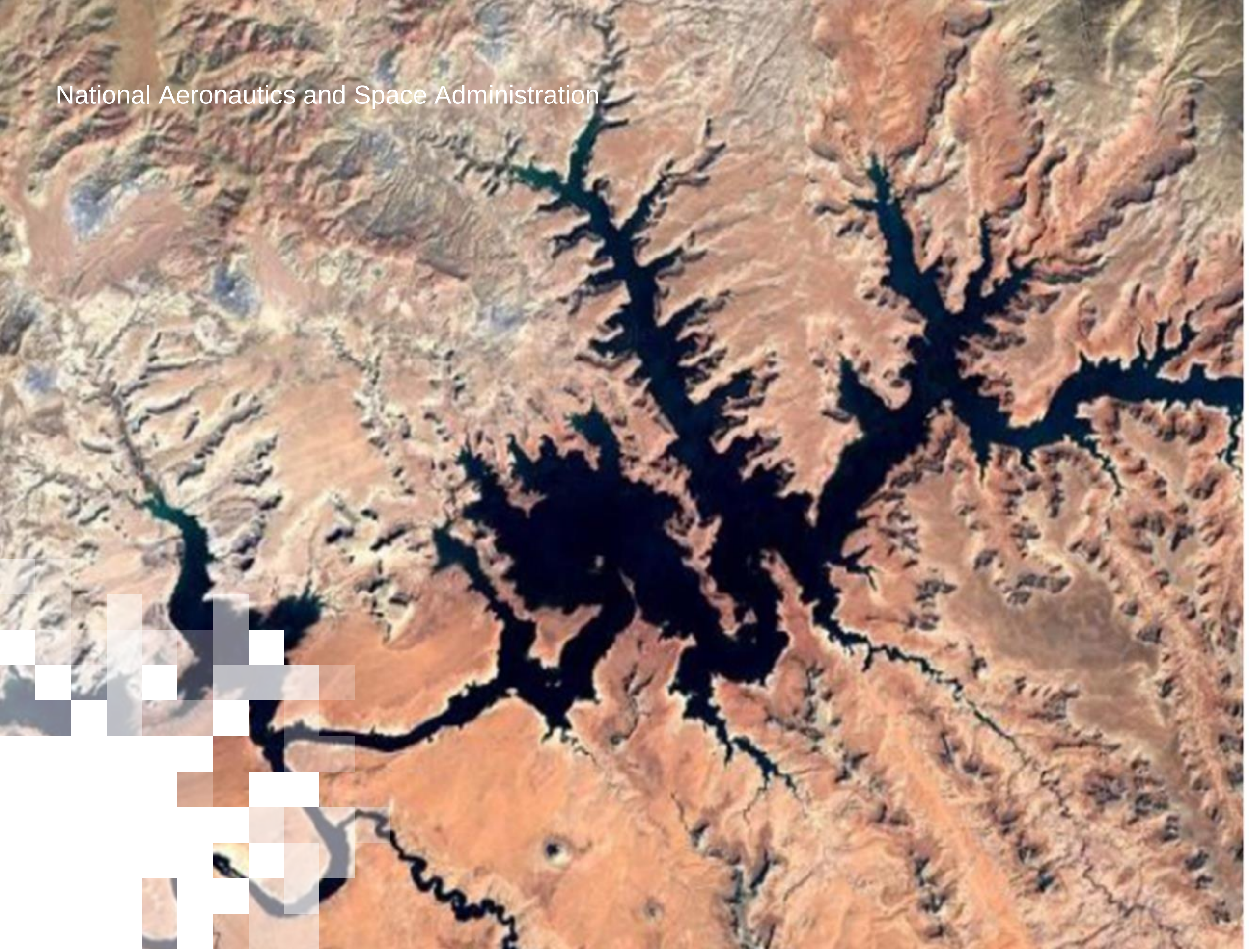
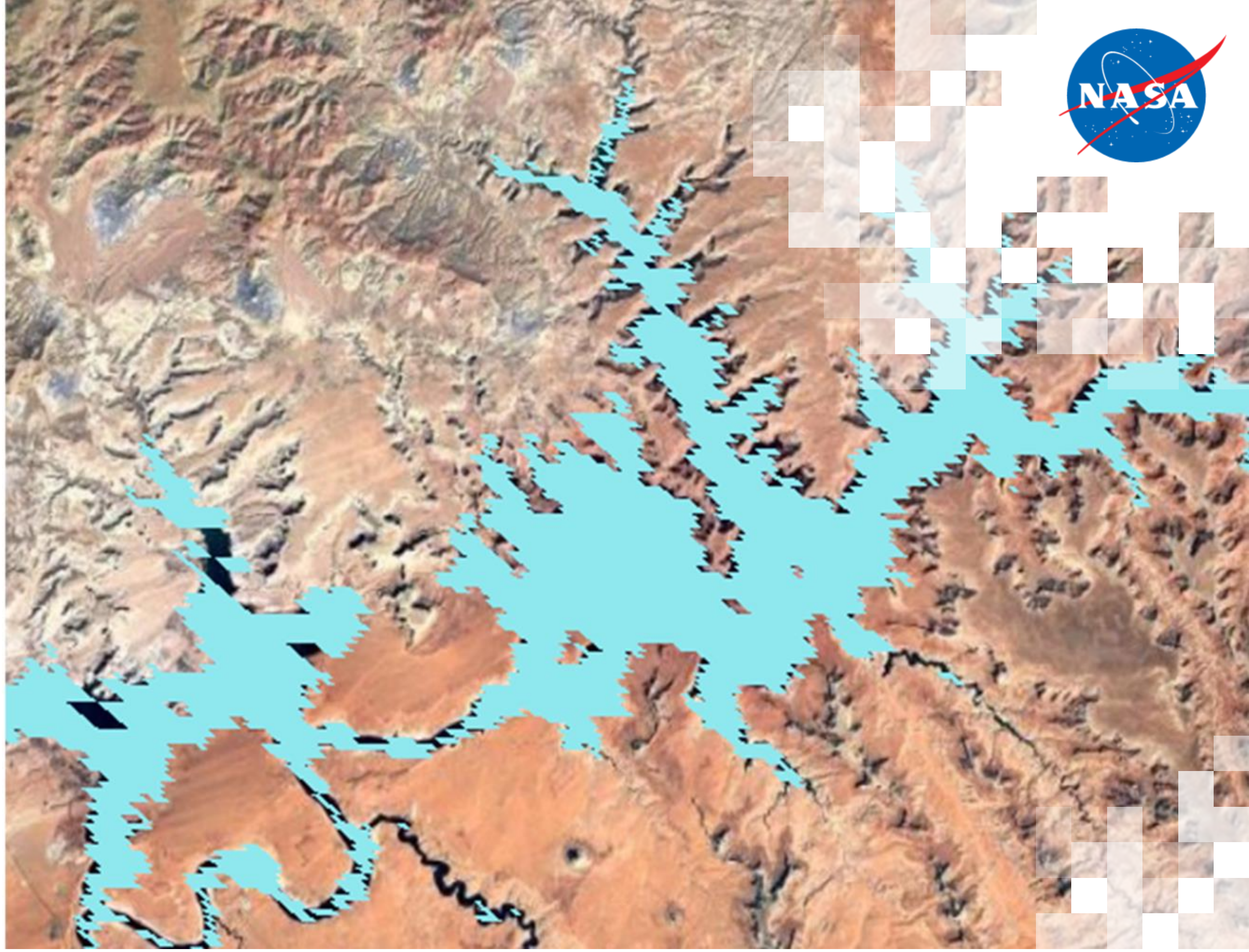
National Aeronautics and Space Administration





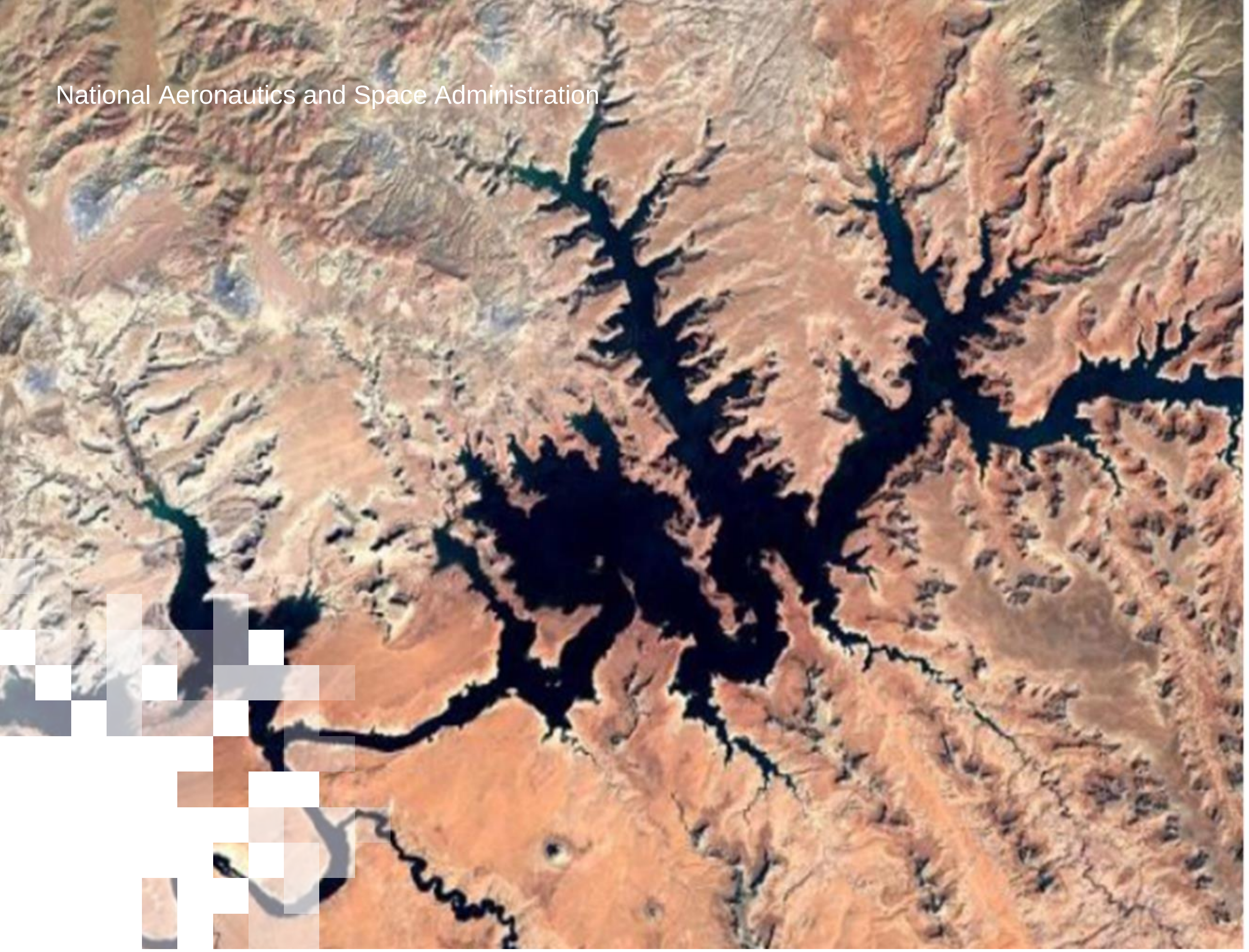
Fundamentals of Machine Learning for Earth Science

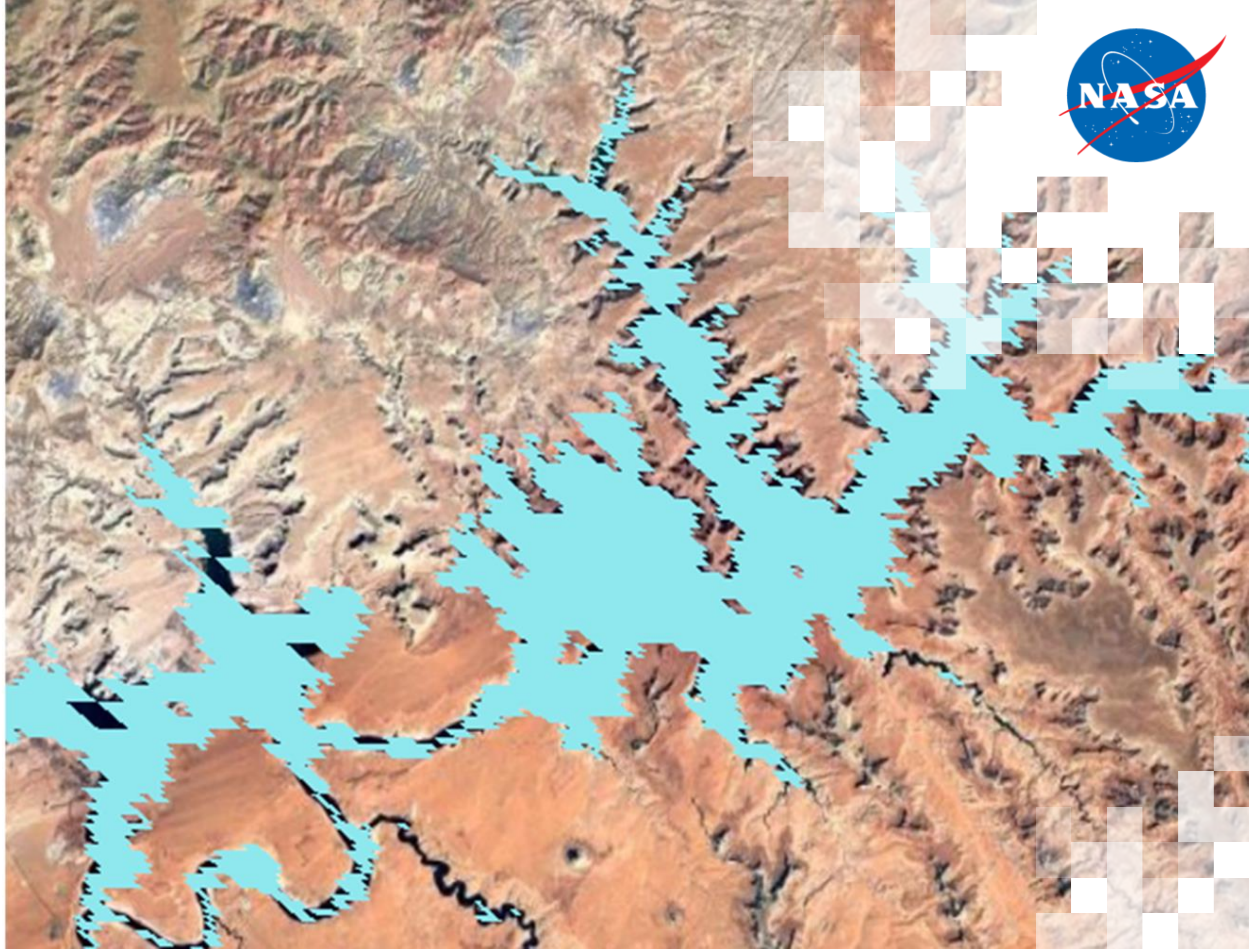
Part 3: Model Tuning, Parameter Optimization, and Additional Machine Learning Algorithms

Trainers: Jordan A. Caraballo-Vega, Mark L. Carroll, Jules Kouatchou, Jian Li, Caleb S. Spradlin

May 4, 2023

National Aeronautics and Space Administration





NASA Applied Remote Sensing Training (ARSET)

Brock Blevins, Training Coordinator

Training Objectives

At the end of the training, participants will be able to:

- Recognize the most common machine learning methods used for processing Earth Science data
- Describe the benefits and limitations of machine learning for Earth Science analysis
- Explain how to apply basic machine learning algorithms and techniques in a meaningful manner to remote sensing data
- Use an analysis-appropriate training dataset to evaluate conditions and solutions for a given case study
- Complete basic procedures to interpret, refine and evaluate the accuracy of the results of machine learning analysis

Reminder of Prerequisites

- Prerequisites:
  - Session 1 of our on-demand Fundamentals of Remote Sensing series or have equivalent experience.
  - Attendees will need access to Google Drive and Google Colab. To access these resources, users must use an email ending in ‘gmail.com’.
  - We will have the video of this demonstration within the training recording available within 48 hours after the presentation for you to go through at your own pace.

Training Schedule

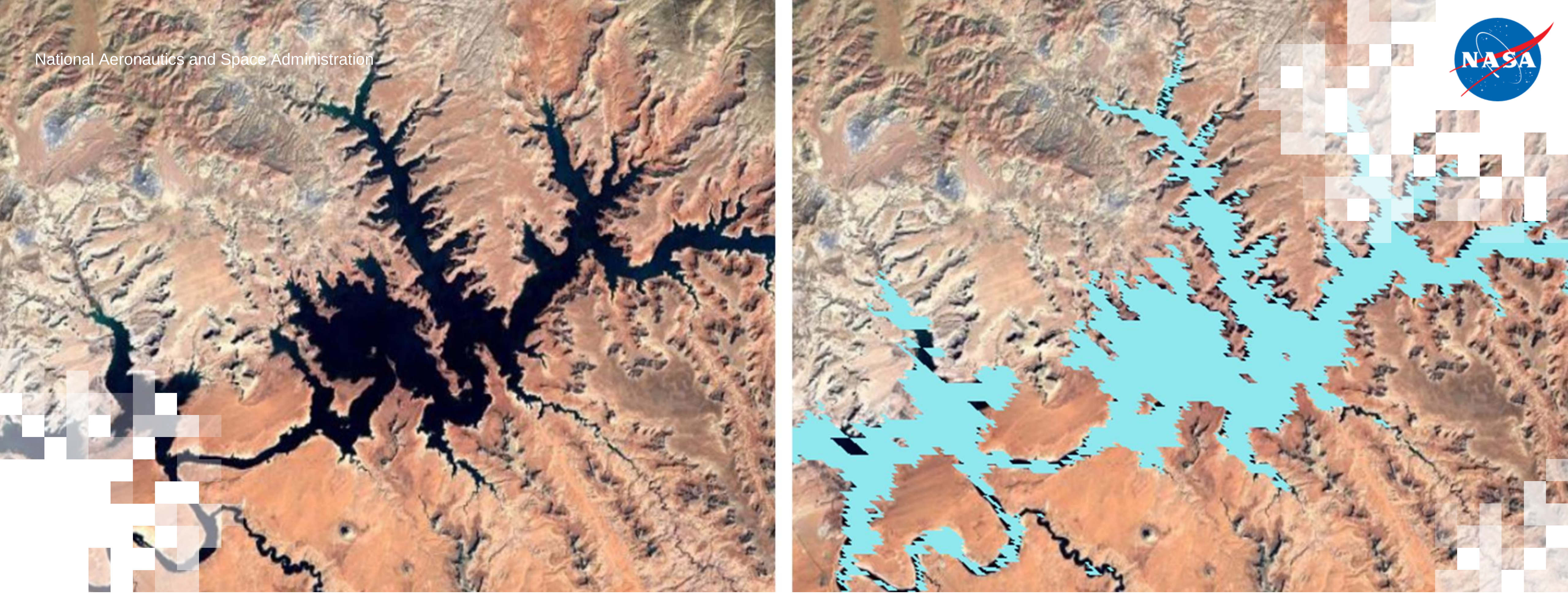
Part 1: Overview of Machine Learning
April 20, 2023

Part 2: Training Data and Land Cover Classification Example
April 27, 2023

Part 3: Model Tuning, Parameter Optimization, and Additional Machine Learning Algorithms
May 4, 2023

Homework
Independent practice and application
Due May 19

Optional opportunity to earn a certificate of completion



Fundamentals of Machine Learning for Earth Science

Part 3: Model Tuning, Parameter Optimization, and Additional Machine Learning Algorithms

Trainers: Jordan A. Caraballo-Vega, Mark L. Carroll, Jules Kouatchou, Jian Li, Caleb S. Spradlin

May 4, 2023

APPLIED REMOTE SENSING TRAINING PROGRAM
TRAIN - EMPOWER - ADVANCE

Session 3 Outline

- Overview of model tuning
- Overview of parameter optimization
- Exercise to optimize existing model
- Overview of model explainability and interpretability
- Overview of additional machine learning algorithms
- Hands-On Jupyter Notebook Exercise: Improvements to MODIS Water Classification Model
- Post-Session Assignment
- Q&A Session

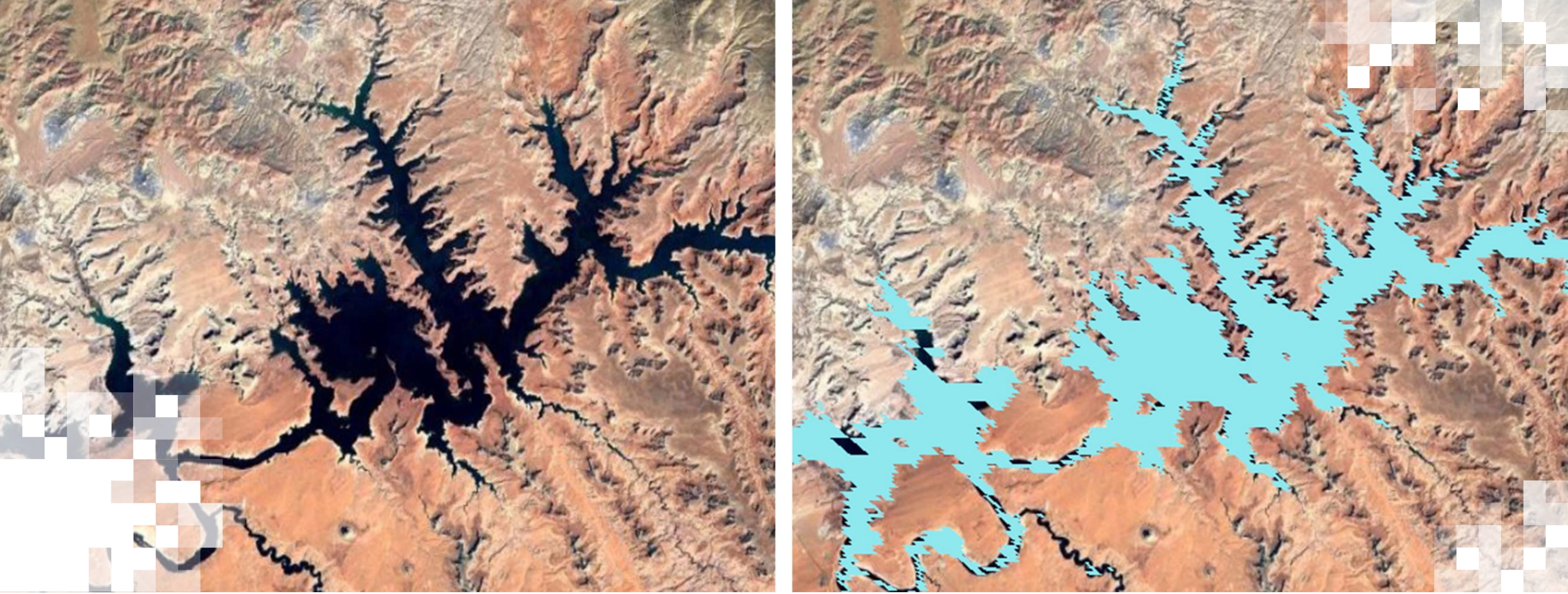
Resources for this Training

https://github.com/NASAARSET/ARSET_ML_Fundamentals

Training Objectives

After participating in this training, attendees will be able to:

- Recognize the most common optimization methods for machine learning algorithms used for processing Earth Science data
- Describe the benefits and limitations of machine learning algorithms optimization for Earth Science analysis
- Explain how to apply basic explainable artificial intelligence techniques for machine learning algorithms in a meaningful manner to remote sensing data

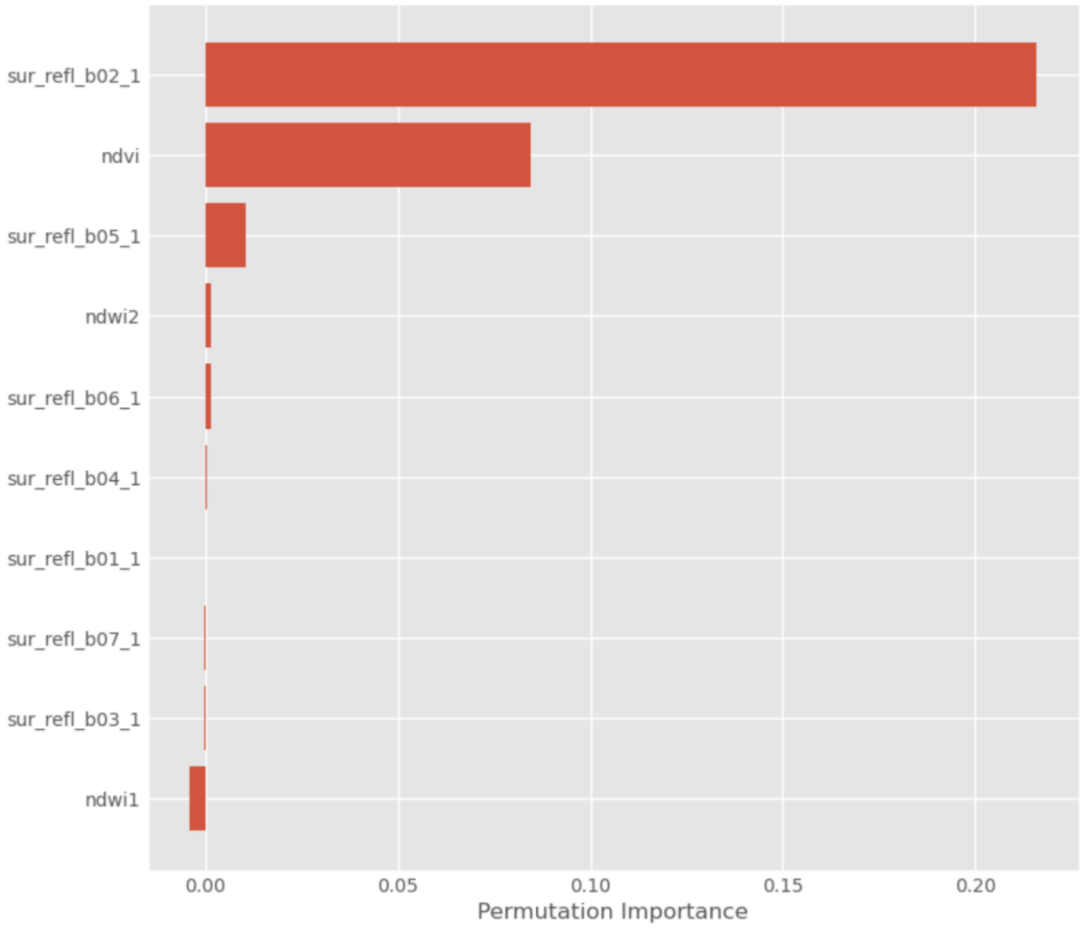


Overview of Model Tuning and Optimization

Trainer: Jordan A. Caraballo-Vega

Summary from Session 2

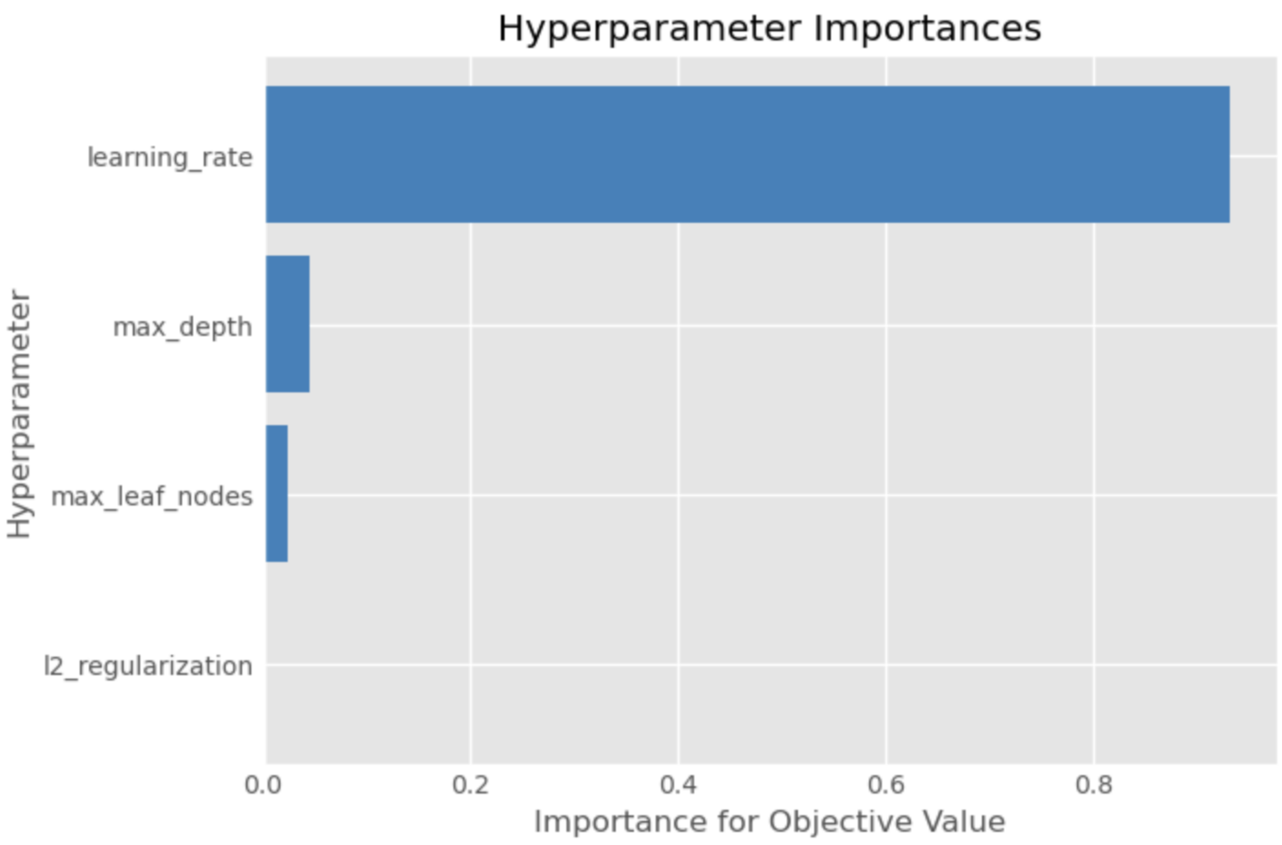
- We trained a Random Forest model that performs binary classification of water and no-water pixels using MODIS surface reflectance data.
- We evaluated our model using our test dataset, and we performed inference using raster tiles.
- Our model was able to successfully identify water pixels, but we also identified locations where the model needs improvement.



Input Bands used in Model Training

Model Tuning and Optimization

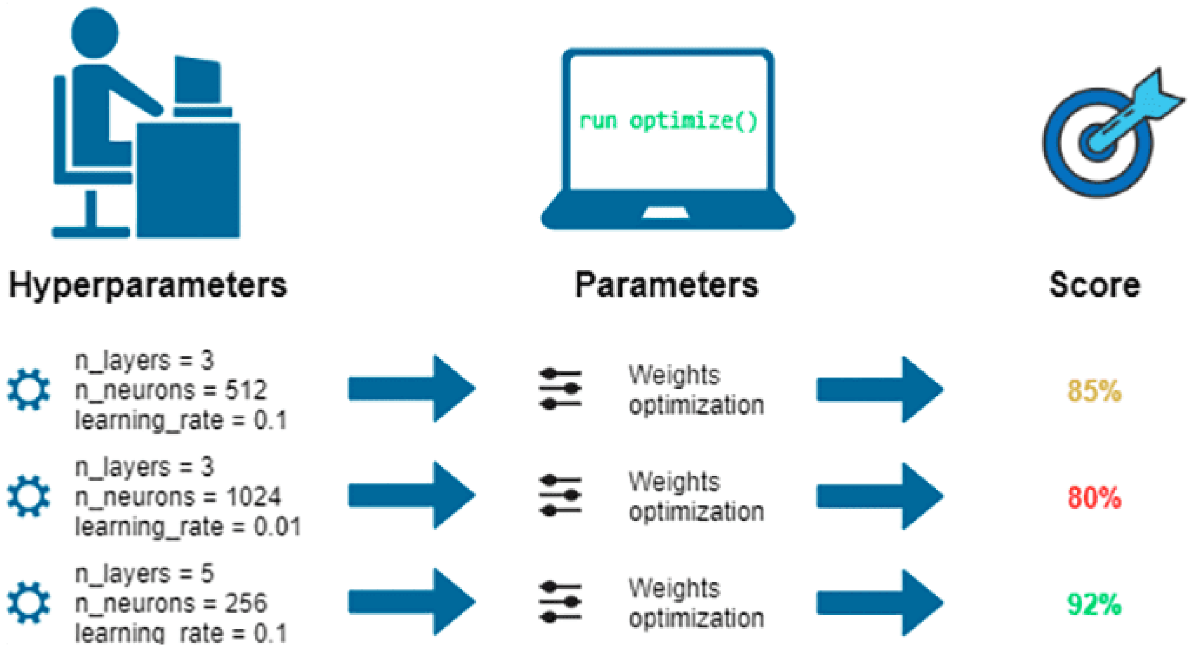
- Your first model will not always be the most accurate.
- Most ML model Application Programming Interface (APIs) will include default hyperparameters to initialize your model, which you can use as a starting point.
- These hyperparameters are set before the learning process begins and will influence model convergence.
- We have several options to improve our model before adding or generating more training data.



Hyperparameter Importances

Model Tuning and Optimization, Cont.

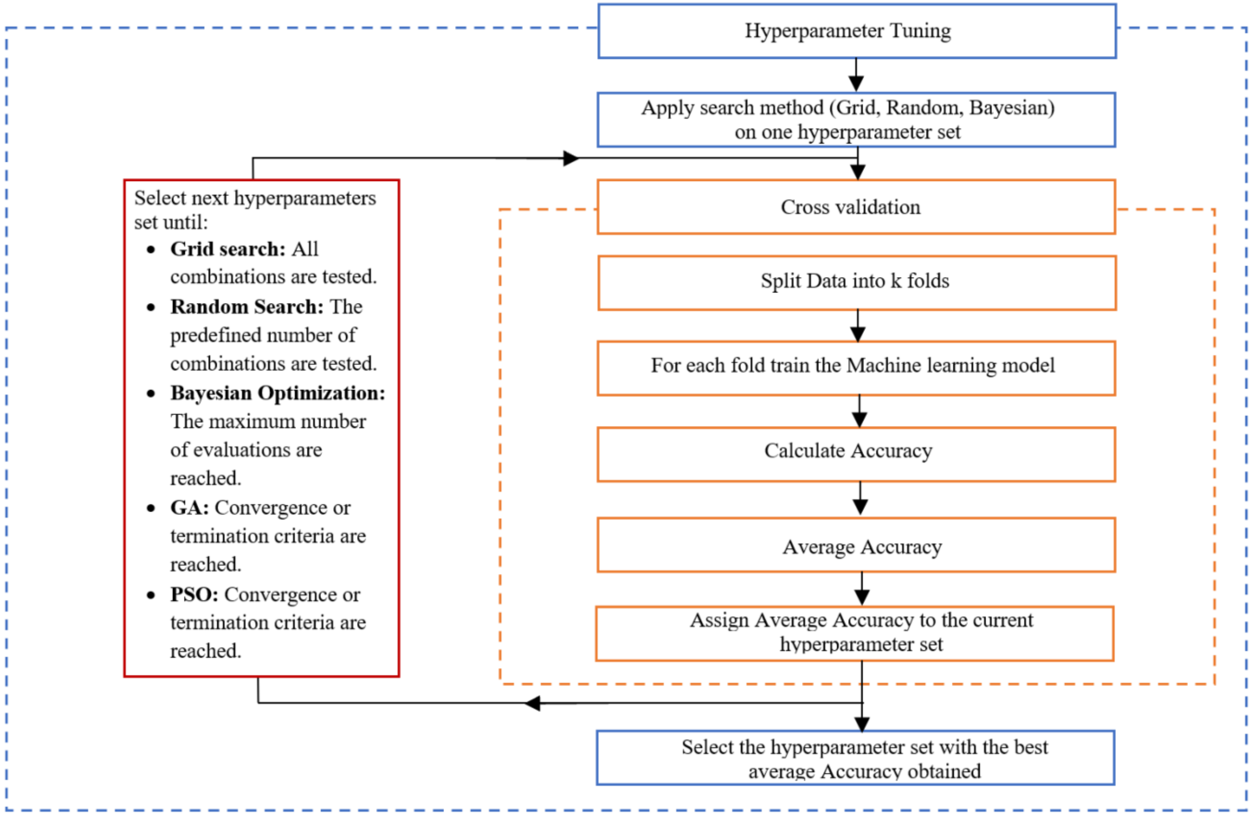
- Most of the time we are not optimizing our models for the most common features, we are attempting to improve our model so it can generalize across uncommon features.
- By fine-tuning the model, we can maximize its performance and get the highest rate of performance possible.
- Model tuning in general is the experimental process of finding the optimal values of hyperparameters to maximize model performance.



There is a metric we will want to optimize for. Image Source: medium.com

Techniques for Model Tuning

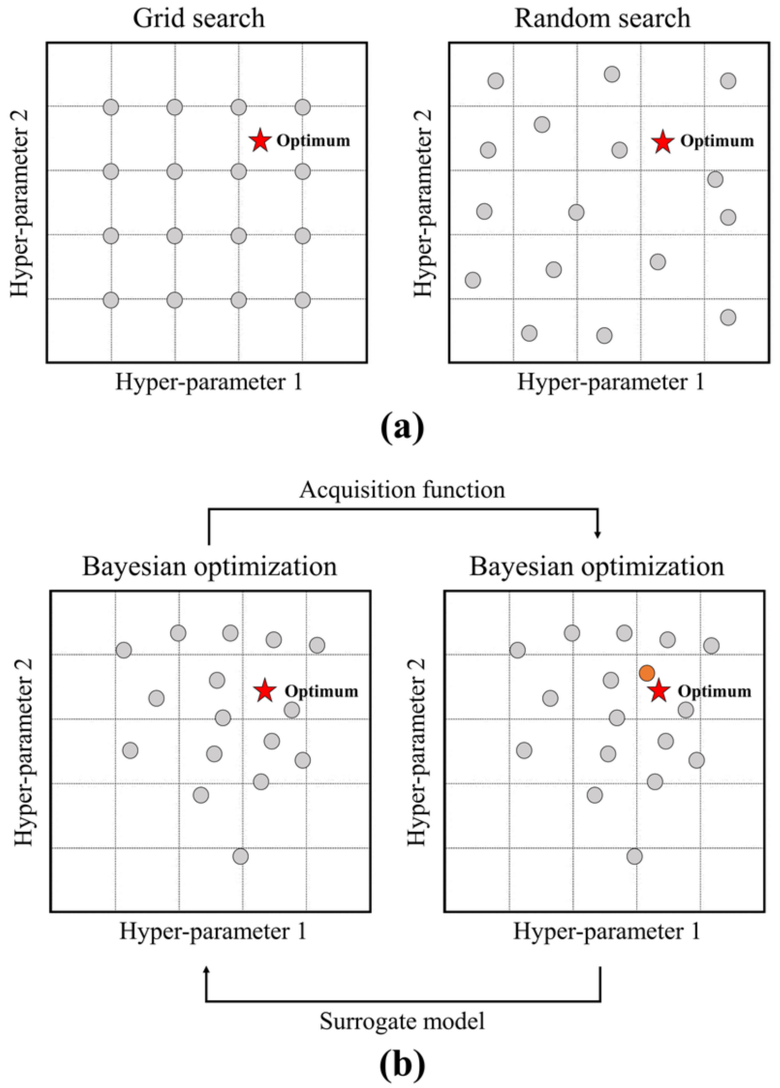
- A robust evaluation criterion should be identified and set before model tuning to optimize the tuning parameters towards the specific goal.
- **Manual Model Tuning:** Hyperparameter values are set based on intuition or past experience. The model is then trained and evaluated to determine the performance using the respective set of hyperparameters.
- **Automated Model Tuning:** optimal hyperparameter values are found using algorithms. Here, we define a hyperparameter search space from which the optimal set of hyperparameter values is selected.



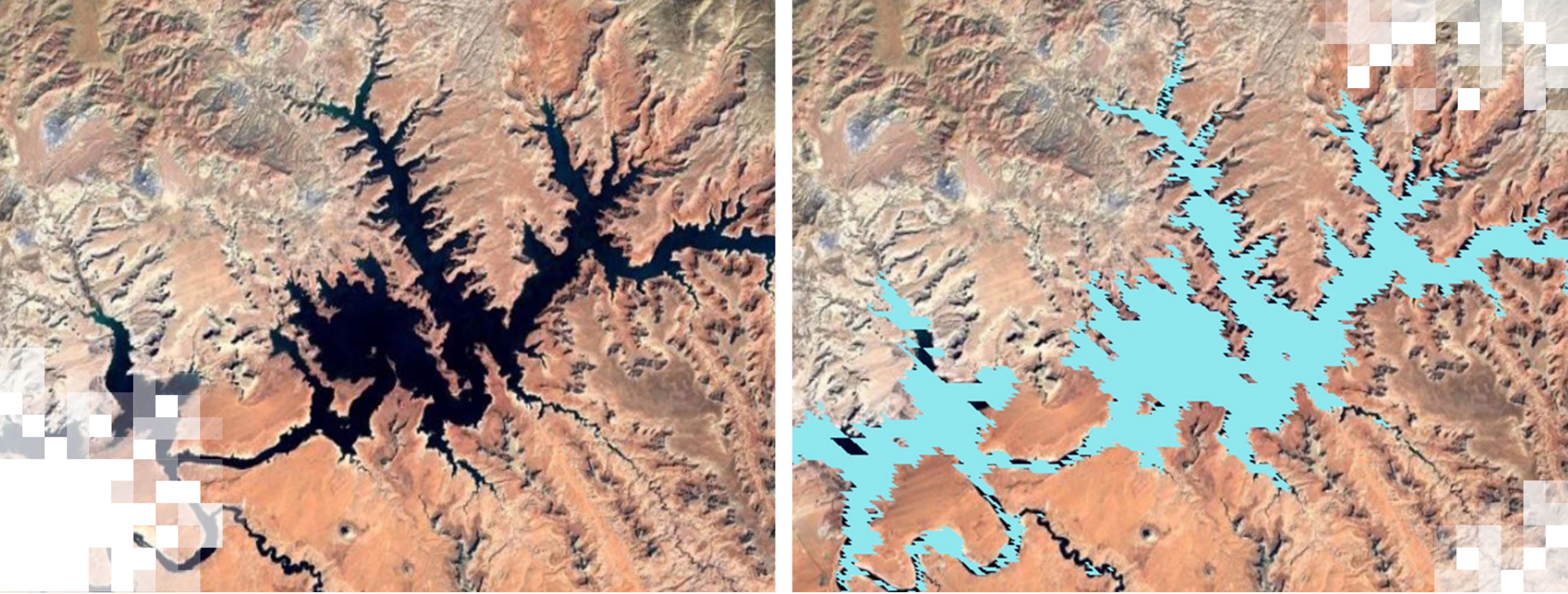
Hyperparameter tuning workflow. Hyperopt is an excellent tool for automated model tuning.

Techniques for Model Tuning

- **Grid Search:** The user defines a set of values for each hyperparameter to form a grid. Different combinations of these hyperparameter values are tried and the combination which yields the best result is selected as the final set of optimal hyperparameters.
- **Random Search:** The algorithm will only try random combinations of hyperparameter values rather than every possible combination.
- **Bayesian Search:** Keeps track of past evaluation results to form the information used to make future decisions in selecting future hyperparameter values.
- At the end, the combination that yields the best result from this is selected as the optimal set of hyperparameters.

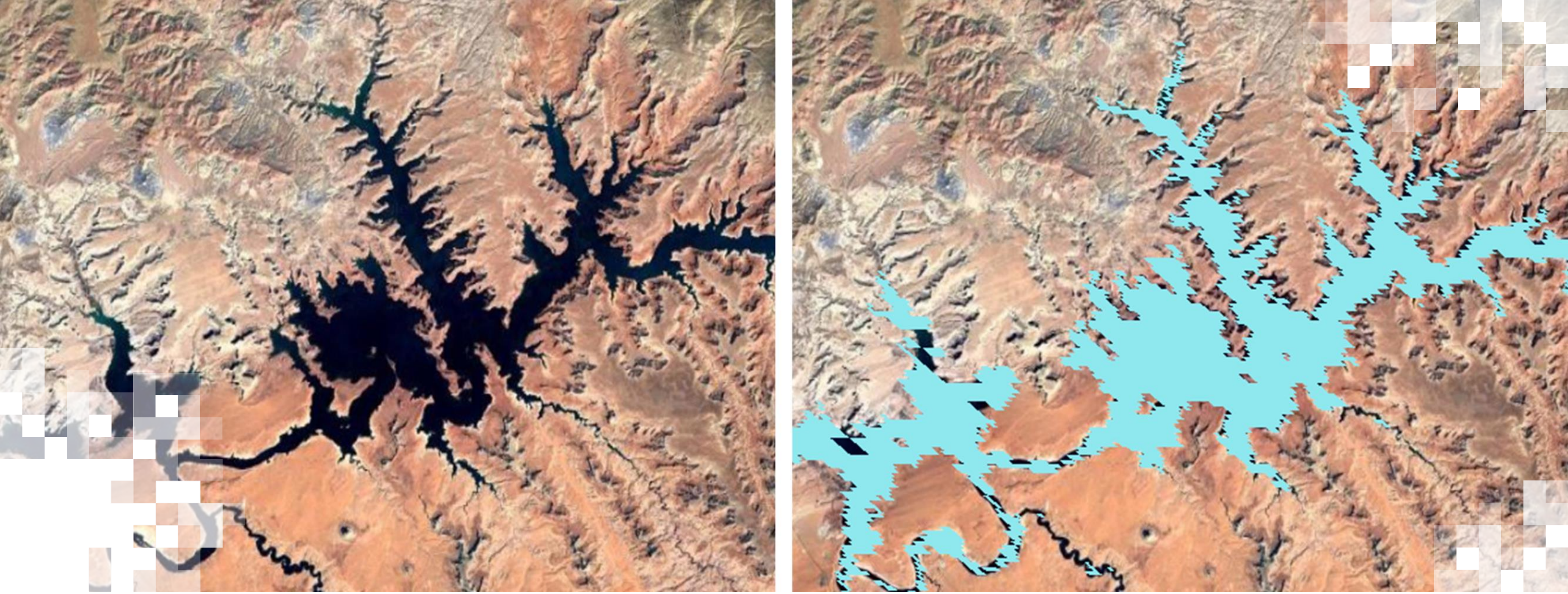


Grid, Random, and Bayesian optimizations illustrated. Kim et al. (2021), https://doi.org/10.1109/ACCESS.2021.3051619



Exercise: Tuning and Optimization of the Random Forest Model

Trainer: Jordan A. Caraballo-Vega

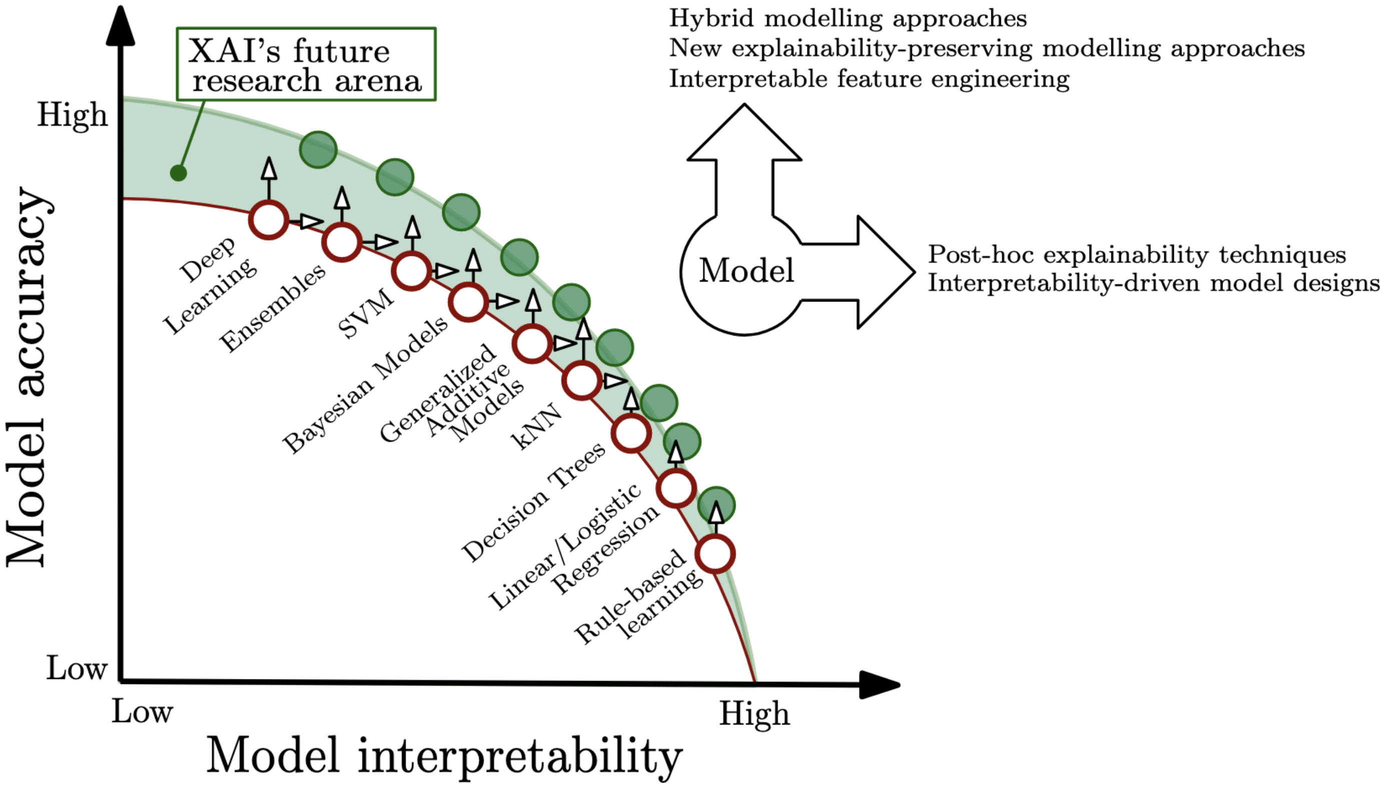


Model Explainability and Interpretability - XAI

Trainer: Caleb S. Spradlin

Model Explainability and Interpretability – XAI

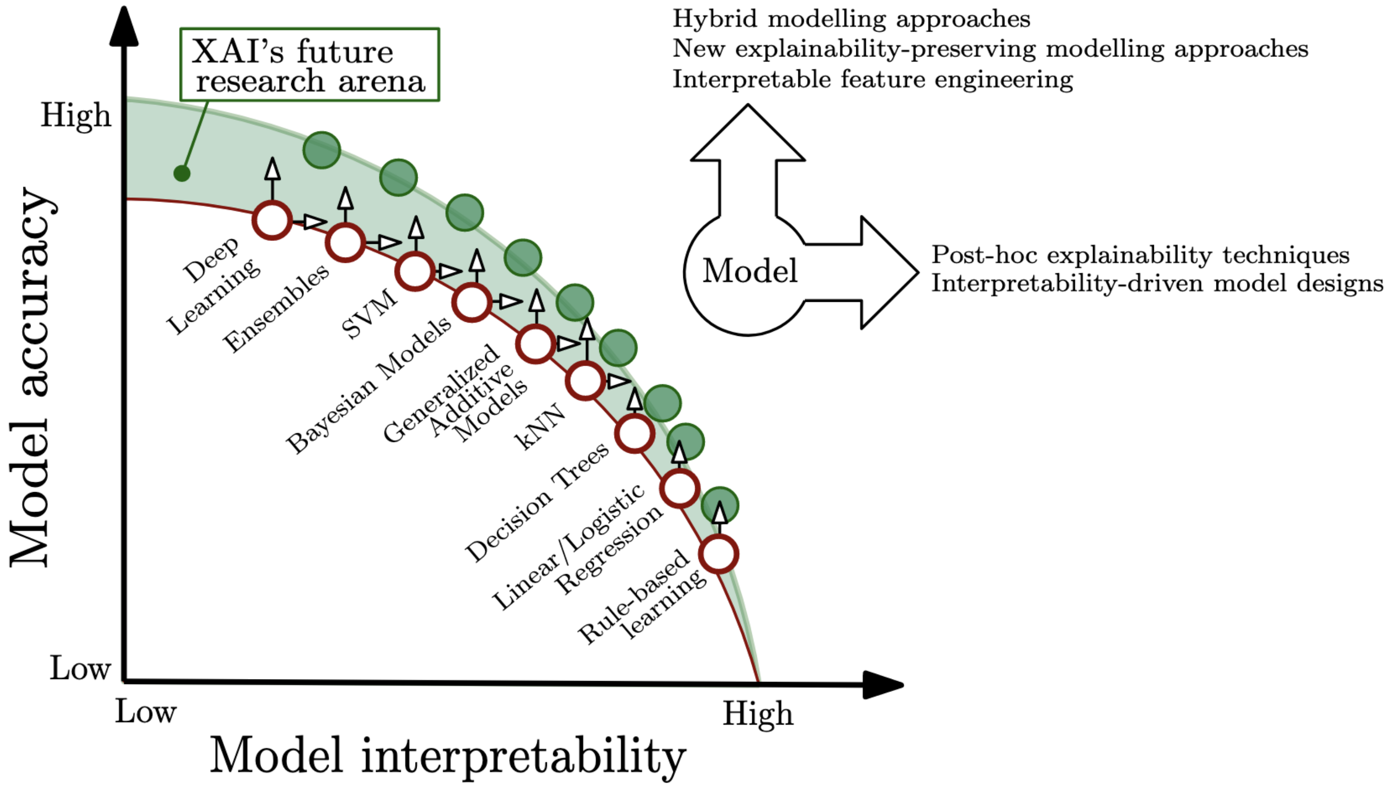
As we come to rely on inferences given by machine learning models, it is important that these models be accurate and interpretable.



Arrieta et al. (2019), https://doi.org/10.3389/fnsys.2021.766980

Why We Need Reliable Models? – XAI

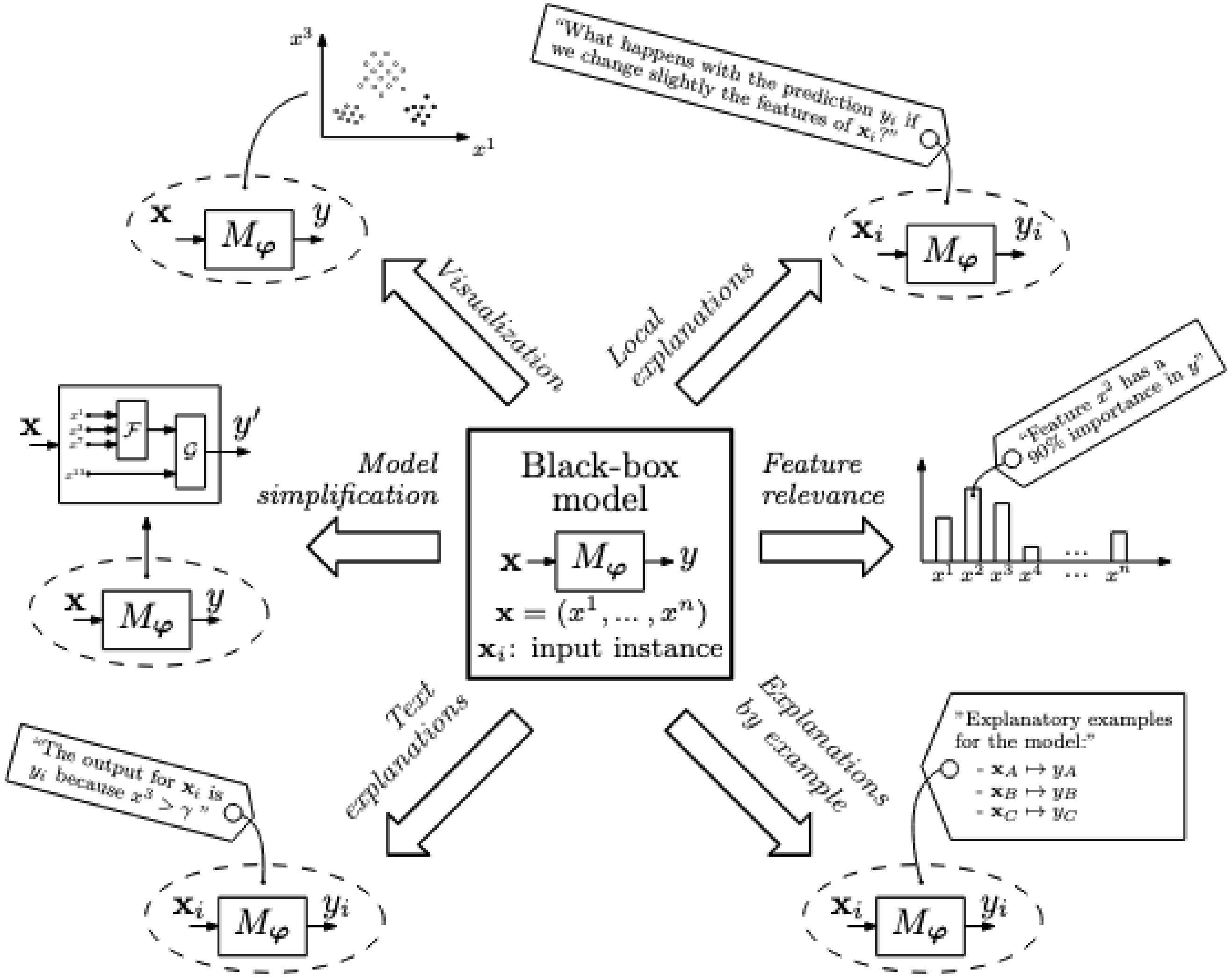
- Accuracy may not be enough.
- Machine learning models need to be reliable.
- Reliability is determined by interpretability and robustness.
- Interpretability: We can explain why a certain outcome was predicted.
- Robustness: Input can be noisy; we still achieve accurate predictions.



Arrieta et al. (2019), https://doi.org/10.3389/fnsys.2021.766980

Post-Hoc Explainability Approaches – XAI

- One of the most common methods of achieving an interpretable ML model is through post-hoc explanation methods (done after the model is trained).
- These methods use the output of the model in conjunction with the inputs to extract information about the model's decisions.



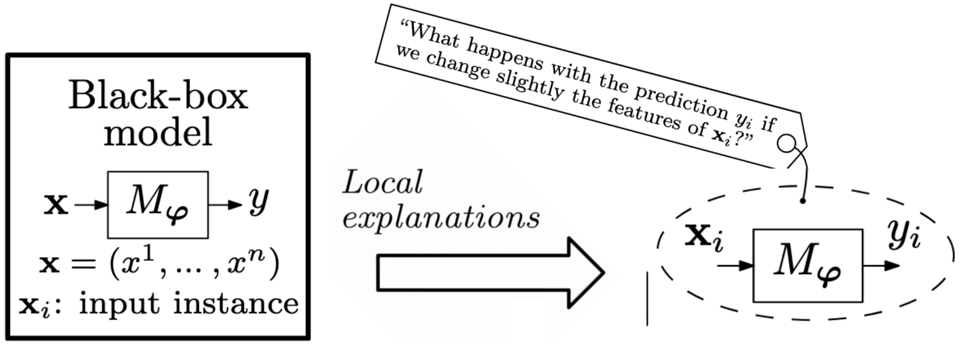
Arrieta et al. (2019), https://doi.org/10.3389/fnsys.2021.766980

Model Explainability and Interpretability – XAI

- A tool used commonly is SHAP (SHapley Additive exPlanations).
- SHAP is a model-agnostic approach which can calculate an additive feature importance score for each prediction.

Using SHAP Values for Local Explanations

- Using Shapely values to provide explanations of single decisions for black box models



Arrieta et al, 2019

sur_refl_b01_1	sur_refl_b02_1	sur_refl_b03_1	sur_refl_b04_1	sur_refl_b06_1	sur_refl_b07_1	ndvi	ndwi1	ndwi2
137	1177	132	236	797	286	7914	1925	6090

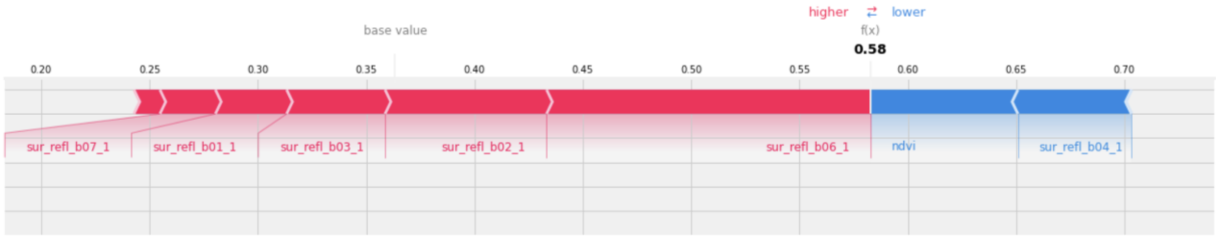
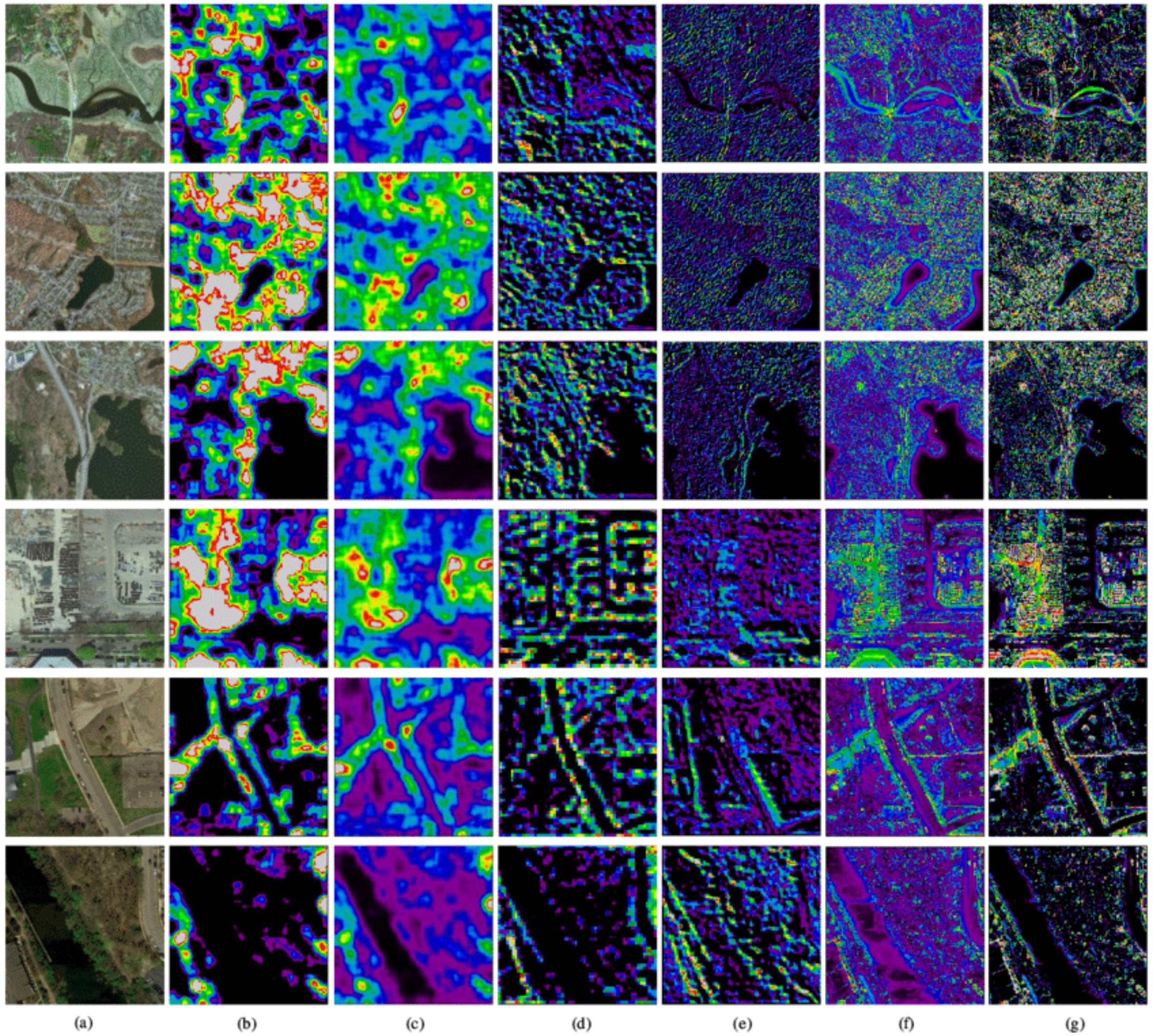


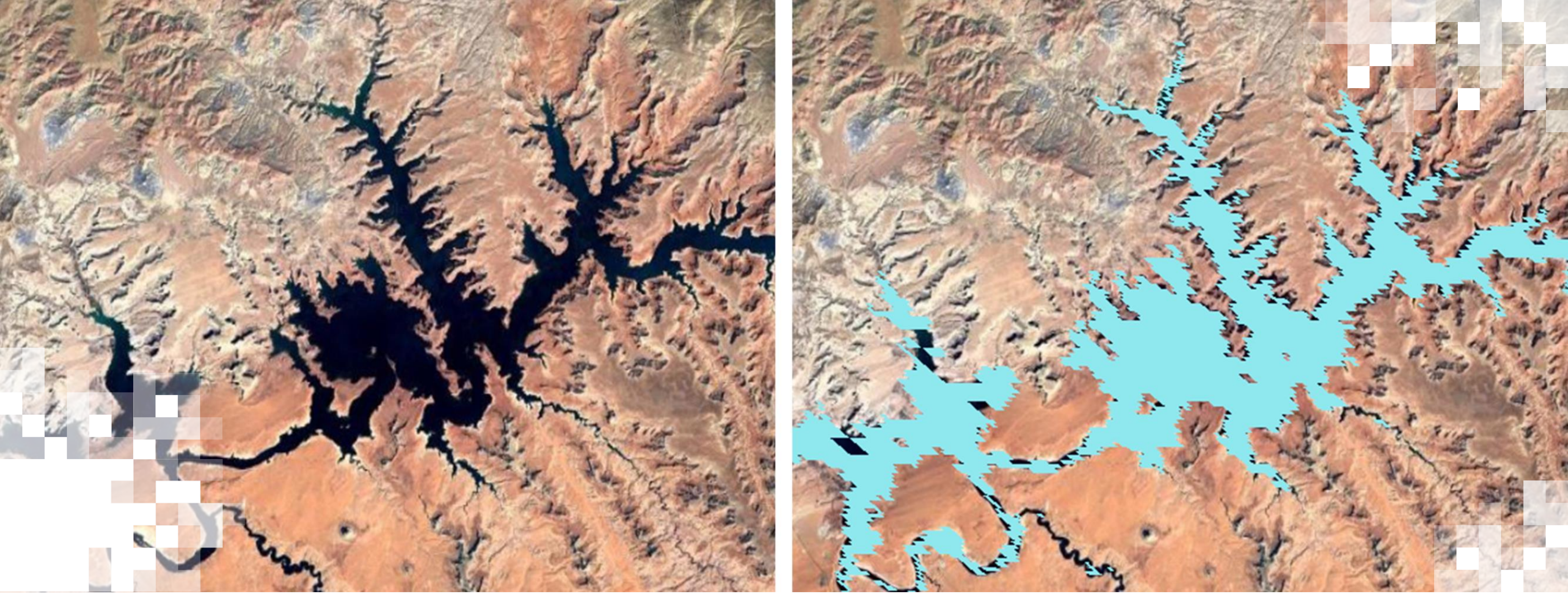
Image Source: https://shap.readthedocs.io/en/latest/index.html

Attention and Explainability – XAI

- Visual Transformers (ViT) can output attention maps.
- Attention maps are the intermediate output of the model that highlights the important region in the image for the target class.
- Visualizing attention maps can lead to a better understanding of how the model is processing the input and which features are most important for the prediction.

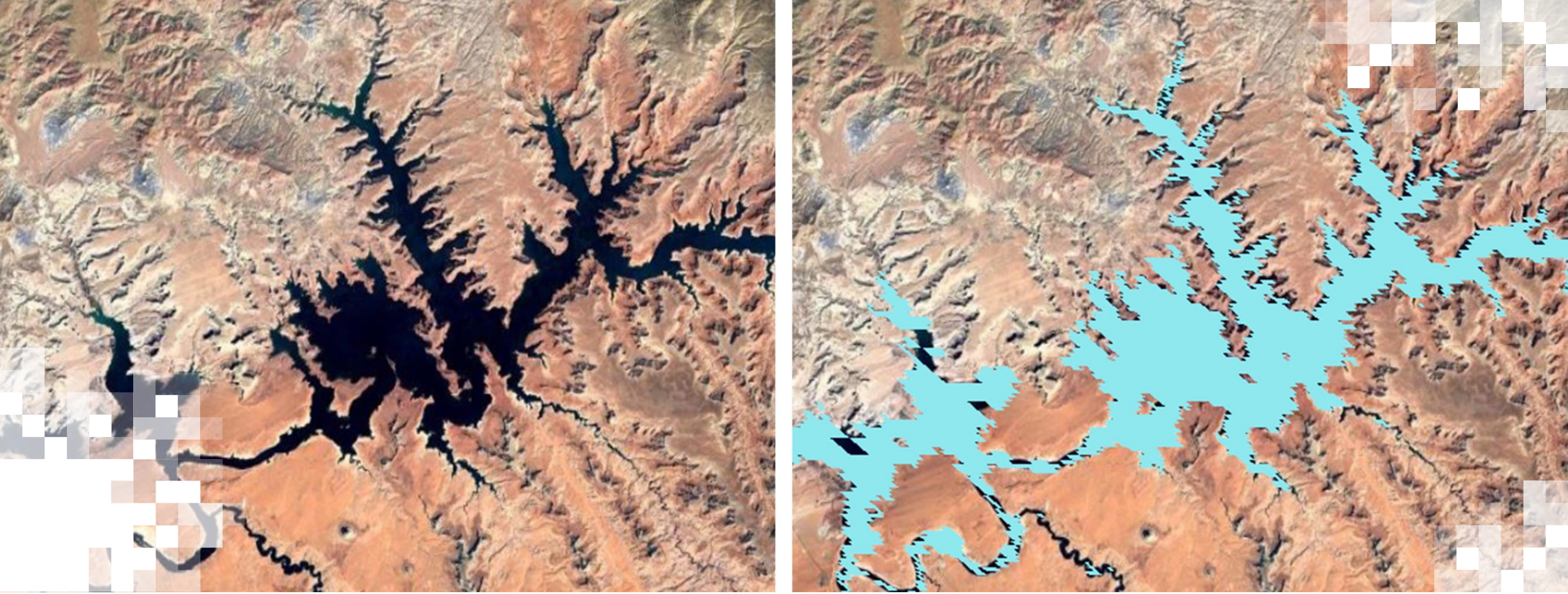


Remote sensing images and visualization of attention maps in different moduls. Shamsolmoali et al. (2020), https://doi.org/10.1109/TGRS.2021.3112481



Exercise: Model Explainability and Interpretability - XAI

Trainer: Caleb S. Spradlin



Closing Remarks

Trainer: Jordan A. Caraballo-Vega

Closing Remarks

- We have:
  - Provided a base on the fundamentals of Machine Learning for Earth Science using a binary classification problem as an example.
  - Introduced the general concept of machine learning and possible scenarios of its benefits across other domains.
  - Provided the base to produce an effective training, validation, and test dataset from both raster and tabular data sources.
  - Provided the tools to train and perform inference of a Random Forest model, including its fine-tuning and XAI analysis.

This is just an introduction to the very broad field of Machine Learning. The fundamentals learned in this training will provide the basis to understand literature and to know when a specific algorithm might be the most applicable.

Contacts

- Trainers:
  - Jordan A. Caraballo-Vega: jordan.a.caraballo-vega@nasa.gov
  - Jules Kouatchou: jules.kouatchou-1@nasa.gov
  - Caleb S. Spradlin: caleb.s.spradlin@nasa.gov
  - Jian Li: jian.li@nasa.gov
  - Brock Blevins: brock.Blevins@nasa.gov

- Training Webpage:
  - https://appliedsciences.nasa.gov/join-mission/training/english/arset-fundamentals-machine-learning-earth-science

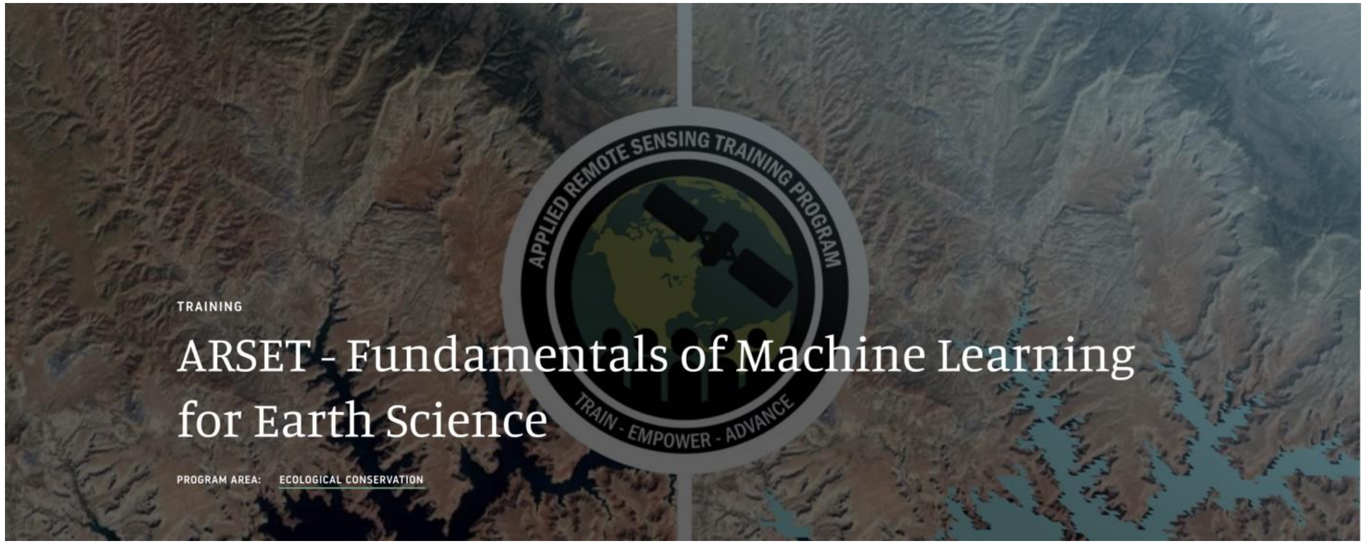
- ARSET Website:
  - https://appliedsciences.nasa.gov/arset

Check out our sister programs:

DEVELOP SERVIR

Questions?

- Please enter your question in the Q&A box. We will answer them in the order they were received.
- We will post the Q&A to the training website following the conclusion of the webinar.



Thank You!

References

- Crankshaw, D., & Gonzalez, J. (2018). Prediction-Serving Systems: What happens when we wish to actually deploy a machine learning model to production?. *Queue*, *16*(1), 83-97.
- Elders, A., Carroll, M. L., Neigh, C. S., D'Agostino, A. L., Ksoll, C., Wooten, M. R., & Brown, M. E. (2022). Estimating crop type and yield of small holder fields in Burkina Faso using multi-day Sentinel-2. *Remote Sensing Applications: Society and Environment*, *27*, 100820.
- Fleming, S. W., Watson, J. R., Ellenson, A., Cannon, A. J., & Vesselinov, V. C. (2021). Machine learning in Earth and environmental science requires education and research policy reforms. *Nature Geoscience*, *14*(12), 878-880.
- Prša, A., Kochoska, A., Conroy, K. E., Eisner, N., Hey, D. R., IJspeert, L., ... & Winn, J. N. (2022). TESS Eclipsing Binary Stars. I. Short-cadence Observations of 4584 Eclipsing Binaries in Sectors 1–26. *The Astrophysical Journal Supplement Series*, *258*(1), 16.
- Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., & Carvalhais, N. the National Energy Research Supercomputing Center in Lawrence Berkeley National Laboratory, Berkeley, CA, USA: Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*, 195-204.
- Yu, S., & Ma, J. (2021). Deep learning for geophysics: Current and future trends. *Reviews of Geophysics*, *59*(3), e2021RG000742.

Contributors

- Jordan A. Caraballo-Vega
- Mark L. Carroll
- Jules R. Kouatchou
- Jian Li
- Caleb S. Spradlin
- Brock Blevins
- Melanie Follette-Cook
- Erika Podest
- Brian Powell
- Akiko Elders