Overview of NYSDOT & SUNY Albany Project:

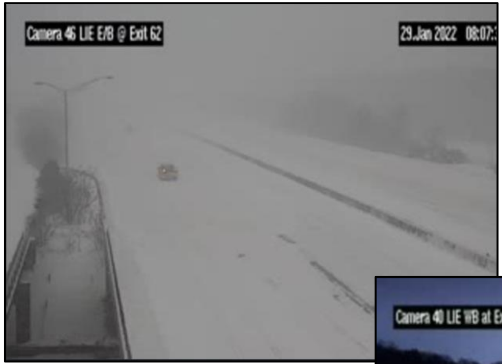
Automated Detection of Road Conditions from DOT Cameras with AI

WINTRE-MIX Workshop
May 22, 2023

Carly Sutter csutter@albany.edu, Nick Bassill nbassill@albany.edu,
Kara Sulia ksulia@albany.edu, Chris Thorncroft

Motivation

Identify road surface conditions from images using machine learning methods



← ***Severe snow***

← ***Wet***

Snow →

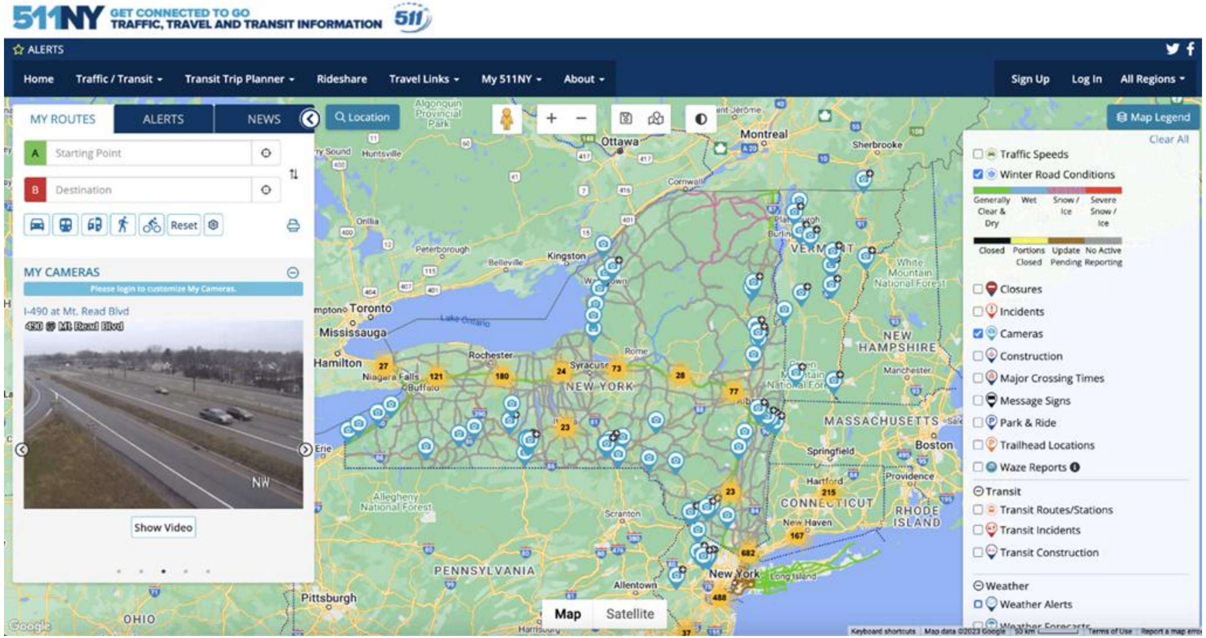
Dry →

Data source: camera images above are from the New York State Department of Transportation (NYSDOT), publicly available at 511ny.org

Data Source

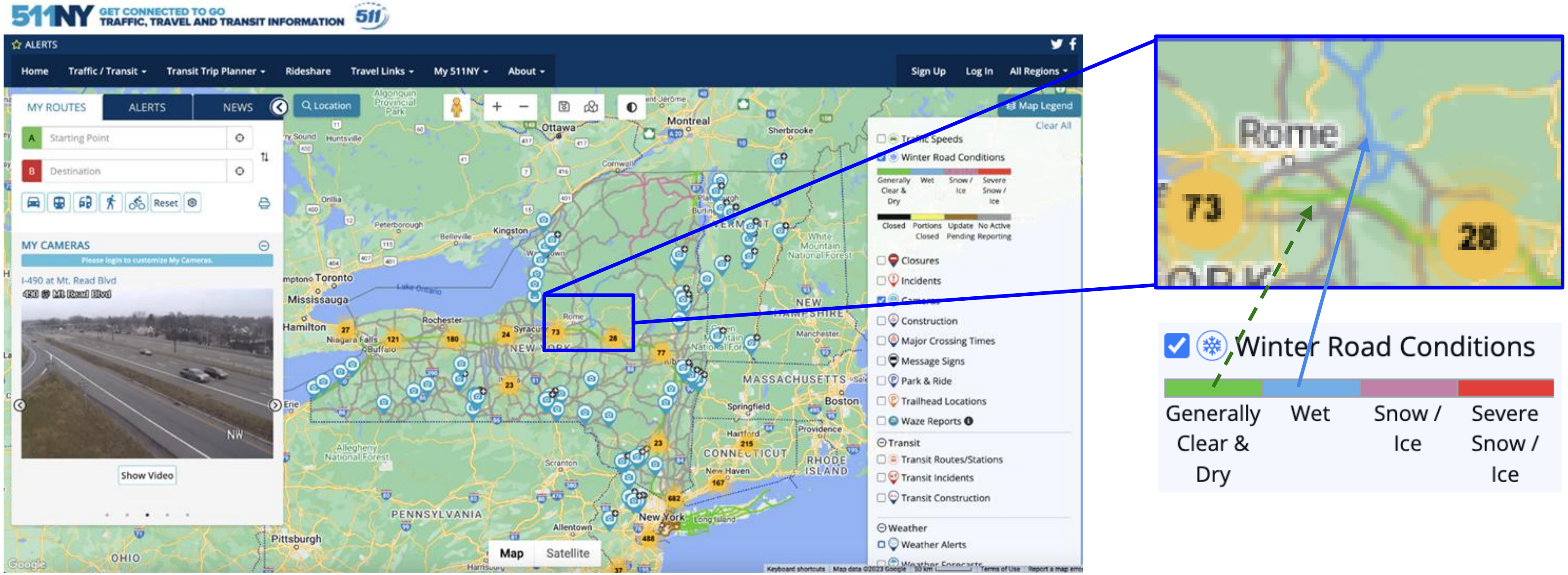
Data source: New York State Department of Transportation (NYSDOT)

Public live cameras: available at 511NY.org



Classification of Winter Road Conditions

Machine learning (ML) models can help make **predictions** about road surface conditions to **aid** current classification approaches



Data Archive

Archive of camera images starting late January 2022 in UAlbany's xCITE lab

Image snapshots saved out every 5 minutes from 2400 camera sites

Example of two Buffalo images 5 minutes apart

13:35:54 UTC

13:40:55 UTC

Modeling - Big Picture

Gather image data ✓

↓

Label the data (categorize data into distinct categories, human labeled)

↓

Use the labeled data to build a model (AI/ML)

↓

Determine effectiveness of model

Classification of Road Surface Condition - 6 classes

Severe Snow

Snow

Wet

Dry

Poor Visibility

Obstructed

The 4 main classes are highlighted and align with DOT WTA (Winter Travel Advisory): https://www.dot.ny.gov/wta/status-definitions

Hand-Labeled Dataset

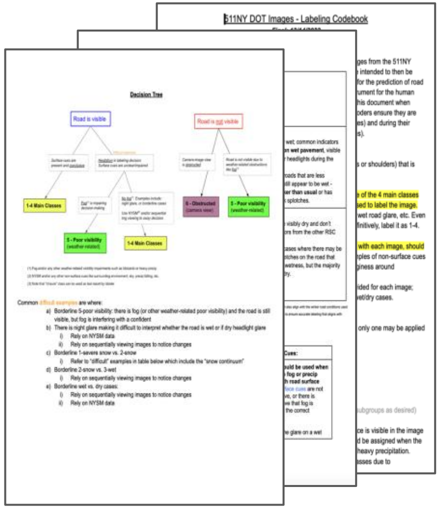
Labeling **“codebook”** developed with **social scientists**:

⇒ Carefully **define & set rules** for labeling to ensure **consistency** within this project and **reproducibility** for future projects

⇒ Works to increase **trustworthiness** in the ML model development process

Collaboration **with NYSDOT**:

- **4 main WTA classes** as used on 511ny.org
- Labeling rules established based on what can **reasonably be determined in images**



NYS Mesonet UNIVERSITY AT ALBANY ⟵ ***High quality weather stations*** *used to* ***aid in labeling decision making***

Labeled Images

Location	Num sites	Num imgs	Num classes
Bronx	2	4000	3 class
Ontario	1	2100	3 class
Rochester	1	2900	3 class
Buffalo (a)	1	1700	3 class
Queens	1	1400	3 class
Rensselaer	1	2000	3 class
Chatham	1	2000	3 class
Long Isl Expy.	20	10,000	6 class
Buffalo (b)	1	3,400	6 class(1)
Total	***29***	***29,000 (2)***	***multiple***

Selected some high quality cameras where road surface is visible at night and day

Image labeling iterations:

Past codebook versions with 3-classes

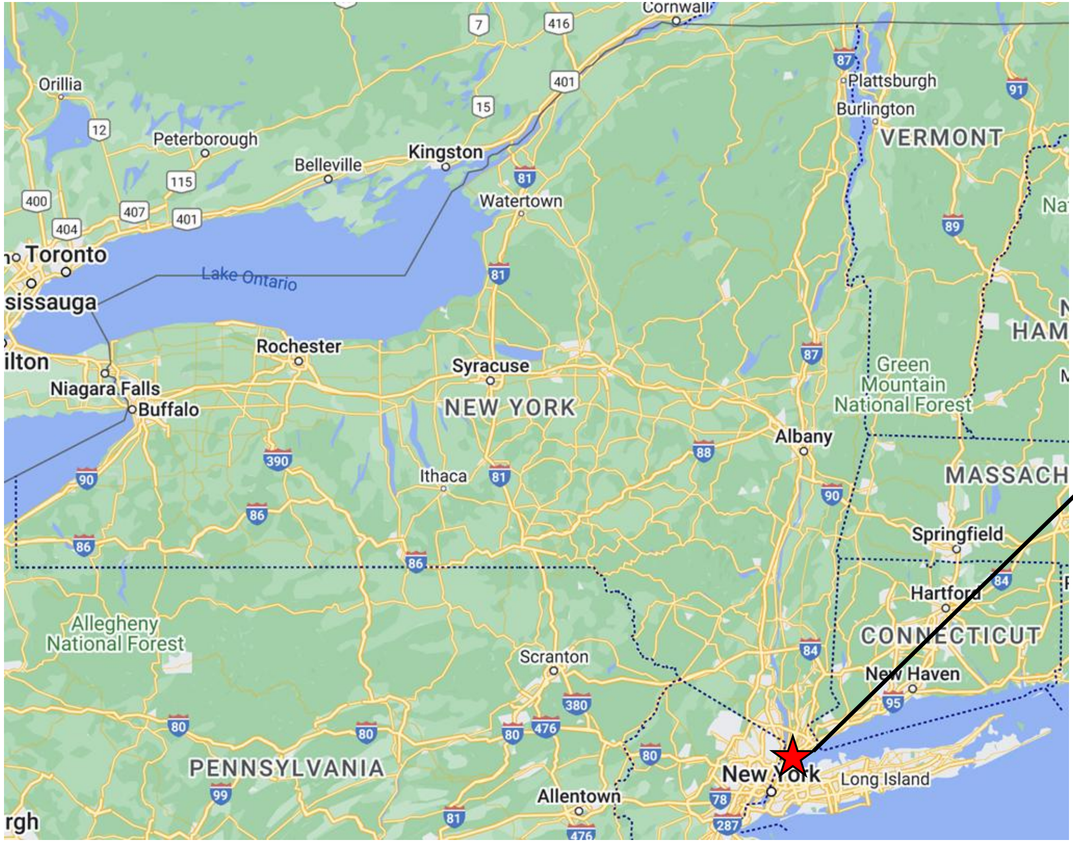
Final codebook with 6-classes

(1) no poor visibility or obstructed occurred yet in samples
(2) 28000 when including only 3-class: snow, wet, dry

NEW YORK STATE OF OPPORTUNITY. Department of Transportation

UNIVERSITY AT ALBANY State University of New York

Experiment Iterations

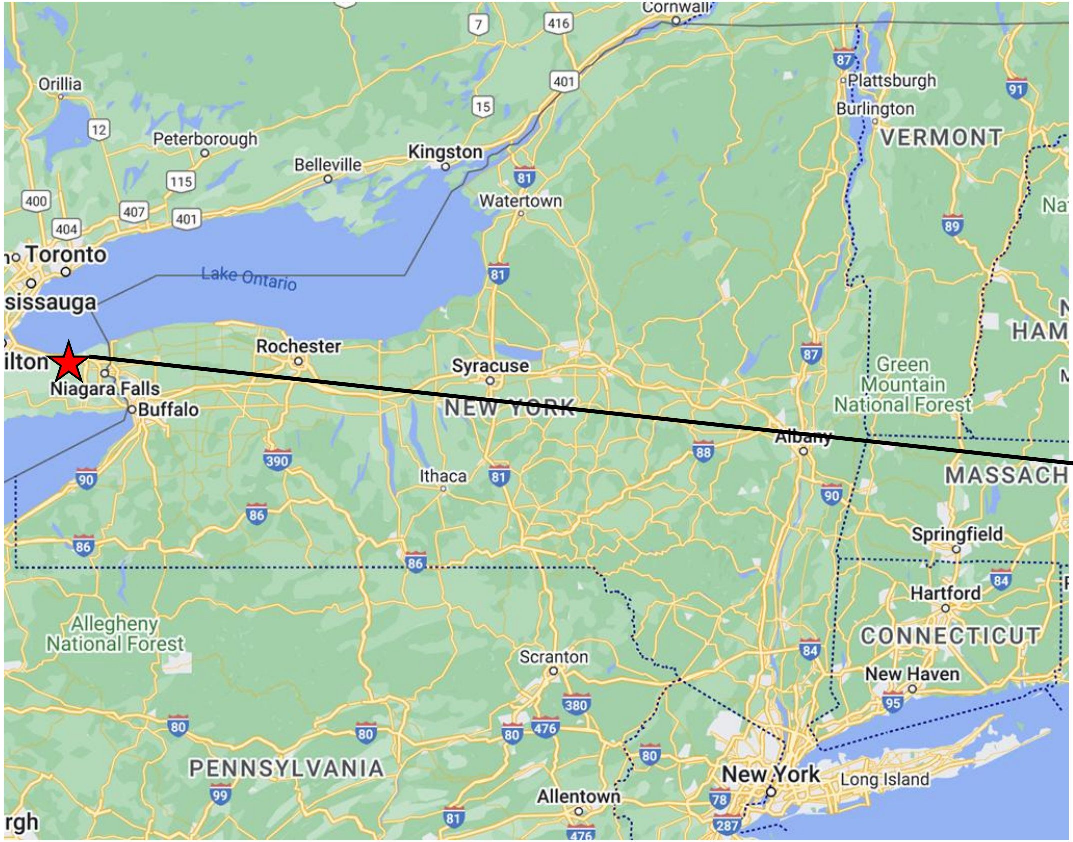


Bronx (2 cameras)

~4,500 labeled images

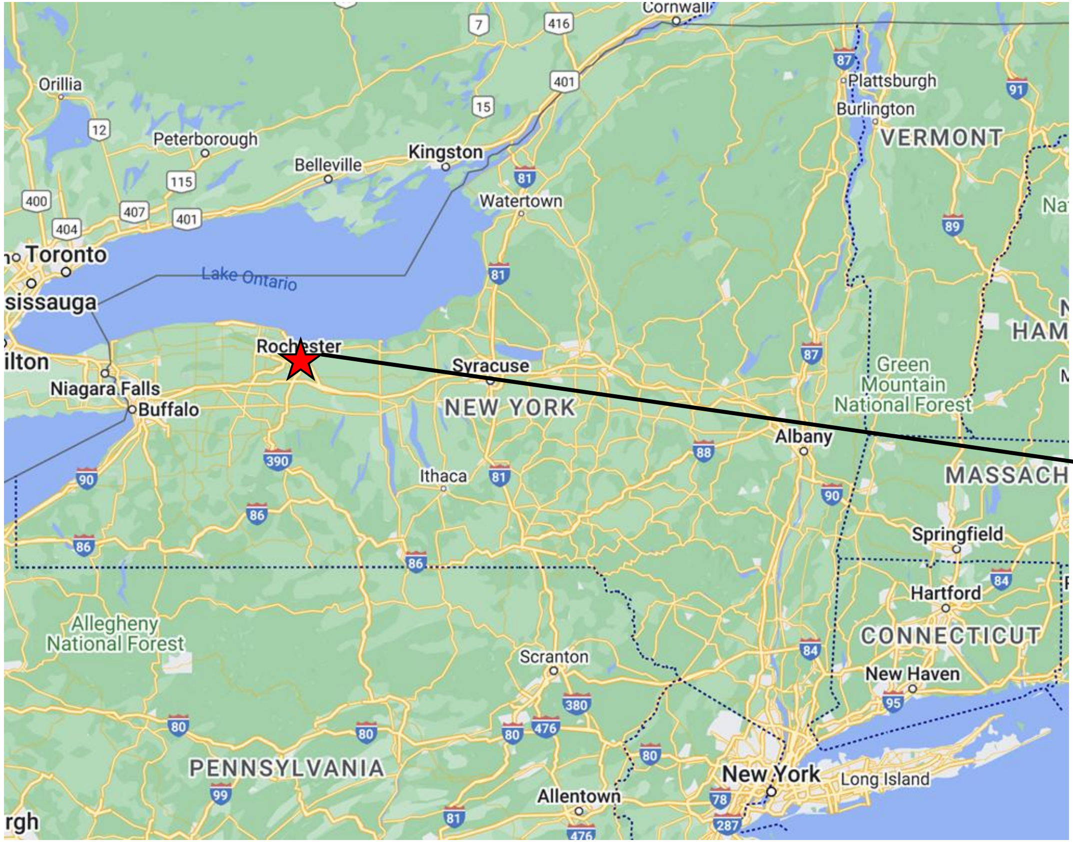
3 classes: snow, wet, dry

Experiment Iterations



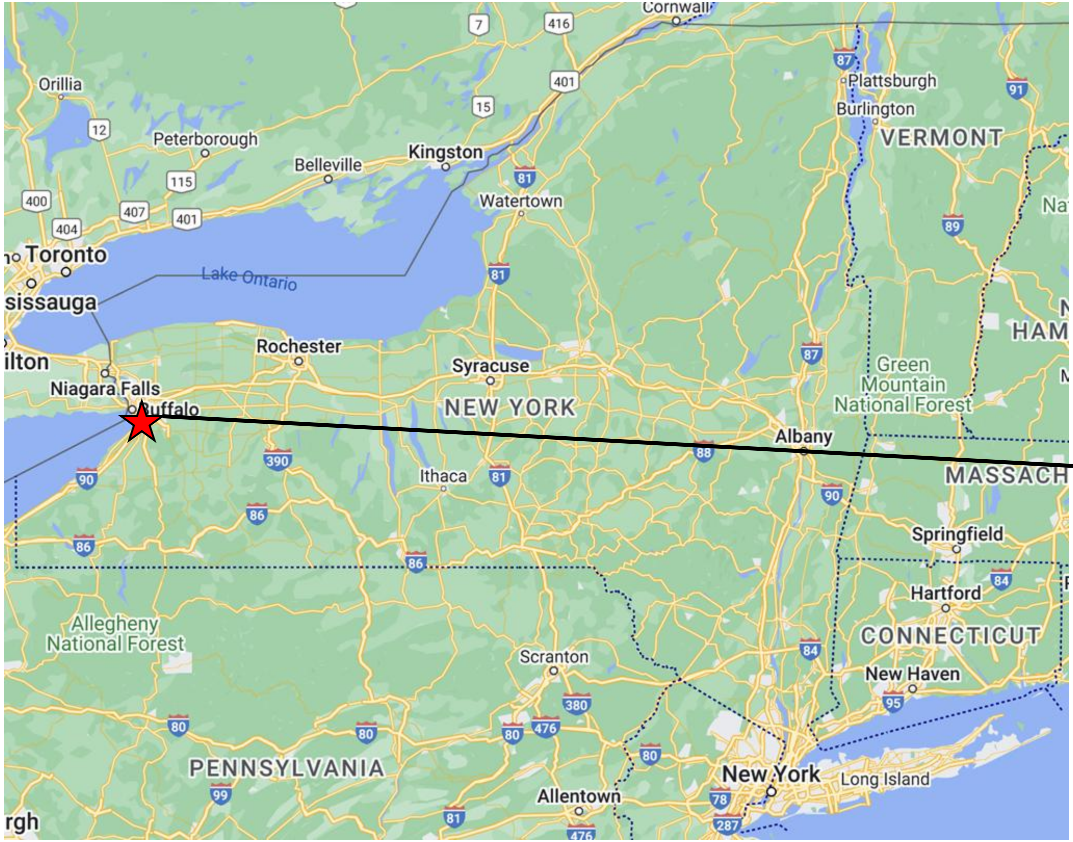
Ontario (St. Catherines)
~2100 labeled images 3 classes: snow, wet, dry

Experiment Iterations



Rochester
~2000 labeled images 3 classes: snow, wet, dry

Experiment Iterations

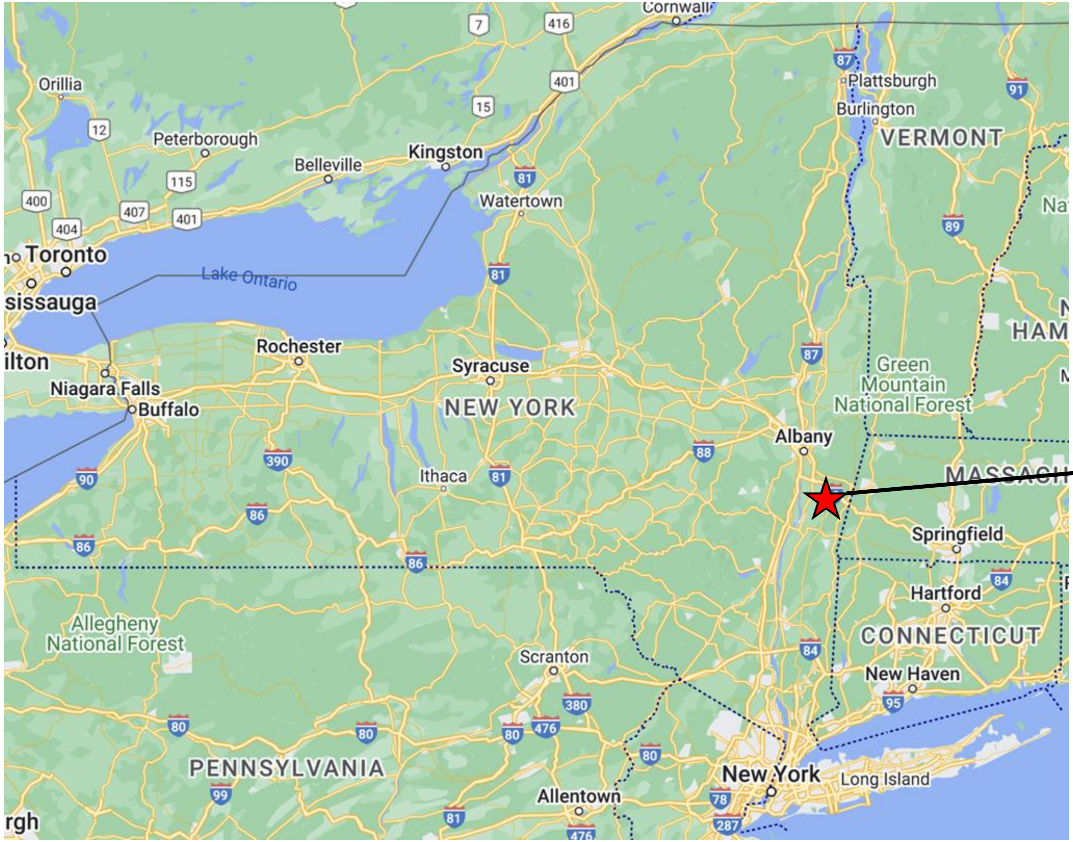


Buffalo

~2900 labeled images

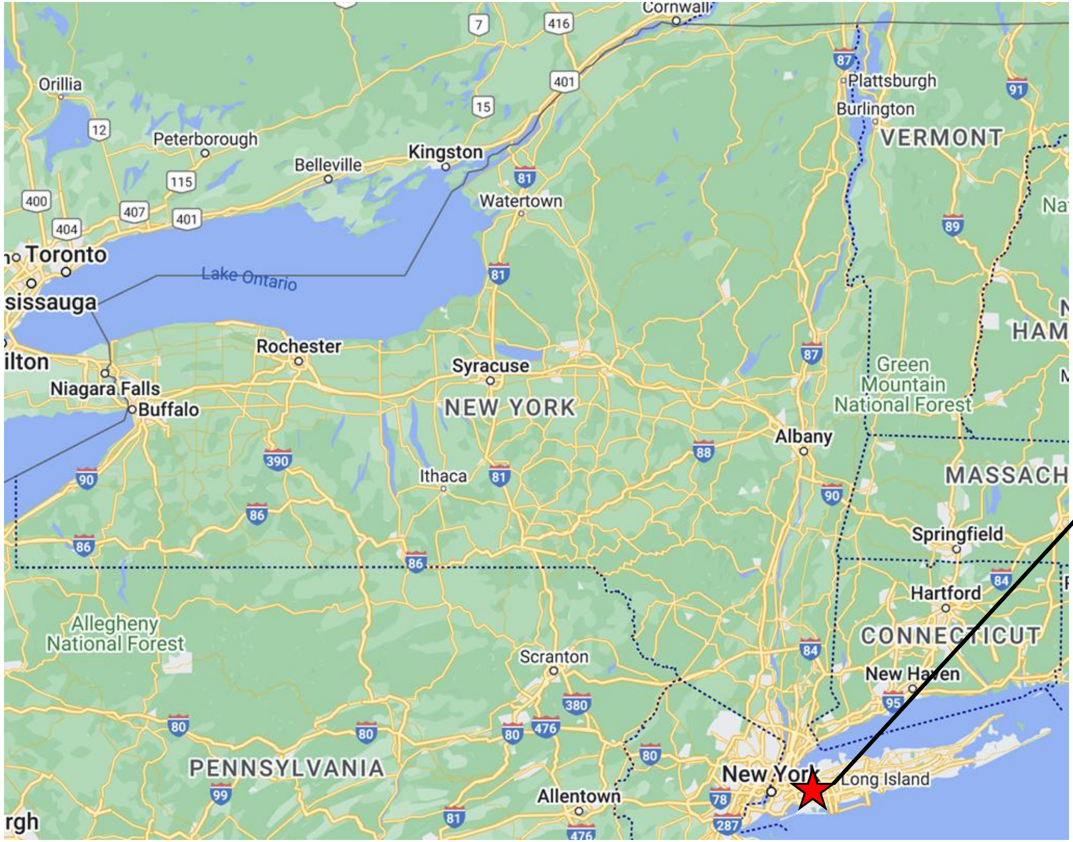
3 classes: snow, wet, dry

Experiment Iterations



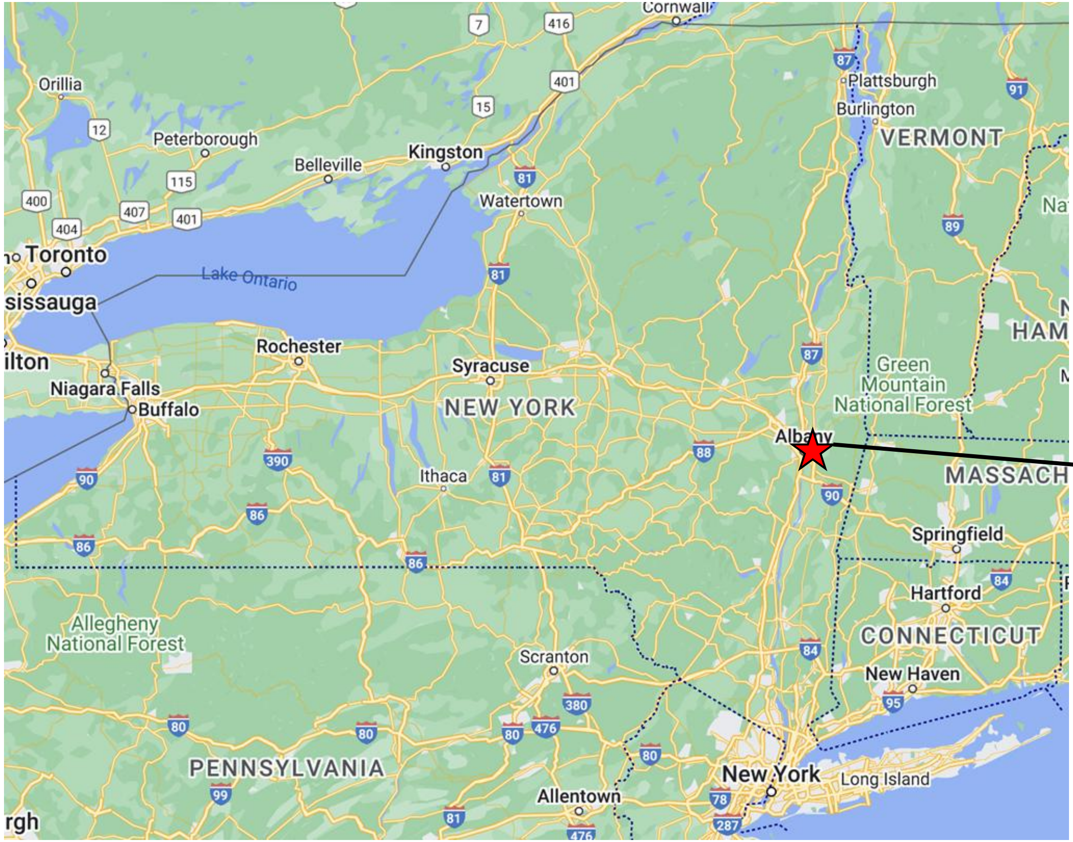
Chatham
~2000 labeled images 3 classes: snow, wet, dry

Experiment Iterations



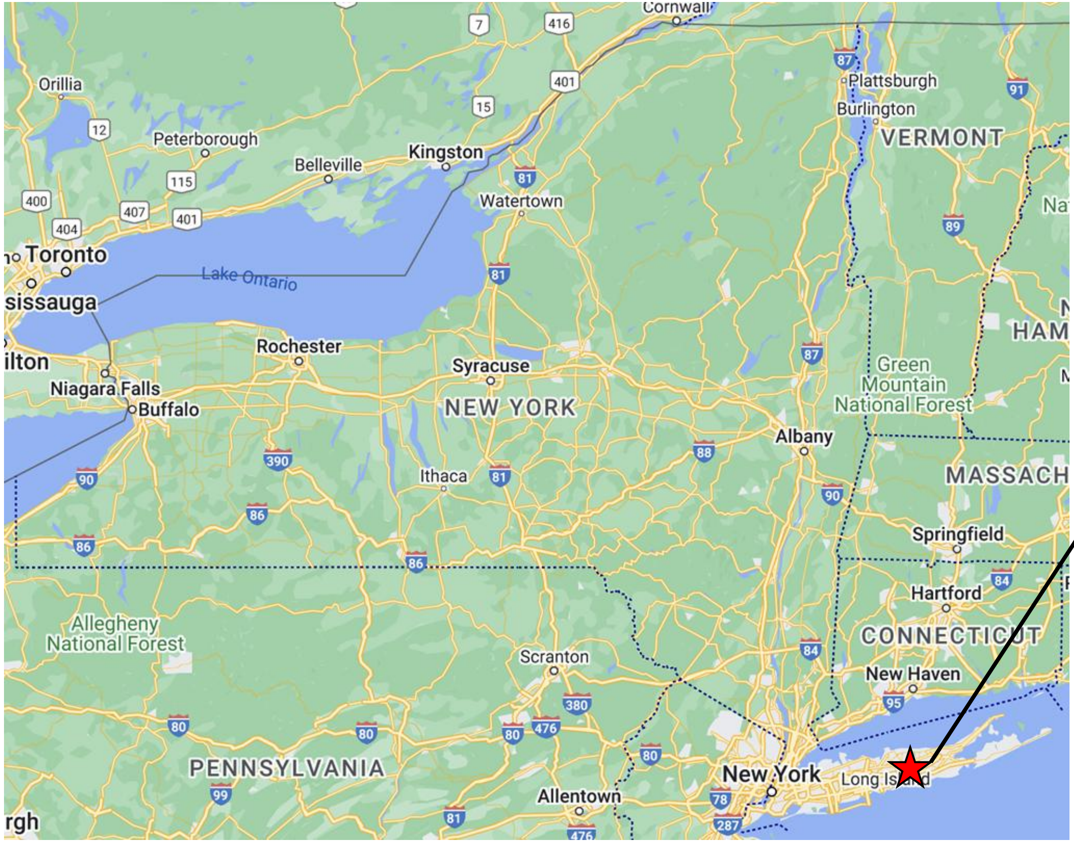
Queens
~1400 labeled images 3 classes: snow, wet, dry

Experiment Iterations



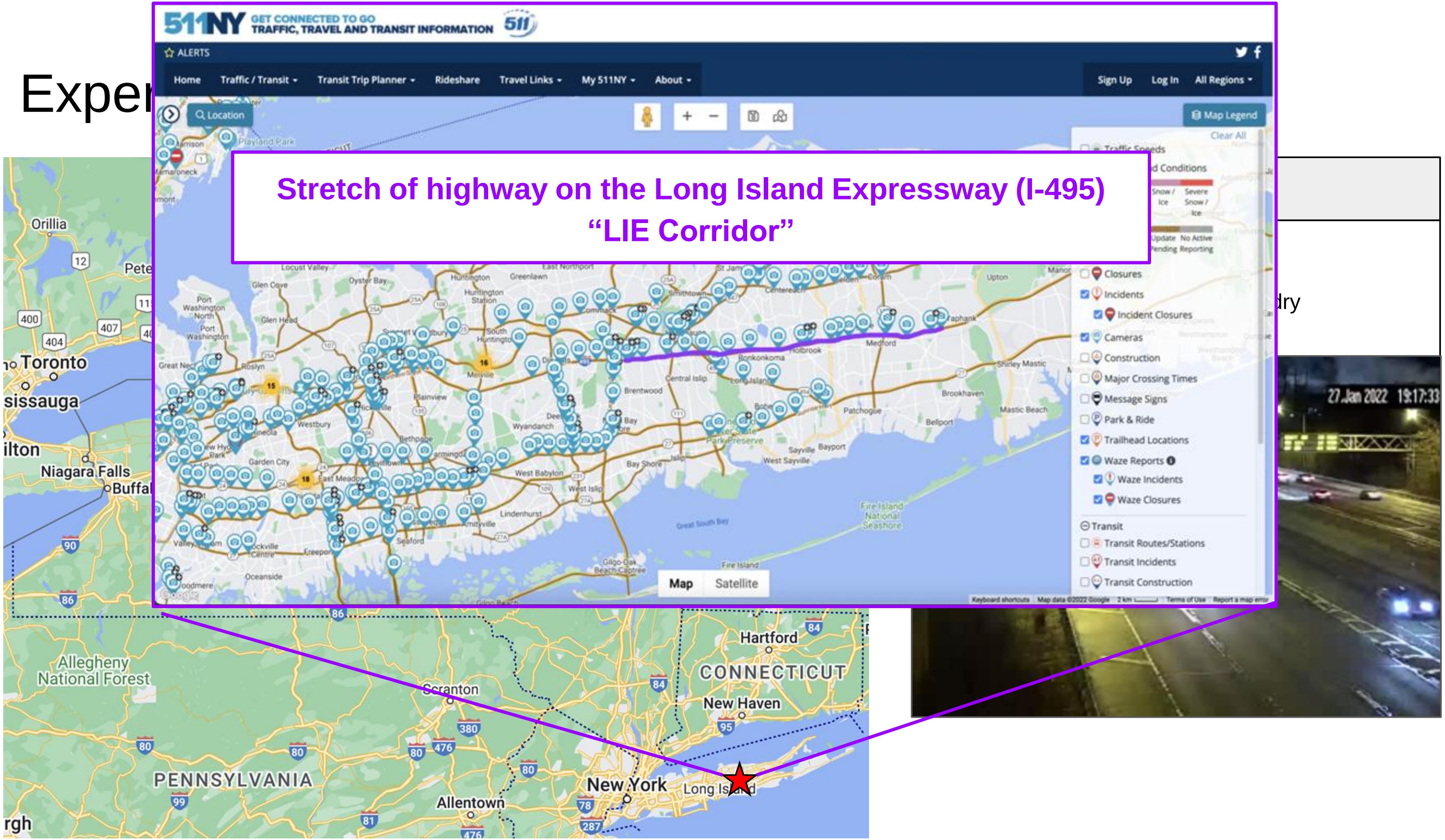
Rensselaer
~2000 labeled images 3 classes: snow, wet, dry

Experiment Iterations



Long Island, 20 cameras
~10,300 labeled images 6 classes: severe snow, snow, wet, dry, poor visibility, obstructed

Exper



Stretch of highway on the Long Island Expressway (I-495)
"LIE Corridor"

dry

Long Island Expressway (LIE) Labeled Images

Four main classes

Class	# images
Severe Snow	~900
Snow	~1,100
Wet	~3,800
Dry	~3,300
Obstructed	~700
Poor visibility	~500
Total	**~10,300** (1)(2)

(1) Class weights used to deal with class imbalance
(2) Augmentation (flip & rotate) to make robust dataset

From NYSDOT WTA (Winter Travel Advisory)

Severe snow

Wet

Poor
Visibility

Snow

Dry

Obstructed

Modeling - Big Picture

Gather image data ✓

↓

Label the data (categorize data into distinct categories, human labeled) ✓

↓

Use the labeled data to build a model (AI/ML)

↓

Determine effectiveness of model

Preprocessing & Modeling

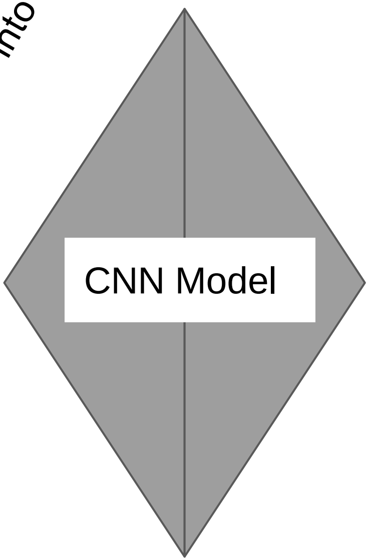
Convolutional Neural Networks (CNNs) are ML algorithms commonly used for image classification problems

Original Image
320x240

Preprocess images →

Preprocessed Image
Resized 224x224
Cropped

Input image into model →



Model outputs a **road surface prediction** →

Snow

Modeling - Big Picture

Gather image data ✓

↓

Label the data (categorize data into distinct categories, human labeled) ✓

↓

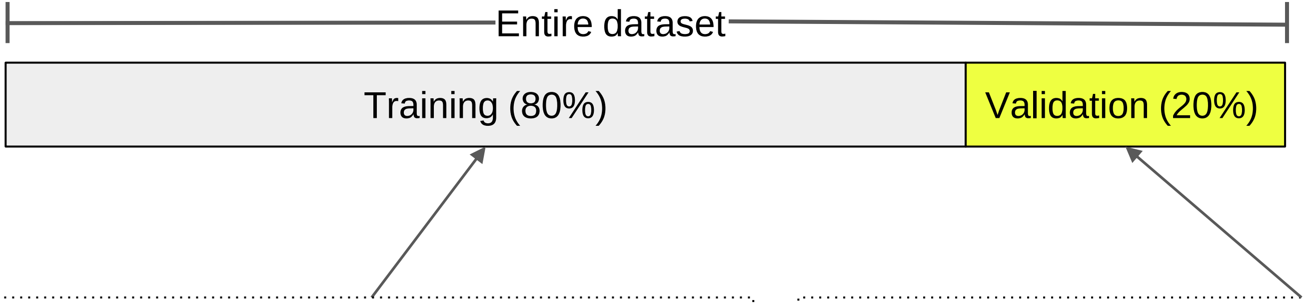
Use the labeled data to build a model (AI/ML) ✓

↓

Determine effectiveness of model

Data splits

Assess performance of models using a **validation** (aka “holdout”) dataset



Dataset used to build the model

Dataset used to evaluate model performance

The goal is to have a model that makes accurate predictions on data it hasn’t seen before (i.e. the holdout data)

Model Successes

Data split method 1: **all 20 sites are represented in both training and validation**

Class	Severe snow	Snow	Wet	Dry	Poor Visibility	Obstructed	***Average Accuracy***
Correct[(1)] % (model label = human label)	94.70%	91.80%	91.60%	94.10%	89.00%	92.90%	***92.60%***

(1) Metric shown is recall (out of total labeled in that class). Calculation is recall = probability of detection (POD) = True Positive / (True Positive + False Negative)

Most commonly misclassified as ***snow***

Most commonly misclassified as ***severe snow*** or ***wet***

Most commonly misclassified as ***poor visibility*** or ***dry***

Most commonly misclassified as ***wet***

Most commonly misclassified as ***wet*** or ***obstructed***

Most commonly misclassified as ***poor vis*** or ***dry***

Examples

Correct Prediction | Incorrect

Severe snow

actual label snow_severe
model predicted snow_severe
dry 0.1%, wet 0.6%, snow 12.7%, sev snow 71.0%, obs 0.8%, viz 14.8%

actual label snow_severe
model predicted snow
dry 0.0%, wet 0.1%, snow 72.3%, sev snow 27.5%, obs 0.0%, viz 0.0%

Snow

actual label snow
model predicted snow
dry 0.3%, wet 0.1%, snow 76.2%, sev snow 22.8%, obs 0.6%, viz 0.0%

actual label snow
model predicted snow_severe
dry 0.0%, wet 0.2%, snow 40.8%, sev snow 58.9%, obs 0.0%, viz 0.0%

Examples

Correct Prediction | Incorrect

Wet

actual label wet
model predicted wet
dry 0.0%, wet 100.0%, snow 0.0%, sev snow 0.0%, obs 0.0%, viz 0.0%

actual label wet
model predicted poor_viz
dry 0.4%, wet 28.7%, snow 0.0%, sev snow 0.0%, obs 0.2%, viz 70.8%

Dry

actual label dry
model predicted dry
dry 38.4%, wet 26.2%, snow 35.1%, sev snow 0.1%, obs 0.1%, viz 0.0%

actual label dry
model predicted wet
dry 38.5%, wet 61.4%, snow 0.0%, sev snow 0.0%, obs 0.0%, viz 0.0%

NEW YORK STATE OF OPPORTUNITY. Department of Transportation

UNIVERSITY AT ALBANY State University of New York

Examples

Correct Prediction | Incorrect

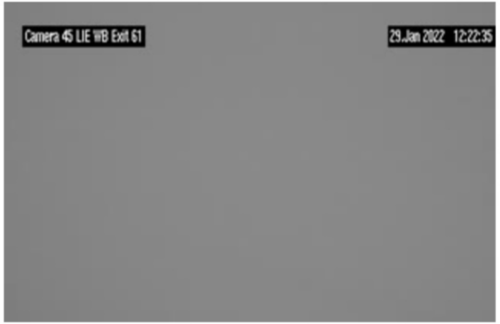
Poor Visibility

actual label poor_viz
model predicted poor_viz
dry 0.8%, wet 17.6%, snow 0.2%, sev snow 1.5%, obs 27.5%, viz 52.5%

actual label poor_viz
model predicted wet
dry 5.7%, wet 92.8%, snow 0.0%, sev snow 0.0%, obs 0.3%, viz 1.1%

Obstructed

actual label obs
model predicted obs
dry 0.0%, wet 0.1%, snow 0.0%, sev snow 0.3%, obs 89.6%, viz 10.1%



actual label obs
model predicted poor_viz
dry 0.0%, wet 0.4%, snow 0.0%, sev snow 0.0%, obs 47.7%, viz 51.8%

Model Difficulties

Data split method 2: **unique sites in training vs validation**

- 16 sites in training
- 4 sites in validation (holdout)

Model accuracy about ~70%

⇒ An important **problem to fix** because we want the model to **generalize** well to unseen sites across the state

Modeling - Big Picture

Gather image data ✓

↓

Label the data (categorize data into distinct categories, human labeled) ✓

↓

Use the labeled data to build a model (AI/ML) ✓

↓

Determine effectiveness of model ✓

Summary

ML model **performs well on data it's "seen"**

Example:

If model is trained on Site A and Site B, it will perform well on Site A and Site B

⇒ A matter of having labeled data

ML model **needs improvement on data it's never seen**

Example:

If model is trained on Site A and Site B it won't perform very well on Site C

⇒ An open-ended algorithm development/computer science question

Future Work

Drivin by goal of model generalizability:

- Labeling
  - Have winter 2022-2023 season images to include
  - Labeling working group to label images, adding thousands more images across new sites
- Improving model architecture

Thank you!

This material is based upon work supported by the National Science Foundation under Grant No. ICER-2019758

Contact for questions or comments:
Carly Sutter csutter@albany.edu,
Nick Bassill nbassill@albany.edu,
Kara Sulia ksulia@albany.edu

ai2es.org

Appendix

Two Methods for Train/Val Data Splitting

Cam #	**All sites in training and validation** *High accuracy model >92%*		**Site-specific validation** *Model accuracy ~70%*
1	Training	Val	Training
2	Training	Val	Training
3	Training	Val	Training
4	Training	Val	Training
5	Training	Val	Training
6	Training	Val	Training
7	Training	Val	Training
8	Training	Val	Training
9	Training	Val	Training
10	Training	Val	Training
11	Training	Val	Training
12	Training	Val	Training
13	Training	Val	Training
14	Training	Val	Training
15	Training	Val	Training
16	Training	Val	Training
17	Training	Val	Val
18	Training	Val	Val
19	Training	Val	Val
20	Training	Val	Val

Preprocessing & Model Details

Convolutional Neural Network (CNN)

Using VGG16 architecture

Original Image
320x240

Preprocessed Image
Resized 224x224
Cropped

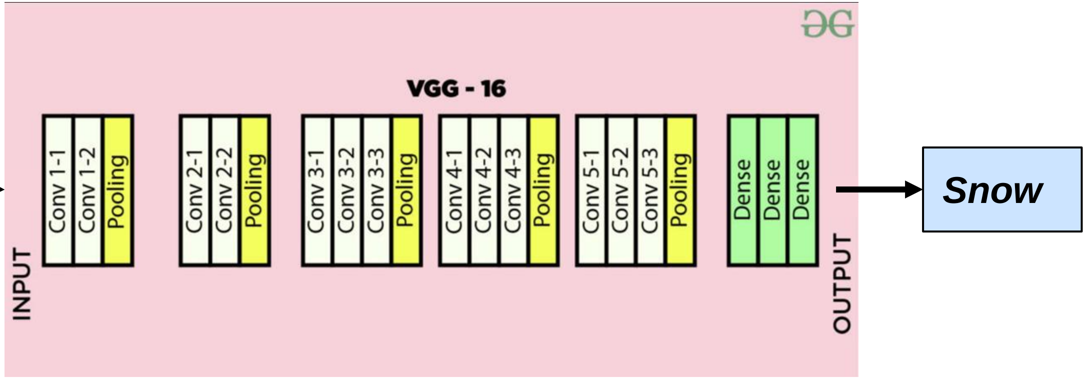


Image source: https://www.geeksforgeeks.org/vgg-16-cnn-model/

More Model Details

Python Package:	***Loss function:***	***Optimizer:***	***Learning Rate:***	***Activation function:***	***Epochs:***
Tensorflow	*Categorical Cross Entropy*	*SGD*	*0.01, exponential decay rate 0.99*	*Relu* *Softmax (output layer)*	*Max 50* *Early stopping 10*

Other modeling details

Image data generator

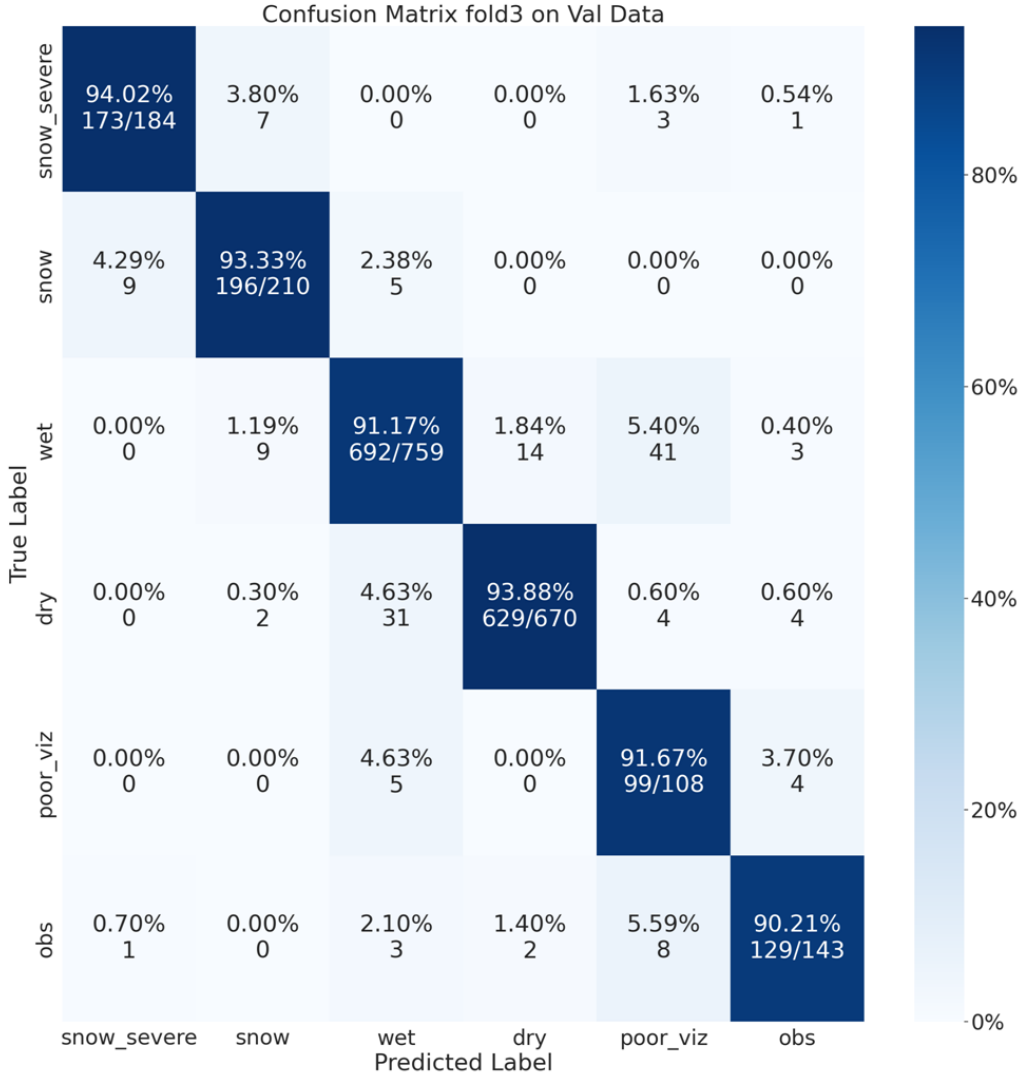
5-fold cross validation used

Updates made from last meeting with DOT in November 2022:

- Regularization
- Dropout
- Additional image augmentation (brightness temperature

Confusion Matrix

One of the folds from 5-fold CV



Example of Incorrect Label

Human error in labeling in which case the model did better!

actual label poor_viz
model predicted wet
dry 0.0%, wet 100.0%, snow 0.0%, sev snow 0.0%, obs 0.0%, viz 0.0%

Experiments to Optimize Image Labeling

Goal of experiments: **understand how to optimize time spent labeling images, especially under the lens of generalizability**

- Secondary goal: provide perspective on claim that model just “needs more data”

Findings:

1) the focus should be on adding **more sites** rather than more images of existing sites (experiment 3)

2) **~400 images per site** is optimal (experiment 2)

3) **~12 training sites per 3 validation sites** (15 sites) is ideal (experiment 1b)

Identifying Images for Labeling