FEDERATED CAUSAL INFERENCE FOR PREDICTING RADIATION EXPOSURE EFFECTS IN OUT-OF-DISTRIBUTION SETTINGS

G. Ughi*[1], L. Scheibenreif*[2], P. Duckworth*[1], O. O'Donoghue*[1], K. Khezeli[3], A. Hoarfrost[4], S. Budd[5], N. Chia[3], P. Foley[6], G. Mackintosh[4], J. Kalantari[3], F. Soboczenski[7], and L. M. Sanders[4]

[1]University of Oxford, [2]University of St. Gallen, [3]Mayo Clinic, [4]NASA Ames, [5]Imperial College London, [6]Intel, [7]King's College London

BACKGROUND

The effects of radiation exposure are a health risk for multiple populations, including cancer patients undergoing radiotherapy, and astronauts in space. In order to mitigate these effects, we must identify the specific biological responses which cause adverse outcomes. However, there are several challenges facing researchers in the field. First, it is often difficult to differentiate causal effects from merely correlated ones. Second, much relevant radiation exposure data is in rodent models rather than human patients, with ongoing research on appropriate model-human translation[1]. Third, human medical or astronaut data is scarce and existing data can be difficult to obtain due to security restrictions or bandwidth limitations. Overall, data limitations lead to combining heterogeneous data from various sources with different generating distributions. In particular, genomics or multi-omics datasets often have very few samples compared to the high dimensionality, making robust statistical inference difficult.

APPROACH

In order to address these challenges, we leveraged and expanded upon the Causal Relation and Inference Search Platform (CRISP), an ensemble learning platform with demonstrated ability to identify candidate causal biomarkers of disease in heterogeneous multi-omics data[2].

Synthetic Datasets

We developed a set of synthetic datasets with multi-class targets and Bernoulli random variables, including hidden confounders, to assess CRISP's performance on data from different distributions. We formally evaluated the efficacy of new methods in the ensemble, and the efficacy of dimensionality reduction with observations coming from different frameworks.

Human-Mouse Comparative Analysis

We evaluated the compatibility of mouse and human radiation exposure transcriptomic data from NASA GeneLab and the Gene Expression Omnibus, and assessed the consistency of causal features identified across different divisions of mouse and human data.

Federated Learning Implementation

In order to allow CRISP to run on datasets in private, secure or remote environments, we implemented federated causal inference through Intel's OpenFL project. Due to the limitations surrounding human and astronaut data, effective identification of causal drivers of radiation exposure effects benefits from training on data that are inaccessible due to privacy, security or bandwidth considerations.

RESULTS

We present an early view of the updated CRISP pipeline with federated learning capabilities for causal inference, and we report that across multiple divisions of combined mouse and human radiation exposure data, CRISP was able to consistently identify biomedically-relevant causal features, including genes previously linked to cancer pathophysiology[3].

REFERENCES

[1] McGonigle P., Ruggeri B. (2014) *Biochem Pharmacol* 2014;87: 162–171. [2] Budd S. et al (2021) *LifeTech* 517–521. [3] Jin H. et al (2021) *Head Face Med* 17–20 .