working to advance road weather information systems technology

Automated Extraction of Weather Variables from Imagery

http://aurora-program.org

Aurora Project 2021-06

Final Report
February 2023

About Aurora

The Aurora program is a partnership of highway agencies that collaborate on research, development, and deployment of road weather information to improve the efficiency, safety, and reliability of surface transportation. The program is administered by the Center for Weather Impacts on Mobility and Safety (CWIMS), which is housed under the Institute for Transportation at Iowa State University. The mission of Aurora and its members is to seek to implement advanced road weather information systems (RWIS) that fully integrate state-of-the-art roadway and weather forecasting technologies with coordinated, multi-agency weather monitoring infrastructures.

The preparation of this report was financed in part through funds provided by the Iowa Department of Transportation through its "Second Revised Agreement for the Management of Research Conducted by Iowa State University for the Iowa Department of Transportation" and its amendments.

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
Aurora Project 2021-06		

4. Title and Subtitle
Automated Extraction of Weather Variables from Imagery

5. Report Date
February 2023

6. Performing Organization Code

7. Author(s)
Gerry Wiener, William Petzke, Tom Brummet, and Curtis Walker

8. Performing Organization Report No.

9. Performing Organization Name and Address
National Center for Atmospheric Research
P.O. Box 3000
Boulder, CO 80307

10. Work Unit No. (TRAIS)

11. Contract or Grant No.

12. Sponsoring Organization Name and Address
Aurora Program
Iowa Department of Transportation
800 Lincoln Way
Ames, IA 50010

Federal Highway Administration
1200 New Jersey Avenue, SE
Washington, DC 20590

13. Type of Report and Period Covered
Final Report

14. Sponsoring Agency Code
Part of TPF-5(435) and Federal SPR Part II, CFDA 20.205

15. Supplementary Notes
Visit https://aurora-program.org/ for color pdfs of this and other research reports.

16. Abstract
Department of transportation (DOT) maintenance supervisors utilize a variety of tools including maintenance decision support systems (MDSS) to gain a better understanding of current and future road surface conditions (RSC) during winter weather. MDSS automatically attempt to deduce current RSC based on road weather information system (RWIS) and other data to a greater or lesser degree of accuracy. Although current MDSS implementations present highway camera imagery, they typically do not incorporate automated camera image recognition in order to improve the MDSS assessment of winter RSC. Thus, there can be discrepancies between the road weather conditions in camera images and MDSS RSC assessments. For example, an MDSS assessment may determine that a highway is clear whereas associated camera images show snow or vice versa. Such discrepancies can lead to a loss of confidence, system criticism, and noncompliance with MDSS recommendations. From that point of view, the integration of automated RSC camera image recognition into MDSS implementations can have a number of benefits:

- Better RSC assessment performance
- Better road treatment recommendations owing to better RSC identification
- Improved MDSS use and compliance with system recommendations owing to user confidence

Recent research in RSC identification has applied convolutional neural networks (CNN) and related techniques to the winter RSC identification problem. The National Center for Atmospheric Research (NCAR) is interested in transitioning these automated RSC identification techniques from the research community to the DOT community. Since NCAR has significant experience in the research-to-operations arena, the NCAR team worked with Aurora Program members to develop a set of recommendations for transitioning the relevant technology to MDSS applications.

Even though recent research efforts seem quite successful, the NCAR team was also interested in the potential to improve the initial CNN RSC identification by incorporating additional relevant data such as the following:

- RWIS precipitation/temperature data
- Vehicle speed/volume data

17. Key Words	18. Distribution Statement
road surface conditions—road weather information systems—winter maintenance—weather conditions	No restrictions.

19. Security Classification (of this report)	20. Security Classification (of this page)	21. No. of Pages	22. Price
Unclassified.	Unclassified.	54	NA

Form DOT F 1700.7 (8-72) **Reproduction of completed page authorized**

AUTOMATED EXTRACTION OF WEATHER VARIABLES FROM IMAGERY

Final Report
February 2023

Principal Investigator
Gerry Wiener, Software Engineer IV
National Center for Atmospheric Research

Research Assistants
William Petzke, Thomas Brummet, Curtis Walker

Authors
Gerry Wiener, Bill Petzke, Thomas Brummet, and Curtis Walker

Sponsored by
Federal Highway Administration Aurora Program
Transportation Pooled Fund
(TPF-5(435))

Preparation of this report was financed in part
through funds provided by the Iowa Department of Transportation
through its Research Management Agreement with the
Institute for Transportation

A report from
Aurora Program
Iowa State University
2711 South Loop Drive, Suite 4700
Ames, IA 50010-8664
Phone: 515-294-8103 / Fax: 515-294-0467
https://aurora-program.org/

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

This research was conducted under the Federal Highway Administration (FHWA) Transportation Pooled Fund Aurora Program. The authors would like to acknowledge the FHWA, the Aurora Program partners, and the Iowa Department of Transportation (DOT), which is the lead state for the program, for their financial support and technical assistance.

This research would not have been possible without the support of the Aurora Pooled Fund technical advisory committee members: Zach Hans, Neal Hawkins, Jamie Yount, Jason Norville, Dale Kirmer, Mike Adams, Kevin Duby, Steve Hancock, James Roath, Jon Bjorkquist, Paul Denkler, Richard “Todd” Miller, Colby Fortier-Brown, Aidan Neeley, and Iowa DOT representative Tina Greenfield Huitt.

The researchers would like to acknowledge the following organizations for sharing data and expertise: the North Dakota Department of Transportation and the Alaska Department of Transportation and Public Facilities (DOT&PF). Additional thanks are due to Khyle Clute of the Iowa DOT.

EXECUTIVE SUMMARY

Machine learning (ML) techniques can be used to automatically identify winter road surface conditions (RSC) using highway camera images. These techniques can then be integrated into maintenance decision support systems (MDSS) and other tools, resulting in a number of benefits:

1. Better RSC assessment performance
2. Better road treatment recommendations owing to more accurate RSC identification
3. Improved MDSS use and compliance with system recommendations owing to user confidence

The overarching goal of the research described in this technical report was to identify and evaluate ML approaches that can automatically identify RSC during winter weather and be transitioned to the department of transportation (DOT) community. The ML approaches utilize camera images of highways and other relevant sensed data at highway locations. The research team investigated technology available from cloud vendors as well as open-source machine learning software used by the research community. One important question was whether cloud ML technology had similar performance to open-source ML software.

The research involved three phases:

1. Data collection and processing
2. Labeling of camera images to support the automatic recognition of RSC
3. Evaluation of ML techniques that identify the winter RSC using camera images

Road weather information system (RWIS) camera images from the state of North Dakota were collected from January through April and from October through December in 2022. The research team used cloud storage for image archiving. The bulk of the processing was performed on National Center for Atmospheric Research (NCAR) workstations and graphical processing nodes. Both Amazon Web Services (AWS) Rekognition and Microsoft Azure Custom Vision ML technology were evaluated and compared to open-source ML algorithms. The research team found that efficient and accurate labeling played a key role in the successful application of ML to the automatic RSC identification problem. The research team also found that cloud-based ML technology can be successfully used for RSC identification.

Key Findings

The key findings from this research were as follows:

- A clear view of the road in a camera image is important for labeling accuracy and ML performance. Images in which the main road is off in the distance are more challenging to classify.
- For successful ML, it is important to have an adequate number of accurately labeled images for each RSC category.

- Dirty camera lenses, glare, darkness, low visibility, and obstructions all have a negative effect on identifying snow conditions.
- Using a time series of camera images at a single site is an effective way to quickly and accurately label the RSC.
- It is important to establish a set of clear, unambiguous, and consistent rules for performing the RSC labeling.
- Cloud vendor image classification technology exhibited similar performance to open-source image classification technology and was easier to apply.
- Camera images from multiple cameras can be combined together to train ML models that perform well on different camera views of roads.
- ML classification technology performed extremely well in many of the experiments performed by the research team and shows promise for integration into MDSS and other tools.

INTRODUCTION

State department of transportation (DOT) road maintenance supervisors need to properly diagnose snow and ice impacts on highways and roads during the winter season. They use a wide variety of information sources to support their decision making. Such information sources include DOT plow operator reports, road weather information systems (RWIS), highway cameras, microwave vehicle radar detectors (MVRD), incident reports, weather nowcasts/forecasts, radar mosaics, etc. Supervisors often amalgamate the information from these different information sources during winter storms, often with the support of meteorologists, in order to gain a better understanding of the current and future road and winter weather conditions at key highway locations. They then make decisions with regard to appropriate road maintenance responses.

With the advent of computerized maintenance decision support systems (MDSS) for winter weather, supervisors have begun to rely on the conclusions these systems provide with regard to the current and future road surface conditions (RSC). These systems automatically attempt to deduce current road weather conditions based on the data mentioned above to a greater or lesser degree of accuracy. Although current MDSS implementations present highway camera imagery, they often do not incorporate automated camera image recognition in order to improve the MDSS assessment of the winter RSC. Thus, there can be discrepancies between the road weather conditions in the camera images and the MDSS RSC assessments. For example, an MDSS assessment may determine that a highway is clear whereas the associated camera images show snow or vice versa. Such discrepancies typically lead to a loss of confidence in the system, system criticism, and noncompliance with system recommendations. From that point of view, the integration of automated RSC camera image recognition into MDSS implementations can have a number of benefits:

1. Better RSC assessment performance
2. Better road treatment recommendations owing to more accurate RSC identification
3. Improved MDSS use and compliance with system recommendations owing to user confidence

There has been recent research in the RSC identification area (Pan et al. 2019, Zhang et al. 2021), and such research has focused on applying convolutional neural networks (CNN) and related image classification techniques to the winter RSC identification problem. Because these research efforts seem quite successful, the National Center for Atmospheric Research (NCAR) team is interested in the potential to transition these automated RSC identification techniques from the research community to the DOT community. Since NCAR has significant experience in the research-to-operations arena, the NCAR team developed a set of recommendations for transitioning the relevant technology to MDSS applications.

NCAR was also interested in whether the initial CNN RSC identification can be improved by incorporating additional relevant data such as the following:

- RWIS sensor data including surface temperature, precipitation rate, and intensity

- Vehicle speed/volume data

For example, NCAR has found that the incorporation of RWIS precipitation rate and intensity into machine learning (ML)-based friction models has led to significantly better road friction predictions compared to the use of air temperature, dew point, relative humidity, pressure, and wind speed measurements alone. In the same way, NCAR expected that incorporating additional in situ information should lead to improved RSC identification performance.

Goals and Objectives

The overarching goal of this project was to identify ML approaches that automatically identify RSC during winter weather using camera imagery and other relevant sensed data at highway locations. It was determined that the ML approaches should be sufficiently accurate to be useful to DOT maintenance staff and should be capable of being effectively implemented in MDSS or standalone implementations.

In achieving these goals, the project set forth the following objectives:

1. Determine what ML techniques have the best performance in identifying road conditions through the use of camera imagery and other relevant observation data
   a. Evaluate techniques that researchers have found to be beneficial in RSC identification
   b. Evaluate cloud-based image identification techniques to see how they compare with the techniques used by researchers
2. Identify approaches that can be implemented in MDSS to identify road surface conditions impacted by weather using camera imagery and other relevant observation data
3. Determine how the ML techniques found in the first two objectives can be implemented within the context of an MDSS or in a standalone manner

The subsequent chapters of this report provide detailed descriptions of the data sources and data processing methods used, the image labeling process used, the ML techniques applied, the models' performance, and recommendations for implementation.

DATA COLLECTION AND PROCESSING

Objectives

The objective of the data collection and processing was to create data sets for a number of different states that could then be used for RSC identification in this study and potentially future studies. It was determined that the particular data sets collected for each state should cover at least one complete winter season. Note, however, that it is preferable to have multiple winter seasons of data. Ideally, the data sets should contain data for all RWIS sites in a state since winter conditions can vary significantly on a per site basis, especially in states that have both mountainous and non-mountainous regions.

Methods

Each state typically provides RWIS camera, sensor, and speed data using a state-specific application programming interface (API). As a result, it is necessary to implement customized approaches for each state in order to ingest the RWIS image and sensor data. Note that the National Oceanic and Atmospheric Administration (NOAA) Meteorological Assimilation Data Ingest System (MADIS) (https://madis.ncep.noaa.gov) does provide a uniform API to ingest RWIS sensor data. MADIS, however, does not currently provide camera imagery, and the RWIS sensor data provided by MADIS can be incomplete. The research team therefore implemented new ingest techniques to obtain RWIS image; sensor; and speed, volume, and occupancy (SVO) data from the North Dakota DOT (NDDOT).

RWIS Camera Data

NDDOT image data are stored on the web at specific URLs in JPEG-formatted files. A map that specifies the actual URLs is available as a JavaScript Object Notation (JSON) file at https://travelfiles.dot.nd.gov/geojson_nc/cameras.json. At regular 20-minute intervals during daylight hours, NCAR used the Linux scheduler, crontab, to download the NDDOT imagery. All the current images listed in the cameras.json file were downloaded and then stored in appropriate directories on disk as described below. The entire process was automatic and required little manual intervention.

The image data were stored in daily subdirectories, each named with a yyyymmdd string such as 20220108. Each daily subdirectory contained images organized by site identifier. For example, NDDOT uses five-digit strings such as 08516 as site identifiers. Within the site-based subdirectories were files named with the site identifier, a camera angle identifier, a date string, and an hour-minute string. So, for example, the file, 08516_02_20220817.1540.jpg corresponds to site identifier 08516, camera angle 02, date 20220817, and hour-minute 1540. With this file naming policy, it was easy to collect time series of image files for a particular site and camera angle for labeling and ML.

The research team had previously archived camera images from Alaska and used those images to supplement the NDDOT image data archive.

RWIS Sensor Files

NDDOT RWIS sensor data are available in a summary JSON file at a specific URL, https://travelfiles.dot.nd.gov/geojson_nc/ess.json. The ess.json file was downloaded at 10-minute intervals and renamed using an appropriate date and time. File names used a template, rwis_raw.yyyymmdd.hhmm.json. For example, rwis_raw.20221229.0750.json is an RWIS file downloaded on December 29, 2022.

SVO Data

SVO data are not available through the NDDOT website. Instead, the research team identified three sites that were impacted by snow, and NDDOT supplied the SVO data for these three sites:

1. Grand Forks, I-29S, covering a time period from January 1, 2022, through May 29, 2022
2. Rubgy, US-2E, covering a time period from January 1, 2022, through March 10, 2022
3. Buffalo, ND-38N, covering a time period from January 1, 2022, through May 31, 2022

LABELING CAMERA IMAGERY TO SUPPORT AUTOMATIC RECOGNITION OF RSC

At this time, the majority of deep learning techniques used for image identification typically involve CNNs. CNNs use multiple-layer neural networks to learn hierarchical representations of the image data and have been successfully employed in object and scene identification.

Cloud vendors such as Amazon Web Services (AWS), Microsoft Azure, and Google Cloud have developed image recognition technology that employ CNNs and other approaches to identify objects in images and in video. The models implemented by the different cloud vendors have been previously trained using millions of images covering a wide variety of domains. Starting from scratch, the training process for CNN models requires significant time, resources, and expertise. However, once the models have been implemented, they can subsequently be leveraged for customized applications. In leveraging the technology for customized applications, the user can provide a custom labeled set of images that cover a particular use case and then retrain an existing model. Cloud vendor technology stages the computational resources and processing to perform the model retraining automatically. The user typically reviews the retraining performance and then decides whether further labeling and retraining is required.

Objectives

In preparing a custom set of labeled images for RSC, the research team sought to answer a number of questions:

1. How does the camera's view of the road affect image identification?
2. What conditions and associated labels should be used for RSC?
3. What camera images should be used for identifying RSC?
4. What can be done to expedite image labeling?
5. What can be done to improve the accuracy of image labeling?

Methods

Manual Labeling

The research team manually labeled approximately 20,000 images. The majority of images were from North Dakota. Note that the proportion of "snow" images versus "no snow" images can differ by state. For example, there is a significantly smaller proportion of "snow" images versus "no snow" images in North Dakota than in Alaska. As a result, additional "snow" images from Alaska were used to supplement the images from North Dakota to create more robust training and test sets.

Challenges in Classifying Images Using Two Labels – "No Snow" Versus "Snow"

The labels "snow" and "no snow" can be ambiguous and open to interpretation. This is especially true when considering actual impacts to the traveling public and the appropriate responses from road maintenance staff. For example, if there is a light dusting of snow on a highway, strictly speaking that highway could be given the label "snow." In practice, however, the traveling public and DOT maintenance staff would not be overly concerned about driving on a highway that has a light dusting of snow (assuming there is no ice underneath). Similarly, if there is some snow on a highway's shoulders but no snow on its traffic lanes, it may make sense to classify the highway with the label "no snow."

In order to get a better grasp of the "snow"/"no snow" classification problem, it's helpful to look at a time series of images taken at the same location using the same viewing angle. For example consider Figure 1 through Figure 7. Should Figure 2 be labeled "snow"? What about Figure 3? What percentage of the traffic lanes or shoulders have to be covered with snow before giving the highway the label "snow"? Will the "snow" labels prescribe any particular action to be taken by DOT maintenance staff?

Figure 1. I-29 at Grand Forks without snow

Figure 2. I-29 at Grand Forks with minor traces of snow on the shoulders

Figure 3. I-29 at Grand Forks with snow on the shoulders and minor traces of snow in the lanes

Figure 4. I-29 at Grand Forks exhibiting low visibility and problematic snow on both the shoulders and traffic lanes

Figure 5. I-29 at Grand Forks exhibiting snow conditions

Figure 6. I-29 at Grand Forks exhibiting snow primarily on the shoulders but traces on the traffic lanes

Figure 7. I-29 at Grand Forks Exhibiting Some Snow on the Shoulders but No Snow in the Traffic Lanes

The “snow” label by itself should be associated with conditions that prompt a cautionary response. That response to snowy road conditions would entail some of the following driver actions:

1. Being more observant of the road environment and traffic situation
2. Slowing down due to less friction and potentially reduced visibility
3. Increasing the trailing distance
4. Avoiding sudden responses, including abrupt steering and braking changes
5. Maintaining traction when driving on curves and when going uphill or downhill
6. Avoiding the use of cruise control
7. Keeping headlights and taillights on

From this point of view, it is important to establish a set of rules that guide the labeling process so that the above labeling concerns can be directly addressed. The set of labeling rules should have the following characteristics:

1. The rules should be simply stated and easy to understand.
2. The rules should aim at being consistent and unambiguous.
3. The rules should be easy to apply by someone who is quickly reviewing camera images of roads.

In practice, an individual labeling the images at a single location with a given camera angle should not have to spend more than 1 to 2 seconds viewing an image before deciding on a label. In creating the labeling rules, it was decided that it would be helpful to identify images that pose labeling challenges. Establishing the labels for such challenging images should prove beneficial in establishing a consistent labeling policy and be helpful in resolving potential labeling ambiguities.

Ultimately, a successful labeling policy depends on the consensus and support of the DOT maintenance staff.

How Can Visibility Conditions Affect Identifying Snow on the Road?

Snowy RSC can be associated with different visibility conditions. For example, see Figure 17, Figure 27, and Figure 28, which illustrate different visibility conditions. Moderate- or low-visibility conditions can be confused with snow, so the research team wanted to explore this particular issue by performing multiple classifications:

1. First, classify images into high and low/moderate visibility categories.
2. Second, classify high-visibility images into “snow” and “no snow” categories.
3. Third, classify low- and moderate-visibility images into “snow” and “no snow” categories.

Note that low-visibility conditions can obfuscate snowy RSC, as demonstrated in Figure 17, making it extremely difficult to see what is actually happening on the road.

Challenges in Classifying Images Using Multiple Snow Categories

The two categories, “no snow” and “snow,” do not capture the wide variety of snowy conditions that are possible on highways. It is clear that additional categories would provide a more precise picture of the winter weather condition that is impacting the road. For example, it is important to get an idea of the snow amount on the road, such as light, moderate, or heavy. Knowing whether ice or slush is impacting the road is also extremely helpful. One might think that classifying the RSC into a more detailed set of conditions would be easier than using a simplified set of categories. However, if there are challenges labeling “no snow” versus “snow” road conditions, such challenges are compounded when attempting to use a more detailed breakdown of road conditions. Consider the following categories:

- Light snow (up to 1/3 of the road segment covered by snow)
- Moderate snow (1/3 to 2/3 of the road segment covered with snow)
- Heavy snow (over 2/3 of the road segment covered with snow)

- Slush
- Wet (no snow)
- Dry (no snow)

Since there are more categories to use in labeling, there are now more boundary cases to consider. For example, what is the dividing line between light and moderate snow? What is the dividing line between dry and wet roads? When does a wet road transition into a road with slush? Since more decisions have to be made when labeling an image with more categories, the labeling process becomes more challenging and more time-consuming.

In order for the ML to perform well in classifying images using multiple labels, it is also important that there be an adequate number of images per label to train on. For example, for training purposes it is often recommended that the user have, at a minimum, 50 to 100 images per label, and it's advisable to have 1,000 images or more per label (Shahinfar et al. 2020). If there is an inadequate number of images per label, the ML modeling will generally suffer degraded performance.

The set of North Dakota images collected by the research team did not include a sufficient number of images to adequately characterize each of the above categories. As a result, the research team focused on two different categorizations:

1. Snow versus no snow
2. High visibility versus low to moderate visibility

Note that the labels "snow" and "no snow" are mutually exclusive. The labels "snow" and "high visibility" are not mutually exclusive. Classification of a camera image using a set of mutually exclusive categories such as "snow" and "no snow" is called single-label classification. Classification into categories that are not mutually exclusive such as "snow" and "high visibility" is called multi-label classification.

Results and Conclusions

1. How Does the Camera's View of the Road Affect Image Identification?

As a general rule of thumb, if a person looks at a camera image and can easily and accurately identify winter RSC for a well-defined road segment in the image, the better an AI image recognition algorithm will perform on that same well-defined road segment. Image recognition algorithms attempt to mimic the labeling logic performed by humans. The algorithms are looking for patterns in the images that are associated with the labels. In that sense, image recognition algorithms generally do not outperform the individuals who are labeling the images in the training set, since the algorithms are simply mimicking the labeling results presented. If the labeling is uncertain or inconsistent, the algorithms will only mimic that behavior, leading to further uncertainty and inconsistency. From that point of view, it is important to strive for consistent, clear-cut labeling performance from the individuals who are manually reviewing the images in order to optimize the performance of the AI image recognition algorithm.

Figure 8 through Figure 12 show five camera images taken at different viewing angles of I-29 at Grand Forks. Figure 9, Figure 11, and especially Figure 12 have excellent views of the highway. It was generally easier to label images corresponding to these three views than the views shown in Figure 8 and Figure 10. Along these lines, the farther a road segment is off in the distance in an image, the more challenging it is to label RSC for that segment.

Figure 8. I-29 at Grand Forks Camera 01

Figure 9. I-29 at Grand Forks Camera 02

Figure 10. I-29 at Grand Forks Camera 03

Figure 11. I-29 at Grand Forks Camera 04

Figure 12. I-29 at Grand Forks Camera 05

2. What Conditions and Associated Labels Should Be Used for RSC?

As mentioned earlier, it is easier to start out with two labels and progressively add new labels than to start out with a larger number of labels. So, for example, the labels “no snow” and “snow” are a good place to start. There can be a label of “none of the above” for images that are uncertain.

Transportation Canada provides two-class, three-class, and five-class systems for categorizing winter RSC. The two-class RSC consists of classes for bare and snow covered. The three-class RSC consists of classes for bare, partly snow covered, and fully snow covered. The five-class RSC consists of classes for bare, <25% covered, 25% to 50% covered (both wheel paths are clear), 50% to 75% covered (one wheel path is clear), and fully snow covered.

In North Dakota, the following RSC categories are presented to the public on the NDDOT website:

1. Ice/compacted snow
2. Scattered ice
3. Snow covered
4. Scattered snow drifts
5. Frost
6. Scattered frost
7. Wet/slush
8. Scattered wet/slush
9. Seasonal/good
10. Water on/near road

These capture a number of important RSC. In the above list, snow conditions are captured in categories 1, 3, and 4. In order to have success in automatically identifying these conditions, it is important that humans be able to easily distinguish the different conditions when viewing camera images. This requires establishing clear and unambiguous dividing lines between categories in each of the following lists:

1. Compacted snow, snow covered, and scattered snow drifts: For example, what is the dividing line between scattered snow drifts and snow covered? Should the images in Figure 22 or Figure 23 be considered snow covered? What percentage of coverage should be required for a road to be considered snow covered?
2. Ice and scattered ice: Can ice and scattered ice be recognized in the camera images?
3. Frost and scattered frost: How do frost and scattered frost differ from ice and scattered ice?
4. Wet/slush, scattered wet/slush, water on/near road: What is the dividing line between slush and snow?

In establishing the dividing lines between these categories, it is advisable to have a wide variety of images at different sites that staff can use to establish the labeling policy. These training images should expose the edge cases that are difficult to classify. Even after clarifying the dividing lines for these categories, there can still be camera images that belong to multiple categories such as “slush” and “snow covered,” as demonstrated in Figure 13. A DOT wishing to have mutually exclusive categories will need to establish a priority scheme to label such images.

Figure 13. Camera image of bridge in Alaska belonging to "slush" and "snow covered" categories simultaneously

3. What Camera Images Should Be Used for Identifying RSC?

As mentioned above, the camera images used for labeling should have a good view of the road, making it easy for individuals to identify RSC. The road segment of interest should be predominant in the image and not a secondary road segment off in the distance. Images that are spotty or blurred owing to the camera lens being dirty should be avoided (e.g., Figure 14 and Figure 16). Note that spots on the camera lens can produce white spots on the camera images, and these spots can be misinterpreted as snow by the AI image recognition algorithms. Nighttime images and images taken at the transitions from day to night and from night to day should be avoided. Images having glare should be avoided (e.g., Figure 15 and Figure 16). Images taken in low-visibility conditions (e.g., Figure 17) can have a poor view of RSC and should be avoided when establishing an initial RSC identification capability. For some ML implementations, it may be helpful to crop the images and increase the size of the training set by flipping and rotating the images prior to training. These operations can be performed after the initial labeling.

Figure 14. North Dakota road image illustrating a dirty camera lens

Figure 15. North Dakota road image exhibiting glare

Figure 16. North Dakota road image with a dirty camera lens and glare

Figure 17. North Dakota image illustrating low visibility and a poor view of the road

4. What Can Be Done to Expedite Image Labeling?

Image labeling can be expedited by gathering a time series of images at the same location having the same view. In a time series of images, it is easy to review the images using a gallery view, such as the gallery view feature in MacOS Finder (Figure 18).

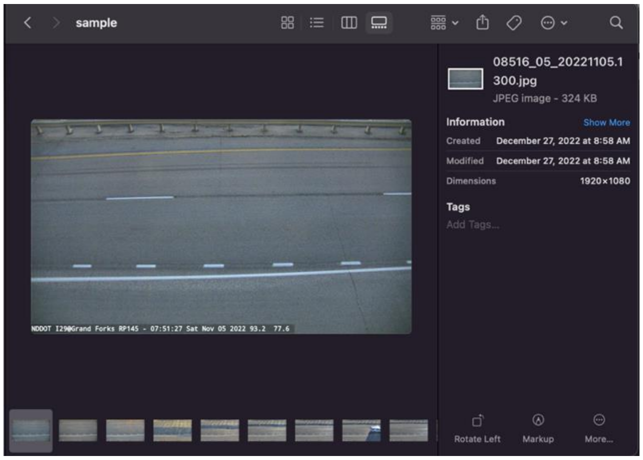


Figure 18. Gallery view highlighting one image in a time series of images at I-29 at Grand Forks

Transitions from "no snow" to "snow" and vice versa can be spotted quickly when fast-forwarding through a time series of images. For example, Figure 19 through Figure 24 illustrate such a transition from "no snow" to "snow" in one of the views of I-29 at Grand Forks.

Figure 19. I-29 at Grand Forks clear of snow

Figure 20. Scattered snow drifts on I-29 at Grand Forks (first image)

Figure 21. Scattered snow drifts on I-29 at Grand Forks (second image)

Figure 22. Scattered snow drifts on I-29 at Grand Forks (third image)

Figure 23. Scattered snow drifts on I-29 at Grand Forks (fourth image)

Figure 24. I-29 at Grand Forks almost totally snow covered

One challenge here is deciding on image criteria that identify conditions requiring an actionable response by maintenance. For example, does Figure 22 indicate RSC that should be addressed by maintenance? In identifying such conditions, it is important to determine whether such conditions can be effectively recognized using ML.

5. What Can Be Done to Improve the Accuracy of Image Labeling?

It is important that criteria are identified that make it easy and natural to label images. For example, in Figure 19 through Figure 24, it is easy and natural to label Figure 19 as “no snow” and Figure 20 through Figure 24 as “snow.” Figure 25 through Figure 28 are challenging to classify because they depict situations that are near the edge of the classification boundary between “snow” and “no snow.” To improve the accuracy of image labeling, it is important to collect a large number of edge cases and then establish a policy on how to classify them. Once such a policy has been formulated, it should then be tested by applying a suitable ML algorithm and evaluating algorithm’s performance. If the ML algorithm performs poorly in correctly identifying edge cases, some tuning of the labeling policy will likely be necessary.

Figure 25. I-29 at Grand Forks “snow” edge case (first example)

Figure 26. I-29 at Grand Forks "snow" edge case (second example)

Figure 27. I-29 at Grand Forks "snow" edge case (third example)

Figure 28. I-29 at Grand Forks "snow" edge case (fourth example)

EVALUATING ML TECHNIQUES THAT IDENTIFY WINTER RSC IN CAMERA IMAGERY

Objective

The main objective for the different labeling experiments performed in this study was to identify ML approaches that can be used effectively in MDSS or other implementations to automatically classify winter RSC primarily using camera imagery of roads. In this regard, an ML approach is not simply characterized by the choice of suitable ML algorithms. Rather, the ML approach involves the image labeling policy, the completeness of the training set in connection with full characterization of the chosen labels, and the choice of the test set.

Methods

ML Algorithm Approaches

The research team chose three different ML approaches for RSC image identification:

1. Open-source CNN algorithms using different layers and connection strategies
2. AWS Rekognition
3. Microsoft Azure Custom Vision

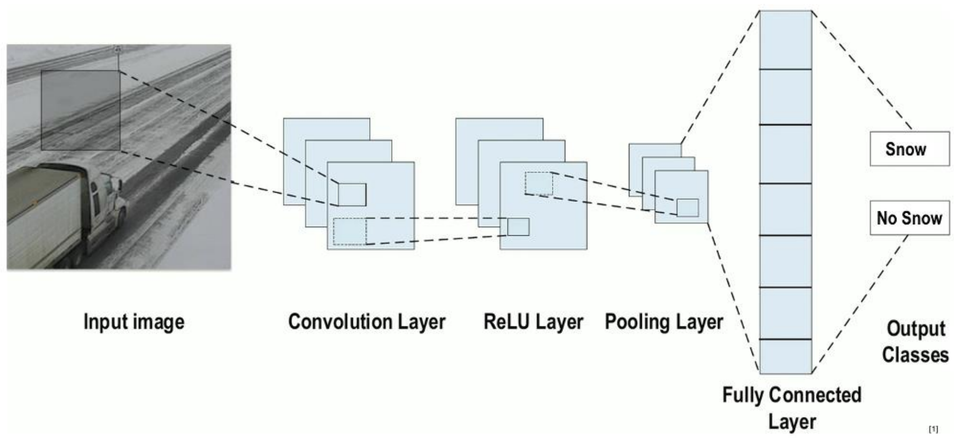
Open-Source CNN Algorithms

The open-source CNN algorithms used in this project were available in PyTorch, an open-source ML software package implemented in Python. PyTorch has support for running CNN models that have been pretrained with extensive image and video datasets. These pretrained models can then be further trained with new images for custom applications. The research team used the following pretrained models:

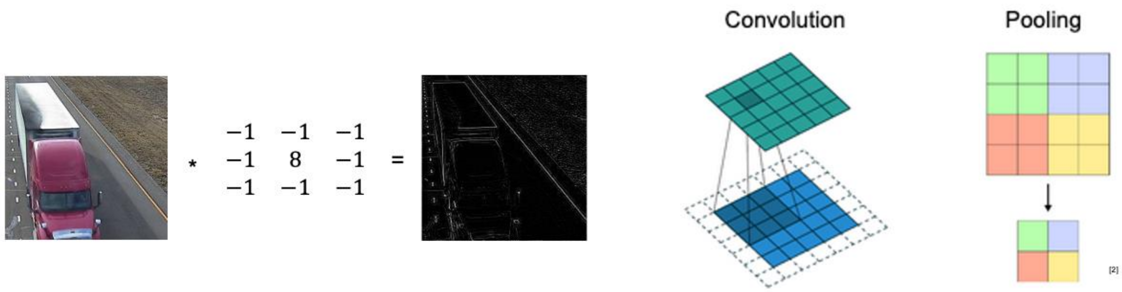
- 101 Layer ResNet (He et al. 2016)
- 152 Layer ResNet (He et al. 2016)
- ResNeXt (Xie et al. 2017)
- EfficientNet v2 (Tan and Le 2021)
- Inception v3 (Szegedy et al. 2016)

The deep neural nets in the models listed above have different numbers of layers and connections between the layers directed toward improving efficiency and accuracy. These models were trained and tested on NVIDIA v100 GPUs at NCAR. Note that most of testing was done on a 101 Layer Resnet. This particular CNN supports the training of deep neural networks with a good tradeoff between runtime and classification performance.

A diagram illustrating some of the CNN processing steps is provided in Figure 29. The original image is represented as one or more grids of numerical values. For instance, a color image has three RGB values per pixel. The convolution layers are produced by multiplying the values of the image grid by small, fixed scalar matrices that are centered at each grid point and then summing the products. The rectified linear unit layer (ReLU) applies a simple nonlinear function *f(x) = max(0, x)* to the previous layer. It maps all negative values to 0 and all positive values back to themselves. The pooling layer downsamples each of the previous layers. The pooling layer typically does this by applying an averaging or maximum operator to a layer's subregions.



Adapted from Alzubaidi et al. 2021 / CC BY 4.0



Adapted from Andreas Maier (Maier et al. 2019) / Wikimedia Commons / CC BY 4.0

Figure 29. Convolutional neural net diagram

AWS Rekognition

Technical details describing the actual ML algorithms used in AWS Rekognition are not generally available. However, AWS Rekognition does make use of deep learning models that have been pretrained on tens of millions of different types of images. Rekognition can leverage the pretraining for particular use cases that involve custom labeling. Rekognition will automatically evaluate different image classification models and will then provide the user with a tuned and optimally trained model.

Microsoft Azure Custom Vision

Azure Custom Vision provides four model types:

1. Standard CNN useful for general image classification
2. Compact CNN for quick training
3. Boosted decision tree (BDT) model
4. Long short-term memory (LSTM) model for predicting the label of the next image in a time series of images

From the four model types, the research team chose the standard CNN model.

Single-Camera Versus Multiple-Camera ML Modeling

The research team performed a set of experiments using single- and multiple-camera ML modeling. Here, the research team was interested in whether ML models trained using camera images from a single camera would outperform models trained using images from multiple cameras. The research team trained five ML models, one for each of five individual cameras located at site 08516 on I-29 at Grand Forks. The research team also trained an ML model using images from all five cameras. Sample images from the five cameras are presented in Figure 8 through Figure 12. Note that the camera images at site 08516 present a variety of different views that avoid focusing on the same road segment. The research team compared the performance of the single-camera ML models with the performance of the multiple-camera ML model.

Results

Precision, Recall, and F1 Metrics for Evaluation

Before discussing the results, it is helpful to review classification evaluation metrics that are typically used in ML known as precision, recall, and F1 scores. In automatically classifying a road segment as “snow” or “no snow,” there are four possibilities:

1. There is snow on the road and the ML model decided that the label for the road is “snow.” This is called a true positive (TP).
2. There is snow on the road and the ML model decided that the label for the road is “no snow.” This is called a false negative (FN).
3. There is no snow on the road and the ML model decided that the label for the road is “snow.” This is called a false positive (FP).
4. There is no snow on the road and the ML model decided that the label for the road is “no snow.” This is called a true negative (TN).

Precision is the fraction of *model-predicted* true values that are correctly classified.

Precision = TP/(TP+FP)

Recall is the fraction of *actual* true values that are correctly classified.

Recall = TP/(TP+FN)

Note that there is a tradeoff between precision and recall. One can increase the precision by reducing the number of false positives. This can be achieved by increasing the certainty threshold for a true classification. For example, a model would have to have over 90% certainty that a road had snow before being willing to classify the road as "snow." But in doing this, the number of false negatives increases because more road segments will be falsely classified as "no snow" and the recall consequently decreases.

There is also a metric called the F1 score that seeks to balance precision and recall. It is defined as the harmonic mean of precision and recall.

F1 = 2 * (precision * recall)/(precision + recall)

Experiments 1 through 5 compare single-camera model results with multiple-camera model results at Cameras 01 through 05 on I-29 at Grand Forks.

Experiment 1. I-29 at Grand Forks Camera 01 (see Figure 8)

Identifier: 08516_01

Classification: Snow versus no snow

Algorithms: Azure Custom Vision, Rekognition

Model Training Sets:

A. Model trained using 08516_03 data
B. Model trained using 08516_01, 02, 03, 04, and 05 data

Model Test Set: All models were tested on the same holdout 08516_01 data set. The timestamps for the data in the 08561_01 test set were chosen to be after the timestamps for the data in the training sets.

Results are shown in Table 1 through Table 4.

Table 1. Rekognition confusion matrix for Model A at Camera 08516_01

	True Snow	**True No Snow**
Predicted Snow	34	7
Predicted No Snow	0	188

Table 2. Custom Vision confusion matrix for Model A at Camera 08516_01

	True Snow	**True No Snow**
Predicted Snow	34	18
Predicted No Snow	0	177

Table 3. Rekognition confusion matrix for Model B at Camera 08516_01

	True Snow	**True No Snow**
Predicted Snow	34	21
Predicted No Snow	0	174

Table 4. Azure Custom Vision and AWS Rekognition scores at Camera 08516_01

	Custom Vision Model A	**Rekognition Model A**	**Rekognition Model B**
Snow Precision	0.65	**0.83**	0.62
Snow Recall	**1.0**	**1.0**	**1.0**
Snow F1	0.79	**0.91**	0.76

Note that the best performing model values are in bold.

Note that the Rekognition single-camera model had the best performance for Camera 08516_01.

Experiment 2. I-29 at Grand Forks Camera 02 (see Figure 9)

Identifier: 08516_02

Classification: Snow versus no snow

Algorithms: Azure Custom Vision, Rekognition

Model Training Sets:

A. Model trained using 08516_02 data
B. Model trained using 08516_01, 02, 03, 04, and 05 data

Model Test Set: All models were tested on the same holdout 08516_02 data set. The timestamps for the data in the 08516_02 test set were chosen to be after the timestamps for the data in the training sets.

Results are shown in Table 5 through Table 8.

Table 5. Rekognition confusion matrix for Model A at Camera 08516_02

	True Snow	**True No Snow**
Predicted Snow	50	0
Predicted No Snow	4	175

Table 6. Custom Vision confusion matrix for Model A at Camera 08516_02

	True Snow	**True No Snow Model A**
Predicted Snow	51	0
Predicted No Snow	3	175

Table 7. Rekognition confusion matrix for Model B at Camera 08516_02

	True Snow	**True No Snow**
Predicted Snow	52	3
Predicted No Snow	2	172

Table 8. Azure Custom Vision and AWS Rekognition scores for Camera 08516_02

	Custom Vision Model A	**Rekognition Model A**	**Rekognition Model B**
Snow Precision	**1.0**	**1.0**	0.95
Snow Recall	0.94	0.93	**0.96**
Snow F1	**0.97**	0.96	0.95

Note that the best performing model values are in bold.

Note that the single- and multiple-camera models had similar excellent performance at Camera 08516_02.

Experiment 3. I-29 at Grand Forks Camera 03 (see Figure 10)

Identifier: 08516_03

Classification: Snow versus no snow

Algorithms: Azure Custom Vision, Rekognition

Model Training Sets:

A. Model trained using 08516_03 data
B. Model trained using 08516_01, 02, 03, 04, and 05 data

Model Test Set: All models were tested on the same holdout 08516_03 data set. The timestamps for the data in the 08516_03 test set were chosen to be after the timestamps for the data in the training sets.

Results are shown in Table 9 through Table 12.

Table 9. Rekognition confusion matrix for Model A at Camera 08516_03

	True Snow	True No Snow
Predicted Snow	34	24
Predicted No Snow	0	171

Table 10. Custom Vision confusion matrix for Model A at Camera 08516_03

	True Snow	True No Snow
Predicted Snow	34	9
Predicted No Snow	0	186

Table 11. Rekognition confusion matrix for Model B at Camera 08516_03

	True Snow	True No Snow
Predicted Snow	34	12
Predicted No Snow	0	183

Table 12. Azure Custom Vision and AWS Rekognition scores for Camera 08516_03

	Custom Vision A	Rekognition A	Rekognition B
Snow Precision	**0.79**	0.59	0.74
Snow Recall	**1.0**	**1.0**	**1.0**
Snow F1	**0.88**	0.74	0.85

Note that the best performing model values are in bold.

Note that the Custom Vision single-camera model and the Rekognition multiple-camera model had similar excellent performance at Camera 08516_03.

Experiment 4. I-29 at Grand Forks Camera 04 (see Figure 11)

Identifier: 08516_04

Classification: Snow versus no snow

Algorithms: Azure Custom Vision, Rekognition

Model Training Sets:

A. Model trained using 08516_04 data
B. Model trained using 08516_01, 02, 03, 04, and 05 data

Model Test Set: All models were tested on the same holdout 08516_04 data set. The timestamps for the data in the 08516_04 test set were chosen to be after the timestamps for the data in the training sets.

Results are shown in Table 13 through Table 16.

Table 13. Rekognition confusion matrix for Model A at Camera 08516_04

	True Snow	**True No Snow**
Predicted Snow	34	15
Predicted No Snow	0	180

Table 14. Custom Vision results for Model A at Camera 08516_04

	True Snow	**True No Snow**
Predicted Snow	34	11
Predicted No Snow	0	184

Table 15. Rekognition results for Model B at Camera 08516_04

	True Snow	**True No Snow**
Predicted Snow	34	6
Predicted No Snow	0	189

Table 16. Azure Custom Vision and AWS Rekognition scores for Camera 08516_04

	Custom Vision A	Rekognition A	Rekognition B
Snow Precision	0.76	0.69	**0.85**
Snow Recall	**1.0**	**1.0**	**1.0**
Snow F1	0.86	0.69	**0.92**

Note that the best performing model values are in bold.

Note that the Rekognition multiple-camera model had the best performance at Camera 08516_04. The Custom Vision single-camera model had slightly worse performance.

Experiment 5. I-29 at Grand Forks Camera 05 (see Figure 12)

Identifier: 08516_05

Classification: Snow versus no snow

Algorithms: Azure Custom Vision, Rekognition

Model Training Sets:

A. Model trained using 08516_05 data
B. Model trained using 08516_01, 02, 03, 04, and 05 data

Model Test Set: All models were tested on the same holdout 08516_05 data set. The timestamps for the data in the 08516_05 test set were chosen to be after the timestamps for the data in the training sets.

Results are shown in Table 17 through Table 20.

Table 17. Rekognition confusion matrix for Model A at Camera 08516_05

	True Snow	True No Snow
Predicted Snow	47	1
Predicted No Snow	0	181

Table 18. Custom Vision results for Model A at Camera 08516_05

	True Snow	True No Snow
Predicted Snow	46	0
Predicted No Snow	1	182

Table 19. Rekognition results for Model B at Camera 08516_05

	True Snow	True No Snow
Predicted Snow	47	1
Predicted No Snow	0	181

Table 20. Azure Custom Vision and AWS Rekognition scores for Camera 08516_05

	Custom Vision A	Rekognition A	Rekognition B
Snow Precision	**1.0**	0.98	0.98
Snow Recall	0.98	**1.0**	**1.0**
Snow F1	**0.99**	0.98	0.98

Note that the best performing model values are in bold.

Note that the single- and multiple-camera models had similar excellent performance at Camera 08516_05.

Table 21 summarizes the results for Experiments 1 through 5.

Table 21. Azure Custom Vision and AWS Rekognition average scores for Experiments 1 through 5

	Custom Vision A	Rekognition A	Rekognition B
Snow Precision	**0.84**	0.82	0.83
Snow Recall	0.98	**0.99**	**0.99**
Snow F1	**0.90**	0.86	0.89

Experiment 6. Multiple-Camera Training and Testing at RWIS 08516 on I-29 at Grand Forks Using All 08516 Camera Images

This experiment uses the union of training sets for Cameras 08516_01 through 08516_05 as the training set. It uses the union of the test sets for Cameras 08516_01 through 08516_05 as the test set. The performance of this experiment can be compared to the performance of the previous five experiments.

Identifier: 08516_05

Classification: Snow versus no snow

Algorithms: Azure Custom Vision, Rekognition

Multiple Road Segment ML Model

Results are shown in Table 22 and Table 23.

Table 22. Rekognition confusion matrix for all 08516 cameras

	True Snow	**True No Snow**
Predicted Snow	201	43
Predicted No Snow	2	899

Table 23. AWS Rekognition scores for all 08516 cameras

	Rekognition
Snow Precision	**0.82**
Snow Recall	**0.99**
Snow F1	**0.90**

Note that the multiple-camera model at 08516 has good performance when tested with all the 08516 cameras.

Experiment 7. Bridge in Alaska

This experiment uses camera images of a bridge in Alaska.

Identifier: 000351053-00-03

Classification: Snow versus no snow

Algorithms: Azure Custom Vision, Rekognition

Single Road Segment ML Model

Results are shown in Table 24 and Table 25.

Table 24. Rekognition confusion matrix for Camera 000351053-00-03

	True Snow	**True No Snow**
Predicted Snow	95	0
Predicted No Snow	6	143

Table 25. AWS Rekognition scores for Camera 000351053-00-03

	Rekognition
Snow Precision	**1.0**
Snow Recall	**0.94**
Snow F1	**0.97**

Note that the single-camera model for Camera 000351053-00-03 has excellent performance.

Experiment 8. Performing a Site Holdout Test

Experiment 8 uses camera images from a number of different sites in both Alaska and North Dakota for training and holds out camera images from site 08516 in North Dakota for testing. This experiment was designed to show how well a model trained on images from multiple locations would perform at a holdout location. In this way, it shows how a model may perform at a new site without specifically training at that site.

All the images in this particular experiment were cropped. The number of images was augmented by flipping the images. These preprocessing steps were performed to help the performance of the 101 Layer Resnet CNN algorithm. Sample images are shown in Figure 30 through Figure 32. The camera lenses were not clean for some of the images, as can be seen in the image in Figure 32. The test set consists of images from site 08516 on I-29 at Grand Forks in North Dakota. None of the 08516 images were used in training. Note that no attempt was made to make the views of the road segments in the camera images in the training set similar to the views of the road segments in the holdout test set.

Figure 30. Sample "no snow" image for training in the Holdout Test

Figure 31. Sample "snow" image for training in the Holdout Test

Figure 32. Image from North Dakota where the camera lens had spots

Identifier: HoldoutTest

Classification: Snow versus no snow

Algorithms: Azure Custom Vision, Rekognition

Multiple Road Segment ML Model

Results are shown in Table 26 through Table 28.

Table 26. Rekognition confusion matrix for HoldoutTest

	True Snow	True No Snow
Predicted Snow	369	72
Predicted No Snow	169	1729

Table 27. Custom Vision confusion matrix for HoldoutTest

	True Snow	**True No Snow**
Predicted Snow	447	235
Predicted No Snow	91	1566

Table 28. Azure Custom Vision, AWS Rekognition, and 101 Layer Resnet scores for 08516 HoldoutTest

	Custom Vision	**Rekognition**	**101 Layer Resnet**
Snow Precision	0.66	**0.84**	0.65
Snow Recall	**0.83**	0.69	0.77
Snow F1	0.73	**0.75**	0.70

Note that the best performing model values are in bold.

Experiment 9. Creating an ML Model for Visibility

Experiment 9 was performed to investigate whether visibility conditions could be accurately classified as a precursor to performing the RSC classification. The research team utilized approximately 6,000 winter images labeled with high-, moderate-, and low-visibility conditions in North Dakota in tuning a 101 Layer Resnet ML model. Note that this model used a random 80/20 training/test set split for testing. To obtain scores that are more likely in practice, the test set should contain images that are separated spatially or temporally from the training set. Still, the scores in Table 29 and Table 30 illustrate a strong capability to correctly categorize pavement conditions in good visibility.

Table 29. Confusion matrix for visibility experiment

	High Visibility	**Moderate Visibility**	**Low Visibility**
Predicted High Visibility	4455	41	1
Predicted Moderate Visibility	36	637	42
Predicted Low Visibility	0	37	564

Table 30. Visibility classification scores

	High Visibility	**Moderate Visibility**	**Low Visibility**
Precision	0.99	0.89	0.94
Recall	0.99	0.89	0.93
F1	0.99	0.89	0.93

Experiment 10. Combining High Visibility and Winter RSC

Experiment 10 was performed to investigate whether high-visibility conditions combined with snowy RSC would lead to changes in the performance of the ML classification model. The research team utilized approximately 4,500 winter images labeled “no snow,” “snow,” and “wet” conditions in North Dakota in tuning a 101 Layer Resnet ML model. Note that this model used a random 80/20 training/test set split for testing. To obtain scores that are more likely in practice, the test set should contain images that are separated spatially or temporally from the training set. Still, the scores in Table 31 and Table 32 illustrate a strong capability to correctly categorize visibility conditions.

Table 31. Confusion matrix for high visibility and RSC

	No Snow	**Snow**	**Wet**
Predicted No Snow	4243	1	4
Predicted Snow	4	184	0
Predicted Wet	6	0	102

Table 32. High visibility and RSC classification scores

	No Snow	**Snow**	**Wet**
Precision	0.99	0.98	0.94
Recall	0.99	0.99	0.96
F1	0.99	0.99	0.95

Experiment 11. Combining Low/Moderate Visibility and Winter RSC

Experiment 11 was performed to investigate whether low- and moderate-visibility conditions combined with snowy RSC would lead to changes in the performance of the ML classification model. The research team utilized approximately 1,200 winter images labeled “no snow,” “snow,” and “wet” conditions in North Dakota in tuning a 101 Layer Resnet ML model. Note that this model used a random 80/20 training/test set split for testing. To obtain scores that are more likely in practice, the test set should contain images that are separated spatially or temporally from the training set. Still, the scores in Table 33 and Table 34 illustrate a strong capability to correctly categorize pavement conditions in good visibility.

Table 33. Confusion matrix for low/moderate visibility and RSC

	No Snow	**Snow**	**Wet**
Predicted No Snow	722	4	3
Predicted Snow	13	311	0
Predicted Wet	10	1	206

Table 34. High visibility and RSC classification scores

	No Snow	**Snow**	**Wet**
Precision	**0.99**	0.96	0.95
Recall	0.97	0.98	**0.99**
F1	**0.98**	0.97	0.97

Note that the best performing model values are in bold.

Conclusions

The research team performed a number of experiments on the image data from site 08516 on I-29 at Grand Forks in North Dakota using both single-camera and multiple-camera ML models.

Table 21 presents the average precision, recall, and F1 scores for Experiments 1 through 5. On viewing the average F1 scores in Table 21, it is apparent that the performance of Azure Custom Vision single-camera Model A is slightly better than the performance of AWS Rekognition single-camera Model A on this particular data set. It is also apparent that the average performance of Azure Custom Vision single-camera Model A is comparable to that of Rekognition multiple-camera Model B on the test image data set.

Experiment 6 provides additional results that support the fact that Rekognition multiple-camera Model B has similar performance to the average performance of Azure Custom Vision single-camera Model A. Experiment 7 shows that the same ML single-camera ML modeling approach can be applied to camera images from a site in another state, the state of Alaska. Experiment 8 was designed to illustrate the performance of models at a holdout site. Experiment 8 suggests that a model developed using a well-cultivated training set can be used on new sites and still retain predictive capability. Still, if one wishes to optimize performance at specific sites within a DOT’s domain, it would be beneficial to train on camera images at those specific sites. Experiment 9 suggests that the CNN ML techniques can be useful in identifying visibility conditions. Further study must be performed to determine whether prescreening for visibility leads to substantive improvement in RSC classification performance in Experiments 10 and 11.

RESULTS AND DISCUSSION

What Role Does Image Labeling Have in RSC Identification Using ML?

Image labeling has a significant effect on the performance of ML algorithms. The research team initially formulated a detailed labeling scheme having numerous categories. Owing to a lack of an adequate number of images to delineate the various categories and clear guidelines for classifying images into the different categories, the initial performance of the ML algorithms was quite poor. The ML algorithm performance improved considerably as the categories were consolidated into "snow" versus "no snow" categories and as the guidelines for distinguishing the categories were made clearer.

How Should Camera Images Having Glare, Dirty Camera Lens Spots, or Low Lighting Be Handled?

Numerous road camera images have glare, camera lens spots, or low lighting issues. One way to handle such images is to actively attempt to classify images having these problems. For example, images can be classified into two sets, those having glare and those that do not. An ML model can then be trained on glare and can then be used to identify glare. Initially, images found to have glare or other common problems can be excluded from further RSC identification. There may be certain cases where images with minor problems could still be used for further classification of RSC.

Was There a Significant Difference in Performance in the ML Algorithms Evaluated?

The research team did not find a significant difference in performance among the three ML algorithms that were evaluated. The experiments, however, were not exhaustive, so performance differences could be exposed by further study.

Key Findings

- A clear view of the road in a camera image is important for labeling accuracy and ML performance.
- Having a sufficient number of accurately labeled images for each RSC category is of key importance for successful ML.
- Dirty camera lenses, glare, darkness, low visibility, and obstructions all have a negative effect on identifying winter RSC.
- Using a time series of camera images at a single site is an effective way to quickly and accurately label RSC.
- It is important to establish a set of clear, unambiguous, and consistent rules for performing the RSC labeling.
- Cloud vendor image classification technology exhibited similar performance to open-source image classification technology and was easier to apply.

- Single-camera ML and multiple-camera ML exhibited similar performance when tested on temporally segregated image data from cameras included in the training set.
- Multiple-camera ML exhibited poorer performance when evaluated using images from a site not included in the training set.

RECOMMENDATIONS

Can ML Technology Available from Cloud Vendors Be Used Effectively for Automatic Classification of RSC?

Based on the research team's findings, cloud technology such as Microsoft Azure Custom Vision and AWS Rekognition can be used effectively for automatic classification of RSC. ML performance is dependent on clear, unambiguous, and consistent image labeling. To improve the outcome of the ML training, it is important that the road segment have a predominant position in the image and not be off in the distance. Also, it is important to have an adequate number of images that cover the winter weather conditions of interest.

Is It Preferable, in Terms of Accuracy and/or Cost, to Use Open-Source Software for Performing the Automatic Classification of RSC?

The research team performed RSC classification using a number of CNN algorithms in the open-source PyTorch package and found that the accuracy of the PyTorch CNN results was comparable to that of the results produced by Azure Custom Vision or AWS Rekognition. The Azure Custom Vision and AWS Rekognition technologies were easier to use for performing the training, testing, and online recognition. These technologies also provided APIs to support the construction of fully automated solutions. Some limited software engineering support is needed to integrate either Azure Custom Vision or AWS Rekognition capabilities into an MDSS or custom solution. In contrast, more significant software engineering support is necessary to create a custom automatic classification solution in the case of the PyTorch package.

The research team did not have the time to perform a complete evaluation of the costs of the different ML approaches. Generally speaking, the cost for the different cloud-based ML offerings seemed reasonable for this application.

What Are the Steps for Incorporating Automatic Identification of RSC Using Camera Imagery in a State DOT Maintenance Decision Support System or Website?

The following is a series of steps that can be taken to incorporate automatic classification of camera images into an MDSS or other system for the identification of RSC:

1. Decide on a set of high-priority camera sites where automatic classification will have clear benefits. These may be sites where snow impacts are particularly problematic and/or require timely maintenance responses.

2. Archive camera images at the camera sites of interest for at least one complete winter season (two or more winter seasons would be preferable). The set of seasons for image archiving could be expanded to non-winter seasons if the DOT is interested in non-winter RSC such as rainy road conditions, water on the road, flooding, low visibility, etc. Note that cloud storage can potentially be used for camera image archiving.

3. While archiving the camera images, perform spot checks on a regular basis to ensure that the camera images do not have critical quality control problems. Ensure that the spot checks adequately cover the image archive set.

4. Formulate a set of rules for labeling the different categories and ensure that the labeling rules support clear, unambiguous, and consistent image labeling. The labeling procedure should take advantage of the image time series labeling discussed in the chapter above on labeling images.

5. Test the performance of the labeling procedure using cloud-based image classification technology such as Azure Custom Vision, AWS Rekognition, or Google Vision AI. This involves formulating image training and test sets at single camera sites. Run the ML training and then test the performance of the ML model on the test sets. Iterate, refining the labeling procedure until the performance of the ML modeling on the test sets is adequate.

6. Formulate combined training and test sets covering multiple camera sites. Generate one or more ML models for performing the RSC image classification. Test the ML models on holdout images that are either temporally or spatially separate from the training images.

7. Integrate the ML models into the appropriate MDSS or website.

Can an RSC Image Classification Model Developed for One State Be Successfully Applied in a Different State?

If the RSC image classification model was carefully and thoroughly developed for one state using steps similar to those detailed above, it could potentially be successfully applied in a different state. Image labeling would still be required using camera images from the new state to evaluate performance. The labeling rules used in formulating the model for the original state would have to be adhered to in the new state. Thus, it is important that the labeling rules associated with the given model be reviewed to determine whether they are a good fit for the new state. It would then be advisable to perform tests to ensure that the original model exhibits adequate performance on the camera images from the new state.

FUTURE WORK AND RESEARCH NEEDS

Which Winter RSC Categories Are Useful for Supporting Road Maintenance Operations?

Ideally, it would be beneficial to know the precise accumulation of snow on the road, the amount of ice deposition, the lanes that are impacted, the road friction in the lanes, etc. Beyond the basic categories of snow versus no snow, what additional categories are reasonable to include in ML? It is important to clarify how additional categories will benefit DOT maintenance personnel.

How Can Automatic RSC Image Classification Be Used Effectively to Provide Road Treatment Guidance?

If cameras have been placed strategically along a plow route, automatic RSC image classification can potentially support the automatic diagnosis of locations on the route requiring treatment. Here it is important to identify the dividing line between winter RSC requiring treatment and RSC that do not require treatment.

How Can RSC Image Classification Be Integrated into MDSS Nowcasting?

MDSS short-term nowcast alerts can be improved through the integration of accurate RSC classifications. Such integration reduces the chance that MDSS nowcasts are contradicted by recent camera images.

How Can Automatic RSC Image Classification Improve General Driver Situational Awareness?

Road camera images contain information regarding snow and ice, rain, visibility, lighting conditions, traffic problems, etc. Finding effective ways to identify these conditions and then convey suitable information to drivers is an area of further research.

How Can RSC Image Identification, RWIS, and Speed Data Be Used to Categorize Road Friction and Establish Virtual Speed Limits?

There are potential connections between the winter RSC depicted in images and road friction. The research team noted that the average recorded speed connected with “snow” conditions was 5 mph lower than the average speed connected with “no snow” conditions. It would be interesting to study these connections more thoroughly and bring in road friction measurements as part of the study.

Can Classifying Visibility Conditions Prior to Classifying RSC Impact RSC Classification Performance?

The research team did not have adequate time to resolve this question. It would be interesting to determine whether prescreening the camera images into visibility categories would have an impact on classifying winter RSC. One challenge in performing this test is to collect a sufficient number of low- and moderate-visibility images during the winter season.

Can Additional Data such as RWIS Sensor Data or SVO Data Have a Positive Impact on RSC Classification Performance?

The research team did not have an adequate amount of labeled data to resolve this question. This question can be investigated after gaining further experience with the performance of ML classification models based on camera images.

REFERENCES

Alzubaidi, L., J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie, and L. Farhan. 2021. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, Vol. 8, No. 53.

He, K., X. Zhang, S. Ren, and J. Sun. 2016. Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV*, pp. 770–778.

Maier, A., C. Syben, T. Lasser, and C. Riess. 2019. A gentle introduction to deep learning in medical image processing. *Zeitschrift für Medizinische Physik*, Vol. 29, No. 2, pp. 86–101.

Pan, G., L. Fu, R. Yu, and M. Muresan. 2019. Evaluation of alternative pre-trained convolutional neural networks for winter road surface condition monitoring. *5th International Conference on Transportation Information and Safety (ICTIS)*, pp. 614–620.

Shahinfar, S., P. Meek, and G. Falzon. 2020. How many images do I need? Understanding how sample size per class affects deep learning model performance metrics for balanced designs in autonomous wildlife monitoring. *Ecological Informatics*, Vol. 57.

Szegedy, C., V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. 2016. Rethinking the inception architecture for computer vision. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV*, pp. 2818–2826.

Tan, M., and Q. Le. 2021. Efficientnetv2: Smaller models and faster training. *Proceedings of the 38th International Conference on Machine Learning*, pp. 10096–10106.

Xie, S., R. Girshick, P. Dollár, Z. Tu, and K. He. 2017. Aggregated residual transformations for deep neural networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI*, pp. 1492–1500.

Zhang, C., E. Nateghinia, L. Miranda-Moreno, and L. Sun. 2021. Winter road surface condition classification using convolutional neural network (CNN): Visible light and thermal images fusion. *Canadian Journal of Civil Engineering*, Vol. 49, No. 4, pp. 569–578.