Activity:

Revisiting the Model Development Process

For each of the following 4 prediction tasks:

- Each group will discuss one component of model development. These will rotate for each task.

  1. The general approach: what model(s) should be explored, and why?
  2. Data management: what data pre-processing or augmentation steps would you recommend?
  3. Training and evaluation process: how should the data be divided, how will each portion be used, and what (if any) hyperparameters should be tuned?
  4. Toward clinical use: what performance measures are most important to report, and will it be important to interpret or explain model predictions?

- The first team listed for each component will present first. The second team listed for that component will add any other relevant details.

Part 1: Personalized treatment response

Your team wishes to predict which patients will respond to a new drug based on their genetic profile. To build your model, you have RNA-Seq data + treatment outcomes (responder or non-responder) in 1000 patients who participated in a recent clinical trial of the drug.

- Teams 1&2: general approach
- Teams 3&4: pre-processing
- Teams 5&6: training process
- Teams 7&8: evaluation and interpretation

Part 2: Identifying clinician fatigue

Your team is concerned about the effect of fatigue on clinician performance. As a first step, you decide to develop a model that can identify fatigue based on clinicians' notes. To build your model, you collect 100,000 clinical notes along with corresponding information about whether the clinician who wrote them had been working for 0-8 hours, 8-16 hours, or >16 hours.

- Teams 5&6: general approach
- Teams 7&8: pre-processing
- Teams 1&2: training process
- Teams 3&4: evaluation and interpretation

Part 3: Surgery for small bowel obstruction

Your team wishes to promptly identify individuals who will require surgery for SBO based on information collected shortly after admission. You believe both imaging data (abdominal series) and clinical variables (e.g. vitals, labs, stool and urine volume) are important. To build your model, you obtain a clinical snapshot (imaging + clinical variables) along with the binary outcome (0: survived without surgery; 1: required surgery or died) in 10,000 patients.

- Teams 4&3: general approach
- Teams 6&5: pre-processing
- Teams 8&7: training process
- Teams 2&1: evaluation and interpretation

Part 4: MS Subtypes

You would like to develop a model that predicts multiple sclerosis disease course (CIS, RRMS, SPMS, PPMS) based on imaging (MRI) at initial diagnosis. However, you only have access to data (MRI + disease course) for 90 patients from your practice.

- Teams 8&7: general approach
- Teams 2&1: pre-processing
- Teams 4&3: training process
- Teams 6&5: evaluation and interpretation