

Temporal Assessment of Drought Stress Progression through Large-Scale Machine- Learning-Based Phenotyping

Jinyoung Barnaby¹, Scott E Warnke¹, and Yonghyun Kim^{1,2}

¹U.S. Department of Agriculture, Agricultural Research Service, Northeast area, U.S. National Arboretum, Floral and Nursery Plants Research Unit

²Oak Ridge Institute for Science and Education

January 14, 2023



Temporal Assessment of Drought Stress Progression through Large-Scale Machine-Learning-Based Phenotyping

Jinyoung Y. Barnaby¹, Scott E. Warnke¹, and Yonghyun Kim^{1,2}

¹*U.S. Department of Agriculture, Agricultural Research Service, Northeast area, U.S. National Arboretum, Floral and Nursery Plants Research Unit, Beltsville, MD 20705*

²*Oak Ridge Institute for Science and Education*

ORCID: [0000-0001-6507-9985 of Jinyoung Y. Barnaby], [0000-0001-9082-6583 of Scott E. Warnke]

Keywords: Turfgrass, High-Throughput Digital Phenotyping, Machine Learning, Drought, Stress Severity, Stress Progression, Mapping Population

Precise assessment of large mapping populations, comprising a few thousand plants including replications (a prerequisite step for breeding) is time-consuming and labor-intensive. Furthermore, phenotyping results tend to be variable and subjective depending on who is doing the scoring. One way to overcome these limitations is by collecting more data in the form of digital images, and precisely evaluating phenotypic variation in stress severity as well as temporal progression of stress symptoms within the population through machine learning methods. 230,400 images representing temporal progression of drought stress symptoms of interspecific turfgrass hybrid mapping population were processed using Python OpenCV and NumPy packages for noise removal, edge-preserving smoothing, color space conversion, contrast enhancement, and identification mapping. Then machine learning-based algorithms and models were developed not only to quantify stress severity but also to monitor temporal progression rate of stress symptoms. Hierarchical clustering was then performed to assess genotypic variation in stress progression. Such machine learning-based high-throughput digital phenotyping platforms can significantly increase the success of quantitative trait locus mapping and candidate gene identification to develop potential molecular markers that will assist in a faster characterization of germplasm to ultimately breed for stress resilient cultivars.