 4th course students in secondary education using the iNaturalist platform and we evaluated their engagement. This outdoor activity to observe and collect plant species, which has been carried out for the last 25 years at the Nuestra Señora del Puy School, gave us the opportunity to integrate the use of new and collaborative technologies to create a virtual herbarium. This virtual herbarium is complementary to the traditional handmade herbarium that the students had to do with 15 specimens of trees and shrubs. Despite the recording of observations of plant species with iNaturalist being voluntary, 45.9% of the students participated with 1.68 records per student, identifying 32 species, which provides evidence of interest created by this activity .

DISCUSSION:

In this work, we implemented a collaborative biodiversity project on the subject of biology for 4th course students in secondary education using the iNaturalist platform and we evaluated their engagement. This outdoor activity to observe and collect plant species, which has been carried out for the last 25 years at the Nuestra Señora del Puy School, gave us the opportunity to integrate the use of new and collaborative technologies to create a virtual herbarium. This virtual herbarium is complementary to the traditional handmade herbarium that the students had to do with 15 specimens of trees and shrubs. Despite the recording of observations of plant species with iNaturalist being voluntary, 45.9% of the students participated with 1.68 records per student, identifying 32 species, which provides evidence of interest created by this activity.

CONCLUSION:

The applications for AI/ML are much wider than just climate and impact forecasting. They are interdisciplinary and holistic, informing society, using and feeding public libraries, and affecting global sustainability. They also easily include disease/pandemic topics. Good AI/ML data analysts will actually develop repeatable and transparent globally applicable workflows for AI/ML models, data, and their applications so that forecasting of events—done in a strategic fashion—can be achieved (Bluhm et al. 2010; Huettmann 2020). It is less the model or the algorithm, as such, but the wider approach to the inherently linked Arctic problem to be addressed where AI/ML can excel and assist human thinking and avoid human errors in an automated fashion, even done online and globally accessible.

ACKNOWLEDEMENT :

An early demonstration of this concept was developed at the British Ecological Society Quantitative Ecology Special Interest Group hackathon, led by T.A. Thanks to members of the hackathon team for their contributions: Nathan Fox, Celia Marlowe, Joseph Millard, Nadia Bystriakova, Elliot Shayle, and Roy Sanderson. T.A. was supported by COST action CA17122 “Increasing understanding of alien species through citizen science” and Natural Environment Research Council award number NE/R016429/1 as part of the UK-SCAPE program delivering National Capability. We thank three anonymous reviewers for their comments.

REFERENCES :

Aycrigg, J., G. Beauvais, T. Gotthardt, F. Huettmann, S. Pyare, M. Andersen, D. Keinath, J. Lonneker, M. Spathelf, and K. Walton. 2015. Novel Approaches to Modeling and Mapping Terrestrial Vertebrate Occurrence in the Northwest and Alaska: An Evaluation. Northwest Science 89:355–381. doi: /dx.doi.org/10.3955/046.089.0405.

Baltensperger A. P. and F. Huettmann. 2015. Predicted Shifts in Small Mammal Distributions and Biodiversity in the Altered Future Environment of Alaska: An Open Access Data and Machine Learning. PLOS One DOI: 10.1371/journal.pone.0132054 .

Bluhm, B., D. Watts, and F. Huettmann. 2010. Free Database Availability, Metadata and the Internet: An Example of Two High Latitude Components of the Census of Marine Life. Chapter 13, pp. 233–244. In: S. Cushman and F. Huettmann. Spatial Complexity, Informatics and Wildlife Conservation. Springer Tokyo, Japan. pp. 233–244.

SUBMITTED BY :

ASLIN JARCITHA.P (960319104009)

ASHLIN JEYA.J (960319104301)

SHEENU.S.L (960319104033)

NEETHU.J.G (960319104024)