Diana Gentry
diana.gentry@nasa.gov
NASA Ames Research Center

2017 ASGRS (American Society for Gravitational and Space Research) Annual Meeting
Oct. 25-28, 2017
Seattle, WA
Conference Website: https://www.asgsr.org/index.php/meetings/2017-meeting

Category: Life Sciences
Abstract Title: Machine Learning Approaches to Increasing Value of Spaceflight Omics Databases

Abstract:
The number of spaceflight bioscience mission opportunities is too small to allow all relevant biological and environmental parameters to be experimentally identified. Simulated spaceflight experiments in ground-based facilities (GBFs), such as clinostats, are each suitable only for particular investigations -- a rotating-wall vessel may be 'simulated microgravity' for cell differentiation (hours), but not DNA repair (seconds) -- and introduce confounding stimuli, such as motor vibration and fluid shear effects. This uncertainty over which biological mechanisms respond to a given form of simulated space radiation or gravity, as well as its side effects, limits our ability to baseline spaceflight data and validate mission science.

Machine learning techniques autonomously identify relevant and interdependent factors in a data set given the set of desired metrics to be evaluated: to automatically identify related studies, compare data from related studies, or determine linkages between types of data in the same study. System-of-systems (SoS) machine learning models have the ability to deal with both sparse and heterogeneous data, such as that provided by the small and diverse number of space biosciences flight missions; however, they require appropriate user-defined metrics for any given data set. Although machine learning in bioinformatics is rapidly expanding, the need to combine spaceflight/GBF mission parameters with omics data is unique.

This work characterizes the basic requirements for implementing the SoS approach through the System Map (SM) technique, a composite of a dynamic Bayesian network and Gaussian mixture model, in real-world repositories such as the GeneLab Data System and Life Sciences Data Archive. The three primary steps are metadata management for experimental description using open-source ontologies, defining similarity and consistency metrics, and generating testing and validation data sets. Such approaches to spaceflight and GBF omics data may soon enable unique insight into which measured phenomena correlate to biological mechanisms that are truly affected by spaceflight conditions; which are most likely to be confounded by other variables; and which are insufficiently characterized, significantly increasing existing and future science return from ISS and spaceflight missions.