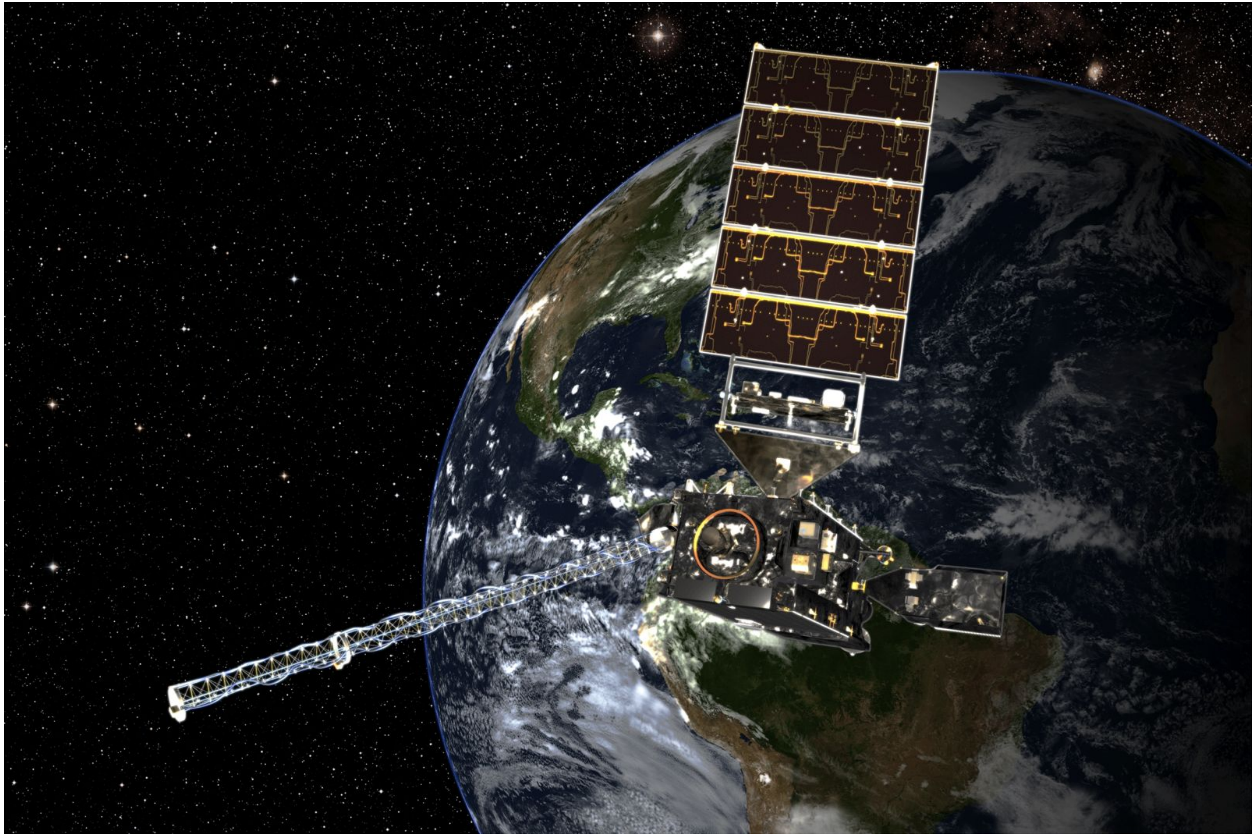
AI4ESS Hackathon: GOES Challenge

David John Gagne, Gunther Wallach, Charlie Becker, Bill Petzke



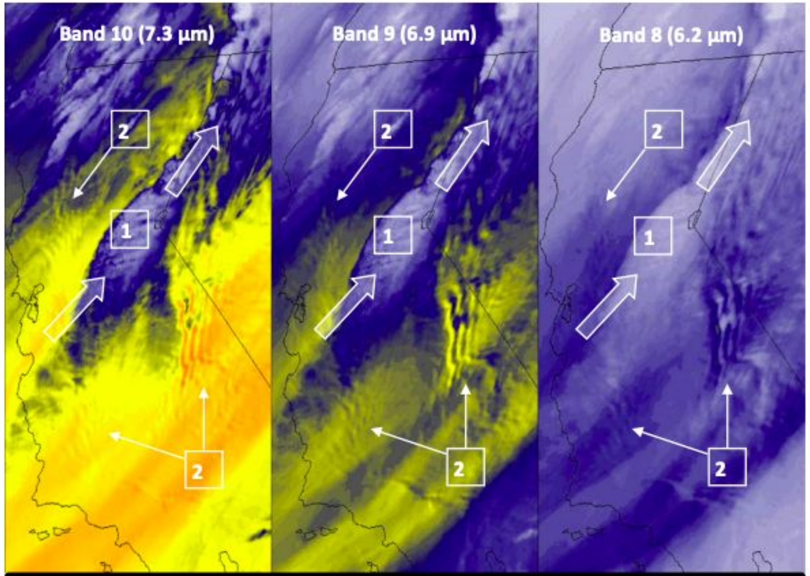
- The Geostationary Operational Environmental Satellite 16 (GOES-16) is a weather satellite that orbits the Earth
- It can provide a hemispheric, multispectral view of cloud patterns at high space and time resolution through its Advanced Baseline Imager (ABI) camera.
- The satellite holds the Geostationary Lightning Mapper (GLM) instrument that records lightning flashes across the hemispheric view of the satellite

Image Interpretation

1 Axis of strong middle / upper tropospheric jet streak

2 Mountain waves downwind of the Coastal Ranges and the Sierra Nevada.

Depending on the topography as well as the atmospheric temperature and moisture profile, mountain waves might show up better on any of the water vapor bands. You may have to apply different enhancements as well.



GOES-16 water vapor Band 10, Band 9 and Band 8 images at 12:02 UTC on 13 April 2017

Dimensions

Dimension Name	Description	Size
Band	ABI Band Number	4 (Bands 8,9,10,14)
Patch	spatio-temporal patch	~3600 per day
X	X-plane	32
Y	Y-plane	32

Potential Input Variables

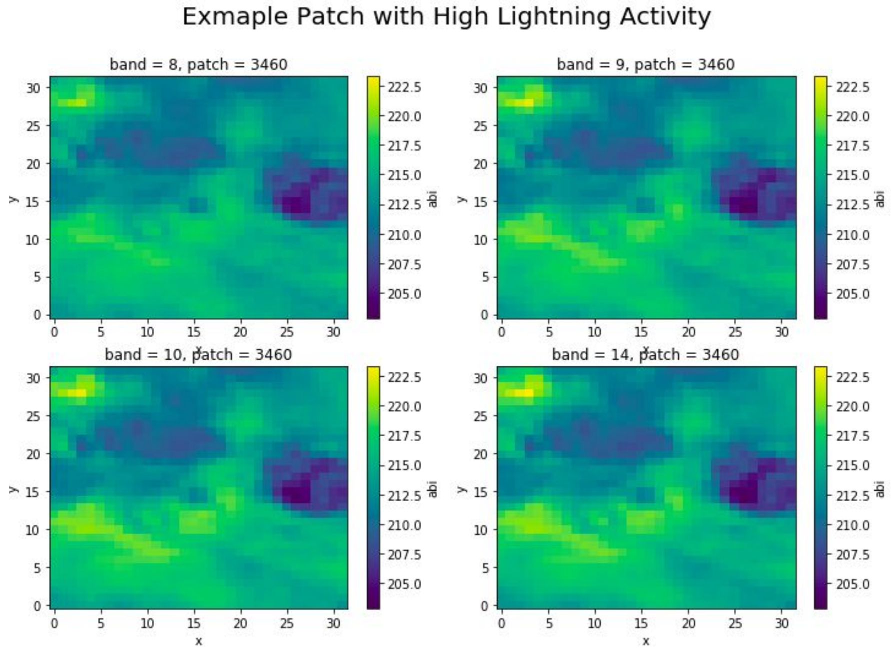
Variable Name	Units	Description
abi (Band 08)	K	Upper-level Water Vapor
abi (Band 09)	K	Mid-level Water Vapor
abi (Band 10)	K	Lower-level Water Vapor
abi (Band 14)	K	Longwave Window

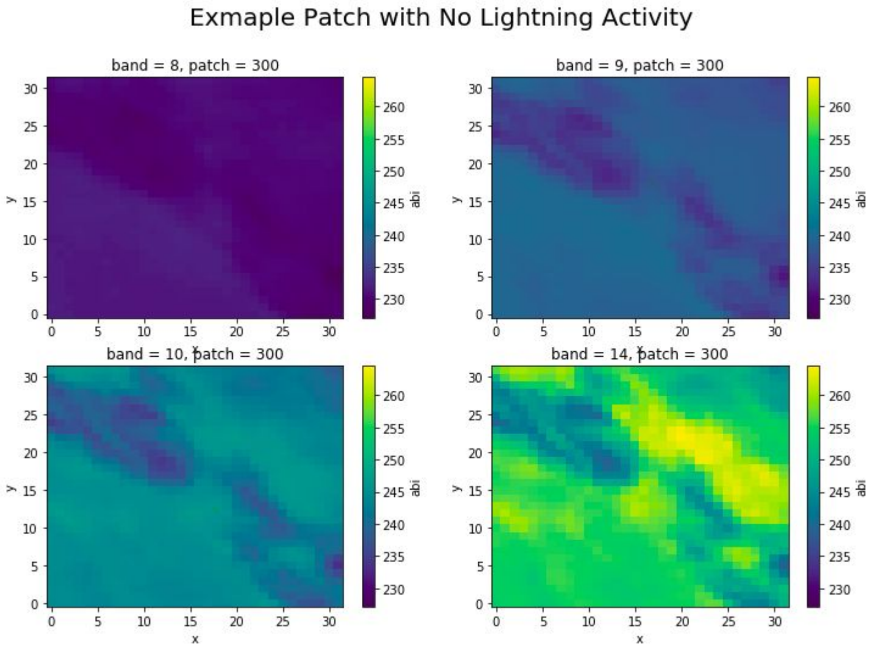
Output Variables

Variable Name	Units	Description
flash_counts	-	Lightning strike count

Convolutional Neural Networks (CNNs) work well to capture spatial structural differences.

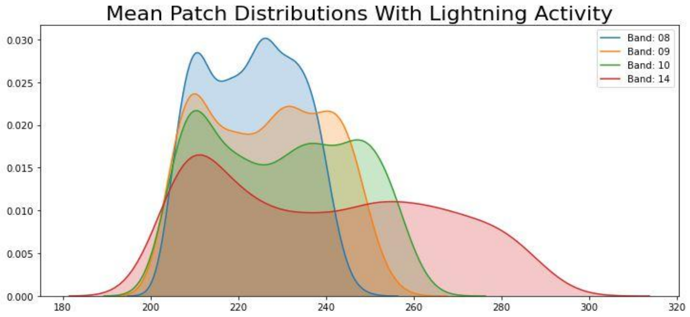
Residual Networks can be used to increase the effectiveness of the depth of the NN.

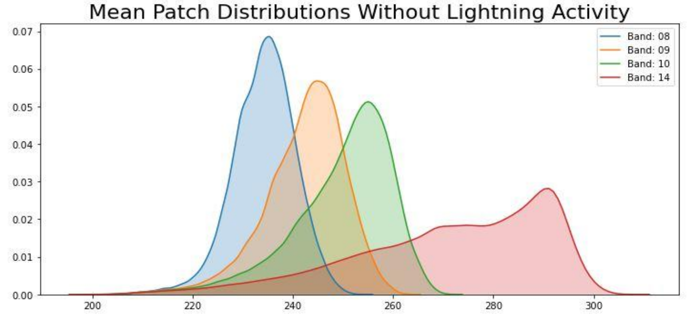




Team 2: GOES

Algorithms Tested: Random Forest, Gradient Boosting, Linear and Polynomial Regression, Densely Connected and Convolutional NN





Kevin Bachmann
(University of Albany)

Adrian Phillips-Samuels
(Boulder, CO)

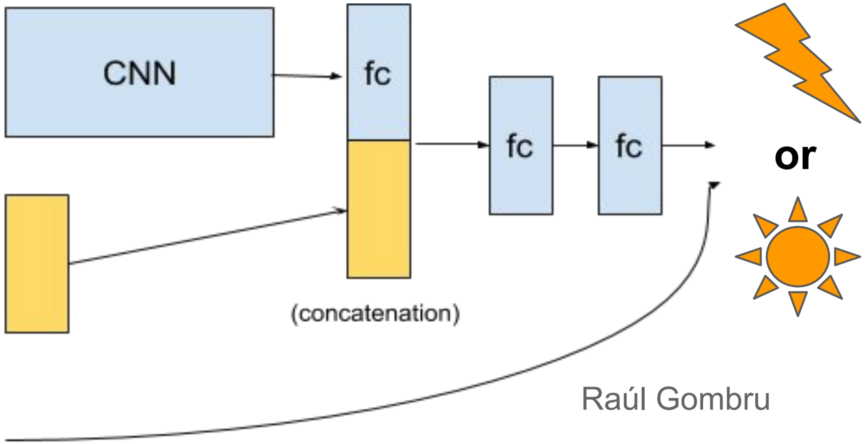
Mostofa Kamal
(University of Saskatchewan)

Anh Pham
(IPSL - France)

CNN with additional metadata to capture areas or times with increased lightning activity

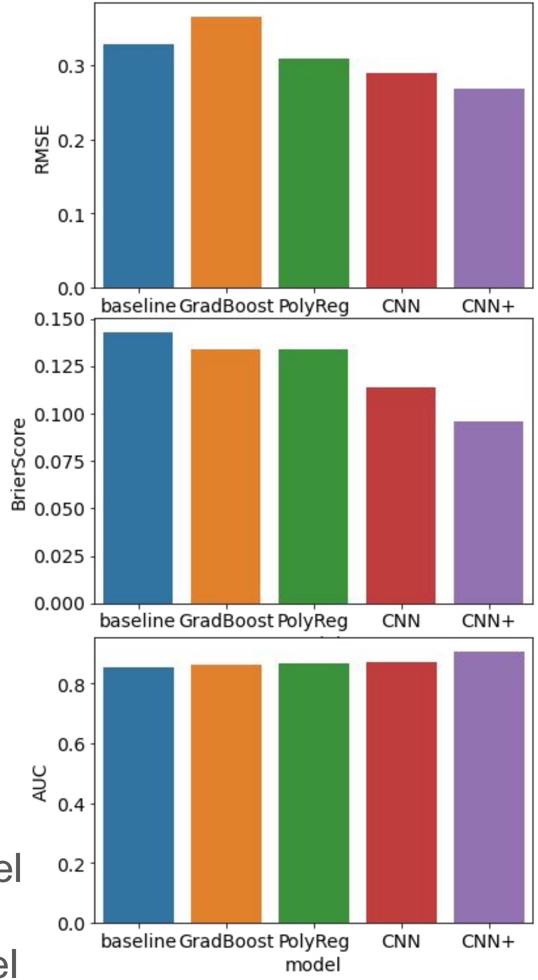
GOES data

Metadata: lat, lon, time of day



Lessons learned/challenges:

- Batch normalization improved training speed but not performance
- Skill quickly saturated for larger and deeper CNNs
- Standard scaling (vs. MinMax) significantly improved the baseline model
- 3 Degree Polynomial Regression out performed Tree Based Methods
- Making metadata available to CNN produced our best performing model



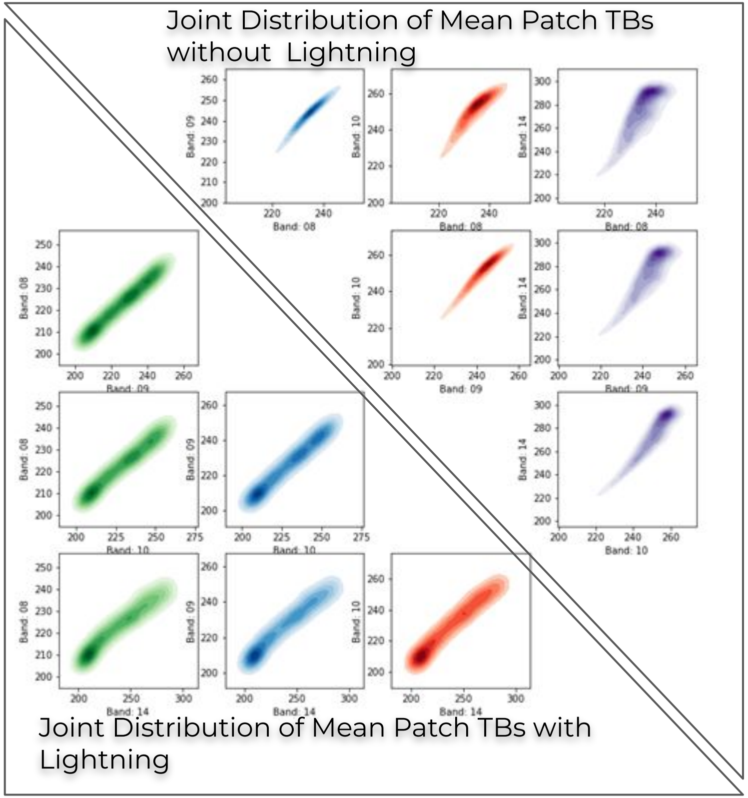
Model Metric Comparisons

Supporting Material

Evaluation Metrics	List of Used Algorithms						
	Random Forest (leaf size = 25)		Gradient Boosting (leaf size = 10)	Simple Linear Regression	Polynomial Linear Regression	Dense (Epochs = 5; optimizer = Adam algorithm)	CNN (Epochs = 3; optimizer = Adam algorithm)
	Mean	Mean + SD					
RMSE	0.367	0.366	0.336	0.346	0.311	0.309	0.272
R squared	0.517	0.518	0.519	0.5	0.595	0.603	0.696
Brier Score	0.134	0.134	0.134	0.134	0.134	0.134	0.134
Brier Skill Score	0.438	0.438	0.441	0.33	0.454	0.452	0.595
AUC	0.86	0.86	0.861	0.83	0.862	0.867	0.903
Hellenger Distance	0	0	0.0	0.757	0.546	0.397	0.407

Team 8: GOES

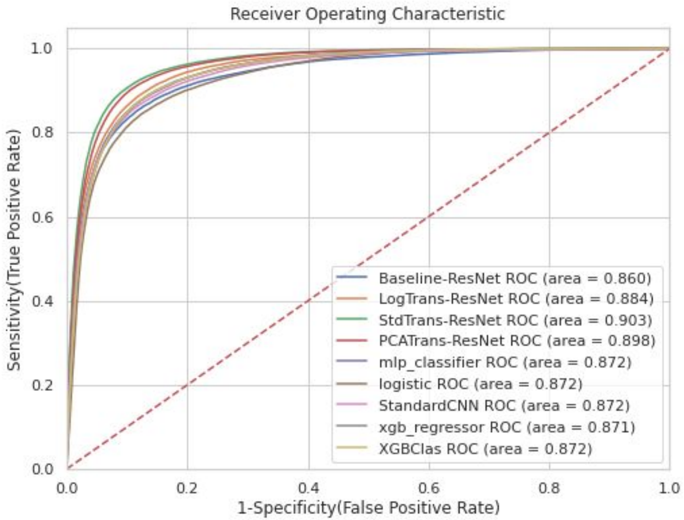
- Abhishek Singh, Yiyi Huang, Eric S. Maddy, Keely Lawrence*
- 2D-image feature methods - ResNet, StandardCNN
  - MinMax, Standardized, PCA compression, MinMax(Log(Tb)) as inputs
- 1D feature methods using patch statistics - XGBoost(Regression and Classification), Linear, Logistic, MLP, GaussianNaiveBayesian
  - Using Patch(Min) or Patch(Min,Q1,Median)... as input



Joint distributions of mean patch Tbs (all Tbs similar) show large degree of correlation in ABI spectral bands in patches with lightning and without lightning. PCA of the raw Tbs in all scenes - 1st and 2nd PC represents 99% of variance of training data - spectrally redundant of Tbs.

Differences in joint distributions with and without lightning (*what enables lightning detection*) likely due to differences in vertical sensitivity of water (mid-upper troposphere) and window (surface or cloud-top) bands and resultant spatial and vertical centroids - cloud sensitivities across the patch and to cloud tops in each of those types of scenes. For instance, lightning patches have colder Tbs (cloud tops) with wider joint distributions, but less skewed across bands. No lightning patches have skewed joint distributions especially with the window channel which senses either the surface (warm temps) or cloud top emission (cold temps) at or below water band weighting functions..

Team 8: GOES - Results On "Test" Data



The skill of 1D methods is not surprising given that there's a strong signal in the mean Tbs (and other statistics) between patches with lightning and without lightning.

Best out of all runs is StandardScaler ResNet with AUC=0.903 and MSE=0.273; however all other methods also have higher AUC and lower MSE relative to baselineResNet. Most other metrics are also better. AUC computed using binary predictions. 1D methods use percentile Predictors as they performed better than min alone.

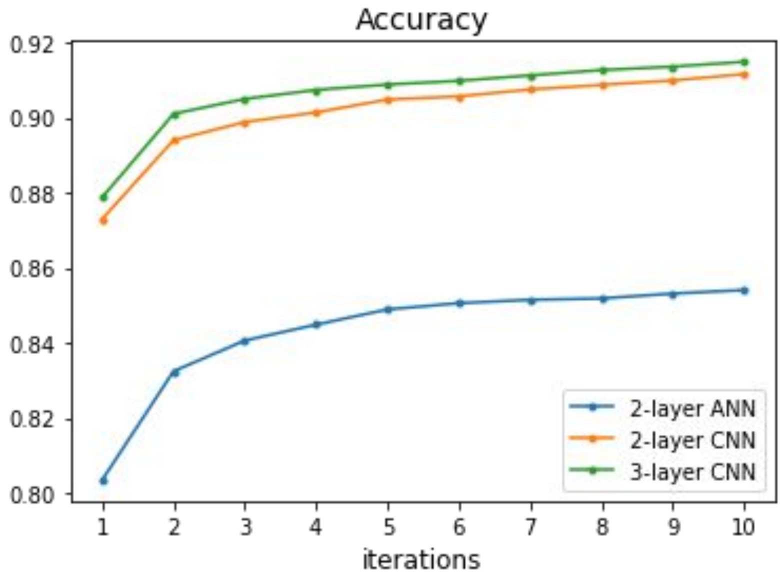
Team 30 - GOES

Team members: Laura Ko, Keith Searight, Yifei Guan* and Irina Melnikova

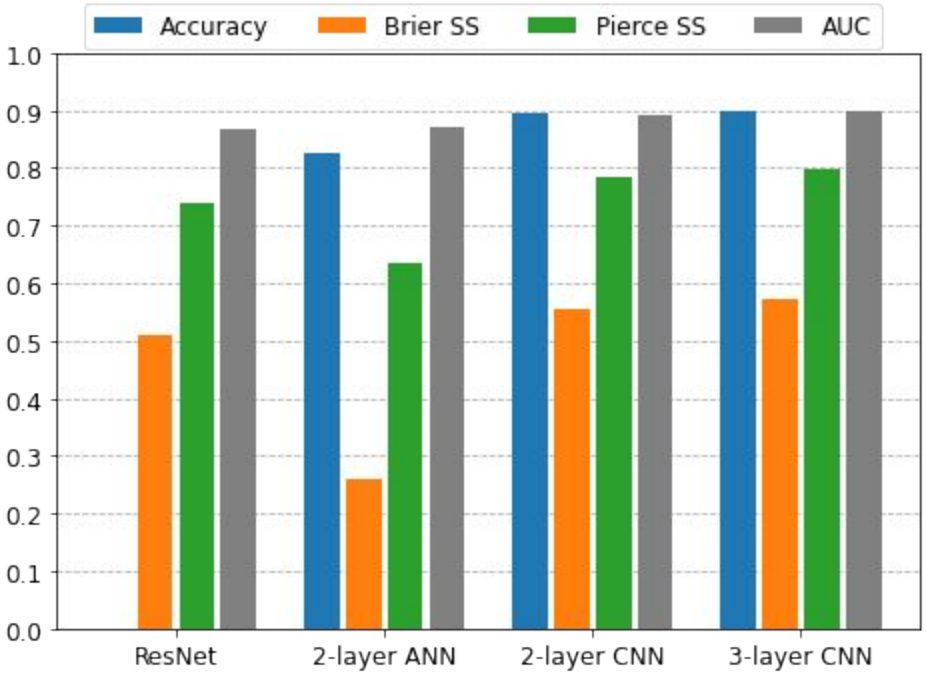
Problem: machine learning solutions for the **short range lightning prediction,** knowing that there is larger brightness temperature range in all bands, especially in band 14

ML models tried (also tried, but didn't measure: LogisticRegression, DecisionTreeClassifier):

1. ResNet with 5 epochs (supplied)
2. StandardConvNet with 5 epochs
3. 2-layer CNN with 10 epochs
4. 3-layer CNN with 10 epochs
5. 2-layer ANN with 10 epochs



Performance metrics:



	Accu	BSS	PSS	AUC
ResNet		0.510	0.739	0.869
Standard ConvNet		0.482	0.744	0.872
ANN2	0.825	0.260	0.634	0.871
CNN2	0.894	0.555	0.783	0.892
CNN3	0.897	0.571	0.797	0.898

Technical challenges:

- Necessity of good knowledge of Python
- Necessity of ML experience
- Lack of basic training (step-by-step explanation)

Lessons learned:

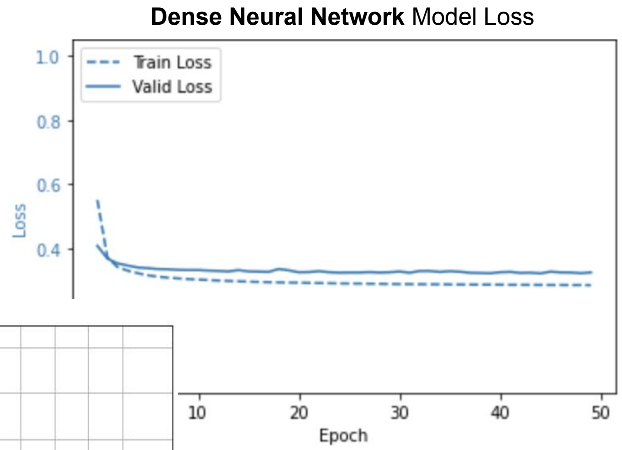
- Good theoretical basis of ML
- Motivation to learn and explore

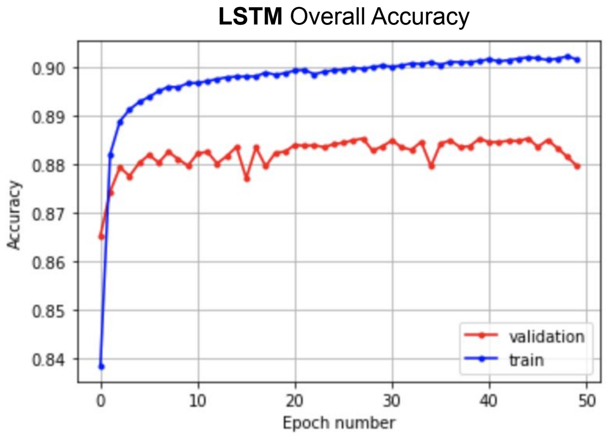
Team 37: GOES

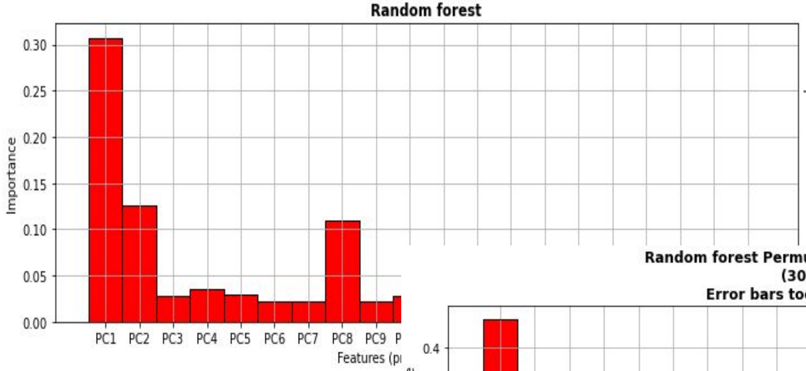
Team Members: Alex Araujo, TC Chakraborty, Dom Heinzeller, Priyanka Rao, Xueying Zhao

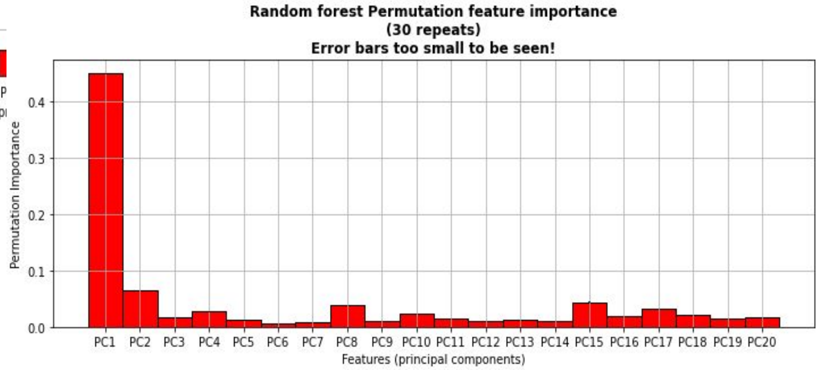
Scores	Linear Model (using 32x32 map means)	Random Forest (PCA with 32 components)	Gradient Boosted Regression Trees (PCA, 32 comp.)	Dense NN (using PCA and selu act.)	Convolutional Neural Network	LSTM (PCA with 20 components)	Resnet with modified scaler (StandardScaler)
RMSE	0.403	0.315	0.321	0.312	0.278	0.286	0.269
R squared	0.324	0.587	0.571	0.594	0.681	0.658	0.7
Hellenger Distance	0.813	0.427	0.642	0.481	0.369	0.647	0.225
Heidke Skill Score	0.472	0.721	0.706	0.718	0.782	0.77	0.803
Pierce Skill Score	0.453	0.716	0.699	0.709	0.775	0.77	0.807
Brier Score	0.242	0.133	0.139	0.111	0.103	0.11	0.094
Brier Skill Score	-0.011	0.445	0.418	0.444	0.569	0.541	0.605
AUC	0.726	0.858	0.849	0.867	0.888	0.885	0.903
Time to Fit. (epochs)	<1s (default params)	556s (default params)	324s (default params)	143s (50)	369s (25)	506s (50)	306s (5)

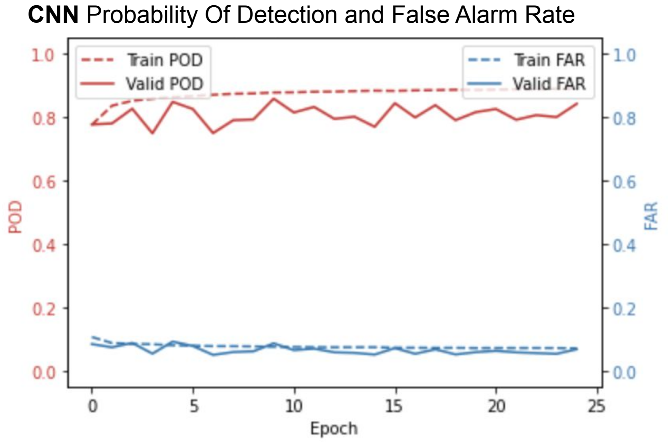
Team 37: GOES





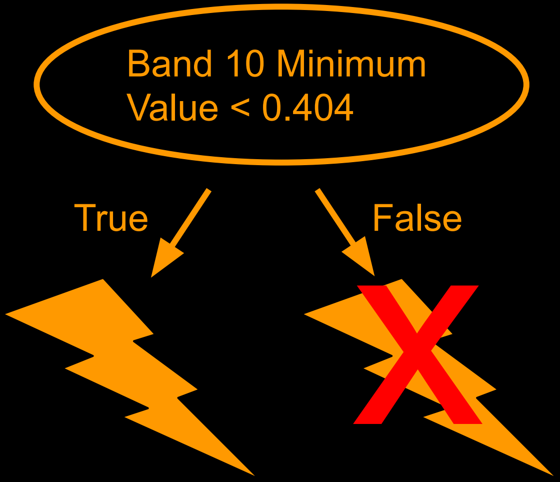




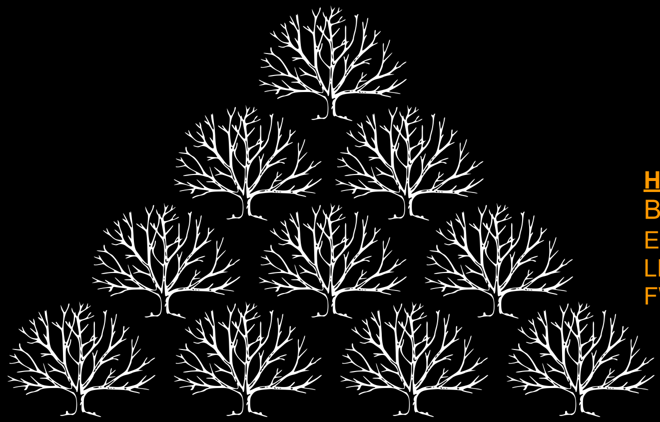


Conclusions: So many opportunities, so much to learn! Huge thanks to the organizers and speakers!

1. Decision Tree (depth=1)

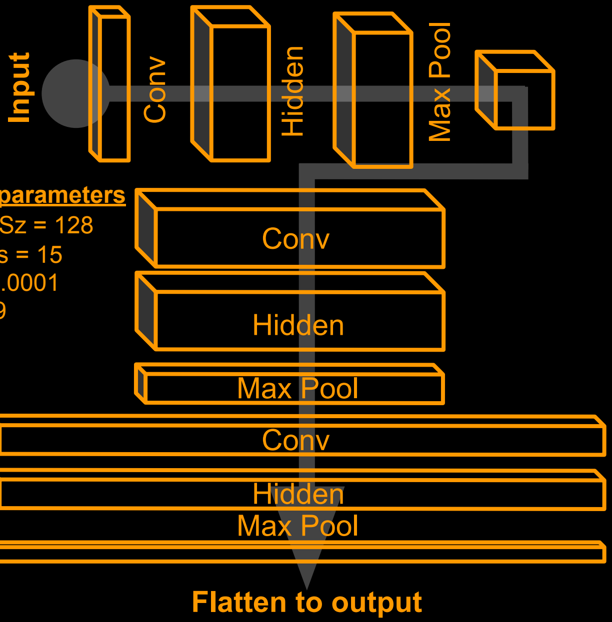


2. Random forest (10 trees)



12 features (**σ**, minimum, median from all 4 bands)

3. CNN (max pooling)



Hyperparameters
Batch Sz = 128
Epochs = 15
LR = 0.0001
FW = 9

Diminishing returns with increasing model complexity:

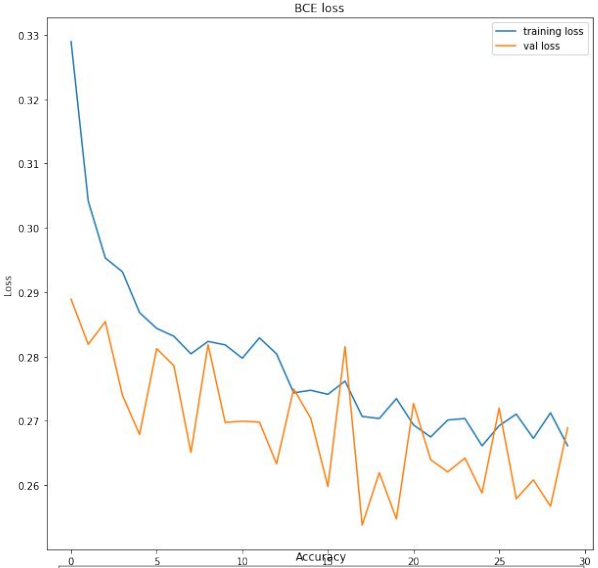
model	% Correct	AUC	Precision	Recall	F1-score	Heidke	Pierce	Brier	Brier Skill
1	84.7	0.839	0.809	0.801	0.805	0.679	0.678	0.153	0.358
2	88.1	0.877	0.847	0.855	0.851	0.752	0.754	0.119	0.504
3	89.5	0.890	0.871	0.862	0.867	0.780	0.779	0.105	0.562

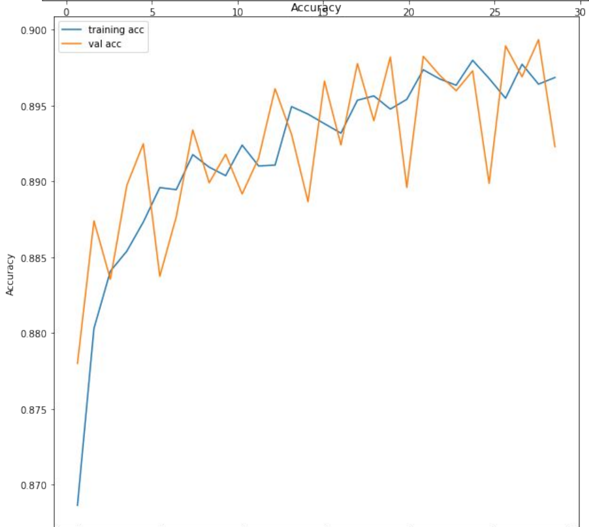
For classification, reconsider the neural net!

Team 39: Rachel Atlas and Andy Barrett (GOES Challenge)

Team 44: GOES

- Mihai Boldeanu, Akila, Fei Luo, Sudhir, Paban Bhuyan*
- Best results with a convnet :
  - RMSE: 0.285
  - R squared: 0.678
  - Hellinger Distance: 0.377
  - Heidke Skill Score: 0.785
  - Pierce Skill Score: 0.792
  - Brier Score: 0.108
  - Brier Skill Score: 0.566
  - AUC: 0.896



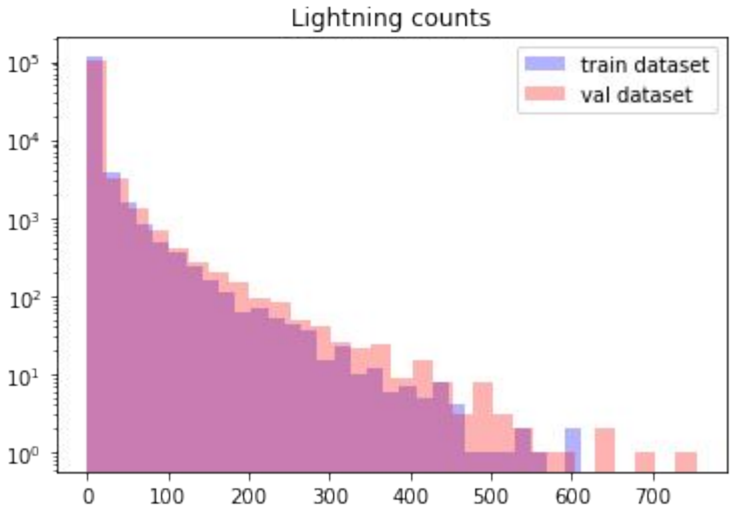


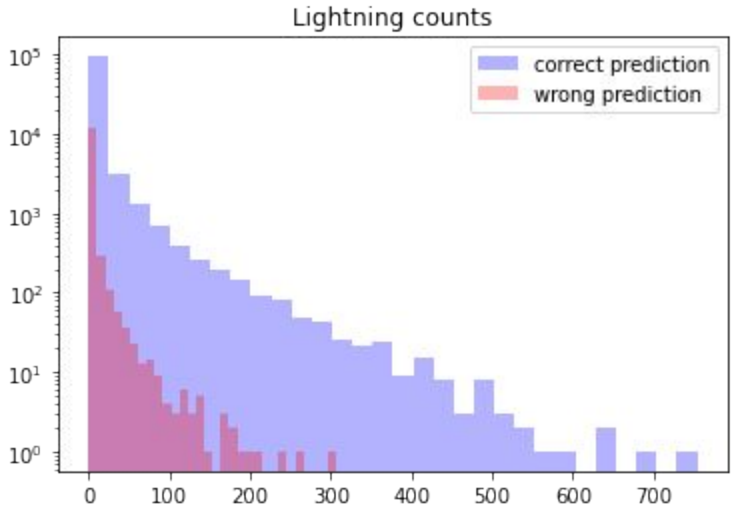
Legend is correct for train and validation. You can have better validation loss than train when you have very large DropOut(0.75) the models acts as an ensemble.

Team 44: GOES

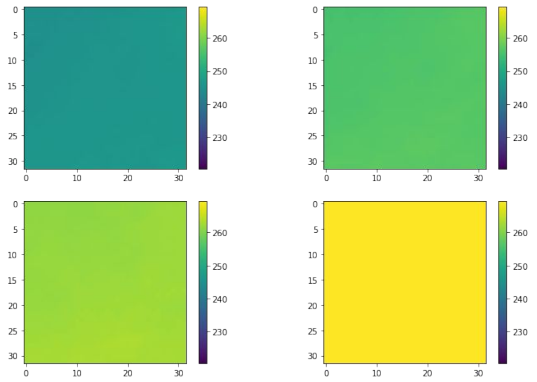
A visualization of your results scores on the problem:

Classifying no lightning versus lightning

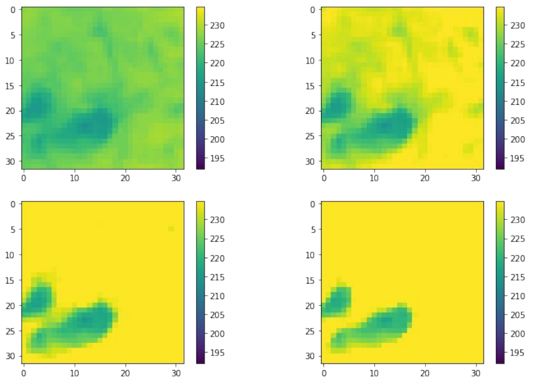




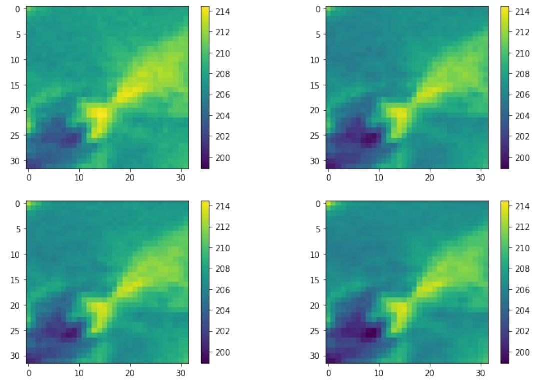
Exmaple Patch with No Lightning Activity



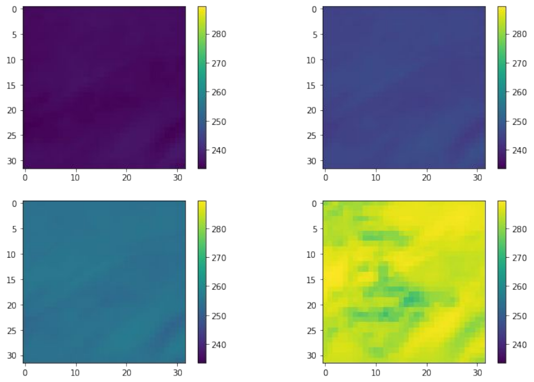
Exmaple Patch with Lightning Activity



Example Patch with 0 Lightning counts classified as Lightning event



Example Patch with 1 Lightning counts classified as no Lightning event

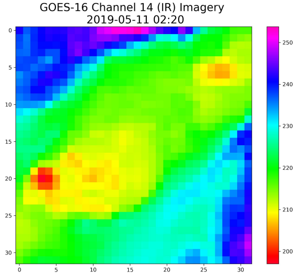


Model learns the general cases. It has problems at low lightning counts or in cases where the images dont look like the general case.

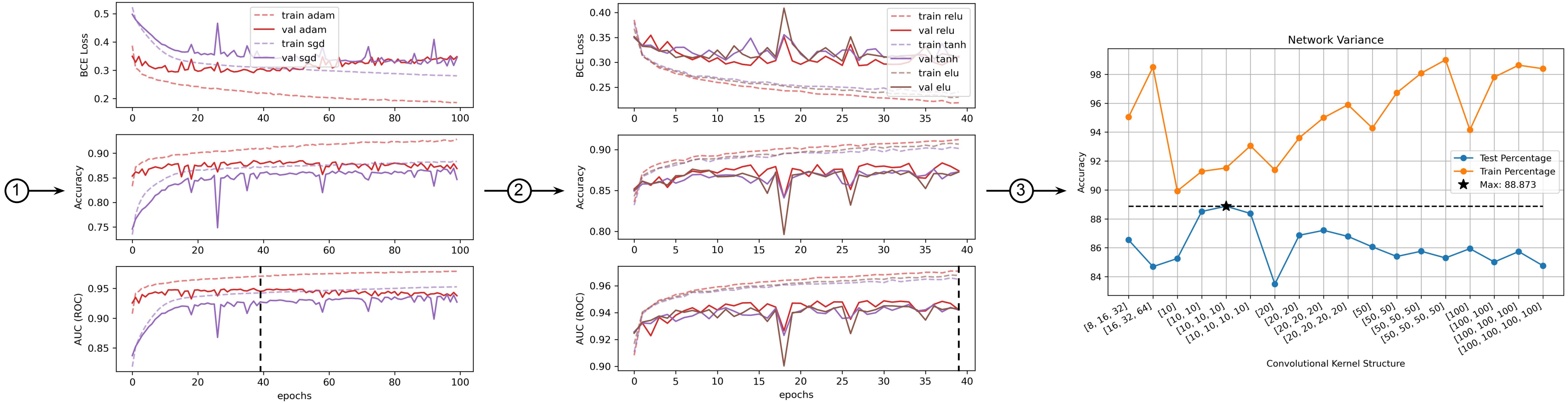
Team 53: GOES Challenge

*Jason Stock / Max Grover / David Mattern**

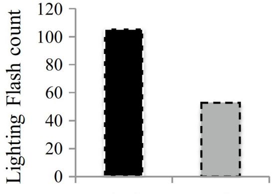
CIRA Connecting Models and Observations / ILLINOIS / NOAA



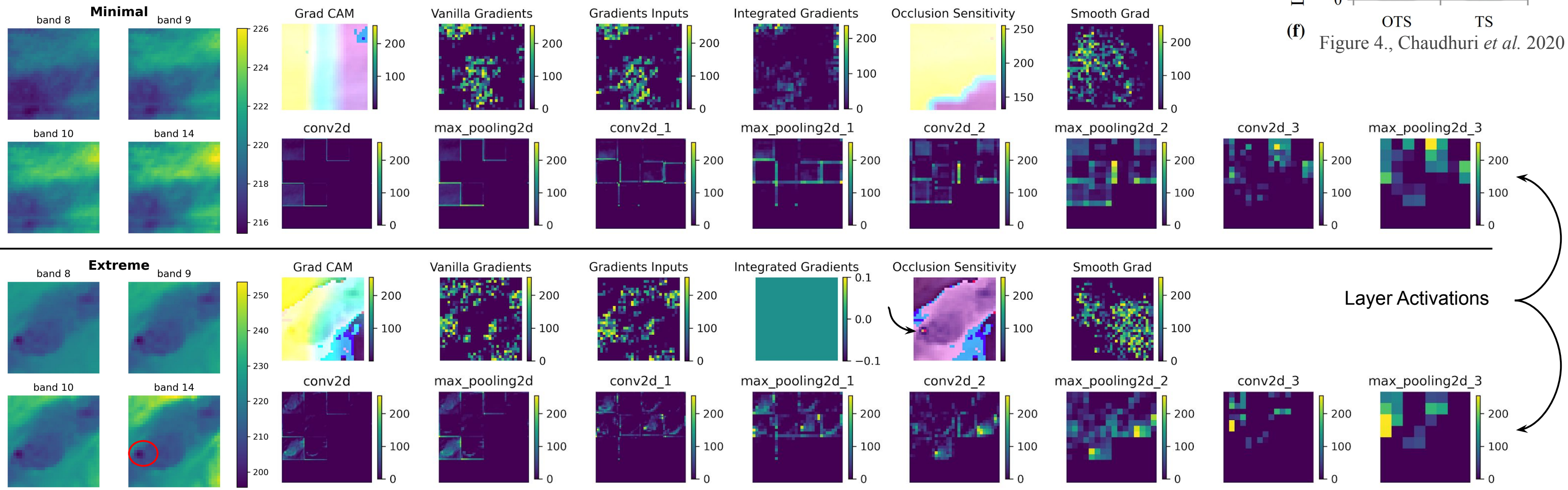
- Explored Decision Trees/Random Forests, Linear Regression, and Neural Networks (NN)
    - Flash Count Regression, Binary Classification (slide 1), **5 Bin Severity Classification** (slide 2)
- Thorough NN **hyperparameter optimization** to find the model that best fits the data
    - 40 epochs, batch size of 64, Adam, ReLU, 3 hidden layers each with 10 kernels and 3x3 convolutions followed by max pooling



Team 53: GOES Challenge



(f) Figure 4., Chaudhuri *et al.* 2020



	None	Minimal	Moderate	Severe	Extreme
None (0)	92.4	7.6	0.1	0	0
Minimal (1-5)	16.2	80.4	3.4	0	0
Moderate (6-50)	11.4	81.1	7.5	0	0
Severe (51-500)	4.5	77.6	18	0	0
Extreme (500+)	0	100	0	0	0

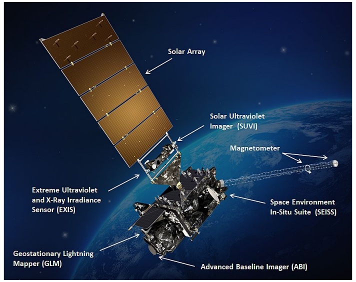
Metric	Value
Accuracy	0.783
Min Normalized Loss:	0.562
Heidke Skill Score:	0.705
Peirce Skill Score:	0.713
AUC ROC:	0.856

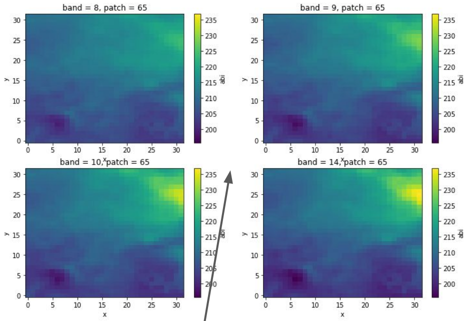
- Divided data into 5 categories representing storm severity
- Investigated misclassification of **Extreme** and **Minimal** events
- Extreme event **occlusion sensitivity** picks up on **overshooting top**
  - Often associated with i**ntense convection** (Figure 4.)
- Challenges encountered:
  - Experienced overfitting on training data
  - Difficulty managing class imbalances

Team 61: Goes Team!

Predict the near-future occurrence of lightning from 4 channels

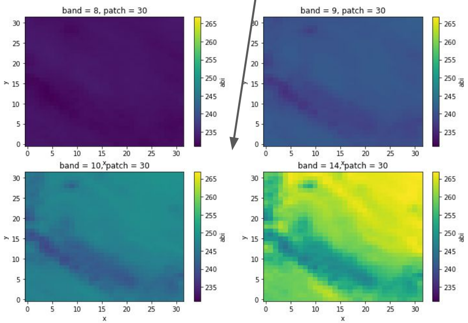
Manh-Hung Le
Ty Ferre
Zhifeng Yang*
Saad Abouzahir
Chiem van Straaten





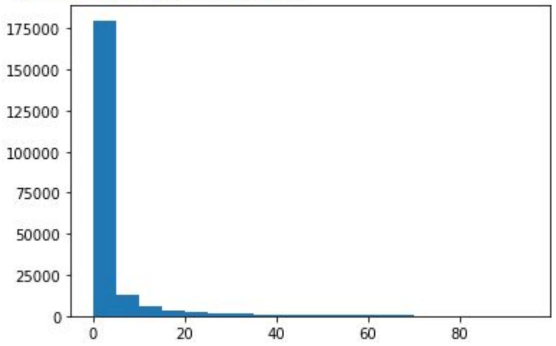
High lightning activity

Low lightning activity



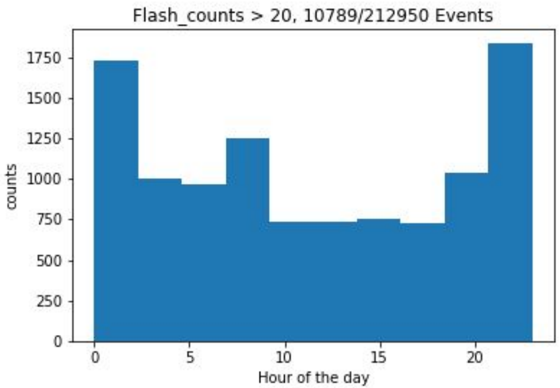
Highly correlated data across channels:
PCA EVRs = 0.977, 0.021, 0.002

Highly skewed lightning counts



Lightning counts

Unevenly distributed over the day



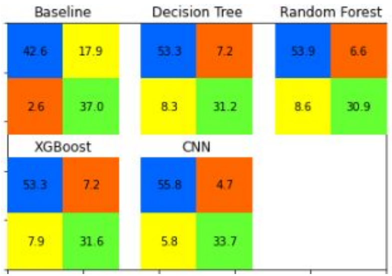
Team 61: Goes Team!

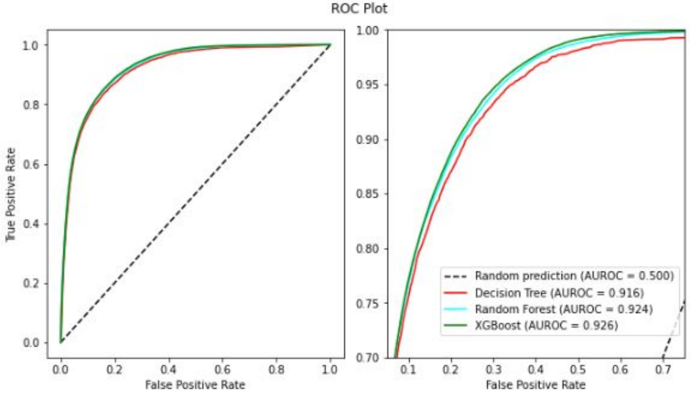
Predict the near-future occurrence of lightning from 4 channels

Manh-Hung Le
Ty Ferre
Zhifeng Yang*
Saad Abouzahir
Chiem van Straaten

Model intercomparison.

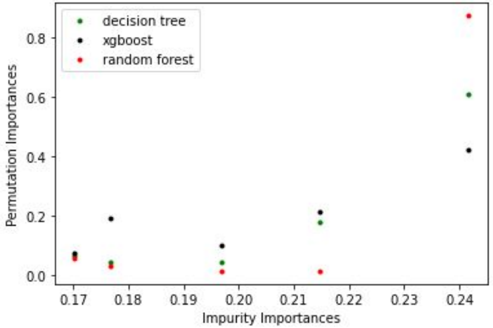
CNN does best, but not by much.
Is it worth the extra effort?





Feature importance.

Two methods gave consistent results across tree-based methods considering mean, std, max, min, and median averaged over grid and all bands. But, degrees of importance varied.



Single pass forward permutation of random forest interpretation showed that minimum over grid on channels 10 & 14 are most important.