Subglacial hydrology connects glacial and oceanic systems, modulates ice dynamics, and remains a major physical uncertainty in future ice-sheet projections. Subglacial water is often stored in subglacial lakes, some of which are *active*, episodically filling and draining on short timescales. Lake drain-fill cycles cause changes in water distribution, grounding-zone stability, freshwater flux, and nutrient and carbon export into the Southern Ocean and sub-ice-shelf cavities. Traditional methods for identifying subglacial lakes can be time-consuming and expensive, so this project proposes an alternative way to identify subglacial lakes using a supervised neural network via the PyTorch package. The goal is to input geological data that covers Antarctica, such as hydropotential and bed elevation, and output a zero or one classification metric, where zero represents a point that is not a lake and one represents a lake point. The neural network will take in these geological features, and perform feature engineering to create new features using linear layers. Then the neural network will use more linear layers to perform feature extraction and reduce the dimensionality of our dataset until we have one feature. This final feature represents the output (logit) of our neural network, but to use this logit value to classify points, we need to apply an activation function that transforms the logit into a probability, and then apply a threshold that determines if a point is a lake or not. For example, if the probability is 0.25, then the point would be classified as not a lake.

I will be using two separate datasets. The first is BedMachine version 3 (Antarctica), a dataset from the NASA MEaSUREs program, which looks at the properties of glaciers. The data set provides a bed topography map of Antarctica along with firn air content (the compacted snow above a glacier), glacier/ice sheet thickness, and ice surface elevation. The second dataset is a collection of features that were used in a random forest to model the likelihood of a sedimentary basin (Li, 2022). The idea behind using this dataset is to have a set of features that relate to bed composition, which might relate to the existence of a subglacial lake. There are 20 features within this dataset, so I will need to perform some data preprocessing and analysis to determine which ones are best. All the features mentioned above have different ranges, for example, firn ranges from 0 to 38, meanwhile, bed elevation ranges from -2394 to 2527, thus the data will need to be scaled before model training and testing. Additionally, it is not possible to get a complete dataset that has values for every point on the Antarctic content, so the above datasets contain NaN values, which we will need to replace with global means. Replacing the NaN values with a global mean is preferred over removing the rows because there are only 131 known subglacial lakes.

For the week of 11/10-11/17, the goal is to have all data read in and processed. The next week, 11/17-11/24, will be spent creating the model and doing some tuning. The final week before presentations are due, 11/24-12/1 will be spent tuning the model and crafting the final class deliverables. The most difficult task with this project is creating code and a collection of datasets that can be easily reproduced, so throughout each week listed above, I will need to ensure that my code can be submitted and run by another individual.

Li, L., Aitken, A.R.A., Lindsay, M.D. et al. Sedimentary Basins Reduce Stability of Antarctic Ice Streams Through Groundwater Feedbacks. Nat. Geosci. 15, 645–650 (2022). https://doi.org/10.1038/s41561-022-00992-5