TuftsMedicine
Tufts Medical Center
NEWDIGS

LEAPS Methods Innovation Team

Case Study Background

For the initial case study, develop a predictive model(s) that can improve decision-making for all stakeholders related to immune checkpoint inhibitor use in patients with advanced NSCLC.

Objectives

Validate the potential of using federated (machine) learning methods, leveraging diverse data types (e.g., EHR, claims, social determinants of health, biologic, clinical trials, patient-generated, etc.) to:

- Identify signals, generate hypotheses about clinically meaningful sub-populations
- Define next step in corroborating/validating promising hypotheses
- Reduce bias in algorithm development through the use of diverse data sets
- Establish federated learning environment (technology enablers, cross-functional expertise, governance) that is scalable

Purpose of the Federated Learning Network Outline

- Discuss the different types & platforms of federated (machine) learning (FL) to assist in the selection of the most appropriate types for initial prototyping
- Explore & define the right structure of the centralized server for successful implementation of a federated learning network prototype for Case Study #1 (immune-checkpoint inhibitors in advanced NSCLC)
- Demonstrate how Federated Learning expands the understanding of the impact of diseases in the broader population beyond structured information from EHR and insurance databases and addresses healthcare disparities by utilizing non-traditional data sources including patient generated information and community initiatives supporting the underinsured.

Application and Approach

The Federated Learning Network Outline seeks to identify and characterize the strengths and limitations of available FL types and platforms that could be selected for application to the Advanced NSCLC Use Case specifically and more generally to other use cases as identified by the LEAPS team. In addition to providing a general

description of the available FL types and platforms, this outline examines both in the context of other key criteria relevant to the implementation of a network prototype.

Criteria for Evaluation

- How do we demonstrate that FL is or could be better than the status quo approaches?

- What are the existing gaps in data analysis from traditional databases that reflect only the insured and limited demographics that can be filled by access and utilization from a broad collection of diverse data types that better reflect the general population and have not been utilized in traditional analysis?

- What are the technical and technology needs for implementing FL among data partners?
    - Is the solution better overall for each institution?
    - Is it that it is more generalizable (ex. performance holds even when moving to a new institution that was not part of the initial federated learning network)?

- Centralized server approach vs. a decentralized approach
    - For the sake of simplicity, is the decentralized approach the best way to start?
    - Are there more legal and privacy issues to be considered and addressed with this approach? Even with de-identified data?

- Data exclusion cycle

- What other criteria for consideration?

Commented [HKA1]: For discussion: Need to add more detail as to what the condition or situation we need to take into consideration here is with respect to data exclusion cycle

Appendix

Table 1: List of Federated Learning types under consideration

Type	Description	Unique Feature(s)	Strengths	Limitations
Horizontal				
Vertical				
Transfer				

Table 2: Aggregator vs. Peer-to-Peer

Type	Description	Unique Feature(s)	Strengths	Limitations
Aggregator				
Peer-to-Peer