Biological Research and Space Health Enabled by Machine Learning to Support Deep Space Missions

Ryan T. Scott[1], Lauren M. Sanders[2], Sylvain V. Costes[3]

[1]KBR, Moffett Field, CA; [2]Blue Marble Space Institute of Science, Moffett Field, CA; [3]NASA Ames Research Center, Space Biosciences Division, Moffett Field, CA

A key science goal of the NASA "Moon to Mars" campaign is to understand how biology responds to the Lunar, Martian, and deep space environments in order to advance fundamental knowledge, reduce risk, and support safe, productive human space missions. Through the powerful emerging computer science approaches of artificial intelligence (AI) and machine learning (ML), a paradigm shift has begun in biomedical science and engineered astronaut health systems, to enable Earth-independence and autonomy of mission operations. We present a decadal view of AI/ML architecture to support deep space mission goals, developed in concert with leaders in the field. We describe current AI/ML methods to support 1) fundamental biology, 2) *in situ* analytics, 3) high performance computing hardware, 4) automated science, 5) self-driving labs, 6) remote data management, 7) integrated real-time mission biomonitoring, and 8) a Precision Space Health system. Cutting-edge AI/ML approaches that can be integrated to support these domains include active learning, explainable AI, adaptive learning, causal inference, knowledge graphs, federated learning, transfer learning, and large language models. Finally, we present results from several current ML projects that are underway in the field to address key challenges of small sample n, high feature count, heterogeneity, and sparse data. These include 1) connecting omics data to phenotypic data using an ensemble model to infer causality of spaceflight rodent liver health disruption, 2) usage of explainable ML to interrogate the muscular underpinnings of spaceflight muscle atrophy, 3) ML models analyzing and determining directed acyclic graphs of human space health risk leveraging rodent bone datasets, 4) usage of large pre-trained models connecting biomedical knowledgebases with small spaceflight datasets to understand gene-to-gene interaction networks, and 5) a suite of benchmarked open science datasets (spaceflight mouse liver; radiation DNA damage) enabling programmers to identify the best ML algorithms to answer space biological science questions.