EAS 4803 / 8803

Machine Learning for Earth & Environmental Sciences

Spring Semester | 3 Credit Hours | Georgia Institute of Technology

Instructor	Dr. Ali Sarhadi	Email	sarhadi@gatech.edu
Office	ES&T 3244	Office Hours	Tuesday, 2:00–3:00 PM, ES&T 3244
Class Meeting	TTh, 12:30–1:45 PM, ES&T L1105	TA Office Hours	TBD

General Course Information

Course Description

This course provides a comprehensive introduction to modern machine learning and deep learning methods, with a particular emphasis on their application to Earth, environmental, and climate systems. Students will learn both foundational and advanced ML/AI algorithms, focusing on how data-driven approaches can be integrated with physical understanding to analyze complex geophysical processes and address critical real-world challenges. Through a combination of theory and hands-on implementation, students will develop the skills to design, implement, and rigorously evaluate models for large and high-dimensional environmental datasets. The course emphasizes prediction, interpretability, and decision-making in Earth systems, preparing students for advanced research and practical applications in environmental science, climate, engineering, and related fields.

Course Learning Outcomes

Upon successful completion of this course, students will be able to:

- **Understand and apply core ML/AI algorithms**, including supervised, unsupervised, and deep learning methods, with a clear grasp of their theoretical foundations and limitations in Earth and environmental systems.
- **Analyze and process large, high-dimensional Earth and environmental datasets**, using modern ML/AI tools and Python-based frameworks for data preprocessing, feature engineering, and model evaluation.
- **Design, implement, and rigorously evaluate ML/AI models**, including deep learning and physics-informed approaches, to improve prediction, forecasting, and understanding of complex geophysical processes.
- **Interpret and explain ML/AI model outputs**, using techniques such as feature attribution and causal inference to ensure transparency and scientific insight in environmental applications.
- **Develop and communicate an end-to-end ML/AI research project**, including problem formulation, data integration, model development, validation, and critical assessment for real-world environmental and climate challenges.

Prerequisites

Students are expected to have prior coursework in statistics, probability, or a related quantitative field, as well as a working knowledge of linear algebra. Familiarity with concepts such as regression, probability distributions, and basic statistical inference is required for successful completion of the course.

Basic proficiency in Python programming is expected. Students without prior experience are strongly encouraged to complete an introductory Python course or tutorial before the start of the semester. Students may use R for assignments if preferred; however, all code must be written and implemented independently.

Course Structure & Format

Classes will consist of lectures, hands-on coding sessions in Python, and collaborative discussions. Lectures will cover the theoretical aspects of machine learning, supplemented with practical coding examples to help students learn key steps in the development of machine learning models. Assignments will assess students' understanding of fundamental concepts and their coding proficiency in solving provided problems. All students will be required to implement code individually. For the final project, students may work individually or in groups of up to two to develop a research project, culminating in a final report and presentation that demonstrates the application of machine learning to an environmental or earth system issue.

Required Course Materials

The course will provide the following materials to support your learning:

- **Lecture slides:** Slides will be uploaded to Canvas after each lecture for reference and review.
- **Practical code examples:** For each topic, practical coding examples will be provided to help students implement the concepts learned in class. These will be available on Canvas and serve as a guide for assignments and projects.

The following textbooks are recommended for students who wish to explore additional resources:

- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- Chollet, F. (2018). *Deep Learning with Python*. Manning Publications.
- Chollet, F., & Allaire, J. J. (2018). *Deep Learning with R*. Manning Publications.

Grading Policy

Student performance will be evaluated based on demonstrated understanding of machine learning concepts, coding proficiency, and the ability to apply ML methods to real-world earth and environmental problems. Final grades are awarded on a scale of A–F with no +/– grades permitted at Georgia Tech.

A>90; B>80; C>70; D>60

Component	Weight
Assignments (multiple throughout semester)	**40%**
In-Class Activities (paper reviews, discussions)	**10%**
Research Proposal	**10%**
Oral Presentation (final project)	**20%**
Final Report	**20%**

Description of Graded Components

Assignments (40%)

Throughout the semester, multiple assignments will be given to students via Canvas. These assignments will cover both the fundamental concepts taught in class and practical coding exercises. Each assignment will focus on applying ML/AI methods to solve real-world issues, offering students the opportunity to practice coding and optimize solutions using the techniques covered in the course.

In-Class Activities (10%)

In-class activities will include group-based paper reviews, where students will collaborate to analyze and discuss key findings from assigned research papers. Each group will then share their insights with the class in a jigsaw format, encouraging peer learning and discussion. Active participation and engagement during these sessions will also contribute to the in-class activity grade.

Research Proposal, Final Report, and Presentation (50%)

Students may work individually or in groups of up to two to develop an original research project on a topic related to the application of machine learning in any discipline of earth, environmental, or atmospheric sciences. Projects must be unique and not part of another course or research project, though proposals related to thesis work are encouraged.

Proposal (10%)

Students will develop an original research proposal that includes: (1) a clear statement of the research question, its machine learning component, and its contribution to the literature; (2) a description of the data required, including data acquisition plans or an actual dataset; and (3) a detailed methodology for data analysis, including the proposed machine learning algorithm(s), justification, alternative approaches, and potential challenges. The proposal should not exceed 1 single-spaced page using 12pt font.

Oral Presentation (20%)

Students will deliver an oral presentation of their project during the final class session. The presentation should summarize the key ideas, methodology, preliminary results, and conclusions. Evaluation will emphasize clarity, technical accuracy, and effective communication.

Final Report (20%)

The final report will build on the oral presentation and include detailed results of the data analysis, including the methods applied and an evaluation of the model’s performance. The report should also address challenges faced during the project and possible improvements. The report should not exceed 8 single-spaced pages.

Bonus Points: Students who complete the Course-Instructor Opinion Survey (CIOS) will earn 1–2 bonus points toward their final grade.

The grading scale will be based on overall class performance. The instructor reserves the right to adjust the scale if necessary to ensure a fair and balanced distribution of grades.

Course Outline

This is a tentative schedule, and topics may be adjusted as the course progresses.

Week	Topic	Details
1	**Introduction to Machine Learning**	• Overview of machine learning concepts and applications in Earth and environmental systems • Supervised, unsupervised learning paradigms • Data preprocessing, feature engineering, and model evaluation fundamentals
2	**Supervised Learning: Regression**	• Linear regression, regularization (Ridge, LASSO), polynomial regression • Bias-variance tradeoff and model selection • Applications to environmental data
3	**Supervised Learning: Nonlinear Methods**	• k-Nearest Neighbors (KNN), decision trees • Overfitting, pruning, and cross-validation • Applying nonlinear classifiers to earth science datasets
4	**Ensemble Learning and Support Vector Machines**	• Random forests, gradient boosting, support vector machines (SVMs) • Hyperparameter tuning and feature importance • Environmental hazard classification applications
5	**Neural Networks**	• Fundamentals of artificial neural networks, feedforward networks, backpropagation • Activation functions, loss functions, and optimization • Training deep networks: batch normalization, dropout, regularization
6	**Deep Learning I**	• Convolutional neural networks (CNNs), recurrent neural networks (RNNs) • CNNs for spatial data and image-based earth observations • Foundation models and transfer learning for environmental data
7	**Deep Learning II**	• RNNs and LSTMs for time-series environmental modeling
8	**Deep Learning III**	• Attention mechanisms and transformer architectures
9	**Representation Learning**	• Autoencoders and latent space learning • Variational autoencoders (VAEs) for data generation and anomaly detection • Dimensionality reduction and feature extraction for earth system data
10	**Spring / Fall Break**	• No class — Break
11	**Generative Models I**	• Generative Adversarial Networks (GANs) • Applications in climate downscaling and synthetic data generation • Evaluation metrics for generative models
12	**Generative Models II**	• Diffusion models and denoising diffusion probabilistic models (DDPMs) • Score-based generative modeling • Applications in weather and climate emulation
13	**Physics-Informed Machine Learning**	• Physics-informed neural networks (PINNs) and hybrid modeling • Embedding physical constraints and conservation laws into ML models

		• Applications in fluid dynamics, hydrology, and climate systems
14	**Reinforcement Learning**	• Deep reinforcement learning for decision-making in environmental systems • Markov decision processes and policy optimization
15	**Interpretable ML & Causal Inference**	• Model interpretability (e.g., SHAP), feature attribution, causal inference • Explainability methods for black-box models • Causal discovery and counterfactual reasoning in earth sciences
16	**Project Presentations**	• Student presentations and discussion of applied ML/AI projects in environmental sciences

Course Policies

Attendance and Participation

Regular attendance is expected. Students are responsible for all material covered in class, including announcements, schedule changes, and in-class activities. In-class activities cannot be made up except in cases of documented illness or emergency. Review Institute expectations related to attendance at catalog.gatech.edu/rules/4/.

Academic Integrity

Georgia Tech aims to cultivate a community based on trust, academic integrity, and honor. Students are expected to act according to the highest ethical standards. Review Georgia Tech's Honor Code and the student Code of Conduct.

Any student suspected of cheating or plagiarism on a quiz, exam, or assignment will be reported to the Office of Student Integrity, who will investigate the incident and identify the appropriate penalty for violations.

Collaboration, Group Work, and Use of Generative AI

AI-based assistance (e.g., ChatGPT, Copilot, Claude) is treated similarly to collaboration with other individuals. Students may use these tools to discuss ideas, debug code, and gain conceptual understanding. However, all submitted work must be the student's own original work.

Students must not include in any assignment (including code, text, figures, tables, or slide content) material that was not created directly by them without proper citation. Submitting uncited material generated by another person or an AI tool constitutes academic misconduct.

Acceptable uses include:

- Debugging code and understanding errors
- Clarifying concepts or algorithms
- Brainstorming approaches

Unacceptable uses include:

- Generating and submitting complete solutions (code, reports, or figures)
- Copying or slightly modifying AI-generated content
- Using AI tools in place of learning and implementation

Students must clearly disclose any use of AI tools in their assignments, including how the tool was used. Students must be able to explain and reproduce all submitted work, including code, models, and results.

Accommodations for Students with Disabilities

If you are a student with learning needs that require special accommodation, contact the Office of Disability Services (404-894-2563) as soon as possible to make an appointment to discuss your special needs and to obtain an accommodations letter. Please also e-mail the instructor as soon as possible to set up a time to discuss your learning needs.

Student-Faculty Expectations Agreement

At Georgia Tech, we believe that it is important to strive for an atmosphere of mutual respect, acknowledgement, and responsibility between faculty members and the student body. The Student-Faculty Expectations articulate basic expectations that you can have of the instructor and that the instructor has of you. All members of this class are expected to contribute to a respectful, welcoming, and inclusive environment.

Extensions, Late Assignments, and Missed Deadlines

Late assignments will be penalized unless an extension is arranged in advance with the instructor. Extensions may be granted for documented illness, emergency, or approved Institute activities. Requests should be made prior to the assignment deadline whenever possible. Final project deadlines (proposal, report, and presentation) cannot be extended without prior written approval.

Inclement Weather and Digital Learning Days

If a weather-related event affects campus operations, the instructor has the discretion to cancel class or pivot to digital instruction. Students will be notified via Canvas of any such changes. Please monitor Canvas and your Georgia Tech email for updates.

Campus Resources for Students

Graduate Student Academic and Professional Success Resources

A list of resources for graduate students is given on the Office of Graduate and Postdoctoral Education website, including academic resources, student resources, and professional development.

Undergraduate Student Academic Success Resources

A list of resources for undergraduate students' academic success and information about advising can be found at Success at Tech.

Student Well-Being

At Georgia Tech, we are concerned about your overall physical, social, and mental well-being. A comprehensive list of wellness-related resources has been compiled and maintained by the Office of the Vice President for Student Engagement and Well-being.

Online Course Evaluation

Students are encouraged to provide professional and constructive feedback on the quality of instruction in this course by completing the online course evaluations via the Georgia Tech Course Instructor Opinion Survey (CIOS). You will be notified when the evaluation period opens, and evaluations can be submitted through the email invitation you receive, directly in Canvas under the CIOS tab, or via the Georgia Tech CIOS website at academiceffectiveness.gatech.edu/surveys/cios. Students who complete the CIOS will earn 1–2 bonus points toward their final grade.