University of Texas at El Paso
School of Sciences
Department of Earth, Environmental and Resource Sciences
Face-to-Face Course Syllabus with the Integration of UTEP Edge Practices

Machine Learning in Geoscience

Course Information

GEOL4315-008, GEOL5324-001, GEOL6324-001: Machine Learning in Geoscience
CRN: 16743, 17498, 17499
Term: Fall 2024
Delivery Method: In-person
Meeting Day and Time: Tuesdays 10:30 am- 12:20 pm
Thursdays 10:30 am – 1:20 pm
Location: Geology Building, Room 308 or 318 (See schedule)

Instructor Information

Instructor: Hernan A. Moreno, Ph.D
Email: moreno@utep.edu
Office: GEOL321
Written Communication: Email
Office Hours: TR 3:00 PM- 4:00 PM at Office or by appointment
Teaching Assistant: Stephanie Marquez
TA Email: snmarquez@miners.utep.edu
Teaching Assistant Hours: T 12:20- 1:20 PM and R 1:20 – 2:20 PM @ Classroom

Course Description

This course provides knowledge, skills, and tools for data science and machine learning applied to earth, environmental sciences, and engineering. Students are expected to have a minimum background in linear algebra, statistics, probability, geospatial methods and computer coding. The course covers a selection of supervised learning methods from the most fundamental (K-Nearest Neighbors, Decision Trees, and Linear and Logistic Regression) to more advanced methods (Random Forests, Boosting, Support Vector Machines, Deep, Convolutional, Recurrent, and Transformer Neural Networks). The course also offers fundamental knowledge of useful concepts like loss and cost functions, maximum likelihood, bias-variance decomposition, ensemble averaging, kernels, Bayesian approaches, and useful techniques such as regularization, cross-validation, evaluation metrics, and optimization. Along with the theoretical part, laboratory exercises and paper discussions will be developed so that students get hands-on experience and see the learned concepts applied to earth science and engineering. Labs will be developed using Python (e.g., Jupiter Notebook on Google Collab) as the main programming language. Graduate students will develop and present a final semester project at the end of the course to apply one or more ML techniques to a science or engineering problem.

Learning Outcomes/Course Objectives

Students completing this course will be able to:

- Describe the fundamentals of classical learning and its applications.
- Understand model validation and testing as crucial steps to creating model confidence.
- Understand deep neural networks and their variants (CNN, RNN, Transformers and Generative).

By the end of the course, students will be able to:

Student Learning Objective	Outcome
Draw on the knowledge presented during lectures to create "new" or "transformed" knowledge during the laboratory sessions and course assignments.	Critical Thinking Skills
Express scientific ideas and solutions: 1. In verbal mode via paper review presentations of a topic in machine learning with follow-up peer discussions. 2. In writing mode via laboratory assignments and final project paper.	Communication Skills
Apply the learned concepts during the course to a new, open problem of self-interest to come up with a solution via ML/AI	Creative Activities
Adapt and use the learned algorithms to their own research for exploration, development, and creation of new knowledge via the final project.	Research and Scholarship

Reference Textbook & Course Materials

The recommended, but not mandatory book is:

- **Geron, Aurelien (2022).** Hands-On Machine Learning with Scikit-Learn & TensorFlow, 3rd Edition. O`Reilly. ISBN: 9781098125974. FREE WEB VERSION TO 2nd EDITION HERE: https://powerunit-ju.com/wp-content/uploads/2021/04/Aurelien-Geron-Hands-On-Machine-Learning-with-Scikit-Learn-Keras-and-Tensorflow_-Concepts-Tools-and-Techniques-to-Build-Intelligent-Systems-OReilly-Media-2019.pdf

Additional textbooks:

- **Prince, Simon (2023).** Understanding Deep Learning. MIT Press. 544p. FREE WEB VERSION HERE: https://udlbook.github.io/udlbook/
- **Lindholm, A., Wahlström, N., Linsdten, F., and Schön, T. (2022)** Machine Learning: A first course for Engineers and Scientists. Cambridge University Press. 338 pp. FREE WEB VERSION HERE: https://smlbook.org/book/sml-book-draft-latest.pdf
- **Goodfellow, I., Bengio, Y. and Courville, A (2016)**. Deep Learning. MIT Press. FREE WEB VERSION HERE: https://www.deeplearningbook.org.
- **Nielsen, M**. Neural Networks and Deep Learning – A FREE ONLINE BOOK http://neuralnetworksanddeeplearning.com/index.html

- **Bishop, Christopher & Bishop, Hugh (2024).** Deep Learning, Foundations and Concepts. Springer, 649pp.
- **Murphy, K.** (2023). Probabilistic Machine Learning. MIT Press, 1319 pp.
- **Molnar, Christoph (2024)**. Interpretable Machine Learning. FREE WEB VERSION HERE: https://christophm.github.io/interpretable-ml-book/

Our UTEP Blackboard course includes lab presentations, weekly readers, laboratory handouts, and additional information. If you encounter problems accessing this course within Blackboard, please contact the UTEP helpdesk (helpdesk@utep.edu).

Attendance, Assignments, and Grading

COURSE ATTENDANCE IS MANDATORY! Students must attend every session for the full allotted time and sign a mandatory attendance list with the instructor.
Assignments for this course are assessed according to rubrics whose points are found before each question item on the laboratory handouts, paper presentations, and long-term projects. You can also find these rubrics by clicking on the appropriate assignment link in Blackboard and choosing "View Rubric" from the button beneath the Points Possible for the assignment. The grade distribution for this course is split into Laboratory assignments conducted in Python, paper reviews and presentations, and final project.

Description	No	Undergraduates	Graduates
Laboratory Assignments	8	60%	48%
Paper Presentations	1	-	12%
Long Term Project	1	40%	40%

Percent grades will be rounded to one decimal place, and letter grades will have the following equivalence:

Letter Grade	Grade Point	Percentage
A	4.0	89.5 to 100
B	3.0	79.5 to 89.4
C	2.0	69.5 to 79.4
D	1.0	59.5 to 69.4
F	0.0	59.4 to 0

Technology Requirements

- Course content, including lectures, lab handouts, and supplementary materials, are delivered via the Internet through the Blackboard learning management system. Ensure your UTEP e-mail account works and you can access the Web and a stable web browser. Google Chrome and Mozilla Firefox are the best browsers for Blackboard; other browsers may cause complications. When having technical difficulties, update your browser, clear your cache, or try switching to another browser.
- You will need access to a desktop computer or laptop, or you can borrow one for free from UTEP Technology Support. The cost-free equipment checkout program is currently enrolled students. Check https://www.utep.edu/technologysupport/tscenter/tsc_eqcheckout.html.
- If you do not have word-processing software, you can download Word and other Microsoft Office programs (including Excel, PowerPoint, Outlook, and more) for free via UTEP's Microsoft Office Portal.

- IMPORTANT: If you encounter technical difficulties beyond your scope of troubleshooting, please contact the UTEP Help Desk. They are trained specifically to assist with students' technological needs. Please do not contact me for this type of assistance. The Help Desk is much better equipped than I am to assist you!

Course Communication

Here are the ways we can keep the communication channels open:

- Office Hours: The instructor will have office hours for your questions and comments about the course. Office hours are in-person; however, you can request a virtual meeting with the instructor, and you will be sent a Zoom link. Please see the days and times at the top of this syllabus.
- Email: UTEP e-mail is the best way to contact us. We will make every attempt to respond to your e-mail within 24 hours of receipt. When emailing us, email us from your UTEP student e-mail account, and please put the course number in the subject line. In the body of your e-mail, clearly state your question. At the end of your e-mail, be sure to put your first and last name and your university identification number.
- Announcements: Check your email for Blackboard announcements for any updates, deadlines, or other important messages.

Course Structure

The course format will be lectures, lab assignments, article presentations/discussions, and a long-term project. Students will gain hands-on experience using Python for script writing and execution. However, the knowledge gained during the practical sessions will easily be transferable to other platforms like R, Julia, etc.

Lectures: The lecture sessions will include more than instructor-led discussions. The instructor expects students to attend classes and only work on course-related materials during that time. To help students gain a better insight into machine learning, the lecture session will also include additional time for application demonstrations, in-class exercises, and more. Attendance will be taken during every class and lab meeting. Participation is essential to the course and will be assessed based on in-class activities. This active learning strategy enhances your learning and allows you to reflect, elaborate, and apply.

Lab Exercises: The laboratory sessions are an essential component of machine learning training since they provide students with hands-on experience in ML coding to consolidate their understanding of theoretical concepts and analytical techniques. In this course, we will use Python (via Jupiter Notebook, Google Collab, or Jupyter Books through Kaggle) so that students can develop their labs in this popular and powerful computer language. The Labs include initial Python installation and crash courses followed by sequential application Labs. Labs 0 (Python Installation) and 1 (Python Refreshing Courses) are optional but highly recommended. Most lab work coincides with skills and concepts learned in lectures and readings. It is, therefore, vital for students to attend class so that they understand the Lab exercises and complete them on time. Students will have time to advance lab assignments during scheduled class meetings and require time outside class to complete them. Lab assignments must be turned in via Blackboard by the due date specified in the lab document, and the submitted assignment must be original work. Students should use the time devoted to the lab activities wisely and not expect to leave class early. Questions outside the classroom are welcomed as long as some effort is made to resolve the problems. Students should not expect to miss lab time and get extra help outside class to compensate.

Paper Presentation and Discussion: Graduate students will be tasked with presenting a scientific article assigned at the beginning of the semester. Presentations will be around 20 minutes long, with 10 minutes for comments and questions from the audience. A list of curated suggested articles is presented at the end of this syllabus. Students will select one of these articles and inform the professor by the second week of classes. The list is not exclusive, and students can select a similar article to present (previous agreement with the instructor).

Long-Term Project: All students will complete a research project due at the end of the semester (final phase). The project builds on skills and concepts learned throughout the entire course. Still, certain components (such as the project "outline" and identified data sources") are due by mid-semester (see tentative schedule). Please plan accordingly and start early; the project should take 20-40 hours to develop. More specific details on the project guidelines will be provided in class. Students should expect to spend time outside scheduled hours to complete the course project. Students are expected to develop straightforward science questions for paper-style work. The final project will assess the student's ability to complete a research project comparable to what a machine learning expert will see working in the field.

Deadlines

Lab Exercises: This course has an absolute due date for completing each lab assignment, as indicated in this syllabus. This is because each lab conceptually builds on previous labs, so students must stay on schedule. Labs must be submitted on time and to the correct blackboard assignment location to be considered for grading. We also encourage students to start working on labs well in advance of the assigned deadline. Note that we will not make any exceptions to this policy for any student unless there is a documented extenuating circumstance. Please plan using the course schedule at the end of this syllabus and the deadlines as posted on Blackboard.

Paper Presentation and Discussion: The assigned dates for the review paper presentation and discussion are compiled in this syllabus and must be strictly followed since each presentation builds upon the concepts discussed each week. Students need to prepare an AGU-style presentation synthesizing the main elements of the assigned article, including motivation, objectives, methods, results, and conclusions.

Long-term Project: Each of the two phases of the final project will have an assigned in-class due date. The first phase will be worth 5% of the total course grade and consist of the data for the final project and the main idea or problem to resolve. In the second phase, which is worth 35%, students will prepare a presentation compiling the work's objectives, methods, results, and conclusions.

Late and Missing Work

Late paper presentations or late laboratory and term project work submissions are accepted, <u>but 20% of the total grade will be deducted each day after an assignment deadline</u>. Make-up assignments or late submissions can only be allowed in the event of a documented medical or family emergency. If you encounter an emergency, you must notify the instructor on or before the day of the assignment or exam due date. Documentation could include a note from a physician, a hospital admittance slip, or correspondence from an academic advisor or the

Dean of students. Foreseeable excused absences, such as participation in university-sanctioned athletic or academic events, require documentation and notify the instructor at least one week in advance. For foreseeable absences, you must turn in work early rather than late. In these situations, the student must communicate and keep the instructor informed.

Alternative Means of Submitting Work in Case of Technical Issues

We suggest you submit your due work via Blackboard with plenty of time to spare if you have a technical issue with the course website, network, and/or your computer. We also suggest you save all your work (answers to discussion points, quizzes, exams, and essays) in a separate document as a backup. This way, you will have evidence that you completed the work and will not lose credit. If you are experiencing difficulties submitting your work through Blackboard, please contact the UTEP Help Desk. You can email us your backup document as a last resort.

Illness Precautions

Please stay home if you have symptoms of a communicable illness. If you are feeling unwell, please let the instructor and the TA instructor know as soon as possible so we can work on appropriate accommodations.

Excused Absences and Course Drop Policy

We will not drop you from the course. However, if you cannot successfully complete it, please let us know and then contact the Registrar's Office to initiate the drop process. If you do not, you risk receiving an "F" for the course.

Incomplete Grade Policy

Incomplete grades may be requested only in exceptional circumstances after you have completed at least half of the course requirements. Talk to the instructor immediately if you believe an incomplete is warranted. If granted, we will establish a contract of work to be completed with deadlines.

Accommodations Policy

The University is committed to providing reasonable accommodations to students with documented disabilities. Students who become pregnant may also request reasonable accommodations in accordance with state and federal laws and regulations and University policy. Accommodations that constitute undue hardship are not reasonable. To make a request, please register with the UTEP Center for Accommodations and Support Services (CASS). Contact CASS at 915-747-5148, email them at cass@utep.edu, or apply for accommodations online via the CASS portal.

Scholastic Integrity

Academic dishonesty is prohibited and is considered a violation of the UTEP Handbook of Operating Procedures. It includes but is not limited to, cheating, plagiarism, and collusion. Cheating may involve copying from or providing information to another student, possessing unauthorized materials during a test, or falsifying research data on laboratory reports.

Plagiarism occurs when someone intentionally or knowingly represents the words or ideas of another as ones' own. Collusion involves collaborating with another person to commit any academically dishonest act. Any act of academic dishonesty attempted by a UTEP student is unacceptable and will not be tolerated. All suspected violations of academic integrity at The University of Texas at El Paso must be reported to the Office of Student Conduct and Conflict Resolution (OSCCR) for possible disciplinary action. To learn more, please visit HOOP: Student Conduct and Discipline.

Notes and Lectures: Any notes, labs, or lecture materials are for personal use only, and their sale or distribution to people outside the class is not permitted.

Guidance on Artificial Intelligence

- AI tools (like ChatGPT or other tools to help programming) are allowed with proper acknowledgment.
- The use of AI technologies or automated tools, particularly generative AI such as ChatGPT or DALL-E, is only allowed with proper attribution given for its use.
- Students must properly cite and give full credit to the program used upon submission of every relevant assignment. For example, text generated using ChatGPT must be cited: Chat-GPT(version). Date of query (year/month/day). "Text of your query."
- Generated using OpenAI. https://chat.openai.com/
- A short paragraph describing how the tool(s) was/were used for the assignment must be included.

Course Resources

UTEP provides various student services and support. For a of list campus resources, please refer to the QR code below or visit https://www.utep.edu/advising/student_resources/student-success-resource-hub.html.

Tentative Schedule (subject to slight changes)

Wk	Date	Topic	Reading	Due
1	27 Aug (T) GEO318	**Introduction, Syllabus and Logistics.** **Unit 1: What is Machine Learning (Part I):** What is Machine Learning. Why use ML?. Examples of ML systems. **Lab#0. Python Installation** **Lab#1. Python Tutorial**	Syllabus Ch.1 Geron Ch.1 Lindhol Lab Handouts	Lab 0 and 1 due 09/04
	29 Aug (R) GEO308	**Unit 1: What is Machine Learning (Part II)**: Types of ML: training supervision, batch Vs online learning, instance-based Vs model-based learning, regression and classification. Main challenges of ML: insufficient quantity of training data, non-representative training data, poor-quality data, irrelevant features, overfitting and underfitting the training data.	Ch.1 Geron	
2	3 Sep (R) GEO308	**Unit 2: End-to-end-ML Project (Part I)**: Working with real data. Looking at the big picture. Getting the data. Exploring and visualizing your data to gain insights..	Ch.2 Geron	
	5 Sep (R) GEO308	**Unit 2: End-to-end-ML Project (Part II)**: Preparing your data for ML. **Lab#2: Data Engineering and Exploratory Analysis**	Ch.2 Geron Lab Handout	Lab 2 due by 5 pm 09/23
3	10 Sep (T) GEO318	**Unit 2: End-to-end-ML Project (Part III)**: Selecting and training a model. Fine-tune your model: grid search, randomized search, ensemble methods, analyzing the best models and their errors, evaluating your system on the test set.	Ch.2 Geron Lab Handout	
	12 Sep (R) GEO308	**Unit 2: End-to-end-ML Project (Part III)**: Selecting and training a model. Fine-tune your model: grid search, randomized search, ensemble methods, analyzing the best models and their errors, evaluating your system on the test set.	Ch.2 Geron Lab Handout	
4	17 Sep (T) GEO308	**Unit 3: Understanding Model Performance**: Expected new data error E_{new}: performance in production. Estimating E_{new}. The training error-generalization gap decomposition of E_{new}. The bias-variance decomposition of E_{new}.	Ch.4 Lindhol Ch.1 Geron	
	19 Sep (R) GEO308	**Unit 4: Classification:** MNIST, training a binary classifier; performance metrics, cross-validation, confusion matrices; precision and recall; the ROC curve. Multiclass classification, error analysis, multilabel classification, multioutput classification **Lab #3: Classification**	Ch.3 Geron Lab. Handout	Lab 3 due by 5 pm 09/30
5	24 Sep (T) GEO318	**Unit 5: Training Parametric Models (Part I):** Linear Regression; the normal equation; computational complexity. Gradient Descent; batch, stochastic and mini-batch. Polynomial Regression. Learning Curves.	Ch.4 Geron Ch.3 Lindhol	

Wk	Date	Topic	Reading	Due
	26 Sep (R) GEO318	**Unit 5: Training Parametric Models (Part II):** Regularized Linear Models; Ridge regression, lasso regression, elastic net regression, early stopping. Logistic Regression; estimating probabilities, training and cost function, decision boundaries, softmax regression. **Lab #4: Training Parametric Models**	Ch.4 Geron Ch.3 Lindhol Lab. Handout	Lab 4 due by 5 pm 10/11
6	1 Oct (T) GEO308 If travel online	**Unit 6. Support Vector Machines (Part I):** Linear SVM Classification; Soft margin classification. Non-linear SVM classification; Polynomial Kernel, similarity features, Gaussian RBF Kernel, SVM classes and computational complexities.	Ch.5 Geron Ch.8 Lindho	
	3 Oct (R) GEO308 If travel online	**Unit 6. Support Vector Machines (Part II):** Non-linear SVM classification; SVM Regression, Under the hood of linear SVM classifiers, the dual problem (Kernelized SVMs). **Paper Presentation Session #1:** Support Vector Machines (Jose Cabral, Paper #3) **Lab #5:** Support Vector Machines	Ch.5 Geron Ch.8 Lindhol Lab. Handout	Lab 5 due by 5 pm 10/14
7	8 Oct (T) GEO308	**Unit 7. Basic Non-Parametric Models (Part I):** K-Nearest Neighbors (KNN); Decision Trees. Training and visualizing. Making predictions. Estimating class probabilities. The CART training algorithm. Computational complexity. Gini Impurity or Entropy?. Regularization hyperparameters.	Ch.6 Geron Ch.2 Lindhol	
	10 Oct (R) GEO308	**Unit 7. Basic Non-Parametric Models (Part II):** Decision Trees; Regression, sensitivity to axis orientation, decision trees have high variance. **Lab#6:** Decision Trees and K-NN	Ch.2 Lindh. Ch.6 Geron Lab. Handout	Lab 6 due by 5 pm 10/23
8	15 Oct (T) GEO308	**Unit 8. Ensemble methods, bagging and boosting (Part I):** Voting classifiers. Bagging and pasting; bagging and pasting in Scikit-Learn; out-of-bag evaluation, random patches and random subspaces. Random Forests; extra-trees, feature importance. **Final Project Handout**	Ch.7 Lindho Ch.7 Geron	
	17 Oct (R) GEO308	**Unit 8. Ensemble methods, bagging and boosting (Part II):** Boosting; AdaBoost; gradient boosting, histogram-based gradient boosting. **Lab #7:** Random Forest & Gradient Boosting **Paper Presentation Session #2:** Random Forest and Gradient Boosting (Pedram Balooch)	Ch.7 Lindho Ch.7 Geron Lab. Handout	Lab 7 due by 5 pm 10/30
9	22 Oct (T) GEO308	**Unit 9. Neural Networks and Deep Learning (Part I):** From biological to artificial neurons; Biological neurons, logical computations with neurons, the perceptron, the multi-layer perceptron and backpropagation, regression MLPs, classification MLPs.	Ch.10 Geron Ch.6 Lindhol	

Wk	Date	Topic	Reading	Due
	24 Oct (T) GEO308	**Unit 9. Neural Networks and Deep Learning (Part II):** Implementing MLP with Keras; building an image classifier and a regression MLP using the sequential API; building complex models using the functional API; using the subclassing API to build dynamic models; saving and restoring a model; using callbacks; using TensorBoard for visualization. **Paper Presentation Session #3** ANN (Farzad Khallaghi)	Ch.10 Geron Ch.6 Lindhol	
10	29 Oct (T) GEO308	**Unit 9. Neural Networks and Deep Learning (Part III):** Fine-Tunning Neural Network Hyperparameters; number of hidden layers, number of neurons per hidden layer; learning rate, batch size, others.	Ch.10 Geron Ch.6 Lindhol	
	31 Oct (R) GEO308	**Unit 10. Training Deep Neural Networks (Part I):** The Vanishing/Exploding Gradients Problems; Glarot and He initialization; better activation functions; batch normalization; gradient clipping. **Invited Speaker: Isabela Suaza**	Ch.11 Geron Lab. Handout	
11	5 Nov (T) GEO308	**Unit 10. Training Deep Neural Networks (Part II):** Reusing Pretrained Layers; transfer learning with Keras; unsupervised pretraining; pretraining on an auxiliary task. Faster Optimizers; momentum; Nesterov accelerated gradient; AdaGrad; RMSProp; Learning Rate Scheduling; Avoiding Overfitting Through Regularization; L1 and L2 regularization; dropout; Montecarlo dropout; max-norm regularization. Summary and Practical Guidelines. Adam; AdaMax; Nadam; AdamW. **Lab #9:** Artificial Neural Networks (ANN)	Ch.11 Geron Lab. Handout	Lab 8: Interactive Learning and Discussion Lab 9 due by 5 pm 11/15
	7 Nov (R) GEO123	**Unit 11. Deep Computer Vision & Convolutional Neural Networks (Part I):** The Architecture of the Visual Cortex; Convolutional Layers: filters; stacking multiple feature maps; implementing convolutional layers with Keras; **Term Project Progress Report** **Paper Presentation Session #4:** NEAT (Alejandro Medina)	Ch.14 Geron	Prepare a short ppt with your main idea and the data sources (Classic ML, ANN)
12	12 Nov (T) GEO308	**Unit 11. Deep Computer Vision & Convolutional Neural Networks (Part II)**: Memory requirements. Pooling Layers. Implementing Pooling Layers with Keras. CNN Architectures (LeNet5, AlexNet, GoogleNet, VGGNet, ResNet, Xception, SENet, Others. Choosing the right architecture	Ch.14 Geron	
	14 Nov (R) GEO308	**Unit 11. Deep Computer Vision & Convolutional Neural Networks (Part III)**: Implementing ResNet-34 CNN using Keras. Using Pretrained models from Keras. Pretrained models for transfer learning. Classification and Localization; Object Detection (Fully convolutional nets, YOLO); Object Training; Semantic Segmentation. **Paper Presentation Session #5:** CNN (Miguel Mendez) **Lab #10:** Convolutional Neural Nets (CNN)	Ch.14 Geron	Lab 10 due by 5 pm 11/27

Wk	Date	Topic	Reading	Due
13	19 Nov (T) GEO308	**Unit 12. Processing Sequences Using RNN and CNN (Part I):** Recurrent neurons and layers: memory cells, input and output sequences. Training RNNs; Forecasting and Time Series: The ARMA model family; preparing the data.	Ch.15 Geron	
	21 Nov(R) GEO308	**Unit 12. Processing Sequences Using RNN and CNN (Part II):** Forecasting and Time Series: Using a simple RNN; using a deep RNN; multivariate time series; several time steps ahead; using a sequence-to-sequence model. **Term Project Progress Report** **Paper Presentation Session #7 and #8:** LSTM & GRU (Alexia Reyes, Aman Wraich)	Ch.15 Geron	Prepare a short ppt with main idea and the data sources (CNN, RNN)
14	26 Nov (T) GEO308	**Unit 12. Processing Sequences Using RNN and CNN (Part III):** Handling long sequences: fighting the unstable gradient problem; tackling the short-term memory problem. **Lab #10:** Reccurrent Neural Networks (RNN) **Paper Presentation Session #9** Wavenets (Saima Zahen)	Ch.15 Geron	
	28 Nov (R)	**No Class Thanksgiving ********************		
15	3 Dec (T) GEO308	Final Project Progress		
	5 Dec (R) Online	Final Project Progress		
	6 Dec (F)	**Final Project Presentations** (10:30 AM – 1:30 PM)		Slides Due 12/06 by 10 AM

Suggested Articles

Support Vector Machines (SV) for Classification and Regression

1. Xiang Yu, Yuhao Wang, Lifeng Wu, Genhua Chen, Lei Wang, Hui Qin. Comparison of support vector regression and extreme gradient boosting for decomposition-based data-driven 10-day streamflow forecasting, Journal of Hydrology, Volume 582, 2020, 124293, SSN 0022-1694.

2. Yu, J.-W.; Yoon, Y.-W.; Baek, W.-K.; Jung, H.-S. Forest Vertical Structure Mapping Using Two-Seasonal Optic Images and LiDAR DSM Acquired from UAV Platform through Random Forest, XGBoost, and Support Vector Machine Approaches. *Remote Sens.* 2021, *13*, 4282. https://doi.org/10.3390/rs13214282.

3. Kamran, K.V., Feizizadeh, B., Khorrami, B. et al. A comparative approach of support vector machine kernel functions for GIS-based landslide susceptibility mapping. Appl Geomat 13, 837–851 (2021). https://doi.org/10.1007/s12518-021-00393-0

4. Chih-Chun Liu, Tzu-Chi Lin, Kuang-Yu Yuan, Pei-Te Chiueh, Spatio-temporal prediction and factor identification of urban air quality using support vector machine, Urban Climate, Volume 41, 2022, 101055, ISSN 2212-0955.

Random Forest (RF) for Regression and Classification

5. Coleen Carranza, Corjan Nolet, Michiel Pezij, Martine van der Ploeg. Root zone soil moisture estimation with Random Forest, Journal of Hydrology, Volume 593, 2021, 125840, ISSN 0022-1694. https://doi.org/10.1016/j.jhydrol.2020.125840

6. Yu Zhan, Yuzhou Luo, Xunfei Deng, Michael L. Grieneisen, Minghua Zhang, Baofeng Di, Spatiotemporal prediction of daily ambient ozone levels across China using random forest for human exposure assessment, Environmental Pollution, Volume 233, 2018, Pages 464-473. https://doi.org/10.1016/j.envpol.2017.10.029

7. Koch, J., Berger, H., Henriksen, H. J., and Sonnenborg, T. O.: Modelling of the shallow water table at high spatial resolution using random forests, Hydrol. Earth Syst. Sci., 23, 4603–4619, https://doi.org/10.5194/hess-23-4603-2019, 2019. https://hess.copernicus.org/articles/23/4603/2019/hess-23-4603-2019.html

8. Svoboda, J.; Štych, P.; Laštovička, J.; Paluba, D.; Kobliuk, N. Random Forest Classification of Land Use, Land-Use Change and Forestry (LULUCF) Using Sentinel-2 Data—A Case Study of Czechia. *Remote Sens.*2022, *14*, 1189. https://doi.org/10.3390/rs14051189

9. Dharumarajan S, Hegde R. Digital mapping of soil texture classes using Random Forest classification algorithm. Soil Use Manage. 2022; 38: 135 149. https://doi.org/10.1111/sum.12668

Gradient Boosting (GB) for Classification and Regression

10. Stavrakoudis, D. and Gitas, I. (2023) Object-Based Burned Area Mapping with Extreme Gradient Boosting Using Sentinel-2 Imagery. *Journal of Geographic Information System*, **15**, 53-72. doi: 10.4236/jgis.2023.151004.

11. Jianchao Cai, Kai Xu, Yanhui Zhu, Fang Hu, Liuhuan Li. Prediction and analysis of net ecosystem carbon exchange based on gradient boosting regression and random forest, Applied Energy, Volume 262, 2020, 114566, ISSN 0306-2619 https://doi.org/10.1016/j.apenergy.2020.114566.

12. C. Candido, A.C. Blanco, J. Medina, E. Gubatanga, A. Santos, R. Sta Ana, R.B. Reyes, Improving the consistency of multi-temporal land cover mapping of Laguna lake watershed using light gradient boosting machine (LightGBM) approach, change detection analysis, and Markov chain, Remote Sensing Applications: Society and Environment, Volume 23, 2021, 100565, ISSN 2352-9385.

13. Elizabeth A. Freeman, Gretchen G. Moisen, John W. Coulston, and Barry T. Wilson. 2016. Random forests and stochastic gradient boosting for predicting tree canopy cover: comparing tuning processes and model performance. Canadian Journal of Forest Research. 46(3): 323-339. https://doi.org/10.1139/cjfr-2014-0562

Artificial Neural Networks (ANN) for Classification and Regression

14. Yangxiaoyue Liu, Wenlong Jing, Qi Wang, Xiaolin Xia, Generating high-resolution daily soil moisture by using spatial downscaling techniques: a comparison of six machine learning algorithms, Advances in Water Resources, Volume 141, 2020, 103601, ISSN 0309-1708, https://doi.org/10.1016/j.advwatres.2020.103601.

15. Sharma, P., Singh, S. & Sharma, S.D. Artificial Neural Network Approach for Hydrologic River Flow Time Series Forecasting. Agric Res 11, 465–476 (2022). https://doi.org/10.1007/s40003-021-00585-5.

16. Gholami, V., Sahour, H. Simulation of rainfall-runoff process using an artificial neural network (ANN) and field plots data. Theor Appl Climatol 147, 87–98 (2022). https://doi.org/10.1007/s00704-021-03817-4

17. Pragya Pradhan, Tawatchai Tingsanchali, Sangam Shrestha, Evaluation of Soil and Water Assessment Tool and Artificial Neural Network models for hydrologic simulation in different climatic regions of Asia, Science of The Total Environment, Volume 701, 2020, 134308, ISSN 0048-9697.

18. Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., & Yang, L. (2021). Physics-informed machine learning. Nature Reviews Physics, 3(6), 422-440.

19. Willard, J., Jia, X., Xu, S., Steinbach, M., & Kumar, V. (2020). Integrating Physics-Based Modeling with Machine Learning: A Survey.

20. Sun, R., Scanlon, B. R., Zhang, Z., Chen, J., & Zhang, J. (2022). Integration of machine learning and process-based hydrological models.

Convolutional Neural Networks (CNN) for Deep Computer Vision

21. Pfreundschuh, S., Brown, P. J., Kummerow, C. D., Eriksson, P., and Norrestad, T. (2022). GPROF-NN: a neural-network-based implementation of the Goddard Profiling Algorithm. *Atmospheric Measurement Techniques*, 15, 5033–5060

22. Sadeghi, M., A. A. Asanjan, M. Faridzad, P. Nguyen, K. Hsu, S. Sorooshian, and D. Braithwaite, 2019: PERSIANN-CNN: Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks–Convolutional Neural Networks. J. Hydrometeor., 20, 2273–2289, https://doi.org/10.1175/JHM-D-19-0110.1.

23. C. Huang *et al*., "Robust Forecasting of River-Flow Based on Convolutional Neural Network," in *IEEE Transactions on Sustainable Computing*, vol. 5, no. 4, pp. 594-600, 1 Oct.-Dec. 2020, doi: 10.1109/TSUSC.2020.2983097.

24. R.J. Pally, S. Samadi, Application of image processing and convolutional neural networks for flood image classification and semantic segmentation, Environmental Modelling & Software, Volume 148, 2022, 105285, ISSN 1364-8152, https://doi.org/10.1016/j.envsoft.2021.105285.

25. Yichen Lu, Thomas James, Calogero Schillaci & Aldo Lipani. (2022) Snow detection in alpine regions with Convolutional Neural Networks: discriminating snow from cold clouds and water body. GIScience & Remote Sensing 59:1, pages 1321-1343.

Recurrent Neural Networks (LSTM) for Sequence Prediction

26. Nearing, G., Cohen, D., Dube, V. et al. Global prediction of extreme floods in ungauged watersheds. Nature 627, 559–563 (2024). https://doi.org/10.1038/s41586-024-07145-1

27. Kratzert, F., Klotz, D., Brenner, C., Schulz, K., and Herrnegger, M.: Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks, Hydrol. Earth Syst. Sci., 22, 6005–6022, https://doi.org/10.5194/hess-22-6005-2018, 2018.

28. De la Fuente, L. A., Ehsani, M. R., Gupta, H. V., and Condon, L. E.: Towards Interpretable LSTM-based Modelling of Hydrological Systems, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2023-666, 2023.

29. Hunt, K. M. R., Matthews, G. R., Pappenberger, F., and Prudhomme, C.: Using a long short-term memory (LSTM) neural network to boost river streamflow forecasts over the western United States, Hydrol. Earth Syst. Sci., 26, 5449–5472, https://doi.org/10.5194/hess-26-5449-2022, 2022.

30. Mohan, A. & Gaitonde, D., 2018. A Deep Learning based approach to Reduced Order Modeling for Turbulent Flow Control using LSTM Neural Networks. Arxiv https://arxiv.org/abs/1804.09269

31. Mohan, A, Daniel Don, Chertkov, Livescu, D. 2019. Compressed Convolutional LSTM: An Efficient Deep Learning framework to Model High Fidelity 3D Turbulence. https://arxiv.org/abs/1903.00033

Recurrent Neural Networks (GRU) for Sequence Prediction

32. Lin, H., Gharehbaghi, A., Zhang, Q., Band, S. S., Pai, H. T., Chau, K. W., & Mosavi, A. (2022). Time series-based groundwater level forecasting using gated recurrent unit deep neural networks. Engineering Applications of Computational Fluid Mechanics, 16(1), 1655–1672. https://doi.org/10.1080/19942060.2022.2104928

33. Yin, B., Zuo, R. & Xiong, Y. Mineral Prospectivity Mapping via Gated Recurrent Unit Model.Nat Resour Res 31, 2065–2079 (2022). https://doi.org/10.1007/s11053-021-09979-2.

34. Wu, L., Zhou, J. T., Zhang, H., Wang, S. R., Ma, T., Yan, H., & Li, S. H. (2022). Time series analysis and gated recurrent neural network model for predicting landslide displacements. Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards, 18(1), 172–185. https://doi.org/10.1080/17499518.2022.2138918

35. X. Chen and W. Huang, "Spatial–Temporal Convolutional Gated Recurrent Unit Network for Significant Wave Height Estimation From Shipborne Marine Radar Data," in IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1-11, 2022, Art no. 4201711, doi: 10.1109/TGRS.2021.3074075.

Convolutional Neural Networks (Wavenet) for Sequence Prediction

36. Jun Chen, Yanhua Huang, Teng Wu, Jing Yan; A WaveNet-based convolutional neural network for river water level prediction. Journal of Hydroinformatics 1 November 2023; 25 (6): 2606–2624. doi: https://doi.org/10.2166/hydro.2023.174

37. Yang S, Zhong S, Chen K (2024) W-WaveNet: A multi-site water quality prediction model incorporating adaptive graph convolution and CNN-LSTM. PLoS ONE 19(3): e0276155. https://doi.org/10.1371/journal.pone.0276155

Generative Models and Learning from Unlabeled Data

38. Botterill, T. E., & McMillan, H. K. (2023). Using machine learning to identify hydrologic signatures with an encoder–decoder framework. Water Resources Research, 59, e2022WR033091. https://doi.org/10.1029/2022WR033091

39. Mohammad Sina Jahangir, John You, John Quilty, A quantile-based encoder-decoder framework for multi-step ahead runoff forecasting, Journal of Hydrology, Volume 619, 2023, 129269, ISSN 0022-1694.

40. Z. Han, N. Lv, X. Ai, Y. Zhou, J. Jiang and C. Chen, "Water Gauge Image Augmentation based on Generative Adversarial Network," 2022 IEEE International Conference on Smart Internet of Things (SmartIoT), Suzhou, China, 2022, pp. 154-160, doi: 10.1109/SmartIoT55134.2022.00033.

41. Ji, Y., Gong, B., Langguth, M., Mozaffari, A., and Zhi, X.: CLGAN: a generative adversarial network (GAN)-based video prediction model for precipitation nowcasting, Geosci. Model Dev., 16, 2737–2752, https://doi.org/10.5194/gmd-16-2737-2023, 2023.

Neural Evolution of Augmented Topologies (NEAT)

42. Kenneth O. Stanley, Risto Miikkulainen; Evolving Neural Networks through Augmenting Topologies. Evol Comput 2002; 10 (2): 99–127. doi: https://doi.org/10.1162/106365602320169811+ VIDEOS

Interpretable Machine Learning

43. Makke, N., Chawla, S. Interpretable scientific discovery with symbolic regression: a review. Artif Intell Rev 57, 2 (2024). https://doi.org/10.1007/s10462-023-10622-0