

UNIVERSITY OF PUERTO RICO

*Abstract*Faculty of Natural Sciences
Department of Mathematics

Master of Science

Classifying phenotypic traits from genomic data using convolutional deep learning methods

by Heriberto A. CARBIA

Identifying the genomic changes that control morphological variation has major importance to studying genetic disease as well as understanding evolutionary change. Deep learning (DL) approaches have the power to significantly improve the identification of complex genomic variation that is associated with morphological variation. Over the past decade DL has revolutionized entire fields (i.e., speech recognition, natural language processing, image classification, and bioinformatics), however, its application to problems in medical and evolutionary genetics is still in its early stages. In this work, we aimed to develop a deep learning approach with the purpose of identifying specific, complex patterns in genetic variation responsible for morphological change. More specifically, we proposed and compared several convolutional deep learning architectures for classifying phenotypic characteristics from genotypes using genomes from different color pattern variants of a group of butterflies (i.e., *Heliconius spp.*). Results from the proposed 2D and 1D convolutional architectures were then compared in terms of predictive performance. For model interpretation, gradient-based visualization techniques were able to provide key positional information on the parts of the genomic input that are used by each model to make a specific class prediction.