Using Federated Learning to Overcome Data Gravity in Space

James Casaletto[1], Graham Mackintosh[2], Lauren M. Sanders[3], Ryan T. Scott[4], Collin Estes[4], Mark Fernandez[5], Patrick Foley[6], Nate Haynes[7], Shashi Jain[6], Sylvain V. Costes[8]

[1]Computational Genomics Lab, University of California Santa Cruz, Santa Cruz, CA; [2]Bay Area Environmental Research Institute, Moffett Field, CA; [3]Blue Marble Space Institute of Science, Moffett Field, CA; [4]KBR, Moffett Field, CA; [5] Hewlett-Packard Enterprise, [6]Intel Corporation, [7]Amazon, [8]Space Biosciences Division, NASA Ames Research Center, Moffett Field, CA

jcasalet@ucsc.edu

James Casaletto is a graduate student researcher at the University of California in Santa Cruz (UCSC). His primary research interest is in the development of analytical methods on privacy-sensitive data sets for health and medical analysis. He's also interested in statistical and machine learning algorithms to predict health outcomes. James has a bachelor's in mathematics and a master's in computer science. Prior to starting graduate school at UCSC, he worked in Silicon Valley as a software engineer and as an adjunct professor in the computer science department at San Jose State University.

Abstract. Humans intend to take longer missions to outer space. Understanding the impact that space has on human health is paramount to the success of these missions. Controlled experiments with model organisms are run to infer the impact of space conditions on human health, but data these experiments generate may be too large to transfer to Earth for building models. The same is true for space-relevant data generated on Earth. Ideally, these data should be combined to improve statistical power and model accuracy without having to transfer data. Federated learning is such a method which trains an algorithm across decentralized computing systems, each of which has their own local copy of training and testing data. In this research, made possible by NASA@Work, the AI for Life in Space group at NASA demonstrates the use of federated learning to train an ensemble of causality inference models on a combination of data residing on the International Space Station (ISS) and in the cloud. Our work leverages CRISP, a causal inference platform developed during the 2020 Frontier Development Lab's "Astronaut Health Challenge." We also leverage the OpenFL federated learning library which was collaboratively developed at Intel and UPenn. We used publicly available data from the NASA Ames Life Sciences Data Archive to identify features in ionizing radiation experiments as causal of changes in cardiac blood velocity. This research demonstrates, for the first time, the possibility of running machine learning algorithms on datasets separated by astronomical distances.

In this experiment, all the data were generated *in terra*, half of which were transferred to the ISS and analyzed on the Spaceborne Computer. In the future, our research will leverage federated learning on data generated *in situ* on the ISS with data generated terrestrially to predict the impact of spaceflight on mammalian female reproductive capacity.