

MAPPING INVASIVE PLANT SPECIES USING MACHINE LEARNING

**ENVIRONMENTAL COSTS OF MACHINE LEARNING ALGORITHMS: REDUCING
UNNECESSARY COMPUTATIONS**

An Undergraduate Thesis Portfolio
Presented to the Faculty of the
School of Engineering and Applied Science
In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Computer Science

By

Surbhi Singh

April 28, 2020

SOCIOTECHNICAL SYNTHESIS

The spread of invasive plant species is currently one of the greatest epidemics facing the agricultural industry, producing a loss of \$137 billion per year in the United States alone. Due to the growth of human factors such as global travel and foreign imports, the prevalence of invasive species has risen substantially. The technical research aims to map invasive plants using remote sensing satellite imagery and machine learning algorithms. The STS research analyzes the environmental effects of training complex Machine Learning models and methods of reducing unnecessary computations. Complex models which involve iterative training and parameters often require large data centers to run on and result in large carbon emissions. The STS research is loosely coupled with the technical research as it could potentially impact the methodology used in the technical research to reduce the amount of computations performed.

The first step to mitigating the presence of invasive plant species is understanding what factors are contributing to the spread. This can be studied using convolutional neural networks, which take an image as input and determine the importance of certain features. Invasive plants species are a global epidemic, but they are especially prevalent in biodiversity hotspots such as the Chitwan Annapurna Landscape of Nepal. Three invasive plant species in this region of Nepal were studied using multiple types of imagery and convolutional neural network based architecture.

The main focus of the research was to experiment with different types of satellite images used to map species distribution. The use of pan sharpened imagery to increase the resolution of the satellite imagery was explored. Pansharpening is the merging of high-resolution panchromatic and lower resolution multispectral imagery for the creation of a single-high resolution image. This technique is used by most mapping software such as google maps. This

process was applied to all of the Nepal satellite imagery and the models were retrained to determine if these images provided better predictions for the spread of invasive plant species. Improvements in prediction accuracy were seen as a result of pansharpener the satellite imagery.

Machine Learning is a subset of Artificial Intelligence (AI), which is the broader concept of intelligent machines. While AI has become a very popular tool in solving today's problems, most people fail to acknowledge the environmental impacts of training such computationally expensive algorithms. These financial and environmental costs are much higher in research which requires retraining of model architecture and parameters. AI models are usually trained using data centers, which are large contributors to carbon emissions. A novel idea was proposed by researchers at the Allen Institute known as Green AI. Green AI includes novel research which is environmentally friendly and is considerate of the amount of resources required. Red AI is the opposite; computationally expensive and often sacrifices large amounts of efficiency for small accuracy gains.

Pacey's Triangle was used to investigate the barriers to developing more efficient and environmentally friendly algorithms which aim to reduce unnecessary computations. Perhaps the largest barrier to Green AI adoption is the cultural aspect of AI conferences. The pressure to get a paper accepted into a top conference or publish drives researchers to excessively train models, disregarding the computational costs. This research explored methods of changing this culture by requiring researchers to report efficiency and computational costs in addition to just accuracy to top AI conferences.

Current Machine Learning researchers are blindly wasting numerous computing resources on redundant training, with little to no accuracy gains. This research does not dismiss

the importance of computing intensive algorithms, but instead urges researchers to take a holistic view and not sacrifice some areas for minor improvements in others. While Machine Learning has endless potential to solve problems such as mitigating the spread of invasive plant species, the environmental impacts of training such large algorithms could soon become dangerous unless new methodology is adopted by researchers.

TABLE OF CONTENTS

SOCIOTECHNICAL SYNTHESIS

MAPPING INVASIVE PLANT SPECIES USING MACHINE LEARNING

Technical advisor: Madhav Marathe, Department of Computer Science

ENVIRONMENTAL COSTS OF MACHINE LEARNING ALGORITHMS: REDUCING UNNECESSARY COMPUTATIONS

STS advisor: Catherine D. Baritaud, Department of Engineering and Society

PROSPECTUS

Technical advisor: Madhav Marathe, Department of Computer Science;

STS advisor: Catherine D. Baritaud, Department of Engineering and Society