# AI for Climate Change
# Laboratory 2

### Dataset Identification

- The dataset used for analysis is *"GLIMS_Full_Dataset.csv"*. It contains information about glacier topography.
- The dataset used in GEE is ***LANDSAT 8*** satellite images. These are multi-spectral images with 7 bands (0.43m to 2.29m, covering visible, near-infrared, and shortwave light).

### Dataset Collection

The dataset was downloaded from a public platform (Global Land Ice Measurements from Space).

### Dataset Understanding

The dataset contains information about glaciers, including their location, size, and other descriptive parameters. Each row represents an individual glacier.

### Exploratory Data Analysis of Data Used as Input for the Project

### Data Description

The `describe()` function in Python was applied to the dataset to obtain descriptive statistics of the numerical columns. The results include:

- Total number of records: 1,186,805
- Mean values, standard deviations, minimum values, quartiles, and maximum values for the numerical columns (e.g., `area`, `db_area`, `width`).

### Handling Missing Data

- Columns with missing values were identified. The strategy applied included removing rows with missing values in key columns.

### Handling Outliers

- **BoxPlot:** Used to visualize the distribution of the data and identify potential outliers in the numerical columns.

### Understanding Relationships and New Insights Through Plots

- **Histogram:** Used to visualize the frequency distribution of values for the numerical columns.
- **Heatmap:** Used to visualize the correlation between the different columns of the dataset.