

Automated classification of a tropical landscape infested by Parthenium weed - Zolo Kiala, University of KwaZulu-Natal, Pietermaritzburg, South Africa

Zolo Kiala

¹University of KwaZulu-Natal, Pietermaritzburg, South Africa

The invasive Parthenium weed (*Parthenium hysterophorus*) adversely affects animal and human health, agricultural productivity, rural livelihoods, local and national economies, and the environment. Its fast spreading capability requires consistent monitoring for adoption of relevant mitigation approaches, potentially through remote sensing. To date, studies that have endeavoured to map the Parthenium weed have commonly used popular classification algorithms that include Support vector machines and Random forest classifiers, which do not capture the complex structural characteristics of the weed. Furthermore, determination of site or data specific algorithms, often achieved through intensive comparison of algorithms, is often laborious and time consuming. Also, selected algorithms may not be optimal on datasets collected in other sites. Hence, this study adopted the Tree-based Pipeline Optimization Tool (TPOT), an automated machine learning approach that can be used to overcome high data variability during the classification process. Using Sentinel-2 and Landsat 8 imagery to map Parthenium weed, we compared the outcome of the TPOT to the best performing and optimized algorithm selected from sixteen classifiers on different training datasets. Results showed that the TPOT model yielded a higher overall classification accuracy (88.15%) using Sentinel-2 and 74 % using Landsat 8, accuracies that were higher than the commonly used robust classifiers. This study is the first to demonstrate the value of TPOT in mapping Parthenium weed infestations using satellite imagery. Its adoption would

therefore be useful in limiting human intervention while optimising classification accuracies for mapping invasive plants. Based on these findings, we propose TPOT as an efficient method for selecting and tuning algorithms for Parthenium discrimination and monitoring, and indeed general vegetation mapping.

Tree-based Pipeline Optimization Tool (TPOT) (Olson and Moore 2016) is a novel AutoML that applies genetic programming (GP) to optimize machine learning pipelines of the sklearn python library for classification and regression problems. The following pipeline operators are implemented in TPOT: pre-processors, decomposition, feature selection, and models. During the optimization process, subsets of the ML pipelines are defined as GP primitives, which are organized as in a tree structure to form individuals. To obtain the optimal combination of processes, GP optimizes the number and order of pipeline operators, as well as each operator's parameters (Sohn, Olson, and Moore 2017). More details on the tool can be found in (Olson and Moore 2016). An example of TPOT workflow is illustrated in Figure 2. In this study, the choice of TPOT parameters was premised on the assumption that better results are achieved with more central processing unit (CPU) time allocated (Hutter, Kotthoff, and Vanschoren 2019). Therefore, parameters such as 'generations', 'population_size' and 'verbosity' were set to 500, 100 and 2 respectively. Furthermore a 'random_state' parameter was added to the python code containing the best pipeline generated by TPOT to allow replication