

# Supervised Learning in Practice

---

March 3, 2026

# Announcements

---

- + Assignment deadlines are now **11:59pm**
- + **Final Project Checkpoint 1** (1-page proposal) due this **Friday, March 6 at 11:59pm**
- + **Lab 3 Notes**
  - + Due March 6 at **11:59pm**
  - + “Free” 3-day late policy can be used to turn in this lab on the Monday after Spring Break (but please email me if you plan to do this)
  - + You are allowed to save your model results to a file/disk and read them in
    - + Code to fit the models should be provided (e.g., set "#| eval: false" for that code chunk)
    - + If you do this, upload your saved results to GitHub (so that I can also read them in)

# Today's plan

---

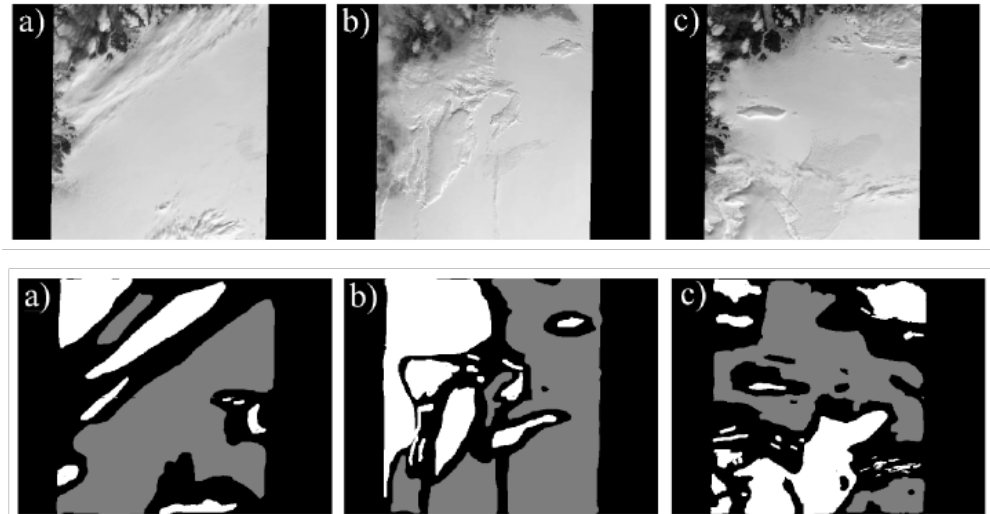
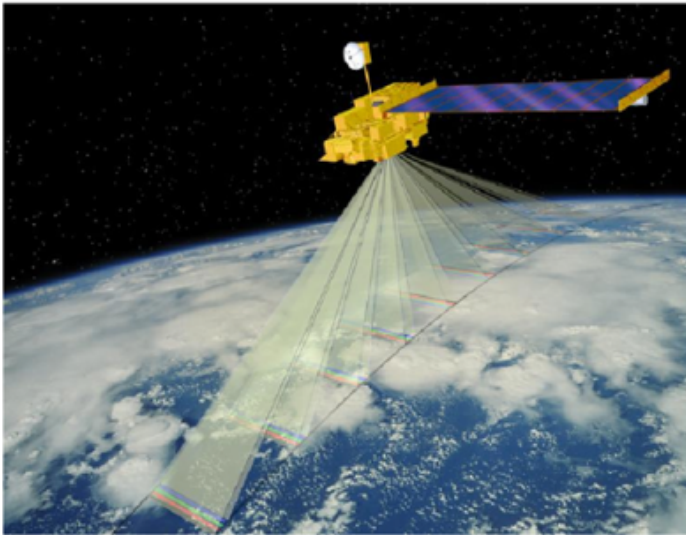
- 1 Lab 3 Data Splitting**
- 2 Hands-on Practice with Cancer Mortality Data**

# Lab 3 Data Splitting

---

# Lab 3: Remote sensing for cloud detection

**Goal:** predict whether each pixel is a cloud or ice (from a glacier) on completely new satellite images

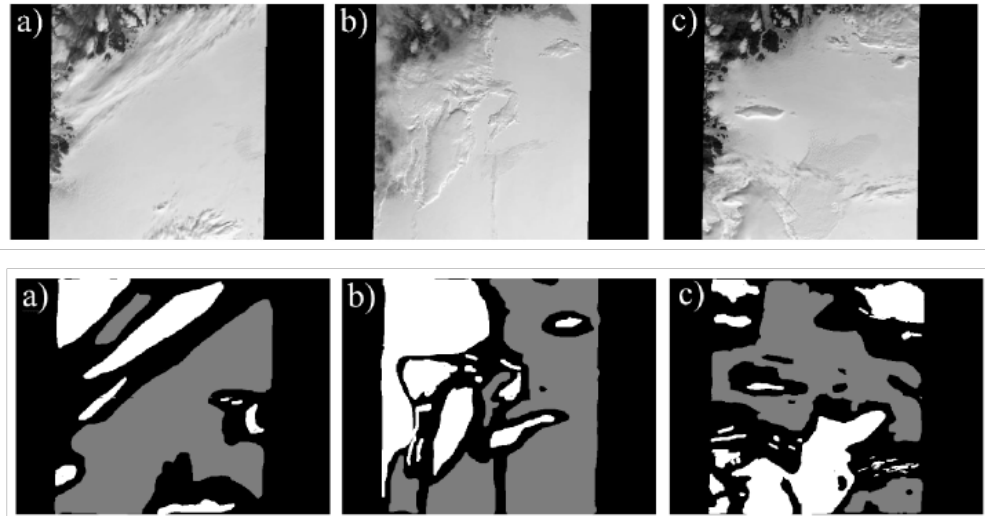


**Original study:** Shi et al. [Daytime Arctic Cloud Detection Based on Multi-Angle Satellite Data with Case Studies](#)

# In-Class Lab 3 Activity Recap

In Lab 3, you are asked to choose between (at least) 2 models, where (at least) one requires hyperparameter tuning. You are also asked to report how well you think your best model would perform on new images.

- + Do we need a training-test split or a training-validation-test split?
- + What data do you plan to use for hyperparameter tuning?
- + Should we allocate data into different splits (or folds) randomly?



# Today: Cancer Mortality Data

---

# In-Class Activity: Cancer Mortality

---

Open:

```
course_materials/cancer_mortality/supervised_learning_*.qmd
```

Work in groups to:

- + Build your best model to predict cancer mortality rates
- + Ultimate objective: output test predictions to file
- + I will release test responses at the end of class so that we can see how well you did.