# Current Topics in Computer Science: Geospatial and Statistical Machine Learning

Fall 2024 at CU Boulder
This is a draft document and information is subject to change

## CSCI 7000 - TBD unit

Tue/Thur 12:30-1:45 pm
Room: TBD
Instructor: Esther Rolf
Office hours: TBD

## Course description

Statistical machine learning has been used to extract patterns and fill in gaps in geospatial data in consequential *applications* ranging from monitoring ecological phenomena (e.g., forest cover, land use) to estimating socio-economic conditions (e.g., population density, well-being). Behind the success of these applications of geospatial machine learning (GeoML) are the *methods* that fuse knowledge and recent research in machine learning, geostatistics, and remote sensing, among other fields. In this course, students will learn fundamentals of GeoML and orient novel research within recent advances in geospatial and statistical ML. Topics include: machine learning with satellite data, geospatial deep learning architectures and libraries, spatial statistics and model evaluation, application areas and the "duality" of potential impacts of GeoML technologies.

## Course objectives

This course will prepare students to:

- Understand the established practices and recent research in geospatial ML, with emphasis on geospatial statistics and ML with satellite data.
- Contribute new research ideas in geospatial and statistical ML, and be able to efficiently and iteratively improve on these ideas. This will include:
    - Working with geospatial deep learning libraries to prototype and implement algorithmic ideas.
    - Working with formal statistical characterisations of geospatial data to describe key aspects of geospatial ML (including spatial autocorrelations and spatial domain adaptation).
- Communicate scientific results through research presentations and written reports.
- Deliver critical and constructive feedback on research papers, relevant to the research skills of paper reviewing, peer-editing, and self editing.

## Prerequisites

- Machine learning (e.g. CSCI 5622)
- Experience with python is highly encouraged

## Deliverables

- **Research Project.** In teams of 1-3, students will plan and execute a research project in a topic of their choosing relevant to the themes of this course. This will consistent of several deliverables:
    - Project proposal, detailing the ideas and goals of the project, so that I can provide feedback on the proposed project and expectations. Due a few weeks into the semester.
    - An intermediate report with preliminary findings, between 2-4 pages. Due mid-semester.
    - A final research report, 4 pages (excluding references), in a "workshop" style template, due at the end of the semester.
- **Peer feedback on mid-semester report.** Students will provide peer-feedback on other teams' mid-semester reports, and suggest avenues to improve the research in the second half of the semester.
- **Homeworks.** In the first half of the semester, there will be two homeworks – one coding based, and one math based. These homeworks are designed to be relatively "lightweight" assignments to refresh the ML skills that will be helpful for the research projects. That said, start early! These may take a while if students need to brush up on coding/mathematical background along the way.
- **Paper presentations.** Throughout the semester, students will deliver 30-45 minute presentations on recent research in geospatial and statistical ML. Papers will be assigned based on student preference, and each student will be expected to present 1 time per semester.
- **Readings/ class participation.** All students are expected to contribute to the discussion of the paper lectures. At a minimum, this means reading the required papers for each week, and submitting at least one question about the research *before class begins.*

*Note that the exact parameters of these deliverables are subject to change, depending on class enrollment.

## Grading

A rough breakdown (subject to change) for grading in:

- Project (40)% broken down into: Proposal (5%), Mid-point report (10%), Final report (25%)
- Homeworks (20%): 10% each
- In-class paper presentation (20%)
    - Graded by quality, we will discuss the rubric in class
- Peer feedback on midterm report (10%)
    - Graded by quality, we will discuss the rubric in class
- Paper readings: sending in questions before class (10%)
    - Due before every class with a student presentation; 2 "freebie" misses