ML4CC: Lecture 6

Sit with your discussion groups (same as last time)!

Assignments reminder

Keep doing your weekly PMIRO+Q

Your second coding assignment was due today before the start of class.

You have an exam on March 6th (8am)

Recap of previous paper

P: No way to say if 2 atmospheric states are “similar”

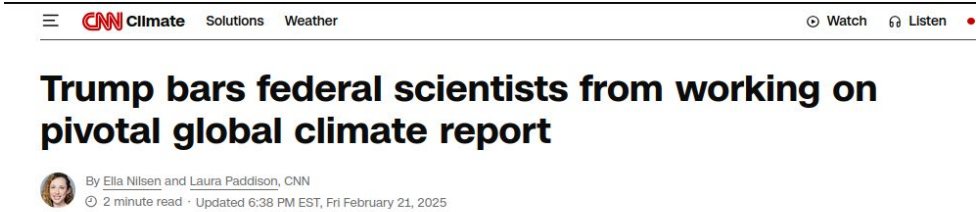
M: Use self-supervised learning (temporal difference prediction) to learn representation of atmospheric states, and use this representation as the basis of a distance metric (AtmoDist)

I: Self-supervised learning for this problem

R: AtmoDist behaves intuitively and better than image-based losses in tasks such as super-resolution

O: Does this definition of similarity capture what matters for climate models (e.g. physical laws)?

Climate Change in the News



CNN Climate Solutions Weather Watch Listen

Trump bars federal scientists from working on pivotal global climate report

By Ella Nilsen and Laura Paddison, CNN
2 minute read · Updated 6:38 PM EST, Fri February 21, 2025

(CNN) — The Trump administration told US government scientists working on a vital global climate report to stop their work, according to a scientist involved in the report – the latest move to withdraw the US from global climate action and research.

The US had been highly involved in planning for the next installment of the report due out in 2029 from the Intergovernmental Panel on Climate Change, the world's leading scientific authority on climate change.

The IPCC assesses how the climate crisis is affecting the planet according to the latest science. Its reports take thousands of scientists many years to produce and are used to inform policymakers across the world of the risks posed by global warming.

In a sense, all of the world's current, accepted knowledge about climate change stems from the IPCC and its reports, the first of which was published in 1990.

An international meeting of IPCC authors that was scheduled to take place in China next week is now in limbo. Kate Calvin, NASA's chief scientist and senior climate advisor, was supposed to co-chair the discussion but was impacted by the stop-work order, according to the scientist involved in the report. The meeting was planned to talk about next steps in the development of the report.

"Dr. Calvin will not be traveling to this meeting," a NASA spokesperson said. NASA denied CNN's request for an interview with Calvin.

The person involved in the report told CNN they were "not sure what this means for the planned work going forward, or if US scientists will participate in the writing of the IPCC reports."

"The IPCC is the backbone of global climate science, providing the world with unbiased, evidence-based insights needed to confront the climate crisis," said Harjeet Singh, a climate advocate and founding director of Satat Sampada Climate Foundation.

"The decision to exclude US scientists significantly undermines this collaborative effort and risks compromising the process at a time when robust climate action is needed more than ever," he told CNN in a statement.

Climate Change in the News

Grist

Donate

The Counter

Rice paddies, like cows, spew methane. A new variety makes them a lot less gassy.

Rice plants are a big source of methane, an extremely potent greenhouse gas. Scientists just developed a strain that cuts those emissions by 70 percent.

In this new study, researchers led by Schnürer and Yunkai Jin, a plant biologist at China's Hunan Agricultural University, compared a non-genetically-modified cultivar of rice, which had average methane emissions, and a genetically modified variety with low emissions. They found that the engineered strain produced significantly less fumarate, an organic compound. And the less fumarate the plants were secreting, the fewer methanogens living in the soil.

The researchers also discovered that the low-methane GMO rice released much more ethanol, an alcohol. When they applied ethanol to soils where rice plants were growing, that alone reduced methane emissions. "That turned out to be an inhibitor for the methanogens," Schnürer said. "So it was two factors: Both fumarate and ethanol play a role in the reduction of methane." They also applied another chemical, oxantel, to the soils, which significantly reduced methane emissions as well.

Paper 5 Discussion

IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, VOL. 60, 2022

Semantic Segmentation Based on Temporal Features: Learning of Temporal–Spatial Information From Time-Series SAR Images for Paddy Rice Mapping

Lingbo Yang, Ran Huang, Jingfeng Huang, Tao Lin, Limin Wang, Ruzemaimaiti Mijiti, Pengliang Wei, Chao Tang, Jie Shao, Qiangzi Li, and Xin Du

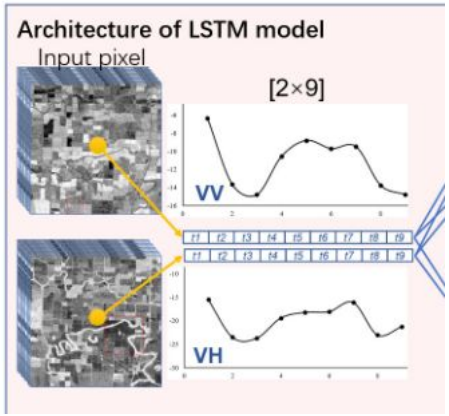
Attendance

Select one person from the group to be the attendance taker and fill out the attendance form (linked to under syllabus in Brightspace)

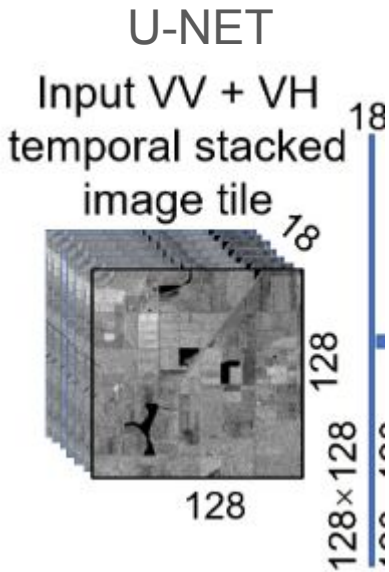
Discussion Question 1

The paper compares 4 different model. What is the same/different about how the input data fed into each of these models?

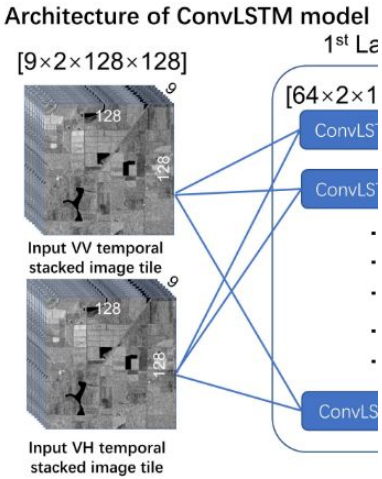
Input dimensions



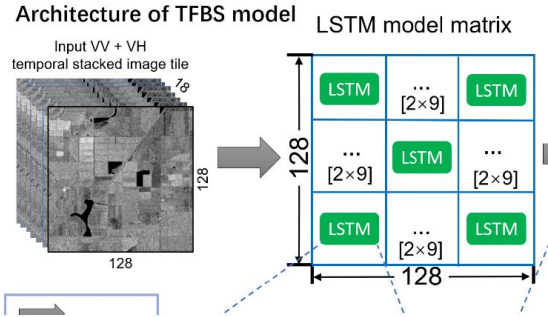
Time-series of two channels at a single location, passed in over 9 time steps.



128x128 spatial images concatenated across both channels and all 9 time steps



128x128 spatial images with 2 channels passed in over 9 time steps

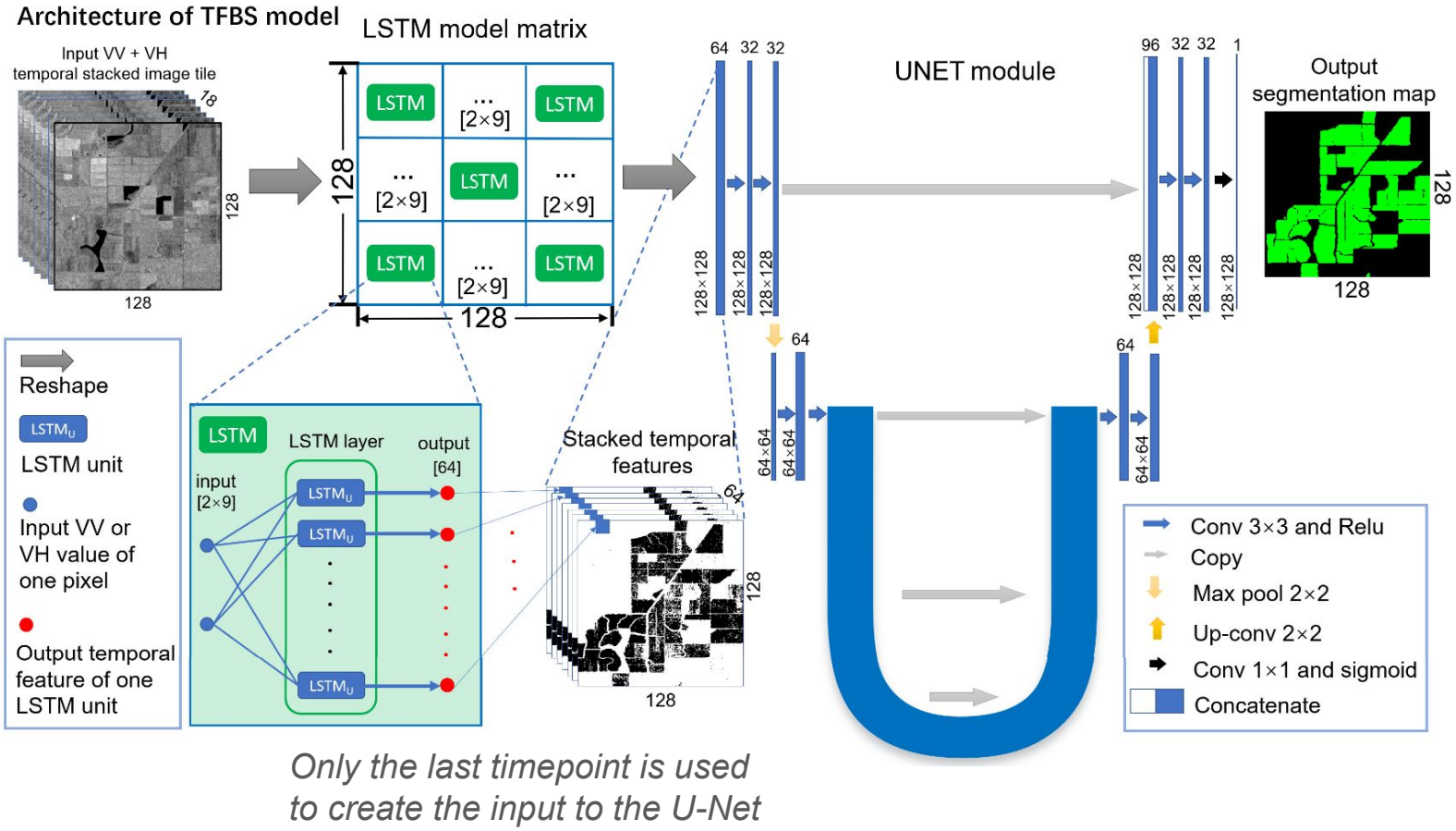


Each location in 128x128 image has 2 channels passed in over 9 time steps

Discussion Question 2

Describe the TFBS model architecture in your own terms. At what point does the model stop having temporal dynamics?

TFBS attaches pixelwise LSTMs to a U-Net



Only the last timepoint is used to create the input to the U-Net

Discussion Question 3

The paper tests 3 different types of generalization performance and one type of transfer learning. What are they and which figures show the results for each?

Within-distribution generalization

training process. The performances of the LSTM, UNET, ConvLSTM, and TFBS models are evaluated using datasets from ARMSMOTN in 2019, with 10-fold cross validation used as the evaluation method. This method randomly partitions the original dataset into ten equally sized subdatasets. Then, a single subdataset is retained as the validation dataset, and the remaining nine subdatasets are used as training data. This process is repeated ten times, with each of the ten subdatasets used exactly once as the validation dataset. A final estimation can then be calculated by averaging the ten results. The

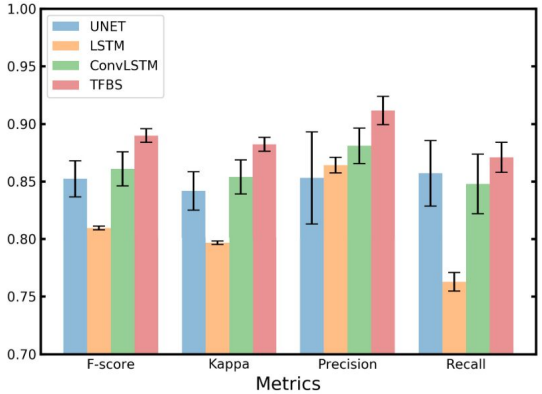


Fig. 7. Final result of the average F-score, kappa coefficient, precision, and recall of UNET, LSTM, TFBS, and ConvLSTM models, assessed by 10-fold cross validation method based on the dataset from ARMSMOTN in 2019. Error bars in the figure represent one standard deviation from the average accuracy.

Spatial and temporal generalization

The growth period of crops varies slightly because the climate conditions vary from place to place and year to year. Therefore, when the model trained in a specific year and place is applied to other years or other places, its accuracy in crop classification is affected by the temporal and spatial generalizability of the model. In this study, all the data from ARMSMOTN in 2019 were used to train four deep learning models. The datasets from ARMSMOTN in 2017 and 2018 were used to evaluate the temporal generalizability of the trained deep learning models. In addition, with a view to further testing the spatial–temporal generalizability of the deep learning models, the datasets from northern LA in 2017–2019 were used as validation data to test the deep learning models trained with the ARMSMOTN 2019 dataset.

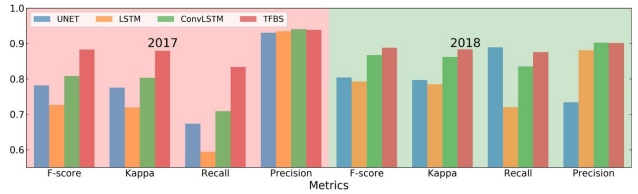
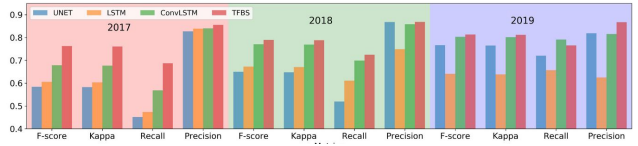


Fig. 10. Illustration of classification accuracies of UNET, LSTM, ConvLSTM, and TFBS models in 2017 (with red background) and 2018 (with green background). All the four models are trained with the ARMSMOTN dataset in 2019. CDL data is used as reference data to assess the accuracies.



Transfer learning

Besides, deep learning models which pretrained using the ARMSMOTN 2019 dataset are applied to the SV region for rice mapping. Due to the sowing date of rice in SV is relatively late than that in ARMSMOTN, it is not practical to map rice in SV by using the pretrained model directly. Therefore, fine-tuning method is employed to adapt pretrained models to new areas [80]. The parameters of the output layer for each pretrained deep learning model are initialized as zero, while the parameters of other layers are retained. An image tile with a spatial size of 128 × 128 pixels and its corresponding CDL image is randomly selected from the SV 2019 dataset. The selections are used to retrain the new output layer for each pretrained deep learning model. Based on the fine-tuned model, the rice classification results in the SV area from 2017 to 2019 are obtained, and the classification accuracy is evaluated by using CDL data as reference. In order to ensure

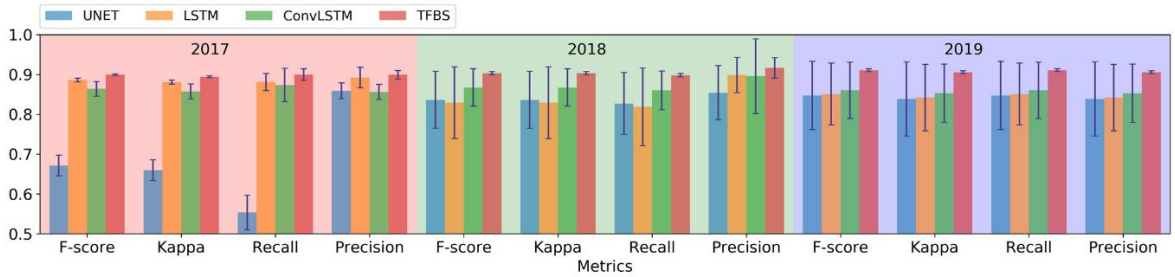


Fig. 13. Rice mapping accuracies of the fine-tuned deep learning models. The models were pretrained by ARMSMOTN 2019 dataset and transferred to SV based on fine-tuning technology. The CDL data of SV from 2017 to 2019 were used as reference data.

Discussion Question 4

What is data augmentation? What kind is done here? Does it help?

Making extra data

DATA AUGMENTATION

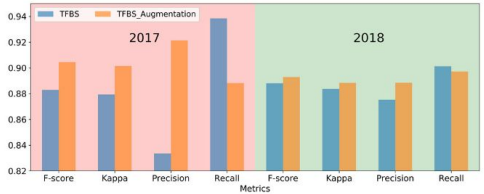


Fig. 15. Accuracies of TFBS models before (blue bars) and after data augmentation (orange bars). Dataset from ARMSMOTN in 2019 is used as training data, and the ARMSMOTN datasets from 2017 (red background) and 2018 (green background) are used as test data.

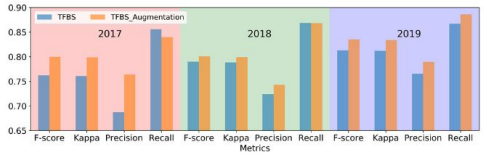
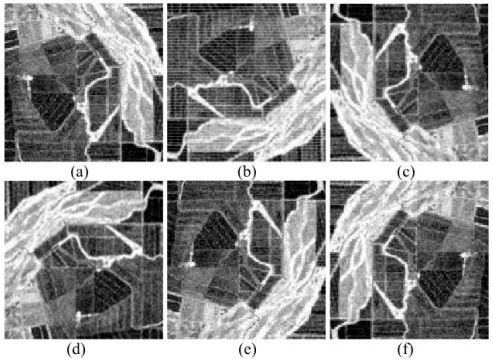


Fig. 16. Accuracies of TFBS models before (blue bars) and after data augmentation (orange bars). Dataset from ARMSMOTN in 2019 is used as training data, and the northern LA datasets from 2017 (red background), 2018 (green background), and 2019 (blue background) are used as test data.

Helps a bit

Data augmentation is a useful technique to increase the amount and the diversity of the training data. The effects of data augmentation on the spatial and temporal generalizability of the TFBS model are evaluated in this study. Several augmentation techniques, including spatial augmentation (rotation and flipping) and temporal augmentation (random scaling of backscatter coefficient), are used to increase the size of the dataset from ARMSMOTN in 2019. The raw tile images are first randomly rotated 90°, 180°, or 270° clockwise [Fig. 6(b)–(d)]; then, the rotated images are multiplied by a random value between 0.9 and 1.1 to increase the fluctuation of the time-series curve of the sample data. In addition, each raw tile image is flipped vertically or horizontally [Fig. 6(e) and (f)], and the flipped image is subsequently multiplied by a random number between 0.9 and 1.1. Thus, the size of the training dataset is tripled.



Discussion Question 5

What is meant by ‘stability’ here? How is it measured and how is it plotted?

D. Rice Mapping Based on Fine-Tuned Deep Learning Models

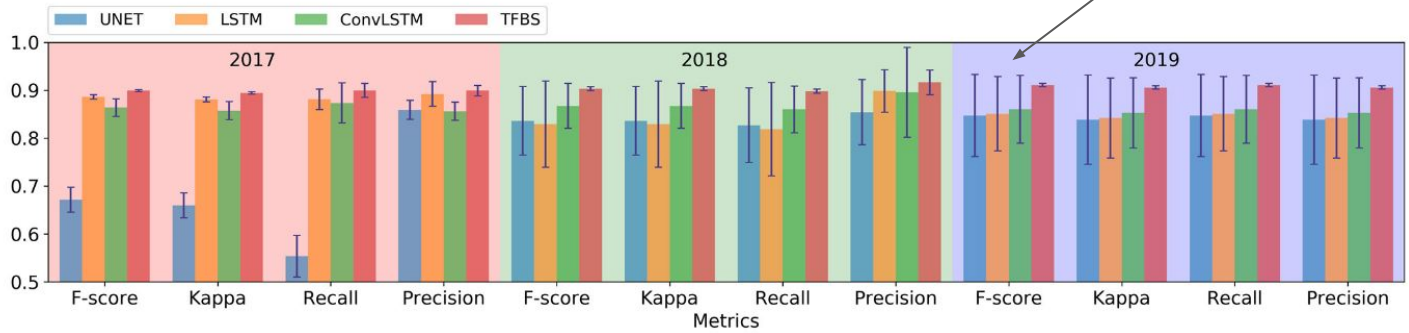
Four deep learning models were pretrained by ARMSMOTN 2019 dataset and then applied to the SV based on fine-tuning method for rice mapping from 2017 to 2019. CDL data were used to evaluate the classification accuracy from each model. Fig. 13 shows that TFBS significantly outperforms UNET, LSTM, and ConvLSTM in all three years in SV’s rice mapping, the average F-score of which is 15.2%, 5.7%, and 4.7% higher than that of UNET, LSTM, and ConvLSTM, respectively. Meanwhile, the stability of TFBS model is also higher than the other three of which shows a much lower standard deviation of classification accuracy. The results show that it is feasible to apply a pretrained model to areas with different rice phenology for rice mapping based on fine-tuning technology and a small number of samples.

Calculating means and standard deviations for processes that involve randomness

each pretrained deep learning model. Based on the fine-tuned model, the rice classification results in the SV area from 2017 to 2019 are obtained, and the classification accuracy is evaluated by using CDL data as reference. In order to ensure the reliability of the results and the stability evaluation of the fine-tuning for each model, this process was repeated ten times. The mean and standard deviation of each classification accuracy metric are calculated and used for model evaluation.

Training a network involves random initialization, random data order, etc

Error bars



Discussion Question 6

What is table two meant to show? How does it relate to concepts from last week's paper?

TFBS learns useful features

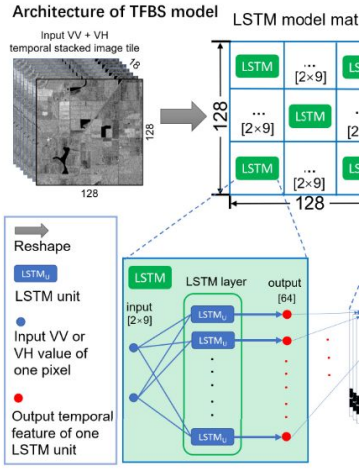


TABLE II

SEPARABILITY INDICATORS OF RICE IN RAW TIME-SERIES SAR IMAGES AND INTERMEDIATE TEMPORAL FEATURES MINED BY TFBS MODEL

	ARMSMOTN			Northern LA			Mean	Std.
	2017	2018	2019	2017	2018	2019		
J-M values of raw time-series images	1.34	1.36	1.22	1.43	1.39	1.36	1.350	0.065
J-M values of intermediate temporal features	2.00	2.00	2.00	2.00	2.00	2.00	2.00	0
TD of raw time-series images	1.46	1.47	1.37	1.57	1.54	1.5	1.485	0.064
TD of intermediate temporal features	2.00	2.00	2.00	2.00	2.00	2.00	2.00	0

J-M represents Jeffries-Matusita and TD represents Transformed Divergence indicator. ARMSMOTN represents Arkansas (AR), Mississippi (MS), south
Missouri (MO) and western Tennessee (TN). LA represents Louisiana.

Comparing how the output of the LSTM layer represents rice vs non-rice. I.e., looking at the representations that were learned.

Discussion Question 7

The authors note that the multi-class models confuse cotton for soybeans. Which model makes this mistake the most? What is another somewhat common mistake made by the models?

Confusion matrices

Soybeans are also commonly classified as others

LSTM makes this error the most

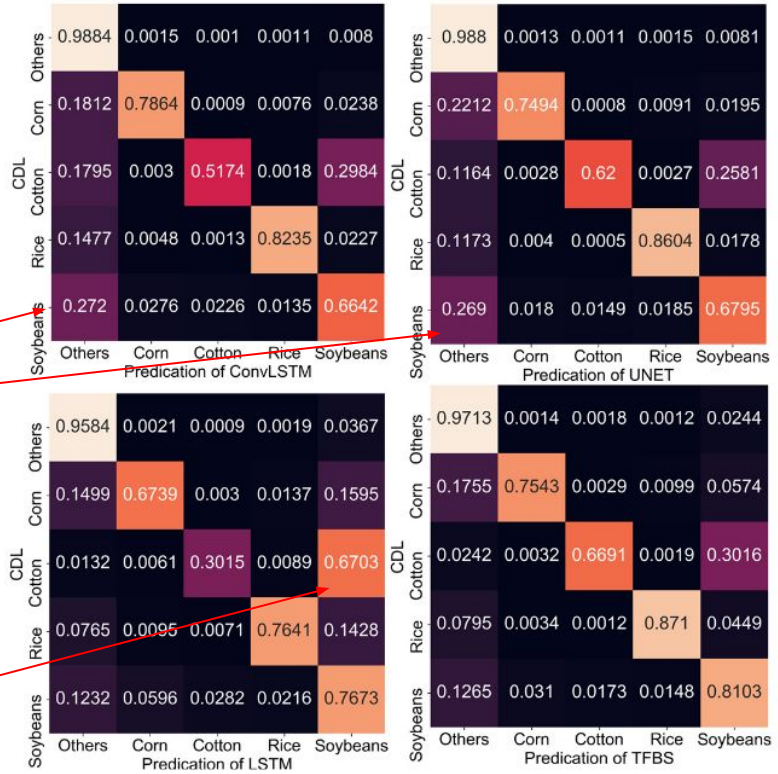


Fig. 18. Confusion matrixes of UNET, LSTM, TFBS, and ConvLSTM models. The models are trained by the augmented ARMSMOTN 2019 dataset and tested by the ARMSMOTN 2018 dataset.

Discussion Question 8

Share what questions you wrote in your PMIRO+Q and decide as a group what you'd like to ask.

Update your PMIRO+Q

Submit a second file to the Brightspace assignment (don’t overwrite the original):

It should:

Update your PMIRO as needed

Answer your own Q

You can be talking with your group during this!

15 min break

Lecture

Career Day, Project Info, Exam Review

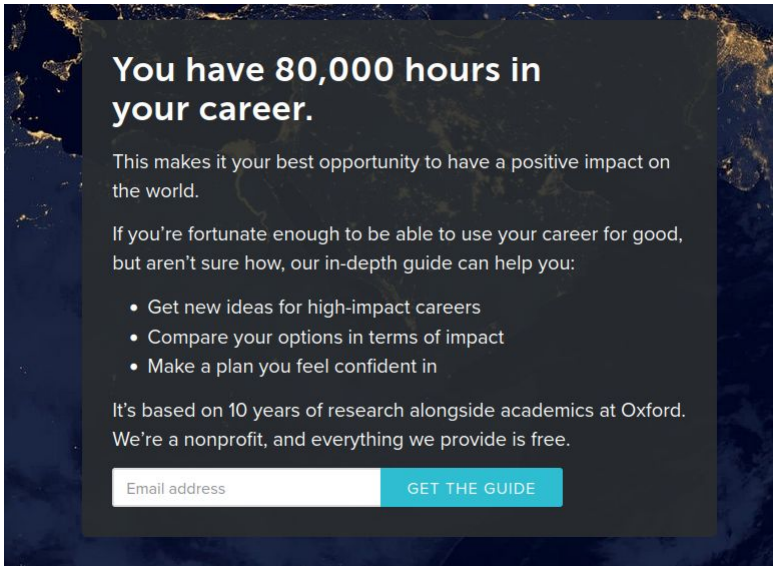
Alex Steffen, Climate Writer - *The Snap Forward*

“We don’t get to choose whether the context of all our work is an unprecedented, all-encompassing planetary crisis. It simply is. There is nowhere to stand outside of it. We can pretend that’s not true — insist that our lives or work or special concerns will remain untouched by ecological catastrophe and societal upheaval — but in the long run, we’re just fooling ourselves.”

Climate career pathways

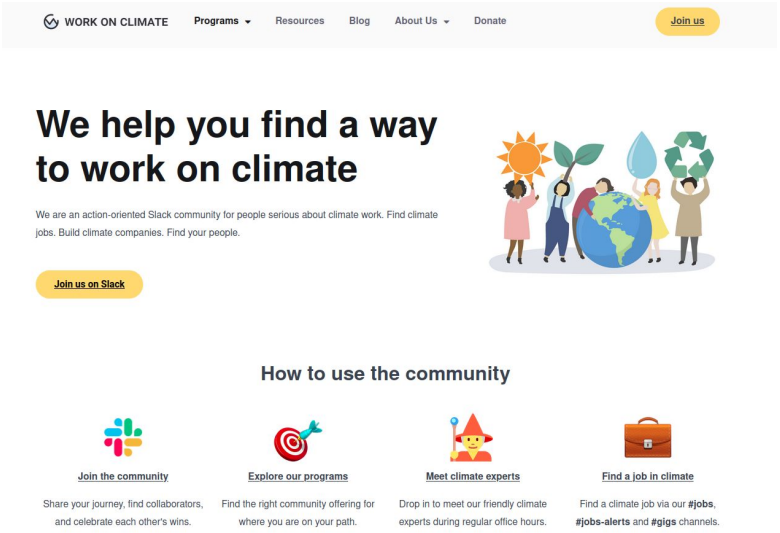
- Work directly on climate change
- Work directly on something else but bring awareness and action on climate to your job
- Regardless of your job, engage in climate action and activism

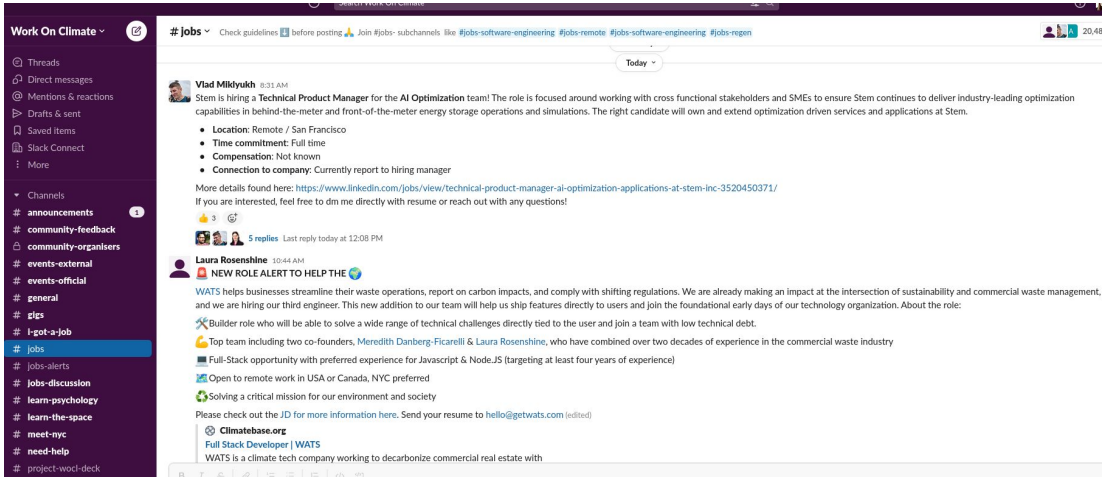
Your career is your biggest chance to make an impact



https://80000hours.org/

Working on climate

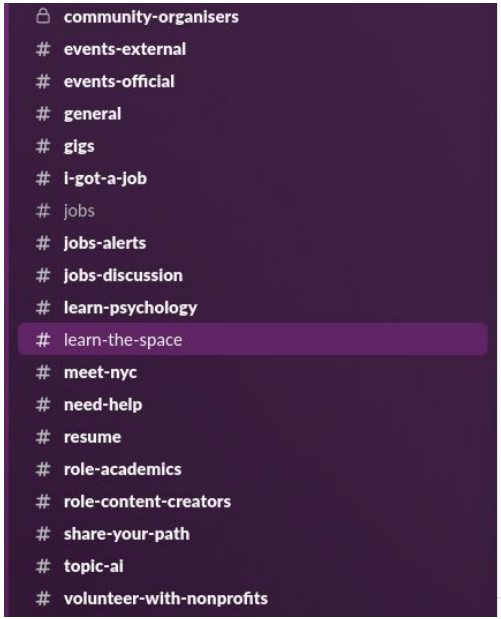


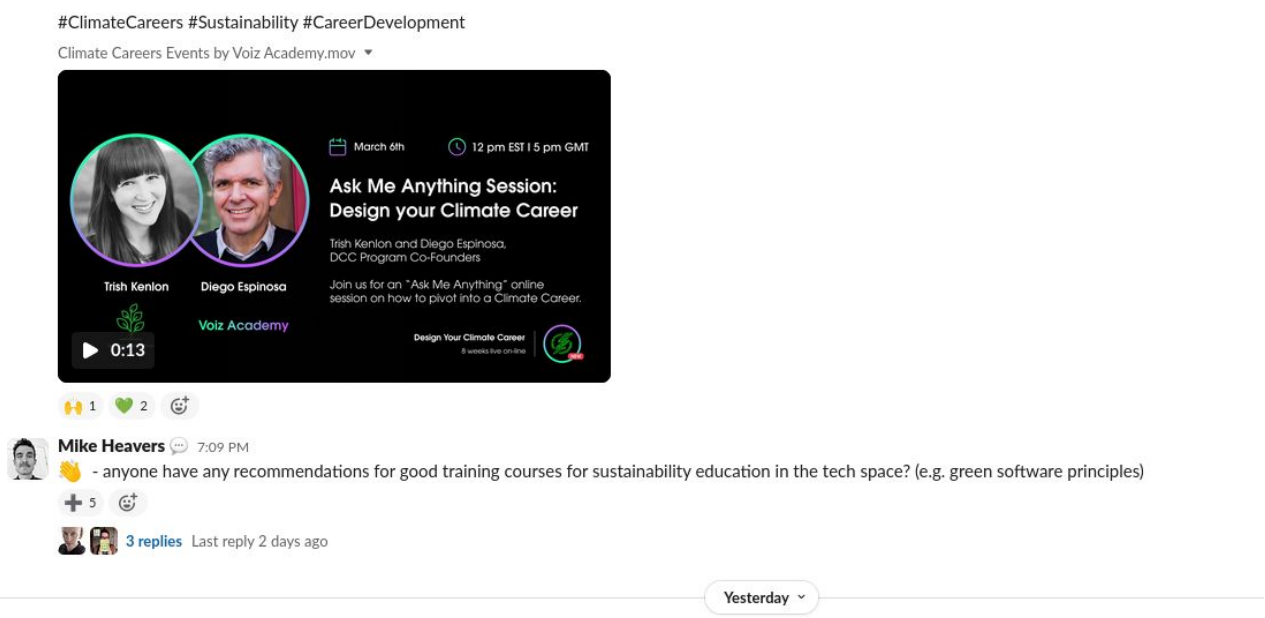


Central Hub: https://workonclimate.org/

Working on climate

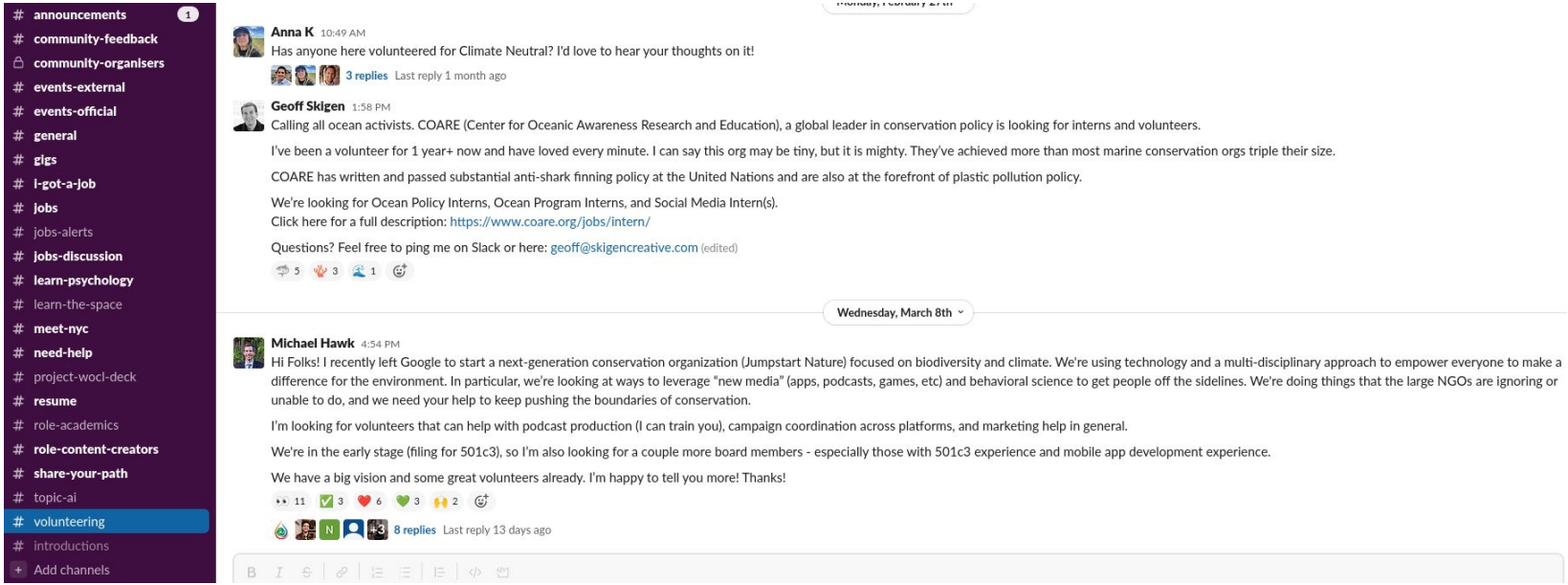
Communities help with resumes, interviews, skill learning etc.





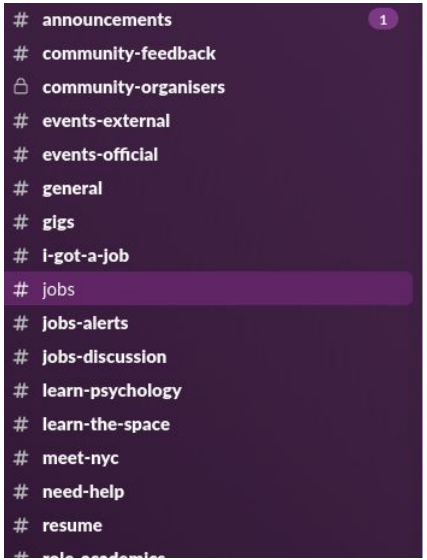
Working on climate

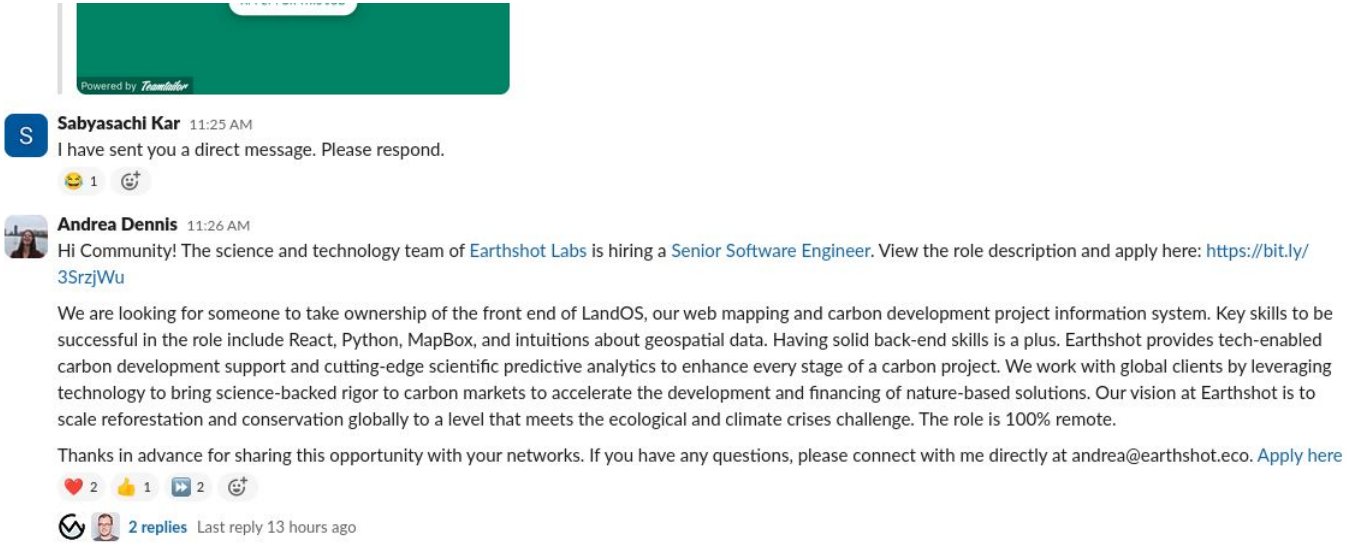
You can also look for volunteer or internship opportunities to skill up



Working on climate

Also job offers





Working on climate

Job boards:

https://www.climatetechcareers.com/#jobboards

https://climatebase.org/

https://terra.do/climate-jobs/job-board/

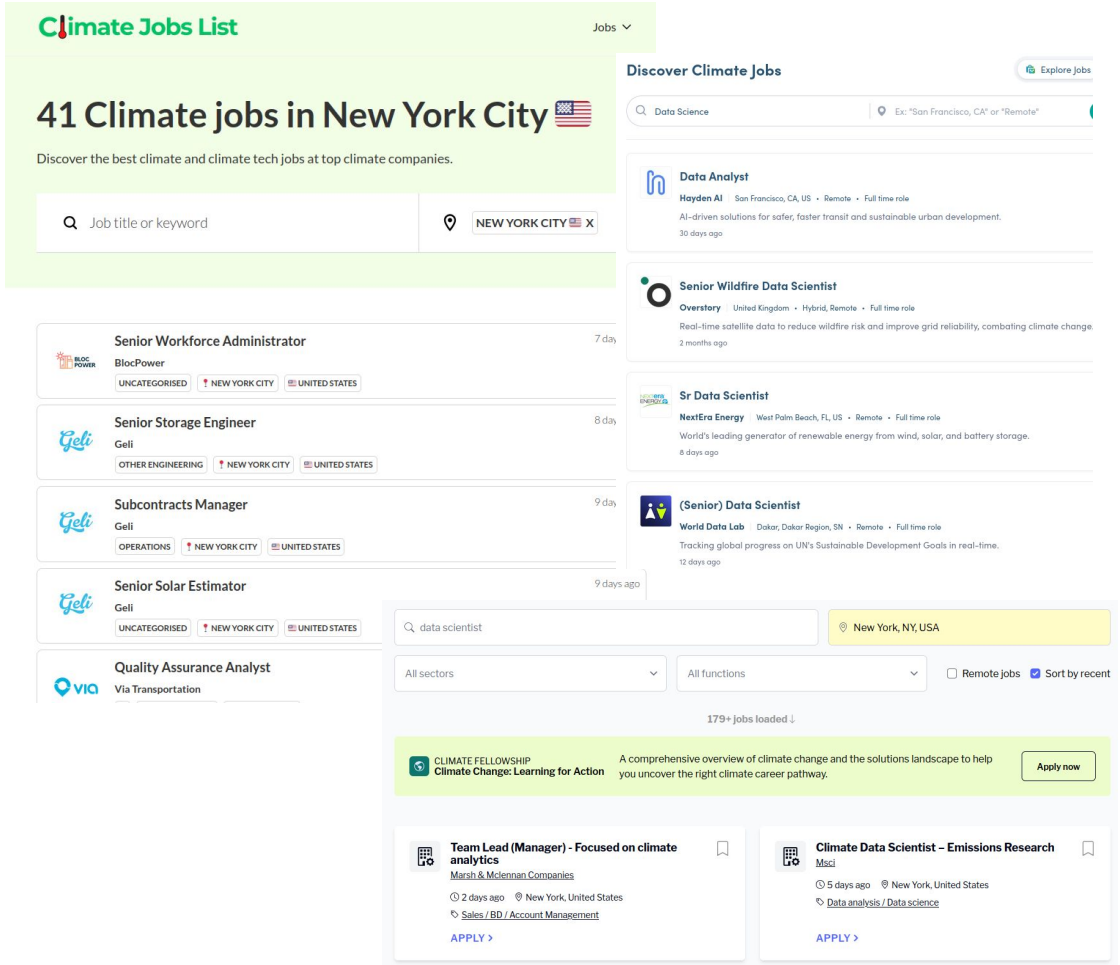
https://www.climatejobslist.com

https://www.climatepeople.com/

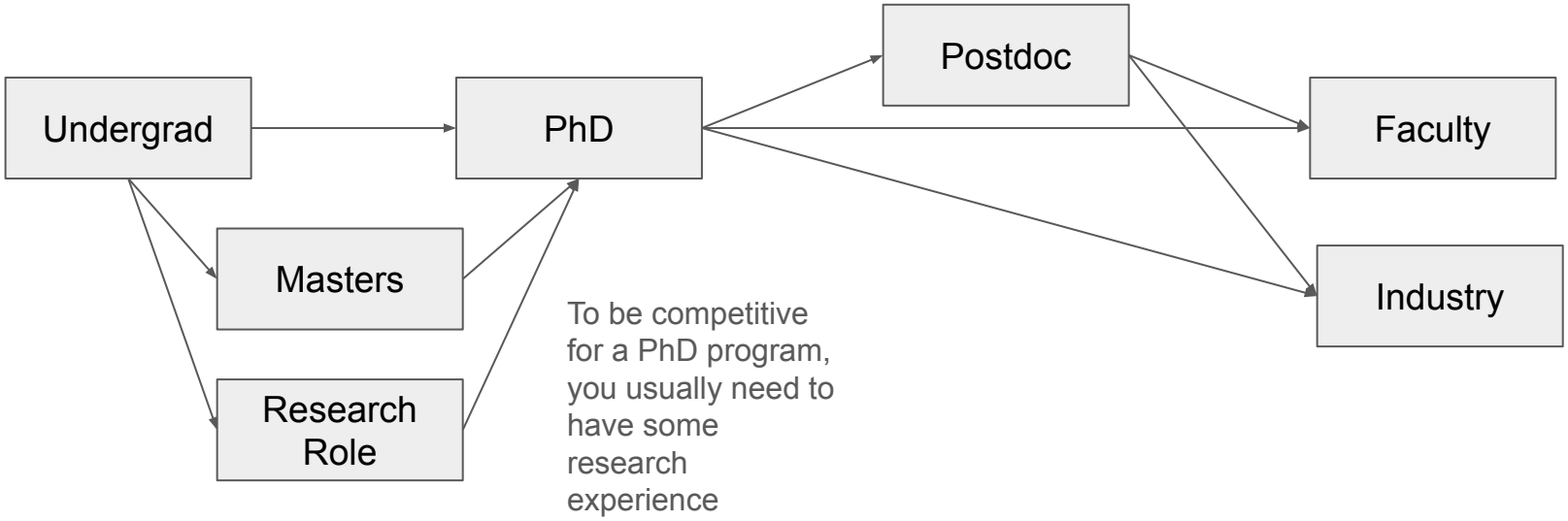
https://climateaction.tech/ (slack)

Freelance: https://www.leafr.work/

and more!



Academic Path



Learning more and networking

Climate Change AI summer school - https://www.climatechange.ai/events/summer_school

https://www.youtube.com/playlist?list=PLpPW7qLmXhdTnd9XSu606n2Qj93_-0O01

Terra.do School - https://terra.do/

One Point Five Academy - https://www.opf.degree/opf-academy

Airminers Boot-up - https://airminers.org/learn

Climatematch Academy - https://academy.climatematch.io/

Starting a company

My Climate Journey community and podcast: https://www.mcjcollective.com/

Resources for all stages: https://climatefounder.org/

Funding:

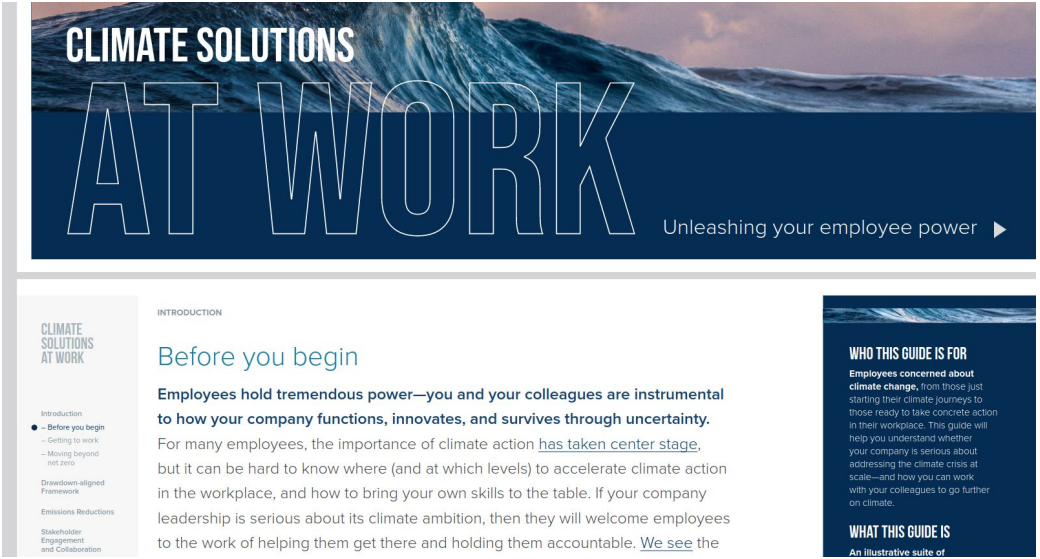
https://www.climate.vc/

https://climatetechvc.org/

https://evergreeninno.org/

Urban Future Lab @ NYU: https://www.ufl.nyc/

“Every job is a climate job”



https://drawdown.org/publications/climate-solutions-at-work

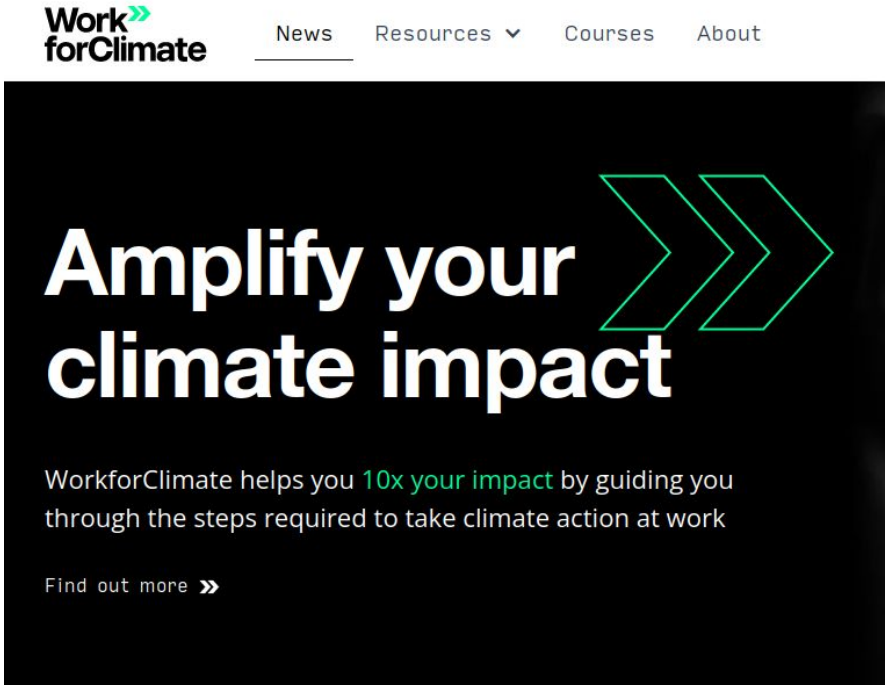
“Every job is a climate job”

	STATUS QUO	DRAWDOWN–ALIGNED
Emissions Reductions	Promises distant “net zero by 2050” commitment with reliance on offsets and no interim targets.	Moves on an accelerated timeline with interim targets and reduced reliance on offsets.
Stakeholder Engagement and Collaboration	Internal climate work is limited to sustainability teams and external engagement is primarily focused on investors.	Employees, communities, board members, and more are included.
Products, Partnerships, and Procurement	Products, partnerships, and procurement (the three Ps) are misaligned with climate commitments.	The three Ps are aligned with a company’s sustainability strategy.
Investments and Financing	Corporate investments and employee retirement plans are misaligned with climate goals.	Business is fully divested from fossil fuels and invested in climate solutions.
Climate Disclosures	A lack of transparency on emissions and climate-related risks is the norm.	Public emissions and climate risk disclosures occur on a regular basis.
Climate Policy Advocacy	Corporate and trade associations are known to lobby against climate policy.	Lobbying and contributions support bold climate policy, including alignment from trade associations.
Business Model Transformation	Sustainability is an under-resourced add-on to the core business of the company.	Business models always shift center on scaling climate solutions.
Long-term Thinking	Business is focused on growth and quarterly returns.	Uses business influence to promote justice and shift economic paradigms to fit within planetary boundaries.

Where we are and where we need to go

Private-sector climate targets need to meet the global climate crisis at scale. Companies can become “drawdown-aligned” by meeting all of the requirements outlined in this guide.

“Every job is a climate job”



CLIMATE CHANGEMAKERS

Case study: How one employee got a climate-friendly option added to her company’s 401(k)

When Camille Smith realised that her company's default 401(k) portfolio (essentially, the American version of a superannuation fund) was lacking in environmentally conscious fund options, she decided to do something about it. Here's how her persistence paid off in making climate-friendly investing available to everyone at her company.

GETTING STARTED

4 pressing climate questions to ask your CEO right now

Ask questions. Get your boss to take climate action seriously. Simple.

https://www.workforclimate.org/

“Every job is a climate job”

Neuron
NeuroView

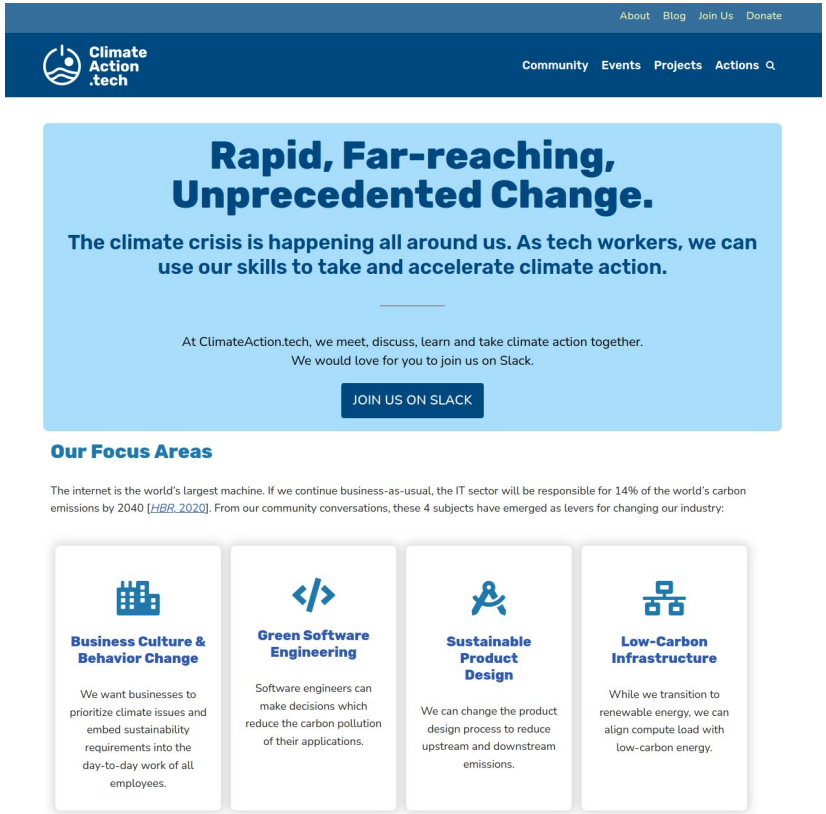
How Can Neuroscientists Respond to the Climate Emergency?

Adam R. Aron,[1,*] Richard B. Ivry,[2] Kate J. Jeffery,[3] Russell A. Poldrack,[4] Robert Schmidt,[5] Christopher Summerfield,[6,*] and Anne E. Urai[7]
[1]Department of Psychology and Neuroscience Graduate Program, University of California San Diego, San Diego, USA
[2]Department of Psychology, University of California, Berkeley, USA
[3]Division of Psychology and Language Sciences, University College London, UK
[4]Department of Psychology, Stanford University, Stanford, CA, USA
[5]Department of Psychology, University of Sheffield, Sheffield, UK
[6]Department of Experimental Psychology, University of Oxford, Oxford, UK
[7]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA
*Correspondence: adamaron@ucsd.edu (A.R.A.), christopher.summerfield@psy.ox.ac.uk (C.S.)
https://doi.org/10.1016/j.neuron.2020.02.019

The world faces a climate emergency. Here, we consider the actions that can be taken by neuroscientists to tackle climate change. We encourage neuroscientists to put emissions reductions at the center of their everyday professional activities.

Profession-specific resources are becoming more common

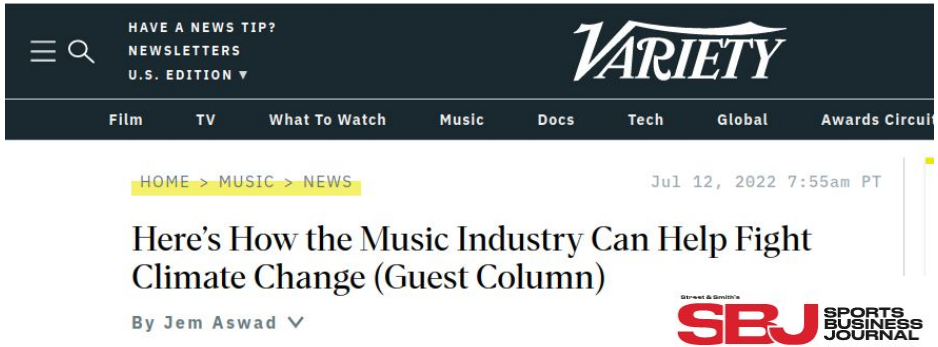
“Every job is a climate job”



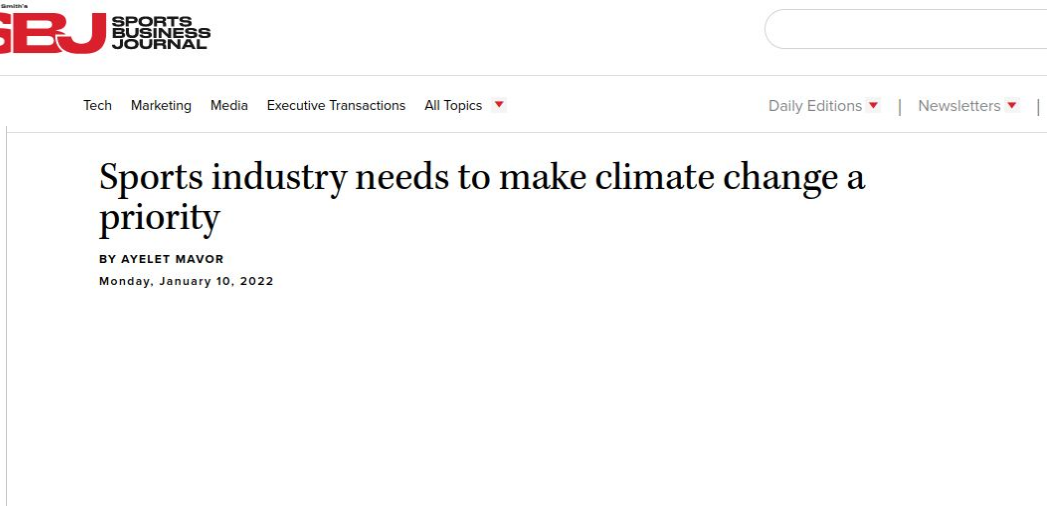
https://climateaction.tech/

Profession-specific resources are becoming more common

“Every job is a climate job”



Profession-specific resources are becoming more common



Climate action outside of work

Stay informed! Newsletters/media sites:

- https://www.carbonbrief.org
- https://grist.org/
- https://www.volts.wtf/
- https://heated.world/

Climate action outside of work

Non-profits working on climate issues

- Clean Air Task Force
- Earthjustice
- Conservation International
- CarbonFund
- Environmental Defense Fund
- many more!

Climate action outside of work

Political organizing

- Citizens' Climate Lobby
- League of Conservation Voters
- Extinction Rebellion
- People's Climate Movement

ANNALS OF A WARMING PLANET

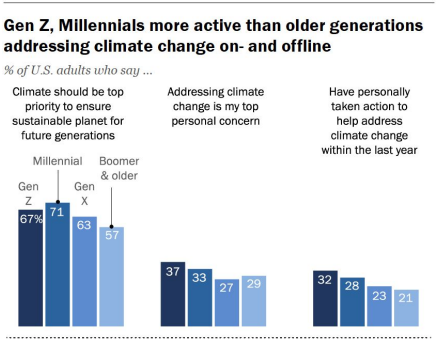
THE ANSWER TO CLIMATE CHANGE IS ORGANIZING

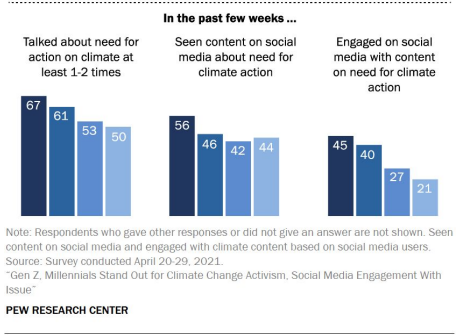
Dealing with global warming is always going to be about the balance of power.

By Bill McKibben
September 1, 2021

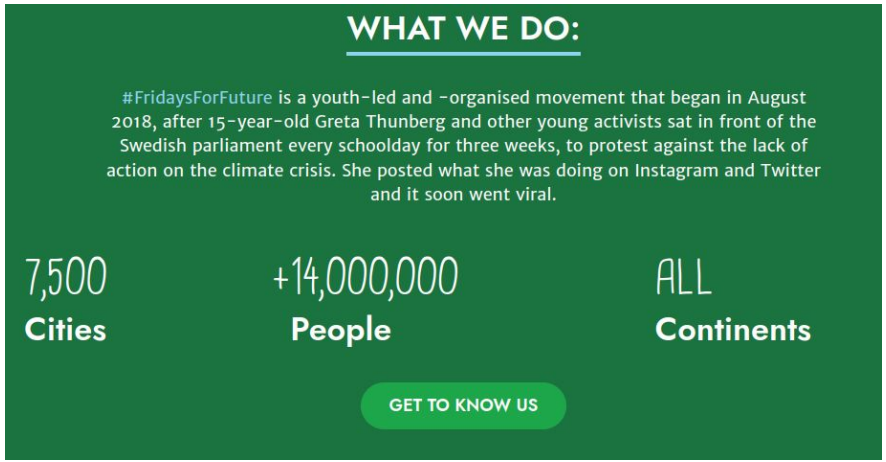
REPORT | MAY 26, 2021

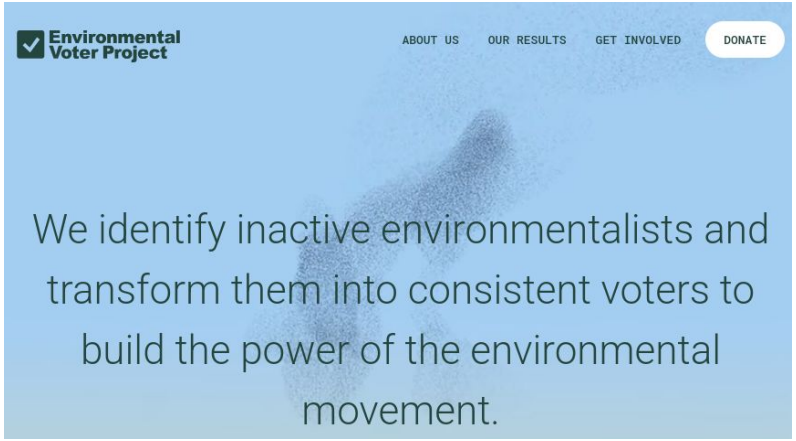
Gen Z, Millennials Stand Out for Climate Change Activism, Social Media Engagement With Issue

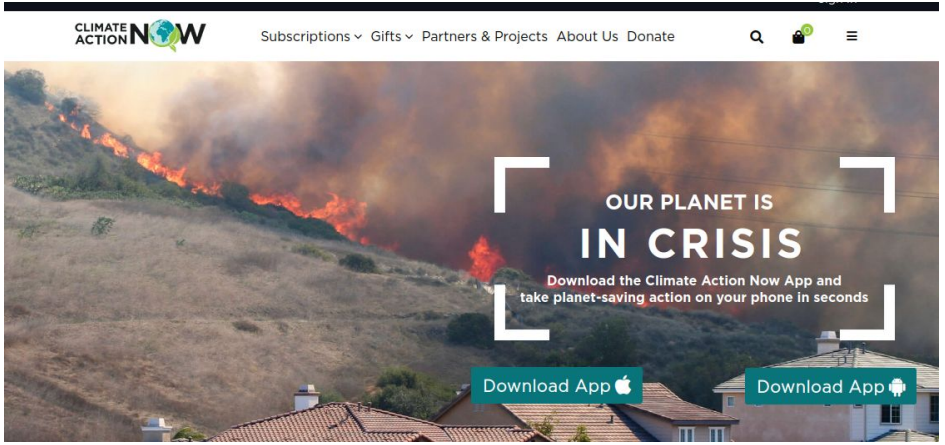




Climate action outside of work







Projects Overview

The class project is your chance to go deeper into a question, dataset, and/or method you are interested in.

You will be evaluated on your ability to pose a good machine learning question in the domain of climate change, the selection and execution of relevant methods, your ability to work on a team, and your presentation skills both written and oral.

Project teams

Everyone must work in a team (no individual projects)

Teams should be 5 people

A single grade will be given to the team.

Projects

Projects need to go beyond what you have done in homework assignments.

Possible approaches:

- replicate/extend an existing paper
- contribute to an established competition/benchmark
- Identify an underexplored problem where ML could help and try to solve it

Projects

Rough guidelines for creating a **substantial** project:

- Compare multiple methods and/or hyperparameters
- Use at least one method that isn't in scikit-learn (Rijul will be going through PyTorch code in labs)
- Explore feature importance methods
- Create your own dataset from multiple sources
- Use multiple evaluation methods
- Include advanced visualizations and/or a dashboard (but it better be really good)
- etc

Possible data sources

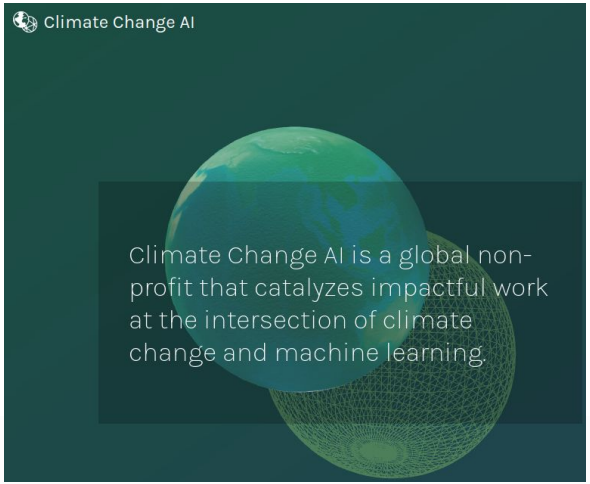
Here are locations to look for relevant data. Not all datasets from these sources will be climate change-related.

Note: these are real datasets. They may have missing data, poor documentation, or other challenges. You will probably have to teach yourselves some new skills to get what you want from them. These are all things data scientists have to face! Look into a dataset before committing to using it.

Starting Points

Climate Change AI

https://www.climatechange.ai/papers



Climate Change AI Workshop Papers

On this page, we show accepted works from all our workshops on "Tackling Climate Change with Machine Learning."

Venue: All Venues | Text Search: Enter search terms... | Subject Areas: Select some options... | Award?

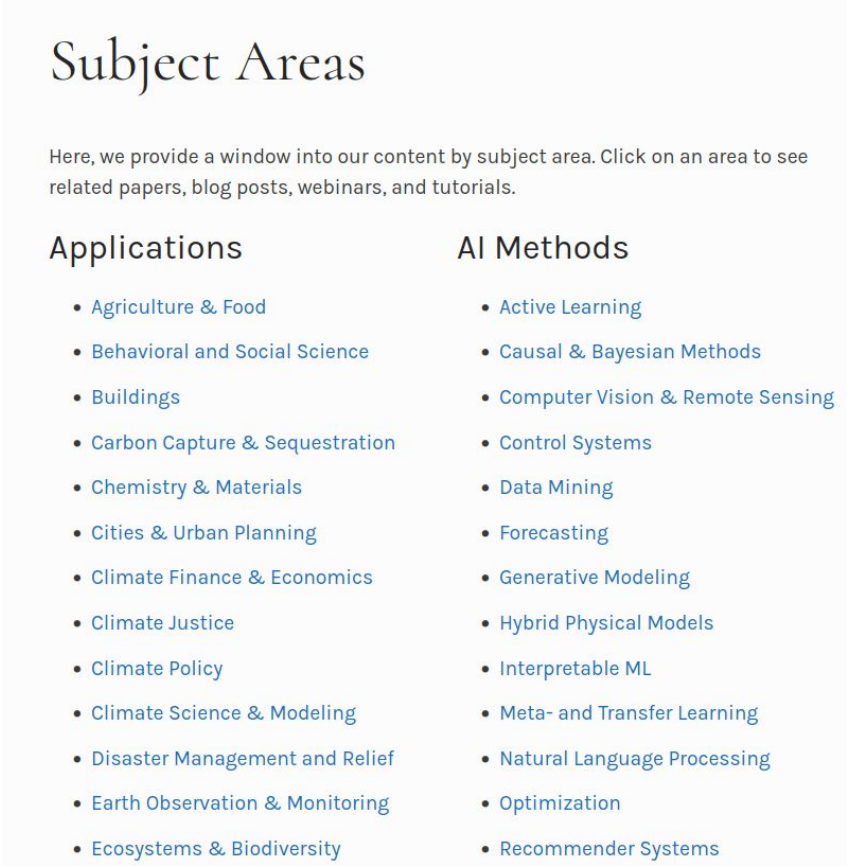
Showing 1 to 728 of 728 entries

Venue	Title	Subject Areas
NeurIPS 2023	Machine learning for gap-filling in greenhouse gas emissions databases (Papers Track) ►Abstract and authors: (click to expand)	Public Policy
NeurIPS 2023	EarthPT: a foundation model for Earth Observation (Papers Track) ►Abstract and authors: (click to expand)	Earth Observation & Monitoring
NeurIPS 2023	Towards Understanding Climate Change Perceptions: A Social Media Dataset (Papers Track) ►Abstract and authors: (click to expand)	Behavioral and Social Science

Starting Points

Climate Change AI

https://www.climatechange.ai/subject_areas



Subject Areas

Here, we provide a window into our content by subject area. Click on an area to see related papers, blog posts, webinars, and tutorials.

Applications

- Agriculture & Food
- Behavioral and Social Science
- Buildings
- Carbon Capture & Sequestration
- Chemistry & Materials
- Cities & Urban Planning
- Climate Finance & Economics
- Climate Justice
- Climate Policy
- Climate Science & Modeling
- Disaster Management and Relief
- Earth Observation & Monitoring
- Ecosystems & Biodiversity

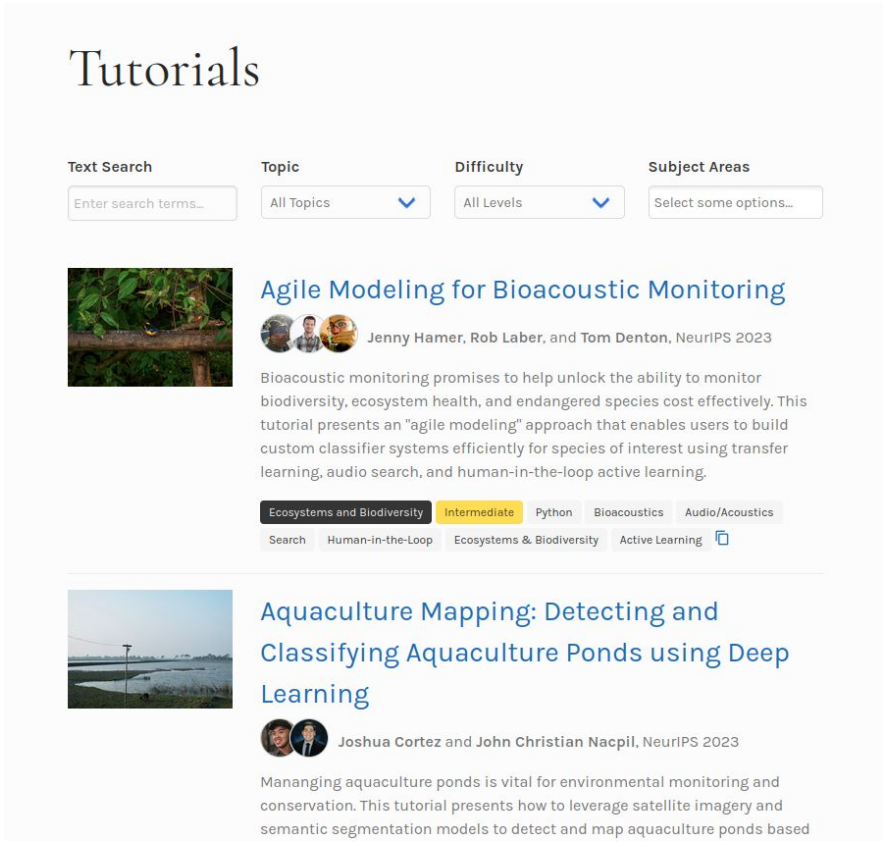
AI Methods

- Active Learning
- Causal & Bayesian Methods
- Computer Vision & Remote Sensing
- Control Systems
- Data Mining
- Forecasting
- Generative Modeling
- Hybrid Physical Models
- Interpretable ML
- Meta- and Transfer Learning
- Natural Language Processing
- Optimization
- Recommender Systems

Starting Points

Climate Change AI

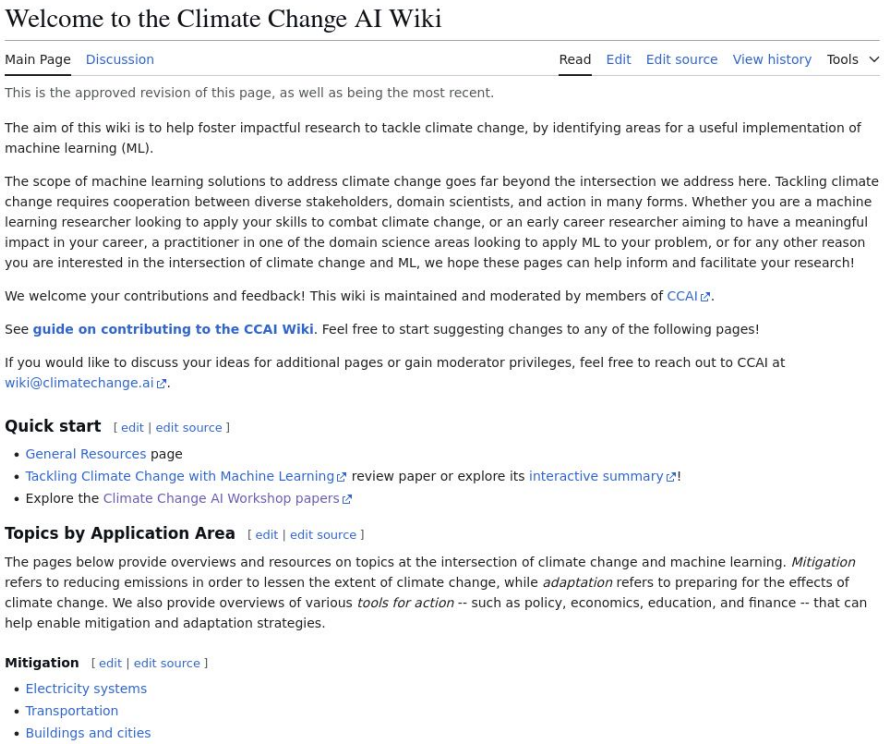
https://www.climatechange.ai/tutorials



Starting Points

Climate Change AI

https://wiki.climatechange.ai/wiki/Welcome_to_the_Climate_Change_AI_Wiki



Welcome to the Climate Change AI Wiki

Main Page Discussion Read Edit Edit source View history Tools

This is the approved revision of this page, as well as being the most recent.

The aim of this wiki is to help foster impactful research to tackle climate change, by identifying areas for a useful implementation of machine learning (ML).

The scope of machine learning solutions to address climate change goes far beyond the intersection we address here. Tackling climate change requires cooperation between diverse stakeholders, domain scientists, and action in many forms. Whether you are a machine learning researcher looking to apply your skills to combat climate change, or an early career researcher aiming to have a meaningful impact in your career, a practitioner in one of the domain science areas looking to apply ML to your problem, or for any other reason you are interested in the intersection of climate change and ML, we hope these pages can help inform and facilitate your research!

We welcome your contributions and feedback! This wiki is maintained and moderated by members of CCAI.

See **guide on contributing to the CCAI Wiki**. Feel free to start suggesting changes to any of the following pages!

If you would like to discuss your ideas for additional pages or gain moderator privileges, feel free to reach out to CCAI at wiki@climatechange.ai.

Quick start [edit | edit source]

- General Resources page
- Tackling Climate Change with Machine Learning review paper or explore its interactive summary!
- Explore the Climate Change AI Workshop papers

Topics by Application Area [edit | edit source]

The pages below provide overviews and resources on topics at the intersection of climate change and machine learning. *Mitigation* refers to reducing emissions in order to lessen the extent of climate change, while *adaptation* refers to preparing for the effects of climate change. We also provide overviews of various *tools for action* -- such as policy, economics, education, and finance -- that can help enable mitigation and adaptation strategies.

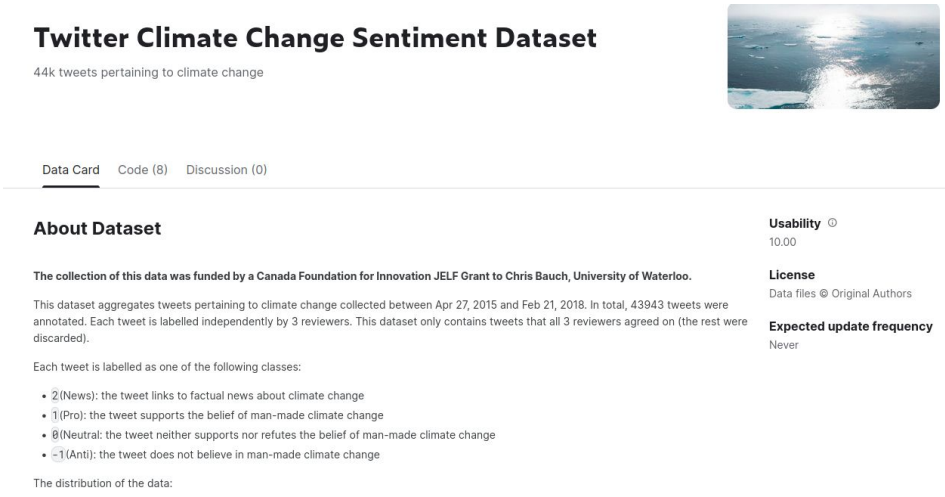
Mitigation [edit | edit source]

- Electricity systems
- Transportation
- Buildings and cities

Places to find data

Kaggle - can look at competitions and data sets (but pay attention to quality!)

www.kaggle.com

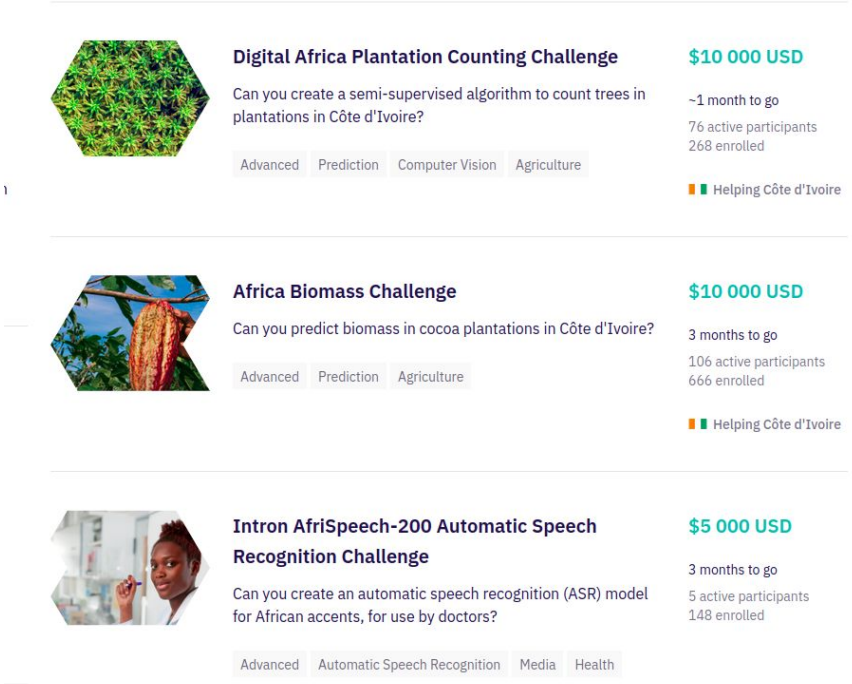


https://www.kaggle.com/datasets/techsash/waste-classification-data

Places to find data

Zindi (can use closed or open competitions)

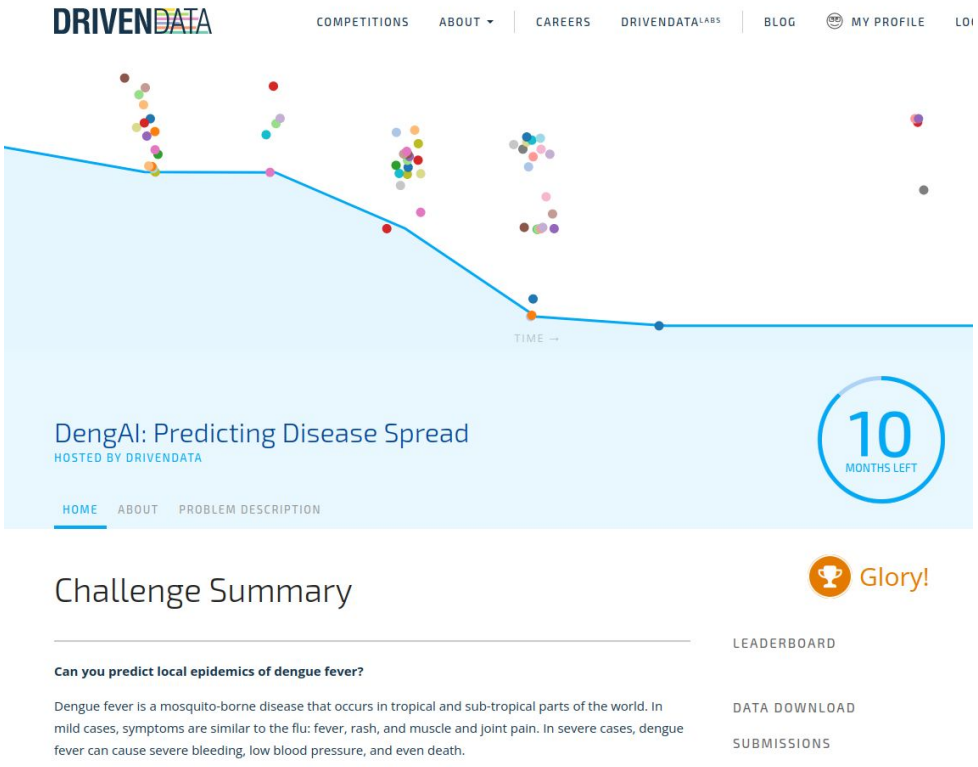
https://zindi.africa/competitions



Places to find data

Driven Data competitions

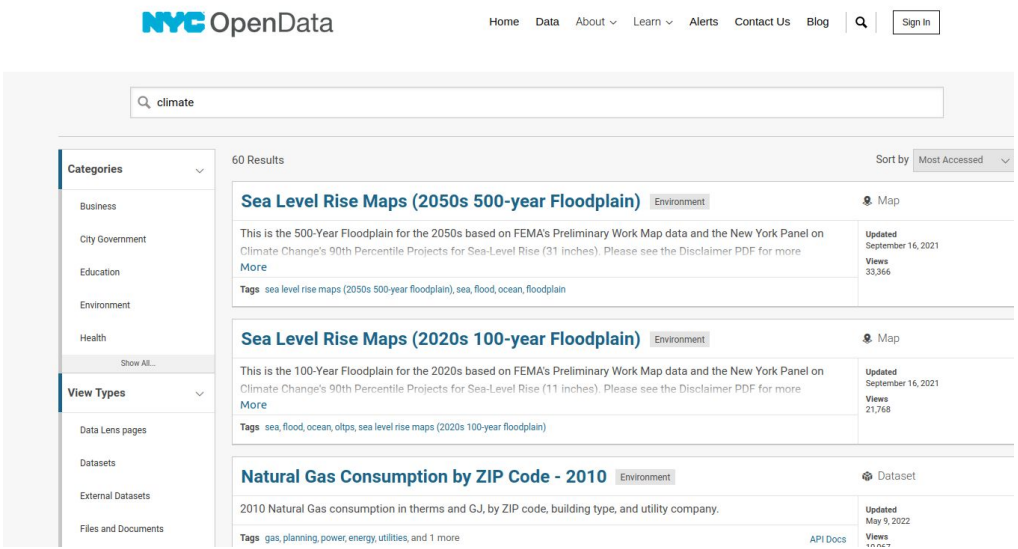
https://www.drivendata.org/competitions/



Possible data sources

https://opendata.cityofnewyork.us/

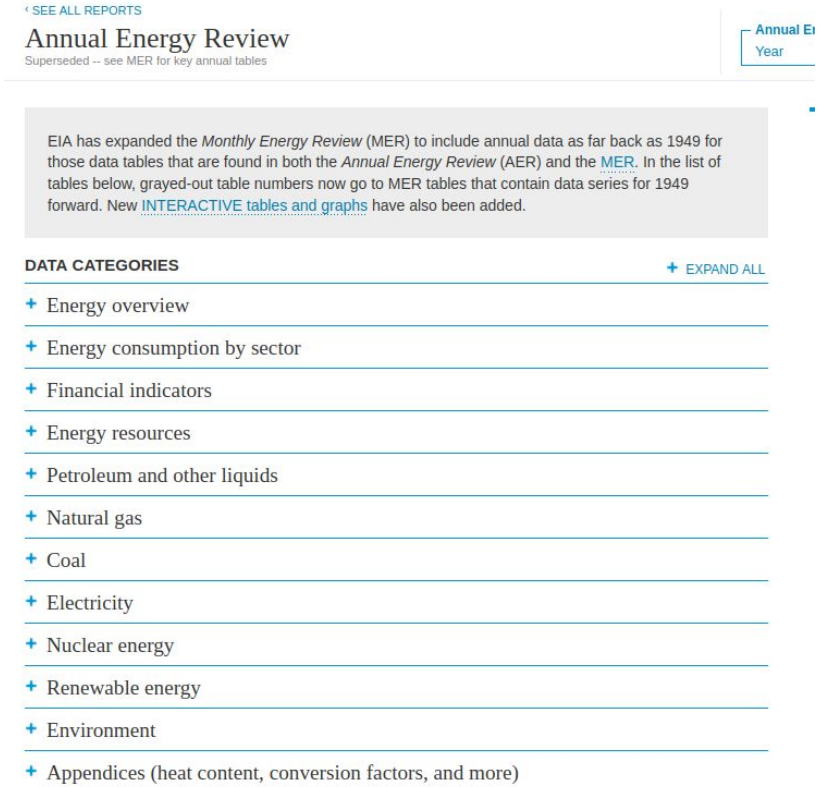
Has an “environmental” section. Contains data on trees, air quality, energy use, sea level, etc.



Possible data sources

Energy information administration

https://www.eia.gov/totalenergy/data/annual/index.php



Possible data sources

Emissions estimations from NEI

https://www.epa.gov/air-emissions-inventories/national-emissions-inventory-nei

ClimateTRACE emissions data

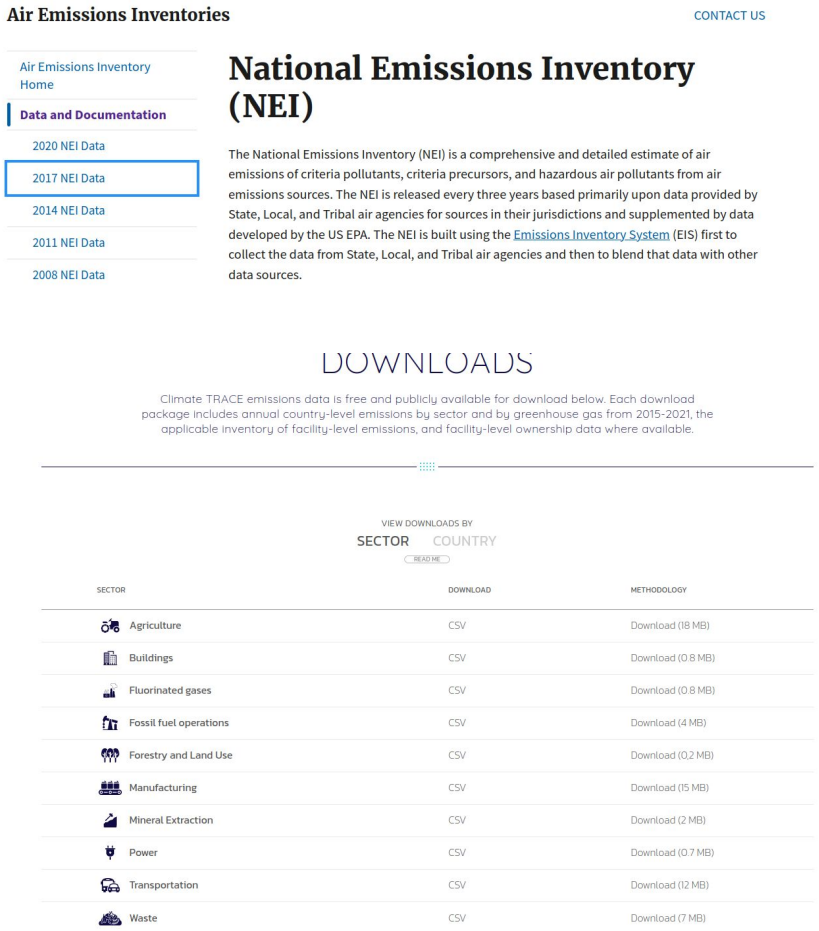
https://climatetrace.org/downloads

Carbon dioxide data from remote sensing

https://disc.gsfc.nasa.gov/datasets/OCO2GriddedXCO2_3/summary?keywords=oco-2

On road emissions data

https://daac.ornl.gov/CMS/guides/CMS_DARTE_V2.html

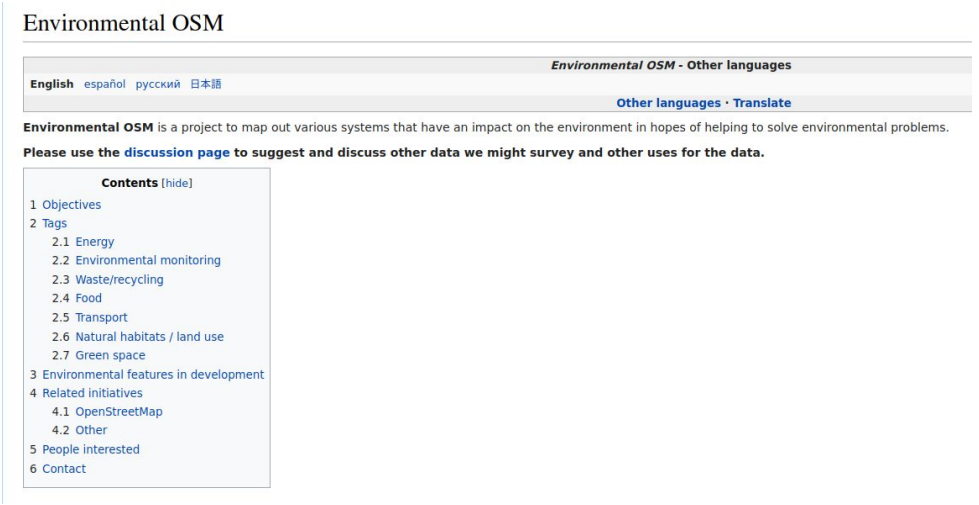


Possible data sources

Open Street Map for general land and road features

https://www.openstreetmap.org

https://wiki.openstreetmap.org/wiki/Environmental_OSM



Environmental OSM

***Environmental OSM* - Other languages**

English español русский 日本語

Other languages · Translate

Environmental OSM is a project to map out various systems that have an impact on the environment in hopes of helping to solve environmental problems.

Please use the discussion page to suggest and discuss other data we might survey and other uses for the data.

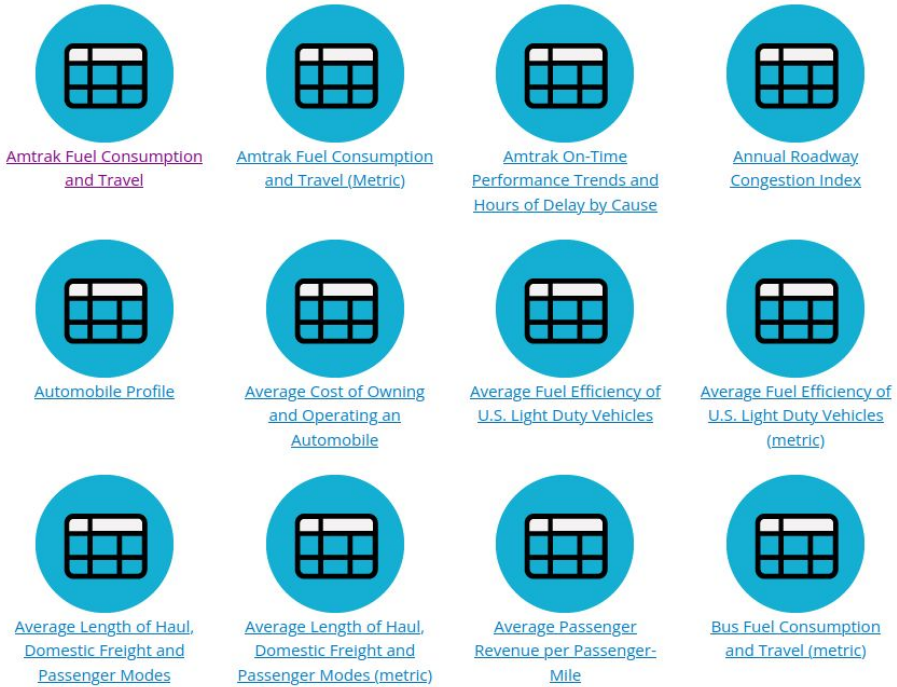
Contents [hide]

1 Objectives
2 Tags
  2.1 Energy
  2.2 Environmental monitoring
  2.3 Waste/recycling
  2.4 Food
  2.5 Transport
  2.6 Natural habitats / land use
  2.7 Green space
3 Environmental features in development
4 Related initiatives
  4.1 OpenStreetMap
  4.2 Other
5 People interested
6 Contact

Possible data sources

Bureau of Transportation Statistics

https://www.bts.gov/product/national-transportation-statistics?keys=miles&field_topic_target_id=All



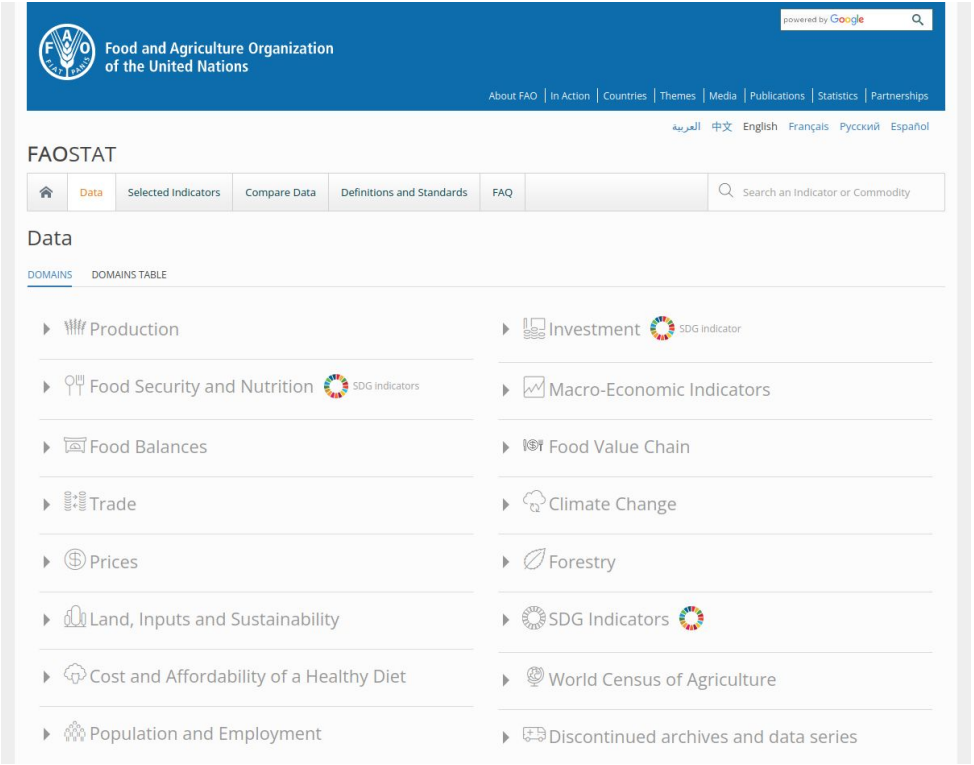
Possible data sources

UN agriculture data (crop and livestock, yields, land area, etc)

https://www.fao.org/faostat/en/#data

Data on what crops are where:

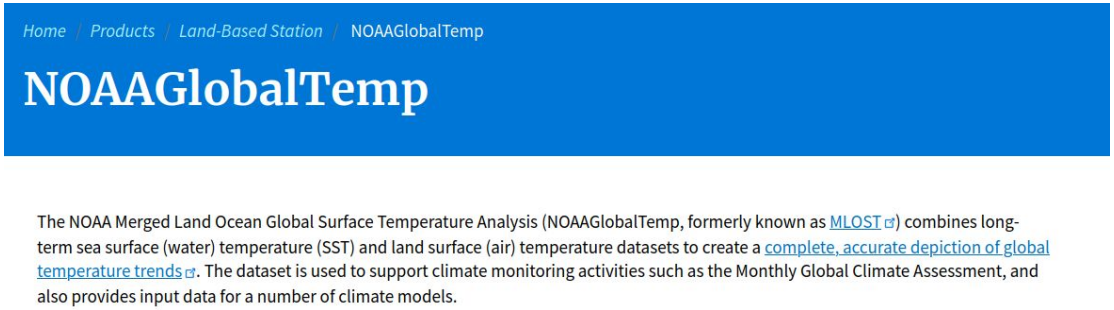
https://www.nass.usda.gov/Research_and_Science/Cropland/sarsfaqs2.php#Section1_1.0



Possible data sources

Global temperatures

https://www.ncei.noaa.gov/products/land-based-station/noaa-global-temp



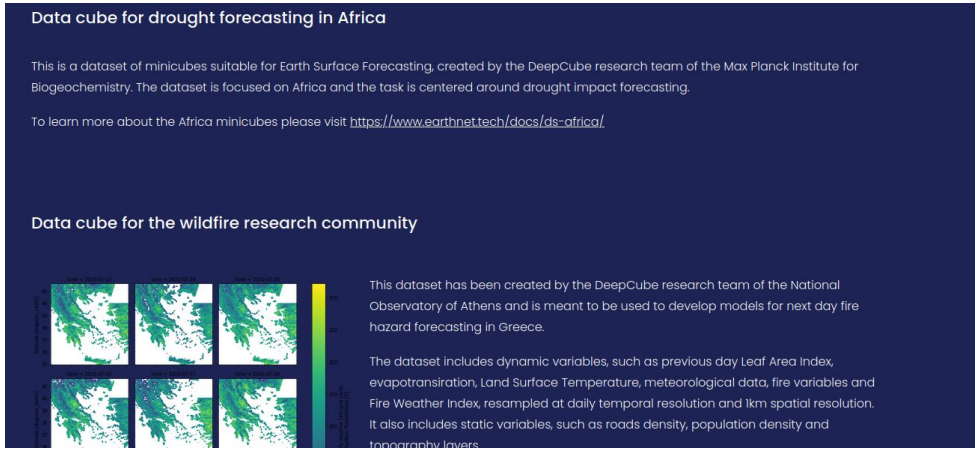
Possible data sources

Satellite datasets

https://deepcube-h2020.eu/technology/earth-system-data-cube/

https://deepcube-h2020.eu/datasets/

https://github.com/satellite-image-deep-learning/datasets



Possible data sources

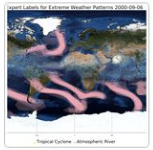
Extreme weather events

https://github.com/andregraubner/ClimateNet

https://gmd.copernicus.org/articles/14/107/2021/

Model description paper | 08 Jan 2021

ClimateNet: an expert-labeled open dataset and deep learning architecture for enabling high-precision analyses of extreme weather



Prabhat, Karthik Kashinath, Mayur Mudigonda, Sol Kim, Lukas Kapp-Schwoerer, Andre Graubner, Ege Karaismailoglu, Leo von Kleist, Thorsten Kurth, Annette Greiner, Ankur Mahesh, Kevin Yang, Colby Lewis, Jiayi Chen, Andrew Lou, Sathyavat Chandran, Ben Toms, Will Chapman, Katherine Dagon, Christine A. Shields, Travis O'Brien, Michael Wehner, and William Collins

Abstract

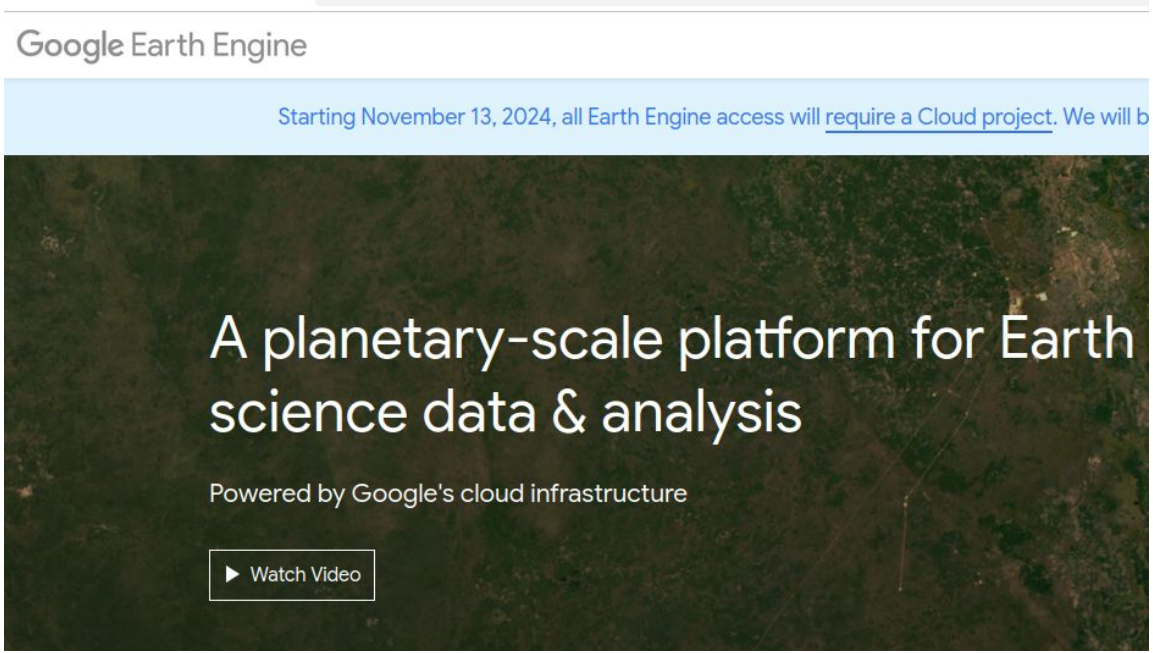
Identifying, detecting, and localizing extreme weather events is a crucial first step in understanding how they may vary under different climate change scenarios. Pattern recognition tasks such as classification, object detection, and segmentation (i.e., pixel-level classification) have remained challenging problems in the weather and climate sciences. While there exist many empirical heuristics for detecting extreme events, the disparities between the output of these different methods even for a single event are large and often difficult to reconcile. Given the success of deep learning (DL) in tackling similar problems in computer vision, we advocate a DL-based approach. DL, however, works best in the context of supervised learning – when labeled datasets are readily available. Reliable labeled training data for extreme weather and climate events is scarce.

We create "ClimateNet" – an open, community-sourced human-expert-labeled curated dataset that captures tropical cyclones (TCs) and atmospheric rivers (ARs) in high-resolution climate model output from a simulation of a recent historical period. We use the curated ClimateNet dataset to train a state-of-the-art DL model for pixel-level identification – i.e., segmentation – of TCs and ARs. We then apply the trained DL model to historical and climate change scenarios simulated by the Community Atmospheric Model (CAM5.1) and show that the DL model accurately segments the data into TCs, ARs, or "the background" at a pixel level. Further, we show how the segmentation results can be used to conduct spatially and temporally precise analytics by quantifying distributions of extreme precipitation conditioned on event types (TC or AR) at regional scales. The key contribution of this work is that it paves the way for DL-based automated, high-fidelity, and highly precise analytics of climate data using a curated expert-labeled dataset – ClimateNet.

Google Earth Engine

Hosts a variety of remote sensing and geospatial datasets

Has a Python API



More paper examples

https://ai4good.org/fragile-earth-2022/

https://www.cambridge.org/core/journals/environmental-data-science

https://www.hadr.ai/accepted-papers

https://openaccess.thecvf.com/CVPR2023_workshops/EarthVision

Want a certain type of data?

Check if there is a government or international agency about it (fingers crossed the data is still there)

Or google (particularly google scholar)

Or ask me

Possible methods

Supervised classification or regression (classic methods, simple artificial neural networks, advanced artificial neural networks)

Computer vision (simple feature extraction, convolutional neural networks)

Self- or Unsupervised methods for representation learning (pca, clustering, artificial neural networks)

Feature importance

Reinforcement learning

Project report

Abstract - a one paragraph summary of the full report

Background - what problem are you addressing, how does it relate to climate change, how has it been approached in the past, how are you approaching it and why

Methods - what data are using, how did you process it, what methods are using, how did you decide hyperparameters, how are you evaluating your model

Results - what did your methods show, what worked and what didn't work, includes pointers to figures

Discussion - what do your results mean, what would you do in the future if you were to continue working on this project, how do your results relate to other studies

References - bibliography of sources you used and cited

Author Contributions - state what work each member of the group did

~3 pages min for first five sections, plus extra pages for references and figures. Plus pre-run python notebook (should replicate your figures from data).

Project Presentations

~10 minutes. Summary of your project report. Should include all the same sections.

All team members must speak in the project presentation.

There will be additional time for questions from other students (required) and me

Advice

Think big (could even result in a publication). This is a growth opportunity.

Choose something meaningful and interesting to you.

Expect things to take a while and be frustrating, especially as you get started with new data or methods

Document what you are doing for yourself and for your final report

Manage your time! Work steadily, don’t wait until the end

Research Quality

How to avoid machine learning pitfalls: a guide for academic researchers

Michael A. Lones*

Abstract

This document is a concise outline of some of the common mistakes that occur when using machine learning, and what can be done to avoid them. Whilst it should be accessible to anyone with a basic understanding of machine learning techniques, it was originally written for research students, and focuses on issues that are of particular concern within academic research, such as the need to do rigorous comparisons and reach valid conclusions. It covers five stages of the machine learning process: what to do before model building, how to reliably build models, how to robustly evaluate models, how to compare models fairly, and how to report results.

https://arxiv.org/pdf/2108.02497.pdf

Contents

Writing evaluation

Your writing should be clear and concise, with minimal grammatical errors.

To write your background, find some example papers that may have tackled a similar problem so that you can situate your work relevant to past work (probably should do this before you start!).

For the methods, be clear and complete enough that someone could roughly replicate what you did just from reading your methods section.

Figures should be easy to understand, with clearly labeled axes and not too many different colors or line styles.

Presentation Evaluation

Slides should have clear and informative titles

Figures should be fully visible with labeled axes

Slides should not have too much written text; rather, you should have one or two main points on the slide with one or two major visuals. Elaborate on what is on the slides as you speak.

Speak loud enough for everyone to hear

Go at the right pace throughout (time yourself beforehand!)

Advice: imagine yourself in the audience listening to a project you know nothing about. How would you want the speaker to communicate to you?

Chat GPT

What are some academic papers that use machine learning to help with building energy efficiency?

Here are some academic papers that explore the use of machine learning for improving building energy efficiency:

1. "A Survey of Machine Learning Techniques for Building Energy Management"

- **Authors**: Hossain, M. S., & O'Brien, W.
- **Published in**: *Energy and Buildings* (2018)
- **Summary**: This paper provides a comprehensive review of how machine learning techniques, such as supervised learning, unsupervised learning, and reinforcement learning, can be applied to various aspects of building energy management. It discusses applications such as HVAC (Heating, Ventilation, and Air Conditioning) optimization, energy consumption prediction, and fault detection in buildings.

2. "Deep Learning for Building Energy Prediction: A Case Study of the Faculty Building in Singapore"

- **Authors**: Chien, S., Lee, J., & Su, Y.
- **Published in**: *Energy* (2020)
- **Summary**: This paper presents a case study where deep learning techniques were used to predict the energy consumption of a faculty building in Singapore. The study used Long Short-Term Memory (LSTM) networks to model and predict energy consumption based on historical data.

3. "Application of Machine Learning in Energy Management Systems"

- **Authors**: Zhang, M., & Wei, H.
- **Published in**: *Energy Procedia* (2019)

Ask anything

Nope

Nope

Nope

Only use AI with extreme caution, and make sure you double check all of its work. *You* are responsible for what you submit.

Writing integrity

Deep Learning for Classifying Food Waste

Amin Mazloumian
Institute of Embedded Systems
Zurich University of Applied Sciences
Zurich, Switzerland
amin.mazloumian@zhaw.ch

Matthias Rosenthal
Institute of Embedded Systems
Zurich University of Applied Sciences
Zurich, Switzerland
matthias.rosenthal@zhaw.ch

Hans Gelke
Institute of Embedded Systems
Zurich University of Applied Sciences
Zurich, Switzerland
hans.gelke@zhaw.ch

***Abstract*—One third of food produced in the world for human consumption – approximately 1.3 billion tons – is lost or wasted every year. By classifying food waste of individual consumers and raising awareness of the measures, avoidable food waste can be significantly reduced. In this research, we use deep learning to classify food waste in half a million images captured by cameras installed on top of food waste bins. We specifically designed a deep neural network that classifies food waste for every time food waste is thrown in the waste bins. Our method presents how deep learning networks can be tailored to best learn from available training data.**

***Index Terms*—deep learning, neural networks, machine learning, food waste**

The automatic classification of food waste in waste bins is technically a difficult computer vision task for the following reasons. a) It is visually hard to differentiate between edible and not-edible food waste. As an example consider distinguishing between eggs and empty egg shells. b) Same food classes come in a wide variety of textures and colors if cooked or processed. c) Liquid food waste, e.g. soups and stews, and soft food waste, e.g. chopped vegetables and salads, can largely hide and cover visual features of other food classes.

In this research, we adopt a deep convoultional neural network approach for classifying food waste in waste bins [9]. Deep convolutional neural networks are supervised machine

My Research Paper

Grace Lindsay

The automatic classification of food waste in waste bins is technically a difficult computer vision task for the following reasons. a) It is visually hard to differentiate between edible and not-edible food waste. As an example consider distinguishing between eggs and empty egg shells. b) Same food classes come in a wide variety of textures and colors if cooked or processed. c) Liquid food waste, e.g. soups and stews, and soft food waste, e.g. chopped vegetables and salads, can largely hide and cover visual features of other food classes. Here, I will test three different neural networks trained to classify food waste

Writing integrity

Deep Learning for Classifying Food Waste

Amin Mazloumi
Institute of Embedded S
Zurich University of Applie
Zurich, Switzerlan
amin.mazloumian@zh

***Abstract*—One third of food**
consumption – approximately 1
every year. By classifying food w
raising awareness of the measu
significantly reduced. In this r
classify food waste in half a mil
installed on top of food waste
deep neural network that clas
food waste is thrown in the w
how deep learning networks ca
available training data.

***Index Terms*—deep learning,**
ing, food waste

Academic Disciplinary offenses include but are not limited to:

- cheating, plagiarism, falsification of data or sources, forgery of academic documents in attempt to defraud;
- destruction, theft, or unauthorized use of laboratory data, or research materials.

The following penalties may be imposed by the faculty for disciplinary infractions:

- Warning
- Disciplinary Probation
- Suspension
- Dismissal

:hnically a difficult
/ hard to differentiate
nsider distinguishing
me in a wide variety
vaste, e.g. soups and
ds, can largely hide
st three different

Writing integrity

Deep Learning for Classifying Food Waste

Amin Mazloumian
Institute of Embedded Systems
Zurich University of Applied Sciences
Zurich, Switzerland
amin.mazloumian@zhaw.ch

Matthias Rosenthal
Institute of Embedded Systems
Zurich University of Applied Sciences
Zurich, Switzerland
matthias.rosenthal@zhaw.ch

Hans Gelke
Institute of Embedded Systems
Zurich University of Applied Sciences
Zurich, Switzerland
hans.gelke@zhaw.ch

***Abstract*—One third of food produced in the world for human consumption – approximately 1.3 billion tons – is lost or wasted every year. By classifying food waste of individual consumers and raising awareness of the measures, avoidable food waste can be significantly reduced. In this research, we use deep learning to classify food waste in half a million images captured by cameras installed on top of food waste bins. We specifically designed a deep neural network that classifies food waste for every time food waste is thrown in the waste bins. Our method presents how deep learning networks can be tailored to best learn from available training data.**

***Index Terms*—deep learning, neural networks, machine learning, food waste**

The automatic classification of food waste in waste bins is technically a difficult computer vision task for the following reasons. a) It is visually hard to differentiate between edible and not-edible food waste. As an example consider distinguishing between eggs and empty egg shells. b) Same food classes come in a wide variety of textures and colors if cooked or processed. c) Liquid food waste, e.g. soups and stews, and soft food waste, e.g. chopped vegetables and salads, can largely hide and cover visual features of other food classes.

In this research, we adopt a deep convoultional neural network approach for classifying food waste in waste bins [9]. Deep convolutional neural networks are supervised machine

My Research Paper

Grace Lindsay

According to Mazloumian et al. (2020): "The automatic classification of food waste in waste bins is technically a difficult computer vision task for the following reasons. a) It is visually hard to differentiate between edible and not-edible food waste. As an example consider distinguishing between eggs and empty egg shells. b) Same food classes come in a wide variety of textures and colors if cooked or processed. c) Liquid food waste, e.g. soups and stews, and soft food waste, e.g. chopped vegetables and salads, can largely hide and cover visual features of other food classes." Here, I will test three different neural networks trained to classify food waste

Writing integrity

Deep Learning for Classifying Food Waste

Amin Mazloumian
Institute of Embedded Systems
Zurich University of Applied Sciences
Zurich, Switzerland
amin.mazloumian@zhaw.ch

Matthias Rosenthal
Institute of Embedded Systems
Zurich University of Applied Sciences
Zurich, Switzerland
matthias.rosenthal@zhaw.ch

Hans Gelke
Institute of Embedded Systems
Zurich University of Applied Sciences
Zurich, Switzerland
hans.gelke@zhaw.ch

***Abstract*—One third of food produced in the world for human consumption – approximately 1.3 billion tons – is lost or wasted every year. By classifying food waste of individual consumers and raising awareness of the measures, avoidable food waste can be significantly reduced. In this research, we use deep learning to classify food waste in half a million images captured by cameras installed on top of food waste bins. We specifically designed a deep neural network that classifies food waste for every time food waste is thrown in the waste bins. Our method presents how deep learning networks can be tailored to best learn from available training data.**

***Index Terms*—deep learning, neural networks, machine learning, food waste**

The automatic classification of food waste in waste bins is technically a difficult computer vision task for the following reasons. a) It is visually hard to differentiate between edible and not-edible food waste. As an example consider distinguishing between eggs and empty egg shells. b) Same food classes come in a wide variety of textures and colors if cooked or processed. c) Liquid food waste, e.g. soups and stews, and soft food waste, e.g. chopped vegetables and salads, can largely hide and cover visual features of other food classes.

In this research, we adopt a deep convoultional neural network approach for classifying food waste in waste bins [9]. Deep convolutional neural networks are supervised machine

My Research Paper

Grace Lindsay

According to Mazloumian et al. (2020): "The automatic classification of food waste in waste bins is technically a difficult computer vision task for the following reasons. a) It is visually hard to differentiate between edible and not-edible food waste. As an example consider distinguishing between eggs and empty egg shells. b) Same food classes come in a wide variety of textures and colors if cooked or processed. c) Liquid food waste, e.g. soups and stews, and soft food waste, e.g. chopped vegetables and salads, can largely hide and cover visual features of other food classes." Here, I will test three different neural networks trained to classify food waste

Technically not plagiarism, but bad writing!

Writing integrity

Deep Learning for Classifying Food Waste

Amin Mazloumian
Institute of Embedded Systems
Zurich University of Applied Sciences
Zurich, Switzerland
amin.mazloumian@zhaw.ch

Matthias Rosenthal
Institute of Embedded Systems
Zurich University of Applied Sciences
Zurich, Switzerland
matthias.rosenthal@zhaw.ch

Hans Gelke
Institute of Embedded Systems
Zurich University of Applied Sciences
Zurich, Switzerland
hans.gelke@zhaw.ch

***Abstract*—One third of food produced in the world for human consumption – approximately 1.3 billion tons – is lost or wasted every year. By classifying food waste of individual consumers and raising awareness of the measures, avoidable food waste can be significantly reduced. In this research, we use deep learning to classify food waste in half a million images captured by cameras installed on top of food waste bins. We specifically designed a deep neural network that classifies food waste for every time food waste is thrown in the waste bins. Our method presents how deep learning networks can be tailored to best learn from available training data.**

***Index Terms*—deep learning, neural networks, machine learning, food waste**

The automatic classification of food waste in waste bins is technically a difficult computer vision task for the following reasons. a) It is visually hard to differentiate between edible and not-edible food waste. As an example consider distinguishing between eggs and empty egg shells. b) Same food classes come in a wide variety of textures and colors if cooked or processed. c) Liquid food waste, e.g. soups and stews, and soft food waste, e.g. chopped vegetables and salads, can largely hide and cover visual features of other food classes.

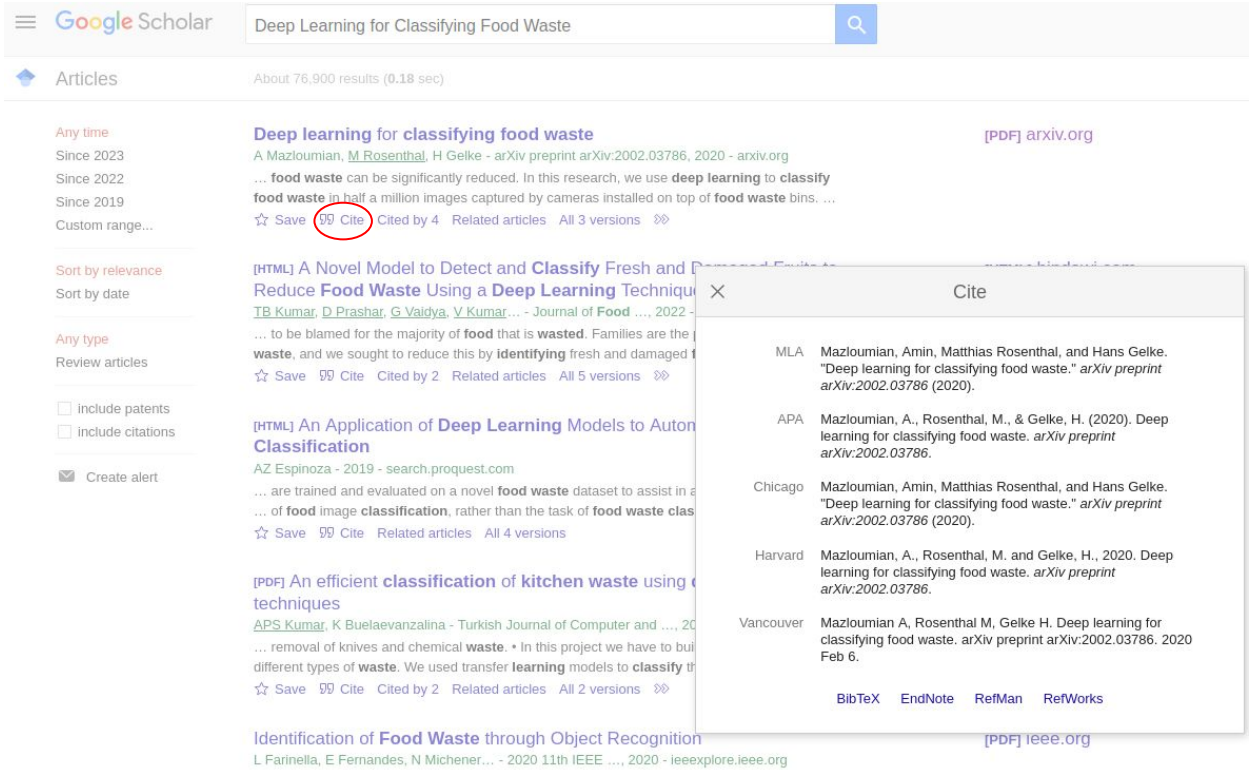
In this research, we adopt a deep convoultional neural network approach for classifying food waste in waste bins [9]. Deep convolutional neural networks are supervised machine

My Research Paper
Grace Lindsay

According to Mazloumian et al. (2020), visually classifying food waste in bins faces three main challenges: distinguishing edible from non-edible waste, overcoming how diverse single food classes can be, and classifying images with liquid and soft waste that hide other waste classes. Here, I will test three different neural networks trained to classify food waste

Lovely!

Citations and references



Put all full citations in the reference section.

Project Timeline

Now - Start thinking about your teams and project ideas, look into data and methods.

Mar 6 - Project Plan Homework assigned

Mar 20 - Project Plan Homework due by the end of the day, feedback the following day (teams cannot be changed after this point)

April 17 - In class project check-in

April 24 - In class project work time

May 1 - In class project presentations

May 8 - Project reports due

Exam Review

Exam will include multiple choice, T/F, and short answer questions

You will be tested on knowledge of climate change topics as well as machine learning topics

There will be questions about the papers we've read

You will not need to write code

To study for the exam:

For climate content, know:

-the basics of what is causing climate change

-where emissions are coming from

-changes we expect to happen as a result of warming

-basics of remote sensing

-SSPs

To study for the exam:

For ML content, know:

-basics of the simple regression models

-the basics of artificial neural networks

-what architectures are used for what tasks and why

-what loss functions are used for different problems

-what balanced data means

-pros and cons of different evaluation metrics

-validation/generalization

-types of data (image, time series, etc)

-Transfer learning, representation learning

To study for the exam:

Make sure for each of the papers we've read, you know:

-PMIRO answers

-The general architecture of the model used, what the inputs to the model were, any preprocessing discussed, what the output was, what the objective function was, how the performance of model was evaluated, what kind generalization did they test for

In all cases, focus on what was covered in class