Optimal RWIS Sensor Density and Location – Phase IV

http://aurora-program.org

Aurora Project 2021-05

Final Report
November 2024

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.

Iowa State University Nondiscrimination Statement

Disclaimer Notice

Quality Assurance Statement

Iowa DOT Statements

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.	
Aurora Project 2021-05			
4. Title and Subtitle		**5. Report Date**	
Optimal RWIS Sensor Density and Location – Phase IV		November 2024	
		6. Performing Organization Code	
7. Author(s)		**8. Performing Organization Report No.**	
Mingjian Wu, Tae J. Kwon, Michael Urbiztondo, and Yong Lee			
9. Performing Organization Name and Address		**10. Work Unit No. (TRAIS)**	
Department of Civil & Environmental Engineering			
University of Alberta			
9211-116 Street NW		**11. Contract or Grant No.**	
Edmonton, Alberta, Canada T6G 1H9			
12. Sponsoring Organization Name and Address		**13. Type of Report and Period Covered**	
Aurora Program	Federal Highway Administration	Final Report	
Iowa Department of Transportation	U.S. Department of Transportation	**14. Sponsoring Agency Code**	
800 Lincoln Way	1200 New Jersey Avenue, SE	Part of TPF-5(435) and Federal SPR Part II, CFDA 20.205	
Ames, Iowa 50010	Washington, DC 20590		

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

16. Abstract

Road weather information systems (RWIS), in both stationary and mobile forms, have become increasingly popular in recent decades for their ability to collect and disseminate road weather and surface data. In addition to meteorological measurements, highway agencies rely heavily on RWIS imagery data to guide winter road maintenance (WRM) operations. However, the analysis of imagery data is still performed manually by trained personnel. Moreover, the limited number of stationary RWIS stations and the infrequent deployment of mobile RWIS units result in significant spatial gaps along the highway network. In our previous project, we developed methodologies based on convolutional neural networks (CNNs) to automatically recognize road surface conditions (RSC) from dash camera imagery and employed regression kriging (RK) to estimate RSC in unmonitored areas using limited point measurements. These methods demonstrated feasibility and robustness in real-world case studies. Building on these efforts, this project aimed to further advance CNN development specifically for stationary RWIS imagery and assess its reliability using explainable artificial intelligence (XAI) techniques, including SHapley Additive exPlanations (SHAP) and class activation map (CAM)-based methods. Additionally, to automatically estimate snow coverage ratios from stationary RWIS imagery, two distinct deep learning-based computer vision techniques, pix-to-pix generative adversarial network and semantic segmentation, were employed. Furthermore, the RK method was revisited to better accommodate a wider range of weather events while considering their variability, and the potential monetary benefits of this approach were also explored. To address the limitations of RK in handling categorical variables, a novel geostatistical method, namely nested indicator kriging (NIK), was developed to interpolate RSC in unmonitored areas directly using CNN classification results. These methods were evaluated using data from two major highways, Interstate 35 and Interstate 80 in Iowa, US, spanning a five-year period and encompassing over 20,000 images. The results demonstrated high accuracy and reliability. Additionally, a web application was developed to integrate these methods, offering real-time monitoring, estimation, and historical data archiving. This project equips decision-makers with a powerful tool to implement WRM activities more swiftly, efficiently, and cost-effectively, ultimately promoting a safer, more mobile, and sustainable winter transportation system.

17. Key Words	**18. Distribution Statement**		
big data—computer vision—deep learning—explainable artificial intelligence—geostatistics—machine learning—road surface conditions—road weather information systems—weather events	No restrictions.		
19. Security Classification (of this report)	**20. Security Classification (of this page)**	**21. No. of Pages**	**22. Price**
Unclassified.	Unclassified.	93	NA

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

OPTIMAL RWIS SENSOR DENSITY AND LOCATION – PHASE IV

Final Report
November 2024

Principal Investigator
Tae J. Kwon, Associate Professor
University of Alberta

Authors
Mingjian Wu, Tae J. Kwon, Michael Urbiztondo, and Yong Lee

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

Work presented in this report has been published or is under review for publication in various journals and conferences in the area of transportation engineering.

Peer-reviewed journal publications (published):

Wu, M., and T. J. Kwon. 2024. Weather event characterization: a catalyst for improved spatial mapping and benefit quantification in winter road maintenance. *Cold Regions Science and Technology*, Vol. 223, Article 104208.

Wu, M., T. J. Kwon, and N. Huynh. 2022. Winter road surface condition recognition using semantic segmentation and the generative adversarial network: A case study of Iowa, USA. *Transportation Research Record*, Vol. 2678, No. 5, pp. 184–195.

Wu, M., T. J. Kwon, and L. Fu. 2022. Spatial mapping of winter road surface conditions via hybrid geostatistical techniques. *Journal of Cold Regions Engineering*, Vol. 36, No. 4.

Wu, M., and T. J. Kwon. 2022. An automatic architecture designing approach of convolutional neural networks for road surface conditions image recognition: Tradeoff between accuracy and efficiency. *Journal of Sensors*, 2022, pp. 1–11.

Book chapter (published):

Wu, M., T. J. Kwon, L. Fu, and A. H. Wong. 2022. Advances in sustainable winter road maintenance and management for future smart cities. *The Rise of Smart Cities: Advanced Structural Sensing and Monitoring Systems*. Butterworth-Heinemann, Oxford, UK. pp. 625–659.

EXECUTIVE SUMMARY

Monitoring road surface conditions (RSC) during the winter season is crucial for transportation agencies responsible for conducting winter road maintenance (WRM) operations. This becomes particularly evident during winter weather events, such as snowstorms, where timely and effective snow clearing is essential to maintaining road safety and mobility. Accurate RSC information is vital not only for determining when and where WRM operations are needed but also for evaluating the effectiveness of various treatment methods and contractors.

This project builds on our previous work, which was aimed at enhancing the application of road weather information systems (RWIS) to improve WRM operations. RWIS, available in both stationary and mobile forms, have gained popularity due to their ability to collect and disseminate critical road weather and surface data. However, highway agencies still rely on manual analysis of RWIS imagery by trained personnel, a process that is time-consuming and limited by the sparse distribution of RWIS stations and the infrequent deployment of mobile units. These limitations result in significant gaps in monitoring RSC across extensive highway networks.

To address these challenges, earlier research developed convolutional neural networks (CNNs) to automatically classify RSC from dash camera imagery (or mobile imagery) and employed regression kriging (RK) to estimate RSC in unmonitored areas using limited point measurements. These methods were validated through real-world case studies, demonstrating their feasibility and robustness.

In this project, several key advancements were made. The developed CNN models were tailored for stationary RWIS imagery, incorporating explainable artificial intelligence (XAI) techniques such as SHapley Additive exPlanations (SHAP) and class activation map (CAM)-based methods, including Grad-CAM, Grad-CAM++, and Score-CAM. As a result, these methods ensured the reliability and robustness of CNNs for both mobile and stationary imagery, enhancing the transparency and trustworthiness of the CNN models in real-world applications.

Additionally, two distinct advanced deep learning (DL)-based computer vision techniques, namely, pix-to-pix generative adversarial networks (pix2pix GAN) and semantic segmentation (SS), were employed to automate the estimation of snow coverage ratios (SCR) from stationary RWIS imagery. These methods provided a more detailed understanding of snow coverage, thereby improving the precision of WRM decisions.

The RK method was also revisited to better accommodate the variability of weather events by integrating a machine learning-based clustering algorithm. This improvement ensured that RK could accurately model the spatial variability of RSC across different weather scenarios, providing more reliable predictions.

Furthermore, a novel geostatistical method, nested indicator kriging (NIK), was developed to handle the categorical nature of RSC data. This method uses CNN classification results to

interpolate conditions in unmonitored areas, addressing gaps in RSC monitoring across highway networks.

These methods were rigorously tested using a comprehensive dataset from two major highways in Iowa, I-35 and I-80, spanning five years and encompassing more than 20,000 images. The results demonstrated high accuracy and reliability. The newly developed CNNs for RWIS imagery achieved an accuracy of up to 98.5%, with the SHAP and CAM-based methods confirming the trustworthiness of CNNs for automatic RSC recognition from both stationary and mobile RWIS imagery. The pix2pix GAN and SS techniques proved capable of estimating SCR by translating winter images to their summer counterparts and detecting drivable areas, achieving up to 99.3% accuracy in detecting drivable areas and a final accuracy for SCR estimation exceeding 93%.

The revisited RK method, enhanced with a K-means clustering algorithm, provided robust RSC estimates across varied weather scenarios. Using road surface temperatures (RST) as a target variable, the RST estimation discrepancies were as low as 0.619°C for the entire study area. The findings also suggest that potential cost savings include the conservation of up to 10 RWIS stations, which can be further translated into reduced RWIS capital costs, fewer collisions, enhanced traffic mobility, and savings in maintenance materials.

The novel NIK method successfully interpolated categorical RSC variables using CNN classification results, achieving accuracy rates ranging from 58.8% to 85.7%, with an average accuracy of 67.5% across the tested dataset.

A web application was also developed to integrate the key methods as mentioned above, providing real-time monitoring, estimation, and historical data archiving for WRM operations. The advancements made in this project equip WRM decision-makers with powerful tools to implement maintenance activities more swiftly, efficiently, and cost-effectively. These innovations are expected to contribute to a safer, more mobile, and sustainable winter transportation system by mitigating the risks and costs associated with adverse winter weather conditions.

LIST OF ABBREVIATIONS AND ACRONYMS

ANN	artificial neural network
ASE	average standard error
AT	air temperature
AVL	automated vehicle location
CAM	classification activation map
CIC	channel-wise increase in confidence
DEM	digital elevation model
DL	deep learning
DNN	deep neural networks
DOT	department of transportation
ESS	environmental sensor stations
FCN	fully convolutional network
FHWA	Federal Highway Administration
FNR	false negative rate
FPR	false positive rate
GAN	generative adversarial networks
GLS	generalized least squares
GPS	global positioning system
IK	indicator kriging
ITS	intelligent transportation systems
MAE	mean absolute error
ML	machine learning
MLR	multiple linear regression

OK	ordinary kriging
PASV	percentage of absolute Shapley value
RF	random forest
RGB	red-green-blue
RK	regression kriging
RMSE	root mean square error
RSC	road surface condition
RSI	road surface index
RST	road surface temperature
RWIS	road weather information system
SCR	snow coverage ratio
SHA	state highway agency
SHAP	SHapley Additive exPlanations
SK	simple kriging
SS	semantic segmentation
SV	Shapley value
SVM	support vector machines
TM	thermal mapping
VQA	visual question answering
WRM	winter road maintenance

1 INTRODUCTION

1.1 Background

In countries with severe winter seasons, transportation agencies often face significant challenges in meeting the safety and mobility needs of road users. Inclement weather events, such as frequent snow, sleet, ice, and frost, result in high variance in road surface conditions (RSC) over space and time (Feng 2013, Kwon et al. 2013, Wang et al. 2018b, Wu et al. 2022b). Poor driving conditions as a result of inclement weather is something that municipalities must contend with, as it is a safety risk for road users. If proper precautions are not taken by municipalities, winter collisions become a constant problem that threatens the lives of motorists.

Given the extensive record of winter incidents, transportation agencies strive to ensure that their roads are in the best condition possible by mobilizing various winter road maintenance (WRM) activities. These activities include plowing, deicing, and sanding operations, which mitigate the chances and severity of collisions by mechanically removing or preventing snow and ice accumulation, or by increasing traction with abrasives. However, due to the vast spatial distances covered by all highway networks and the uncertain nature of weather events it is often hard to monitor and predict the highly variable RSCs, making both public travel and WRM extremely challenging. To prevent or at least reduce the number of weather-related incidents, transportation agencies need to make timely and well-informed decisions on their various WRM activities. These operations are expensive, as they require a great deal of equipment, personnel, and finances to undertake. It is estimated that more than US $2.3 billion is spent annually on WRM by North American transportation authorities, with Canada spending CA $1 billion alone (Transport Association of Canada 2003, Usman et al. 2010). With real-time or near-future RSC information being made available, transportation authorities will be able to make informed decisions that not only improve the safety of roads but will also help find ways to minimize operational costs while maintaining a high level of service.

To help support and facilitate WRM decisions, an advanced intelligent transportation systems (ITS) monitoring technology known as road weather information systems (RWIS) has been deployed throughout many road networks around the world. RWIS can be described as a combination of advanced technologies that collect, transmit, and disseminate road weather and surface condition information to data hubs where it is processed and made accessible to the general public. Road maintenance personnel have access to these data in near real-time and make use of them to effectively plan their WRM activities to shorten response times and reduce material usage (e.g., salt and sand). RWIS collect data using environmental sensor stations (ESS) and live-broadcast cameras to provide real-time and forecasted roadway-related weather and surface conditions. Implementation of this system not only enables cost-effective WRM but also aids motorists in making more informed decisions about their travel. For this reason, RWIS have been widely used in many places in the northern hemisphere. North America alone has more than 3,000 RWIS stations currently in operation, and North American transportation authorities continue to expand their networks to improve their existing WRM services (Kwon et al. 2017, Kwon and Gu 2017). Generally, there are two types of RWIS, namely stationary and mobile.

Stationary RWIS stations, as depicted in Figure 1.1, are generally installed alongside the roadway, with the main purpose of closely monitoring weather/road surface conditions. Each station typically has the following components: cameras, road surface sensors, remote processing units (RPU), and communication hardware. Weather and road surface measurements collected by stationary RWIS often include, but are not limited to, air temperature, road surface temperature, dew point, wind speed and direction, and surface status imagery.

Figure 1.1. Typical stationary RWIS station

A mobile RWIS, on the other hand, as shown in Figure 1.2, is installed on a vehicle. These mobile stations are equipped with a similar suite of innovative sensors and dash cameras that allow them to collect weather and road surface data as they travel along the road. The data collected by mobile RWIS are sent via cellular communication to the maintenance center. In addition to the measurements that are often collected by the stationary RWIS, some mobile RWIS units can provide direct measurements of chemical concentration and pavement friction, which can help maintenance agencies utilize treatment methods that better match the actual weather/road surface conditions.

Figure 1.2. Mobile RWIS unit equipped with spectral road surface temperature sensor

Due to different data collection mechanisms, stationary RWIS provide highly temporal but spatially limited coverage, while mobile RWIS provide spatially continuous but temporally discrete measurements. Regardless of their differences, both versions of RWIS are effective in collecting and disseminating weather and road condition information. Despite their many benefits, the main disadvantage of RWIS lies in their high installation and operation costs. A stationary RWIS station with basic functions costs more than CA $50,000 to install, and this does not include continuous maintenance costs and additional sensors that one might want (Buchanan and Gwartz 2005). Considering their high price tag, municipalities are thus only able to deploy RWIS to a limited number of locations, hampering their coverage and effectiveness. Additionally, the reliability of the point measurements collected diminishes as the distance between RWIS stations increases, resulting in an incomplete map of the surface conditions. Furthermore, contemporary RWIS are equipped with cameras that provide users with a direct view of the road segment; however, determining the RSC via these cameras is still done manually, which prevents the full utilization of these rich image-based road condition data.

1.2 Current Practice of Winter Road Surface Condition Estimation

To mitigate these limitations, several numerical models have been proposed previously in an attempt to quantify the spatial distribution of RSC and estimate snow coverage status on pavement. Sass (1992) developed a prediction model based on heat condition and the surface energy-balance models. Chapman et al. (2001) proposed a multiple regression model to demonstrate that up to 75% of the residual road surface temperature (RST) variation can be affected by surrounding geographical features using thermal mapping techniques. Sokol et al. (2017) applied an ensemble technique for RST forecasting using an energy-balance and heat conduction model where the results tend to underestimate the true values. Perchanok (2002) conducted a discriminant analysis using three friction measurements, namely, peak resistance (F_p), the slip speed at which the peak resistance occurs (V_{crit}), and locked wheel resistance (F_{60}), to classify RSC into categories such as bare wet, bare dry, loose snow, packed snow, or slush. The analysis built a series of linear discriminant functions that could optimally discriminate

different RSC types. Though these prior studies helped provide some insights into how RSC varies over space, they suffered from one major limitation: the models were developed to provide only site-specific RSC information rather than for an entire segment of road. Having continuous RSC information over a road network is critical, not only for road users for improved safety but also for winter maintenance personnel who are responsible for maintaining a good level of service.

Nevertheless, RSC estimation can be challenging, as an ongoing adverse weather event can abruptly change within a short distance and is constantly influenced by many external factors including geographical, topographical, and meteorological features. Likewise, the high degree of uncertainty and randomness associated with inclement weather events, along with their interactions on complex road networks, has made it extremely difficult to accurately estimate conditions between RWIS stations.

In terms of estimating slipperiness or snow status on pavement, there are several methods available. One method is by obtaining RSC (e.g., snow coverage status, roughness) directly via a variety of sensors, such as optical sensors along the roadside and piezoelectric sensors within vehicle tires (Eichhorn and Roth 1992, Erdogan et al. 2010, Yuan et al. 2017). The limitation of this method is that the measurements may be unstable, as these sensors are highly susceptible to environmental and other external factors (e.g., wireless connectivity). Furthermore, installing such sensors is costly and requires manual processing by well-trained personnel. Another method is estimating RSC via numerical methods with auxiliary information as input (Yuan et al. 2017, Alonso et al. 2014, Bo et al. 2014, Wang et al. 2014). This kind of approach often faces struggles with estimation accuracy. For example, Alonso et al. (2014) proposed using tire noise to predict RSC but produced inaccurate results due to noise generated from external variables that did not pertain to the vehicle of interest.

Another type of method is identifying RSC based on computer vision techniques (Bo et al. 2014, Carrillo et al. 2019, Jonsson et al. 2014, Liang et al. 2019, Wu et al. 2022c). Unlike the previous two methods, which require significant effort in maintenance and inspections and suffer from accuracy issues, computer vision techniques have low implementation costs, high estimation accuracy, and the potential to be fully automated. Computer vision itself is a form of machine learning (ML), and in recent years it has advanced gradually from traditional ML to deep learning (DL)-based methods, which allows for the consideration of more information in the RSC recognition process, resulting in higher estimation accuracy. Omer and Fu (2010) investigated the feasibility of classifying winter RSC using images collected from vehicle-mounted cameras. Using red-green-blue (RGB) and gradient features as feature vectors, 400 images were labeled as bare road, snowy road, or tracks. A support vector machine (SVM) model was trained using the extracted features, resulting in a classification accuracy of over 80%. Kawai et al. (2012) proposed a distinction method for nighttime RSC images. They discussed the differences between image features extracted from dry, wet, and snowy road images under different lighting conditions. The K-nearest neighbor algorithm was then used for modeling, displaying an accuracy of 96.1% (dry), 89.4% (wet), and 95.6% (snowy), respectively. Jonsson et al. (2014) used an infrared camera equipped with a set of optical wavelength filters to obtain the brightness of each pixel as features. The images were primarily used to develop multivariate data models and for the classification of RSC in each pixel. The resulting imaging

system could reliably distinguish between dry, wet, icy, or snow-covered sections of the road surface. This system was a vast improvement over existing single-spot RSC classification systems. With the development of DL, some researchers used pretrained deep neural networks to classify the RSC images directly and have achieved results comparable to, if not better than, traditional ML algorithms (Carrillo et al. 2019). Researchers at Technische Universität Braunschweig in Germany created a mixed dataset of RSC images from KITTI (Geiger et al. 2012), Robotcar Dataset (Maddern et al. 2017), and other sources (Nothdurft et al. 2011, Giusti et al. 2015, Smith et al. 2009), which were dry, wet, snowy, etc. They compared two model architectures: InceptionV3 and ResNet50 with long-term and short-term memory (LSTM) units added to improve model accuracy (Nolte et al. 2018). In summary, most of the current methods are developed to select or extract the features manually, such as texture, color, and brightness, before combining them to form a feature database.

Although these developed models display high accuracy in labeling RSC, the important post-process of providing information on hazards has not yet been sufficiently explored. In order for the public to become more aware of current road conditions and for maintenance authorities to make more informed decisions on WRM, the focus should also be on quantifying RSC hazard level (e.g., the snow coverage ratio [SCR]) on top of RSC recognition, which only generates a single predicted category. One recent study tackling this limitation is Yasuno et al. (2021), where a generative adversarial network (GAN) was adopted to translate snow-covered road images to images without snow. The model would then automatically calculate the snow hazard indicator using semantic segmentation (SS). However, the imagery dataset used in that study only contained 500 winter images, which may not be sufficient for developing DL models nor generating reliable snow hazard alerts. Additionally, winter bare road images may not accurately represent the actual road, even via manual inspection.

1.3 Previous Efforts

Acknowledging the limitations of previous studies and current practices in winter RSC estimation, we have developed a range of methods to address existing issues. To automate winter RSC estimation, we created a convolutional neural network (CNN) model that autonomously classifies RSC imagery collected by dash cameras from automated vehicle location (AVL) systems' trucks into four categories: Bare, Partially Snow Covered, Fully Snow Covered, and Undefined. This model achieved an overall accuracy of approximately 95%. Additionally, image thresholding techniques were employed to convert classified images into numerical values (i.e., road surface index [RSI]) as surrogate measurements of road slipperiness (Wu et al. 2022c).

To generate a comprehensive map of RSC, we implemented a geostatistical method called regression kriging (RK) to estimate RSC for all unmonitored highway areas. Both RSC and RSI were used to validate the accuracy of the RK method, demonstrating its feasibility and robustness. With as little as one point of measurement as input, the RK method successfully replicated the general RSC pattern along the highway stretch (Wu et al. 2022b).

However, our previous studies have several limitations that need to be addressed. First, the study area was confined to a single stretch of highway (the southern quarter of I-35 in Iowa), raising

questions about the applicability of our methods to larger regions. Second, the CNN model was developed exclusively for dash camera images, making its effectiveness for RWIS fixed camera imagery (more commonly used by WRM authorities) uncertain. Additionally, whether the CNN model made correct classifications based on intuitive features remains unknown, necessitating an investigation into the model's transparency and reliability. Furthermore, the direct inference of RSC for unmonitored areas via CNN classification results (i.e., the categorical variables) has not been explored. Third, the RK method was tested only on collected hourly event data (228 hourly events for RSC and 33 hourly events for RSI), so its effectiveness for new weather events remains unverified. Lastly, while we believe that our spatial mapping of RSC using RK has benefits for transportation safety, mobility, and sustainability, these potential benefits have not yet been quantified.

1.4 Research Objectives

Building on our previous efforts, this project aims to address the aforementioned limitations by developing highly transferable and universally applicable methodologies, models, and tools for visualizing and inferring RSC using data from RWIS and other road condition monitoring systems. Specifically, our research has the following major tasks:

- Prepare and process a comprehensive training/testing dataset (e.g., RWIS and AVL road weather measurements) that covers a wide range of road, weather, and environmental conditions.
- Test and improve the performance of alternative DL-based image classification models and determine the best model for real-time implementation for both stationary and mobile RWIS.
- Analyze the internal classification process of CNN models via explainable artificial intelligence (XAI) techniques to understand how the accurate results arrive and how the CNNs adapt and refine their classifications with increased training samples.
- Develop a novel geostatistical method, namely nested indicator kriging (NIK), to directly infer categorical RSC variables with DL classification results for all unmonitored areas along the highways.
- Generate a numerical index, namely the SCR that can be used to quantitatively assess the efficiency of existing maintenance programs and the maintenance needs of specific road sections.
- Revisit the continuous RSC spatial mapping method by characterizing winter weather events to make the previously developed RK method more universally applicable, and quantify the potential benefits derived from having spatially rich RSC estimates throughout the developed methods.
- Implement a web-based application for visualizing the spatial mapping and image recognition solutions using real-world scenarios to demonstrate the application's usage.

This project utilizes datasets from two major highways, I-35 and I-80, within the state of Iowa, leveraging the substantial amount of archived stationary RWIS images, dash camera images (i.e., AVL), and geographic information system (GIS) data downloaded and processed during previous projects. The outcomes of this project will provide winter maintenance personnel with newfound knowledge and analytical tools, enabling them to better utilize available resources,

resulting in improved maintenance and enhancements to highway infrastructure, thereby promoting better winter safety, mobility, and sustainability.

This report is structured as follows: The next chapter outlines the proposed methods for automatic RSC recognition using DL-based computer vision techniques and their validation through XAI approaches. It also details the development of the SCR. Additionally, the next chapter introduces methods for the newly developed spatial mapping of categorical RSC (i.e., the NIK) and the refined spatial mapping of continuous RSC (i.e., the refined RK) for unmonitored areas, along with an assessment of the potential benefits of the refined RK method. Subsequent chapters present the results of and discuss a real-world case study. Finally, the conclusion summarizes the main points of this report and suggests potential future research directions to build upon this work.

2 PROPOSED METHODOLOGY

Estimations of RSC have long been a focus for researchers, maintenance authorities, and policymakers. However, accurate estimations are challenging due to the inherent variability of road weather and surface conditions, particularly during inclement weather events. To address this challenge, we previously developed two advanced RSC modeling techniques: CNN (for mobile images) and RK (for continuous RSC variables). Despite these efforts, several limitations remain, including the recognition of stationary RWIS imagery, conversion to numerical performance/RSC indices, categorical RSC variable estimation, and characterization of weather events.

Building on our prior work, this project introduces a new CNN model specifically designed for stationary RWIS imagery, enabling automatic recognition of general RSC at RWIS sites. To validate the accuracy and reliability of CNN predictions for both mobile (from the previous project) and stationary images, we employed two series of XAI techniques: SHapley Additive exPlanations (SHAP) and class activation map (CAM)-based methods. These techniques reveal the internal workings of CNN models by identifying the areas that are most influential in the models' predictions. As mentioned above, a novel geostatistical method (i.e., NIK) was developed to directly estimate categorical RSC for all unmonitored areas with the input of CNN classification results. Additionally, to convert stationary RWIS imagery into a numerical index of pavement snow status (i.e., SCR), we developed and implemented two distinct DL models: pix2pix GAN and SS. To enhance our previously developed RK method for spatially mapping continuous RSC across unmonitored areas, we introduced an ML-based clustering method using the K-means algorithm. This method characterizes weather events based on similar RSC variation patterns. Subsequently, the RK method, including multiple linear regression (MLR) and semivariogram models, was applied to automatically map RSC spatially between RWIS stations. Finally, we analyzed the potential benefits of having spatially rich RSC information using this refined method, exemplified by the RST dataset. These developed methods not only offer novel approaches for accurately estimating winter RSC using existing infrastructure (i.e., RWIS) but also take a pioneering step in quantifying the benefits of refined RK methods. The overall workflow for this method is illustrated in Figure 2.1. Each process is detailed in the following sections, followed by a case study demonstrating a practical application of this methodology.
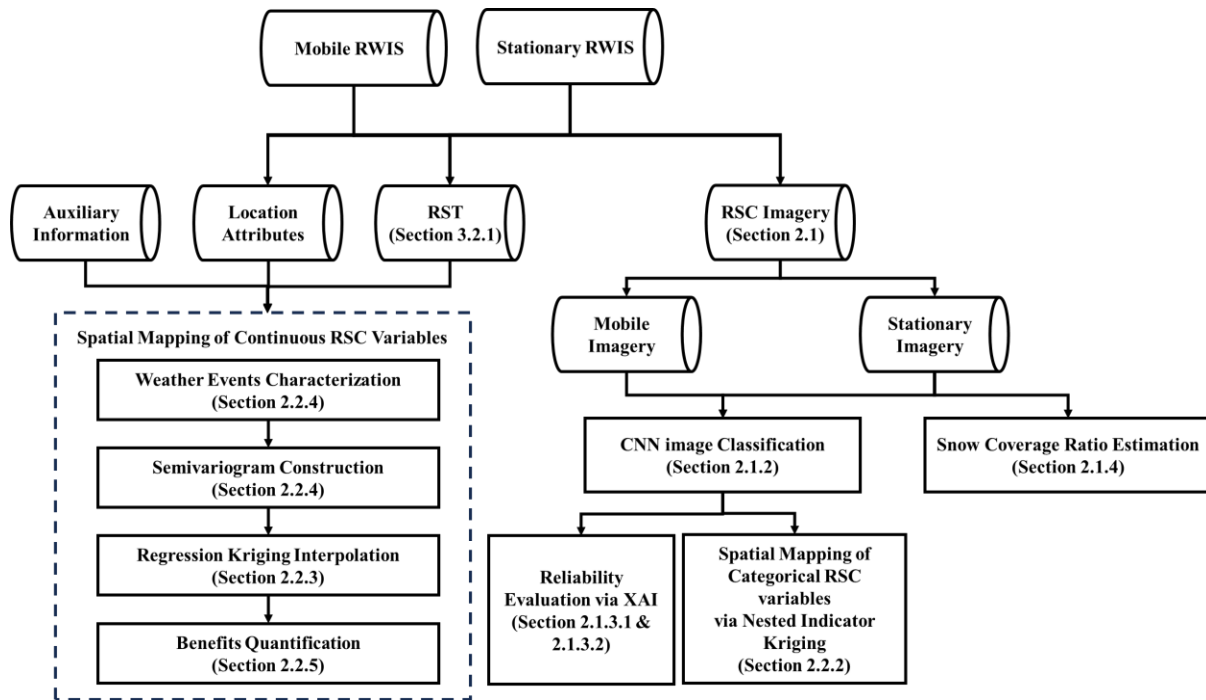
Figure 2.1. Workflow of introduced methods

2.1 Automated Road Surface Conditions Recognition via Deep Learning

2.1.1 The Idea

Due to the numerous advantages of RWIS, this technology has become one of the most widely adopted methods for monitoring winter RSC by highway agencies and maintenance contractors (Boselly 2001). Among the various measurements RWIS can collect, imagery data are particularly critical, as it provides direct visual information on pavement conditions, especially snow coverage, which is essential for determining whether a road section requires immediate maintenance (Fay and Juneau 2007). However, recognizing winter RSC from these images remains a manual process performed by trained personnel. Given the extensive spatial coverage of RWIS networks and the volume of images collected daily (typically one image every 10 to 15 minutes per camera), manual recognition is time-consuming and requires substantial manpower and financial resources. Since winter maintenance is highly time-sensitive, particularly during severe weather events, it is imperative to develop an automated method for RSC recognition using RWIS imagery (Carrillo et al. 2019, Wu and Kwon 2022).

Previous studies, including our own, have proposed the use of CNNs, one of the most widely used DL-based computer vision techniques, to automate the image classification process, a technique that has demonstrated considerable effectiveness (Pan et al. 2019, Pan et al. 2020, Zhang 2021, Ozcan et al. 2019, Khan et al. 2019, Wu et al. 2022c). However, these studies primarily focused on mobile images (e.g., dash camera images) rather than fixed camera images, such as those collected by stationary RWIS cameras. The majority of studies targeting stationary images involve preprocessing steps like cropping pavement areas before inputting them into

classification models or are limited to a few camera sites with similar views where the pavement occupies the majority of the image (Khan and Ahmed 2022, Lyu and Huang 2018). To overcome these limitations, this research investigates the effectiveness of CNNs in classifying stationary RWIS images across the selected study area.

Additionally, while CNNs can classify images into general RSC categories, this may not provide sufficient granularity for assessing finer differences in snow status or road slipperiness. In our previous project, we adopted an image thresholding technique to convert dash camera images into a numerical index known as the RSI (Wu et al. 2022c). Due to the more complex and varied configurations of stationary RWIS images, the same method cannot be directly applied here. Instead, we developed two distinct DL-based computer vision models, namely, pix2pix GAN and SS, to automatically convert stationary RWIS images into numerical indices, which can be used to quantify road conditions. The proposed framework is shown in Figure 2.2. First, the snow-covered RSC image (i.e., winter image) is translated into its dry (i.e., snow-free) condition (i.e., summer image) via pix2pix GAN (Step 1). Second, SS is used to automatically detect the drivable areas (i.e., the exposed pavement without snow cover obtained in the previous step) for winter and summer images, respectively (Step 2). Note that the SS models used for labeling summer and winter images are two separate models. Afterward, to obtain the SCR (numerical value) for assessing the snow hazard level associated with the input winter image, the drivable areas of the winter image are compared to those in its summer image (Step 3). The difference in the detected drivable areas between the winter and summer images is the pavement hidden under the snow. Details of each component involved in this approach are introduced in the following sections.
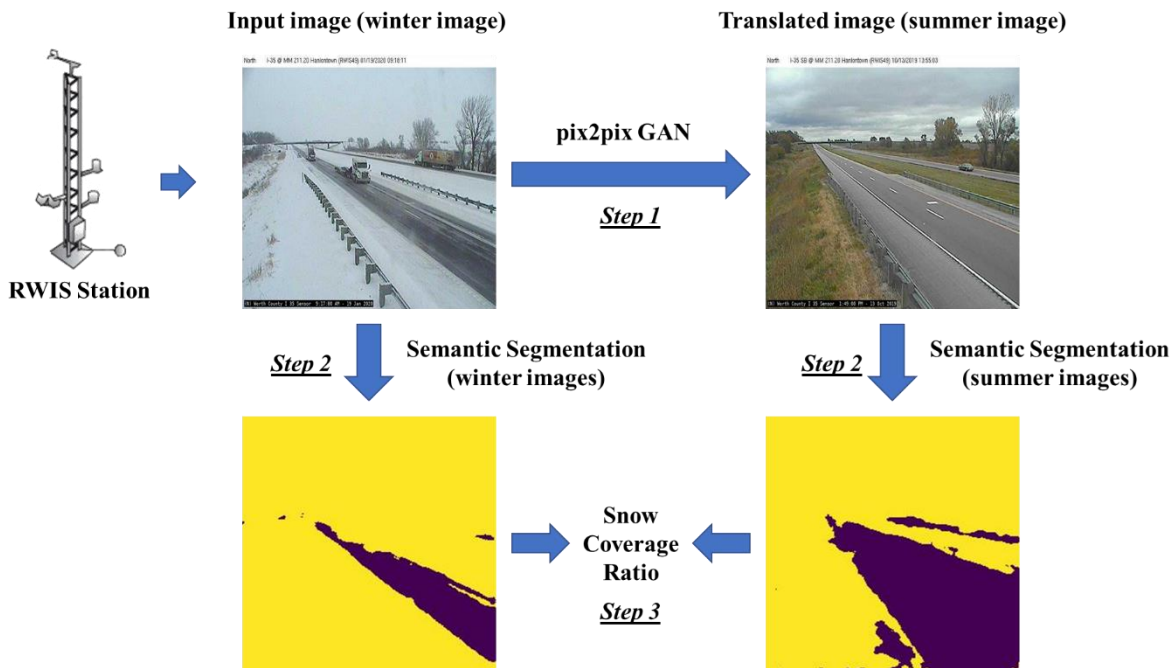
Figure 2.2. Overview of proposed approach for snow coverage ratio estimation

The next subsection introduces all three DL-based computer vision models, details their architectures, and demonstrates their application to stationary RWIS RSC image recognition. Additionally, evaluation methods, including the use of XAI techniques, are discussed to assess model performance and reliability.

2.1.2 Image Classification via CNN

CNNs are a type of DL architecture inspired by the visual perception mechanisms in living organisms. They are particularly effective at image classification tasks due to their ability to automatically and adaptively learn spatial hierarchies of features (Gu et al. 2018). CNNs consist of multiple layers that mimic the neurons in the brain, each with learnable weights and biases. These layers include convolutional layers, pooling layers, and fully connected layers. Each neuron in a layer receives inputs, computes a weighted sum, applies a nonlinearity, and outputs a result (Alzubaidi et al. 2021).

Convolutional layers apply a convolution operation to the input, passing the result to the next layer, mimicking the response of individual cortical neurons to visual stimuli (Lindsay 2021). Filters (or kernels) slide over the input data to produce feature maps, capturing local spatial patterns. The convolution operation helps in detecting edges, textures, and other visual features from the input images (Maggiori et al. 2016). Pooling layers then reduce the spatial dimensions of the feature maps, decreasing the number of parameters and the computational load. This process, known as subsampling or downsampling, provides spatial invariance (Akhtar and Ragavendran 2020). Among different pooling methods, maxpooling is commonly used, which selects the maximum value from a feature map region (Zafar et al. 2022). Fully connected layers, or dense layers, have full connections to all activations in the previous layer. These layers perform high-level reasoning by integrating the features learned by previous layers, making them essential for tasks like classification (Chartrand et al. 2017).

Key components and processes in CNNs include activation functions, training, and regularization. Functions like rectified linear unit (ReLU) introduce nonlinearity into the network, enabling the modeling of complex patterns (Agarap 2018). CNNs are trained using backpropagation and optimization techniques like stochastic gradient descent to minimize a loss function (Bouvrie 2006). Techniques like dropout prevent overfitting by randomly setting a fraction of the input units to zero during training (Srivastava et al. 2014).

CNNs require significantly less preprocessing compared to other algorithms, and their performance improves with larger training datasets. Their ability to identify predictive features from large structures and generate representative vectors makes them effective in various applications, including natural language processing (Wang and Gang 2018). The layered structure of CNNs has been a key factor in their success and widespread adoption (Elngar et al. 2021). As shown in Figure 2.3, the CNN architecture designed for image classification in this study consists of a series of convolutional and maxpooling layers, followed by dense and dropout layers, ending with a softmax output layer.
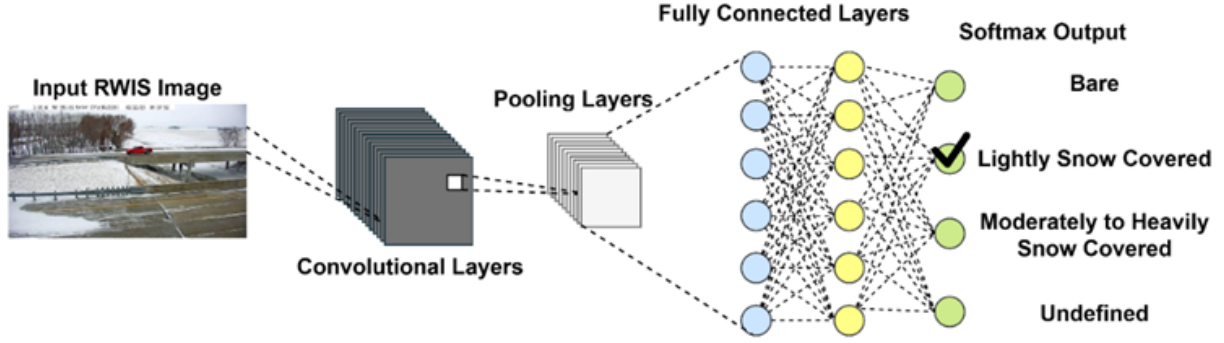
Figure 2.3. Structural overview of CNN for image classification

2.1.3 CNN Evaluation via XAI

As previously mentioned, multiple XAI techniques, including CAM-based methods and SHAP, were employed to enhance the transparency of the CNN classification process, ensuring its reliability before implementation in real-world WRM operations.

2.1.3.1 CAM

CAM is a foundational technique in XAI that helps visualize the regions of an image that CNNs focus on during classification. This method, introduced by Zhou et al. (2016), uses saliency maps to highlight important areas in the input image, providing insights into the model's decision-making process. Building upon CAM, advanced methods such as Grad-CAM, Grad-CAM++, and Score-CAM have been developed to offer more precise and interpretable activation maps, further enhancing our understanding of CNN models. These three Grad-CAM based methods—Grad-CAM, Grad-CAM++, and Score-CAM—were selected to provide a comprehensive and nuanced understanding of the decision-making process of CNNs. Grad-CAM offers a general overview of key regions, Grad-CAM++ improves localization accuracy, and Score-CAM enhances robustness by eliminating gradient dependence. Using all three methods ensures more accurate, reliable, and interpretable insights compared to relying on a single technique, and they serve as practical tools for model analysis and debugging, contributing to the development of more transparent and trustworthy AI systems. The saliency map L_{ij}^c is generated to provide a visual explanation of these XAI techniques. For a particular class c, the saliency map is generated by the weighted summation of feature maps A^k, with i and j indexed as width and height dimensions for the k-th feature map. Specifically, the saliency map is calculated using equation (2.1).

$$L_{ij}^c = \sum_k \omega_k^c . A^k{}_{ij} \tag{0.1}$$

Different weights ω_k^c, calculated for each CAM technique, are used to generate saliency maps. These maps provide valuable information on how various regions of the input image contribute to the classification decision. As a result, they enhance model interpretability and trustworthiness

(Simonyan et al. 2013). Once the saliency map is produced, it is overlaid on the original image to visualize which parts of the image have been activated pixel-wise to make predictions.

Grad-CAM. Grad-CAM generalizes the CAM technique and can be applied to various CNN architectures without requiring changes or retraining. It works with CNNs featuring fully connected layers, structured outputs like image captioning, and tasks involving multimodal inputs such as visual question answering (VQA). Grad-CAM is useful for top-performing models in image classification, captioning, and VQA, offering insights into model decisions and helping to understand model failures. By creating class-discriminative visualizations, Grad-CAM shows which parts of the image the model focuses on for its decisions (Selvaraju et al. 2017).

As shown in equation (2.2), the weights ω_k^c that Grad-CAM uses to generate the saliency maps are computed by normalizing the sum of the gradients of the class score $\frac{\partial y^c}{\partial A^k{}_{ij}}$ (where y^c is the class sore), with a factor of Z, which represents the number of pixels in the feature map. In this context, ω_k^c denotes the weight of the k-th feature map concerning the class (c).

$$\omega_k^c = \frac{1}{Z}\sum_i\sum_j\frac{\partial y^c}{\partial A^k{}_{ij}} \tag{2.2}$$

Grad-CAM++. Grad-CAM++ enhances the visual explanations provided by Grad-CAM, particularly for images with multiple instances of the same class and for more precise object localization. This improvement is achieved by utilizing a weighted combination of positive partial derivatives from the last convolutional layer's feature maps concerning a specific class score (Chattopadhay et al. 2018). The weights ω_k^c in Grad-CAM++ are computed using equation (2.3).

$$\omega_k^c = \sum_i\sum_j \alpha^{kc}{}_{ij} ReLU\left(\frac{\partial Y^c}{\partial A^k{}_{ij}}\right) \tag{2.3}$$

Here, ω_k^c are the weights applied to the k-th feature map A^k to generate the saliency map for class c. These weights are calculated by applying the ReLU function to the gradient of the class score Y^c concerning the k-th activation map $A^k{}_{ij}$, combined with pixel-wise weighting coefficients $\alpha^{kc}{}_{ij}$. The ReLU function ensures that only positive gradients are considered, which contributes to better localization of the regions of interest. The pixel-wise weighting coefficients $\alpha^{kc}{}_{ij}$ are derived from the higher-order derivatives of the class score Y^c with respect to the feature map activations, as shown in equation (2.4).

$$\alpha^{kc}{}_{ij} = \frac{\frac{\partial^2 Y^c}{\left(\partial A^k{}_{ij}\right)^2}}{2\frac{\partial^2 Y^c}{\left(\partial A^k{}_{ij}\right)^2}+\sum_a\sum_b A^k{}_{ab}\left\{\frac{\partial^3 Y^c}{\left(\partial A^k{}_{ij}\right)^3}\right\}} \tag{2.4}$$

The terms $\sum_a \sum_b A^k{}_{ab}$ account for the contribution of all spatial locations in the feature map, $\frac{\partial^2 Y^c}{\left(\partial A^k{}_{ij}\right)^2}$ represents the second-order partial derivative of the class score regarding the activation map, and $\frac{\partial^3 Y^c}{\left(\partial A^k{}_{ij}\right)^3}$ represents the third-order partial derivative of the class score about the activation map. By incorporating these higher-order derivatives, Grad-CAM++ provides a more nuanced and precise weighting of the feature map activations, leading to better localization and interpretability of the model's focus areas. An overview of Grad-CAM and Grad-CAM++ is shown in Figure 2.4.
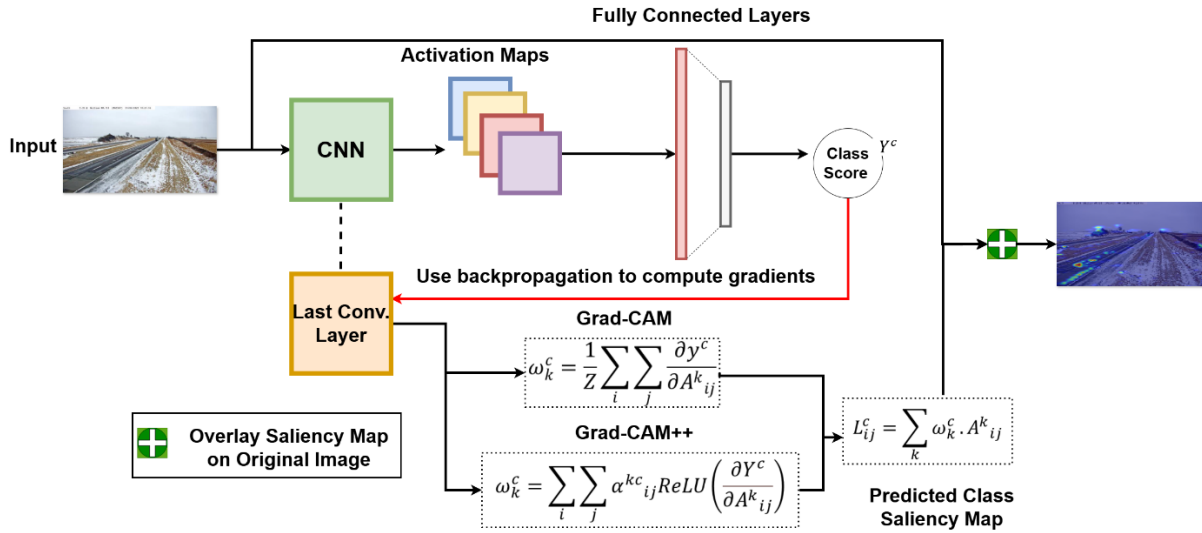
Figure 2.4. Overview of Grad-CAM and Grad-CAM++

Score-CAM. Score-CAM is a post hoc visual explanation method for CNNs that enhances model interpretability by visualizing the importance of input features without relying on gradient information. Unlike gradient-based methods, Score-CAM derives activation map weights based on the forward passing score on the target class, ensuring intuitively understandable and effective visual explanations. To determine the importance of each activation map, Score-CAM uses the channel-wise increase in confidence (CIC) denoted as ΔS^k in equation (2.5).

$$\Delta S^k = f(X \cdot H^k) - f(X) \tag{2.5}$$

In this equation, $f(X)$ is the baseline score function that evaluates the original input image and $f(X \cdot H^k)$ is the score obtained by the element-wise multiplication of the input image X with the upsampled and normalized k-th activation map H^k. This equation measures the change in the model's output score when the input image is masked by the activation map, effectively quantifying the contribution of each activation map to the final prediction. Score-CAM then normalizes these changes in class scores using a softmax function to ensure the weights sum to one.

This normalization step shown in equation (2.6) guarantees that the importance weights of the activation maps are proportional and comparable, where j indexes all activation maps during the normalization step to compute the final weights. The final saliency map is generated by linearly combining the normalized CIC scores with the corresponding activation maps. A ReLU function is applied to this linear combination to ensure that only positive contributions are considered, enhancing the clarity and precision of the visual explanation (Wang et al. 2020). As shown in Figure 2.5, the process begins with Phase 1, where activation maps are extracted. Each activation map functions as a mask on the original image, calculating its forward-passing score for the target class. In Phase 2, which is repeated four times, the final result is generated by linearly combining score-based weights with the activation maps. Both Phase 1 and Phase 2 employ the same CNN module as the feature extractor.

$$\omega^k = \frac{\exp(\Delta S^k)}{\sum_j \exp(\Delta S^j)} \tag{2.6}$$
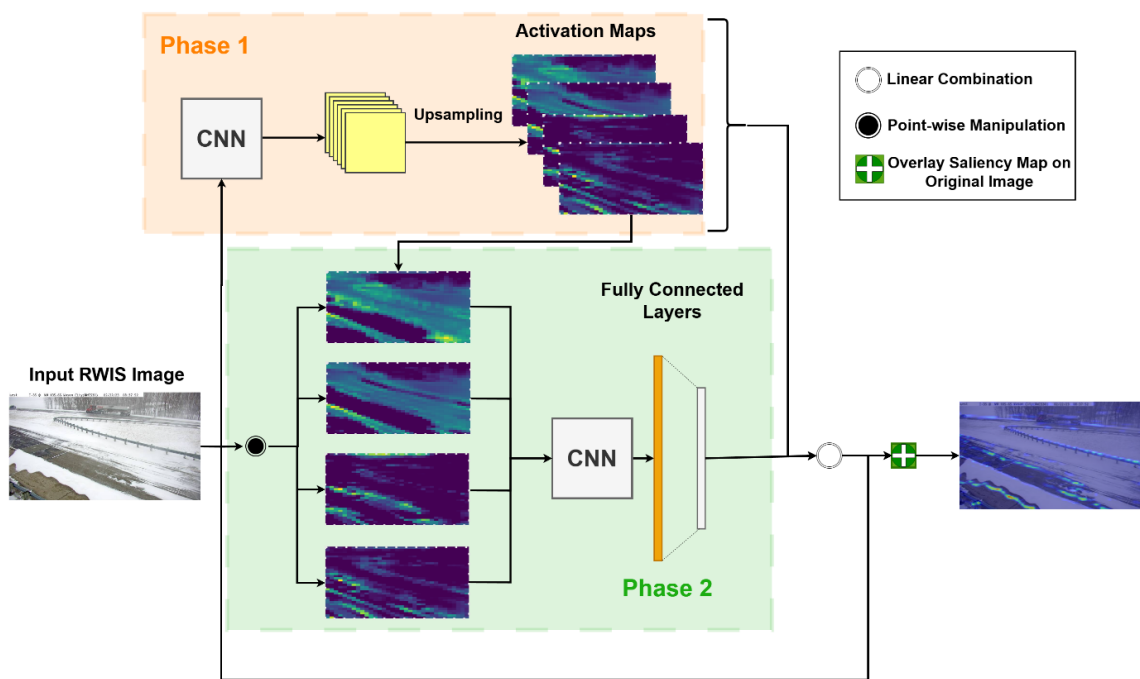
Figure 2.5. Pipeline of Score-CAM

While CAM-based methods provide valuable insights, their effectiveness hinges on the gradients derived from the target class score in relation to the feature maps of the tested convolutional layer (the final convolutional layer in this case). Consequently, they may overlook finer details inherent in earlier network layers. Additionally, there is a propensity in CAM-based methods to produce heatmaps that inadequately align with actual object boundaries (Draelos and Carin 2020). In the domain of winter RSC recognition, this discrepancy can be amplified, as the absence of concrete objects (pavement is the most direct object in our case) makes precise localization challenging. As a result, there may be a risk of imprecise identification of crucial regions, impeding accurate assessment of the CNN's performance in recognizing RSC.

Additionally, the heatmap visualization of CAM-based methods is designed to emphasize regions where the model's activation is high. As a result, areas with low activations, which may be considered as negatively contributing, are not directly highlighted.

2.1.3.2 SHAP

To address the constraints inherent in CAM-based methods and provide a more thorough evaluation of CNN's performance, SHAP was also employed in this study. SHAP is a theoretical game approach designed to provide explanations for the output of any ML model. The fundamental principle of SHAP is to explain the prediction for a specific instance by decomposing it into contributions from individual feature values. These feature values are treated as participants in a cooperative game, where the prediction serves as the game's payout. In this cooperative game setting, Shapley values (SVs), defined as the marginal contribution of each variable value to a prediction across all subsets of features, offer a fair allocation of the payout among the feature values, thus providing insights into their relative importance. It is important to note that in this context, "feature value" refers to the numerical or categorical value associated with a specific feature within the instance. The SV measuring the contribution (feature importance) of the p-th feature can be defined as follows:

$$SV = \sum_{S \subseteq F \setminus \{p\}} \frac{|S|!(|F|-|S|-1)!}{|F|!} \{E[f(X)|X_{S \cup p} = x_{S \cup p}] - E[f(X)|X_S = x_S]\} \tag{2.7}$$

where F is the entire set of features, S denotes a subset, $S \cup p$ is the union of the subset S and feature p, and $E[f(X)|X_S = x_S]$ is the conditional expectation of model f for the subset S. Further details regarding SHAP and SV can be found in Lundberg (2018).

When SHAP is applied to CNN models, the game payout can be interpreted as the image classification accuracies, while each pixel value of an image represents an individual instance. In essence, SVs can be leveraged to ascertain the importance of pixels at a granular level when classifying images. As shown in Figure 2.6, the right-hand side showcases the SVs assigned to each pixel within that image. In this representation, red cells indicate a positive contribution to the prediction, while blue cells represent a negative contribution. The intensity of both red and blue cells corresponds to the level of importance assigned to each pixel.
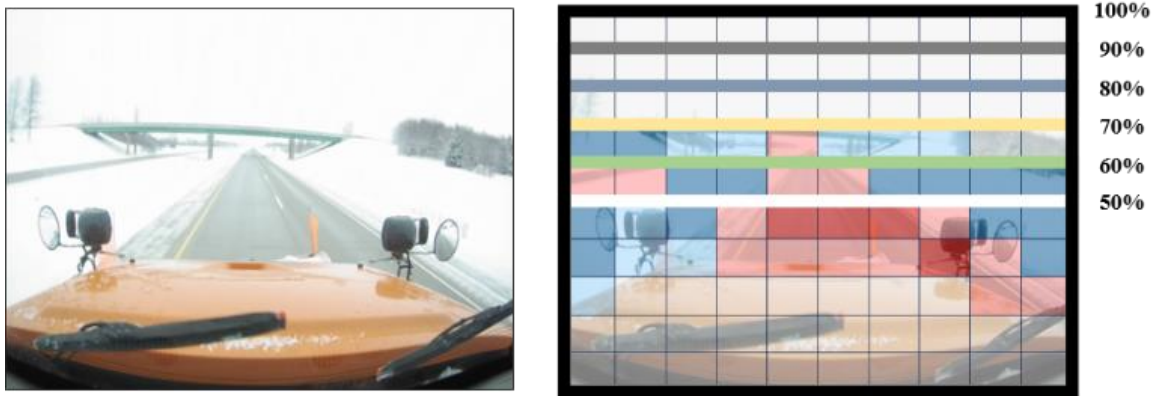
Figure 2.6. Illustration of image area importance analysis

Among all existing XAI techniques, SHAP stands out by leveraging SVs to equitably distribute the contribution of each feature (e.g., pixels of image data in this case) to the model's output. This unique attribute, coupled with SHAP's ability to offer consistent local and global interpretability for model outputs, has made it the preferred XAI technique for delving into and visualizing the rationale behind a CNN's predictions for winter RSC based on imagery data (Burkart and Huber 2021, Haar et al. 2023, Kashifi 2024).

By employing SHAP as the XAI method in this study, we aim to unravel the internal workings of CNN models for winter RSC prediction, gaining insights into the prediction process at the pixel-level. Ultimately, the utilization of SHAP clarifies the underlying decision-making process and provides a comprehensive understanding of the factors influencing the predictions made by the CNN models.

Building on the unique contributions of both Grad-CAM and SHAP, this study leverages the complementary strengths of visual and quantitative interpretability. The combined application of Grad-CAM's intuitive visualizations and SHAP's detailed feature contribution analysis enables a more holistic understanding of the CNN models. This dual approach harnesses both qualitative visual cues and quantitative data, enriching our insights into the model's decision-making processes and justifying the use of both methodologies over relying on a single technique.

2.1.4 Snow Coverage Ratio Estimation

In addition to using CNNs to classify general RSC categories for RWIS imagery, this project also developed two distinct DL-based computer vision techniques to convert RWIS imagery into a numerical value, known as SCR, to quantify the snow hazard level of road sections. The overall approach is demonstrated in Figure 2.2.

2.1.4.1 Pix2pix Winter-to-Summer Image Translation

The first step involved in the proposed approach is to translate the winter images into their corresponding summer images. This translation can be done by using a DL algorithm, namely

pix2pix GAN. Pix2pix GAN was presented by Isola et al. (2017) and is a model designed for general-purpose image-to-image translation. Like all other types of GAN models, its architecture is comprised of a generator (G) model for outputting new plausible synthetic images (i.e., fake/generated images) and a discriminator (D) model that classifies images as either real (from the dataset) or fake (generated by G). The D model is updated directly, whereas the G model is updated via the D model. As such, the two models are trained simultaneously in an adversarial process where G seeks to fool the D, and the D seeks to identify the counterfeit images.

As the pix2pix model is developed specifically for image translation, its generation of the output image is conditional on its input (i.e., the source image x), hence the name conditional GAN (cGAN). cGAN learns a mapping from the input image (x) and random noise vector (z) to the target image (y). The objective of a cGAN can be expressed as equation (2.8).

$$L_{cGAN}(G, D) = E_{x,y}[\, log\ log\ D(x, y)\,] + E_{x,z}[\, log\ log\ (1 - D(x, G(x, z)))] \qquad (2.8)$$

where G tries to minimize the objective against an adversarial D that tries to maximize it, i.e.,

$$G^* = arg L_{cGAN}(G, D) \qquad (2.9)$$

Previous studies (Isola et al. 2017, Yang et al. 2021) also suggest mixing the GAN objective with a more traditional loss (e.g., $L1$ loss). The D's job remains unchanged, but the G is tasked with not only fooling the D but also being near the ground truth output. $L1$ loss is typically chosen as it encourages less blurring than others (e.g., $L2$ loss). Equation (2.10) shows what the $L1$ loss function is.

$$L_{L1}(G) = E_{x,y,z}[\|y - G(x, z)\|_1] \qquad (2.10)$$

Then, by combining these functions, the final objective can be expressed as follows:

$$G^* = arg L_{cGAN}(G, D) + \lambda L_{L1}(G) \qquad (2.11)$$

The D is provided with both a source image (x, i.e., winter image) and the target image (y, i.e., summer image). It must also determine whether the target is a plausible transformation of the source image. A high-level view of the pix2pix GAN architecture is shown in Figure 2.7. The G is trained via adversarial loss, which encourages the G to generate plausible images $G(x)$ in the target domain. The G is also updated via $L1$ loss measured between the generated image and the expected output image. This additional loss encourages the G model to create plausible translations of the source image.
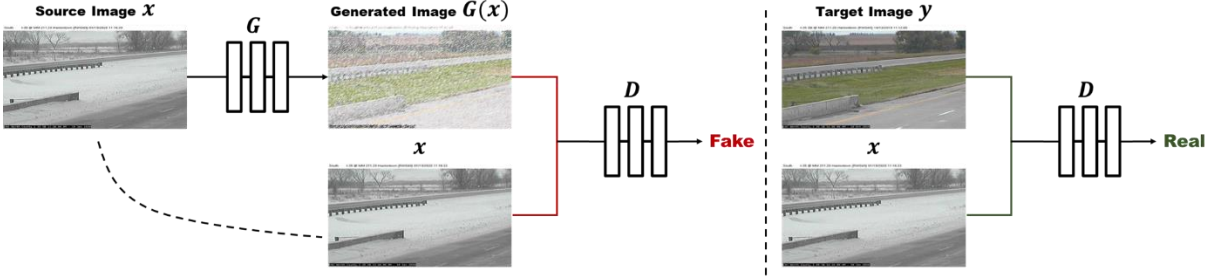
Figure 2.7. Overview of pix2pix GAN architecture

The pix2pix GAN has been utilized in other studies on a wide range of image-to-image translation tasks such as converting maps to satellite photographs, converting black and white photographs to color, and converting sketches of products to product photographs (Ganguli et al. 2019, Dong et al. 2017, Ghosh et al. 2019).

2.1.4.2 Drivable Area Extraction using Semantic Segmentation

With the implementation of pix2pix GAN, the winter images can be translated into their corresponding summer images. However, in order to obtain the SCR, the drivable areas need to be identified by comparing the exposed pavement between the images. For this purpose, another DL algorithm called SS was applied. SS aims to classify each pixel in an image and hence can be thought of as a classification problem per pixel (Wang et al. 2018a). Since this study is only interested in using SS to classify each pixel into drivable or nondrivable areas, a modified binary cross-entropy loss is applied here as the objective function (i.e., the loss function) for SS model development. Equation (2.12) depicts the loss function of SS.

$$L_{SS} = -\frac{1}{n}\sum_n [tlog\,(prob)\, + (1-t)log\,(1-prob)\,] \tag{2.12}$$

where n is the total number of pixels within the segmented mask, t is the ground truth label for drivable area (if drivable $t = 1$, otherwise 0), and $prob$ is the predicted probability for a pixel of being drivable area.

There are generally two types of SS model architectures, namely fully convolutional network (FCN) and U-Net. Both share the core idea of "downsampling-upsampling" or "encoder-decoder." The encoder first reduces the spatial dimensions in every layer and increases the channels, aka downsampling. The decoder then increases the spatial dimensions while reducing the channels, a process that is known as upsampling. But before making a prediction, the spatial dimensions are restored for each pixel in the input image. Since U-Net is less computationally expensive and minimizes information loss, this study utilized a U-Net SS model to detect the drivable areas of the input images. A typical U-Net architecture is shown in Figure 2.8.
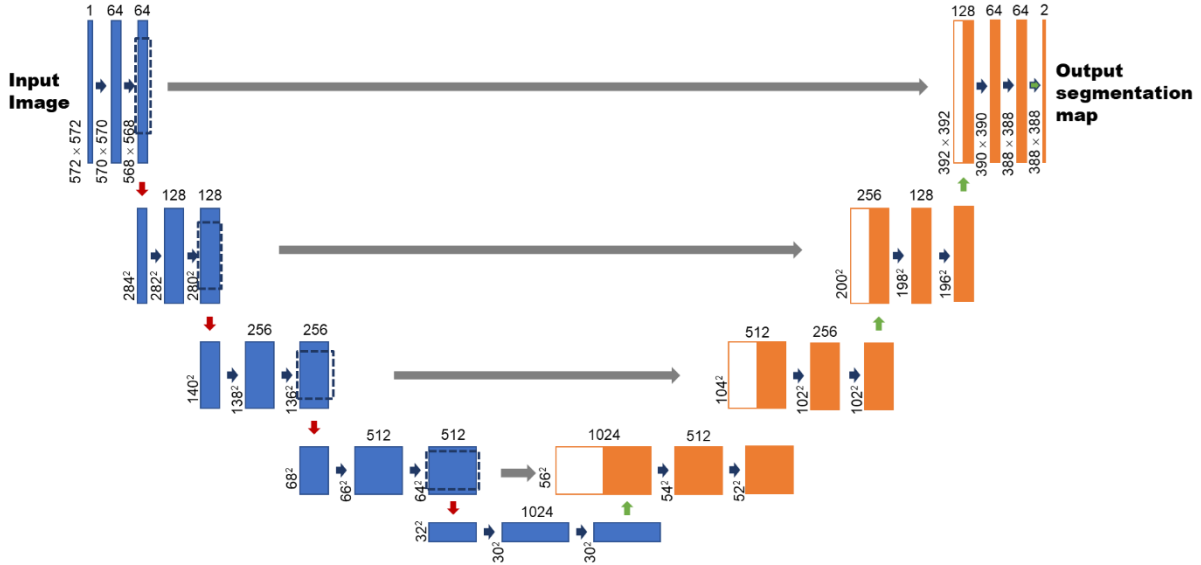
Figure 2.8. Typical U-Net architecture for semantic segmentation

Three metrics are typically used to evaluate the performance of the SS models (model-segmented mask versus manually annotated mask). The first is overall pixel accuracy, defined as the percentage of pixels correctly classified in the validation dataset. The second is intersection-over-union (IoU), also referred to as the Jaccard Index, which measures the number of pixels in common between the target mask (i.e., the manually annotated mask) and the prediction mask (i.e., the model-segmented mask) divided by the total number of pixels present across both masks. Similar to IoU but more sensitive to incorrect annotations is the third metric called the F1 score or Dice coefficient. This metric is calculated by overlapping the areas between two masks divided by the total number of pixels. Equations (2.13) and (2.14) illustrate the computation procedures of IoU and the F1 score, respectively.

$$IoU = \frac{target \cap prediction}{target \cup prediction} \tag{2.13}$$

$$F1\ score = \frac{2 \times (target \cap prediction)}{target + prediction} \tag{2.14}$$

2.1.4.3 Road Surface Condition Recognition via Snow Coverage Ratio

As illustrated in Figure 2.2, the final step is to calculate the SCR by comparing the input image and its corresponding translated summer image to determine the RSC from an input image (e.g., a winter image). During this process, a pixel-level comparison is utilized to calculate how much drivable areas in the segmented mask of the winter image and translated summer image overlap. For example, if pixels in w_1, h_1 and w_2, h_2 are classified as drivable areas in the summer image, while the winter image only classifies w_1, h_1 as drivable areas, then the snow coverage ratio is 1/2=50%.

2.2 Road Surface Conditions Spatial Mapping via Machine Learning and Geostatistics

2.2.1 The Idea

While DL-based computer vision techniques are effective for estimating RSC from both mobile and stationary imagery, they are inherently limited to point measurements. Given the vast expanse of the highway network, combined with the limited number of RWIS stations and the infrequent deployment of mobile sensors, many areas remain unmonitored, leaving RSC entirely unknown. To address this gap, we previously developed a geostatistical interpolation method known as RK to estimate RSC across the entire highway network using a limited number of point measurements (Wu et al. 2022b). Although RK can effectively infer RSC in unmonitored areas, it has some limitations: it is unable to handle categorical variables (such as CNN classification results from RWIS), and our previous RK models were tailored to specific weather events, making them challenging to apply to new, unanticipated weather scenarios.

To overcome these challenges, we developed a novel NIK method, which directly uses CNN classification results to interpolate categorical RSC for all unmonitored areas along the highway network. To ensure the broader applicability of our RK models, we revisited and enhanced them by integrating an ML-based clustering algorithm (i.e., the K-means algorithm) to group similar weather events. For each cluster, we quantified the spatial structures using semivariograms, after which the RK method was applied to complete the interpolation for unmonitored areas. The following subsections elaborate on each component, including their evaluation methods and a quantification of the potential benefits of the refined RK method.

2.2.2 Nested Indicator Kriging for Categorical Road Surface Conditions Estimation

To generate a comprehensive map of real-time (categorical) RSC, NIK is employed to fill in the unmonitored areas using the generated RSC classification results from CNNs. Considering the larger spatial coverage and higher visibility of the road during nighttime, AVL image data are employed over RWIS image data to estimate conditions in unmonitored areas. Within the system, interpolation is applied as one of the last steps prior to final visualization, as shown in Figure 2.1.

Since the variables of interest are categorical (i.e., the RSC), the spatial interpolation method indicator kriging (IK) is employed to estimate these variables at unmonitored points along the highways. IK, a geostatistical approach, operates similarly to ordinary kriging (OK), using the spatial correlation between data points to determine model values (Solow 1986, Arslan 2012). Equation (2.15) expresses how OK and IK estimate a point of interest using a set of observations (n_{obs} in total) at known locations.

$$\hat{Z}(poi) = m(poi) + \sum_{kw=1}^{n_{obs}} \lambda_{kw}[Z(poi_{kw}) - m(poi_{kw})] \qquad (2.15)$$

where Z is the variable of interest (RSC in this case), $\hat{Z}(poi)$ is the estimated value at a location of interest. The terms $m(poi)$ and $m(poi_{kw})$ are the deterministic part or the expected values

(means) of the random variables $Z(poi)$ and $Z(poi_{kw})$, and λ_{kw} is a kriging weight determined by the spatial dependence structure of the residual. A semivariogram quantifies this spatial dependence, as shown in Figure 2.9. The nugget represents microscale variations, measurement errors, or spatial variability at smaller distances. The range indicates the maximum distance where data points are autocorrelated, and the sill represents the variance of the random field (Curran 1988).
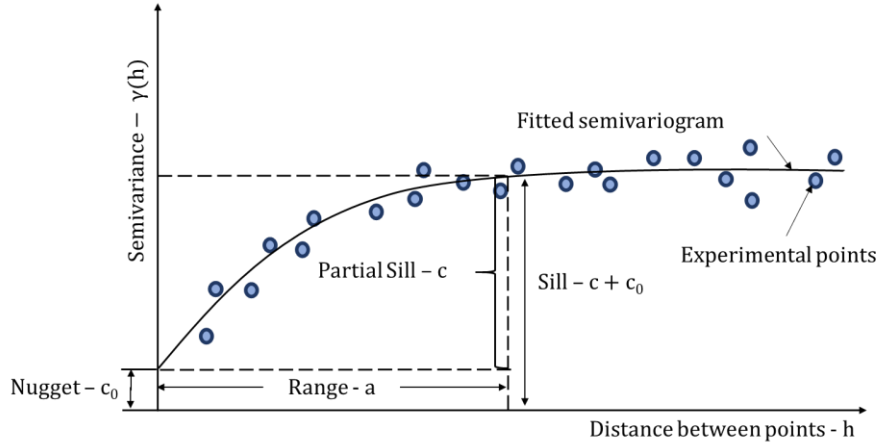
Adapted from Wu et al. 2022c

Figure 2.9. Standard semivariogram

IK is a well-known method for interpolating categorical values in a lot of applications (Smith et al. 1993, Diodato and Ceccarelli 2004, Delbari et al. 2016). Instead of assuming a normal distribution at each estimate location, the standard IK builds the cumulative distribution function (CDF) at each point based on the behavior and correlation structure of surrounding indicator-transformed data points. To achieve this, IK requires a series of threshold values that span the range of data values in the set. These threshold values, known as IK cutoffs, are used to numerically construct the CDF for the estimation point. For each IK cutoff, data in the vicinity are transformed into binary values: 0 if the data exceed the threshold and 1 if they are below it. IK then estimates the probability that the estimation point is less than the threshold value, based on the transformed data and the correlation structure of the IK cutoffs. However, in our study, the CNN outputs categorical values (i.e., Bare, Partially Snow Covered, Fully Snow Covered), lacking continuous numerical values necessary for setting IK cutoffs. This means that these three RSC categories cannot be estimated within a single IK interpolation process. Recognizing the strengths and limitations of standard IK, we proposed a NIK method to directly predict multiple categorical variables. This approach involves feeding categorical variables into a nested structure, allowing for the estimation of more than two categories.

As shown in equation (2.16) the indicator-transformed variable for each point kw is two-dimensional, with each component being a binary variable based on the CNN prediction $Z(poi_{kw})$, as shown in equations (2.17) and (2.18) Consequently, the NIK method employs two layers to estimate RSC for each unknown point. As shown in equation (2.19) the first layer uses standard IK to determine whether the unknown point is Bare conditions. If it is not, the second layer again employs standard IK to distinguish between Partially Snow Covered and Fully Snow

Covered conditions. This approach necessitates constructing two semivariograms for each estimation.

$$F(Z(poi_{kw})) = \{F_1(Z(poi_{kw})), F_2(Z(poi_{kw}))\} \tag{2.16}$$

$$F_1(Z(poi_{kw})) = \begin{cases} 0, & Z(poi_{kw}) = Bare \\ 1, & Z(poi_{kw}) = Partially\ or\ Fully\ Snow\ Covered \end{cases} \tag{2.17}$$

$$F_2(Z(poi_{kw})) = \begin{cases} 0, & Z(poi_{kw}) = Partially\ Snow\ Covered \\ 1, & Z(poi_{kw}) = Fully\ Snow\ Covered \end{cases} \tag{2.18}$$

$$Z(x) = \begin{cases} Bare, & \widehat{F_1}(Z(poi)) < 0.5 \\ Partly, & \widehat{F_1}(Z(poi)) \geq 0.5\ \&\ \widehat{F_2}(Z(poi)) < 0.5 \\ Full, & \widehat{F_1}(Z(poi)) \geq 0.5\ \&\ \widehat{F_2}(Z(poi)) \geq 0.5 \end{cases} \tag{2.19}$$

2.2.3 Regression Kriging for Continuous Road Surface Conditions Estimation

The NIK introduced in the previous section takes CNN classification results as model input to spatially map the categorical RSC along the highway network. However, RSC variables can also be continuous numbers, such as RST and RSI. In a previous study (Wu et al. 2022c), we developed and demonstrated the effectiveness of RK in estimating continuous RSC for unmonitored areas of the highway network. RK is another state-of-the-art geostatistical interpolation technique that combines regression modeling and OK interpolation to estimate the spatial variability of a target variable based on its spatial autocorrelation (Cressie 1988) across the study area. In other words, the main difference between RK and NIK or other kriging-based interpolation methods (e.g., OK) is that RK assumes the determinist part of the target variable has a trend with some external factors. In our previous studies, an MLR model was first calibrated to predict the trend parts (i.e., the deterministic components) of the variables at each unmonitored location. The detrended parts (i.e., the residuals or the stochastic components) obtained from the MLR were then interpolated using OK and preconstructed semivariograms to obtain a spatially continuous estimate of the residuals. This estimate was then added to the predicted values of the MLR to obtain the final RK estimate of the target variable at each unmonitored location in the study area. Equations of RK are shown in equation (2.20), more details can be found in (Wu et al. 2021, Wu et al. 2022b).

$$\hat{Z}(poi) = \hat{m}(poi) + \hat{e}(poi) = \sum_{li=0}^{n_{aux}} \hat{\beta}_{li} \cdot q_{li}(poi) + \sum_{kw=1}^{n_{obs}} \lambda_{kw} \cdot e(poi_{kw}) \tag{2.20}$$

where $\hat{m}(poi)$ is the fitted deterministic part (i.e., the trend part), $\hat{e}(poi)$ is the interpolated residual (i.e., the detrended part), $\hat{\beta}_{li}$ are coefficients of the estimated drift model, $\hat{\beta}_0$ is the estimated intercept, n_{aux} is the number of auxiliary variables, λ_{kw} are kriging weights, and $e(poi_{kw})$ is the regression residual. The regression coefficients $\hat{\beta}_{li}$ can be determined by any appropriate fitting method (e.g., ordinary least squares [OLS]), or, optimally, by using

generalized least squares (GLS) to take the spatial correlation between individual observations into account (Cressie 2015).

RK can be a useful technique to improve the accuracy of predictions for variables with complex spatial patterns and limited data. It does so by incorporating auxiliary information such as geographical and topographical attributes of the target variable and its spatial autocorrelation. However, as previously mentioned, the RSC spatial variation patterns are diverse within different weather events. Constructing an exclusive MLR and semivariogram for each weather event is not feasible. Therefore, a more generalized RK model that can be widely applied to all weather events is needed. This is the main reason weather event characterization is necessary.

2.2.4 Weather Events Characterization via K-means Algorithm

To characterize weather events, the *K*-means algorithm is employed in the proposed framework. *K*-means is an unsupervised clustering algorithm used to group similar data points into *K* clusters. The algorithm starts by selecting *K* random centroids, where *K* is a user-defined parameter. It then iteratively assigns each data point to its closest centroid and recomputes the centroid of each cluster based on the data points assigned to it. This process is repeated until the algorithm converges and the centroids no longer move. The final result is *K* clusters, where each cluster represents a group of data points with similar characteristics. The *K*-means algorithm is widely used in various fields such as data mining (Wu 2012), image processing (Zheng et al. 2018), and pattern recognition (Peng et al. 2013).

In our study, the *K*-means algorithm is used to cluster weather events by their associated RSC variation patterns. Figure 2.10 demonstrates the process using a two-cluster example (i.e., *K*=2). The RSC variation patterns representing weather events obtained from the RWIS stations are grouped. The red lines in the center of the clusters are the centroids, which are the representative RSC variation patterns associated with the weather events within a given cluster. To determine the centroids (i.e., the red lines), the *K*-means algorithm is set to minimize the distance (i.e., difference) between each pair of RSC values of the weather events (i.e., the RWIS point measurements) and their corresponding centroids (determined and assigned by the *K*-means algorithm). The objective function is shown in equation (2.21).
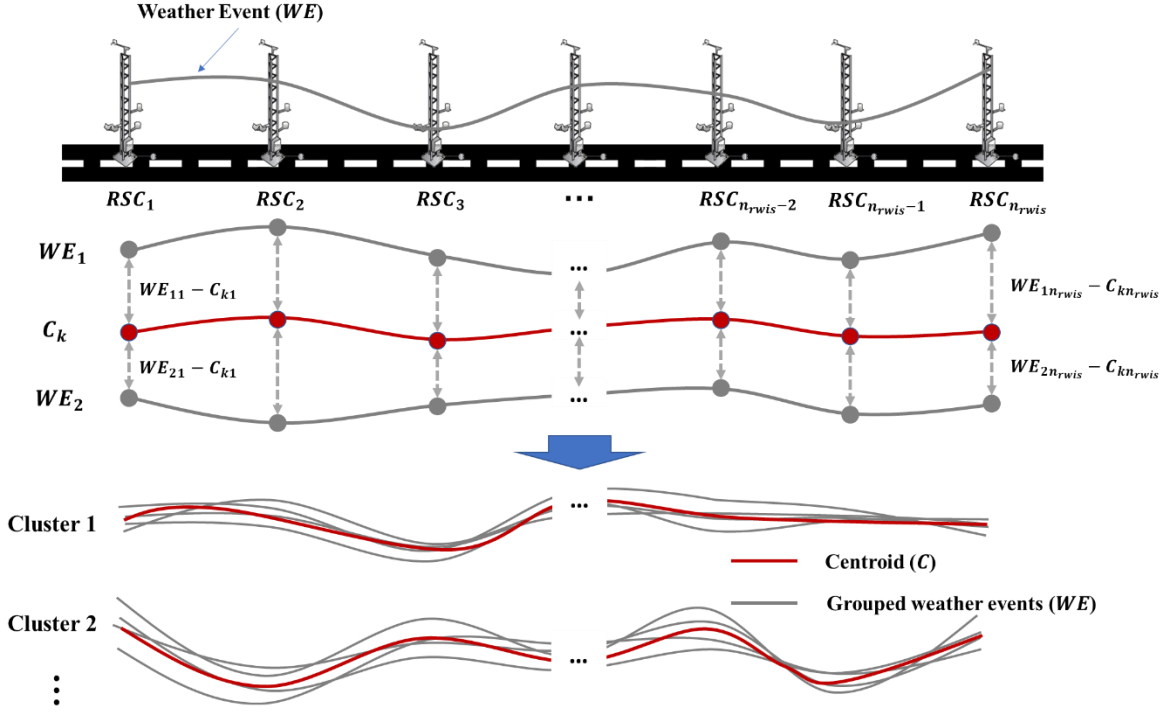
Figure 2.10. Demonstration of _K_-means algorithm for weather event characterization

$$Minimize: J = \sum_{i=1}^{n_{we}} \sum_{k=1}^{K} (\|WE_i - C_k\|)^2 = \sum_{i=1}^{n_{we}} \sum_{j=1}^{n_{rwis}} \sum_{k=1}^{K} (WE_{ij} - C_{kj})^2 \qquad (2.21)$$

where WE_i is the ith weather event represented by n_{rwis} RWIS point measurements of RSC (i.e., $WE_i = \{RSC_{i1}, RSC_{i2}, \ldots, RSC_{in_{rwis}}\}$); C_k is the centroid of the kth cluster and consists of n_{rwis} RSC values; and $\|WE_i - C_k\|$ is the Euclidean distance between weather events, WE_i, and its corresponding assigned centroid, C_k, iterated overall all n weather events in the kth cluster, for all K clusters. This process is demonstrated in Figure 2.10 (i.e., WE_1, C_k, and WE_2).
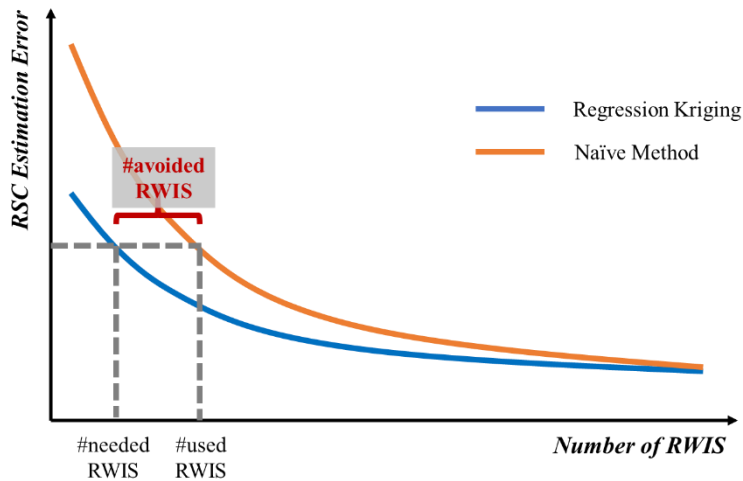
Figure 2.11. Demonstration of benefits quantification process

When a new weather event occurs, its RSC variation pattern is compared with the centroids and subsequently assigned to the cluster with the highest similarity to its centroid. Afterwards, the MLR and the semivariogram models corresponding to that cluster are used to estimate the RSC values through the RK method. This process enables us to accurately estimate RSC values for new weather events using our developed models.

2.2.5 Benefits Quantification

As previously mentioned, this project not only refined the RK method for continuous RSC estimation but also aimed to explore the potential benefits of obtaining spatially rich RSC data from the refined RK method. To quantify these benefits in unmonitored areas where practical implementation has not yet occurred, we employed a simulation-based approach. This approach involved comparing the number of RWIS stations required under our proposed framework with those required by the existing practice (i.e., the naïve method). By analyzing the changing profiles of RSC monitoring levels relative to the number of RWIS stations, we can assess the advantages of our approach. The naïve method assumes that the nearest RWIS measurement represents each unmeasured location. For simplicity, we refer to the refined RK methods as the "RK method" in this report.

As depicted in Figure 2.11, in striving for an equivalent RSC monitoring capability, we explicitly assumed that the actual number of RWIS stations used surpassed the theoretically required number (such a hypothesis can be verified through case-specific assessments, as demonstrated in the subsequent case study). This assumption suggests that implementing our proposed methods can potentially obviate the need for a certain number of RWIS stations. These avoided RWIS stations can be translated into reduced future investments or the possibility of retiring existing RWIS stations. Furthermore, the benefits extend beyond mere cost reduction. Utilizing our methods to estimate RSC in unmonitored areas ensures that the efficacy of the avoided RWIS stations persists. This, in turn, can lead to heightened traffic safety, improved mobility, and a more efficient utilization of maintenance materials.

3 STUDY AREA AND DATA

To assess the feasibility and robustness of our proposed methods, we conducted real-world case studies. This chapter outlines the study area and the data used with the corresponding results presented in Chapter 4.

3.1 Study Area

As noted earlier, this project expands the study area compared to the previous one. The selection of the study area takes into account various factors that may influence the spatial variability of RSC, including geographical, topographical, and traffic-related elements. Additionally, the selection process considers maintenance needs, as well as the cities and populations that can be served. The chosen study area consists of two major highways: I-35 and I-80 in Iowa, US. These highways intersect at Des Moines, with I-35 running from south to north and I-80 traveling from west to east, together covering approximately 845.7 km within Iowa. Given the significance of these highways to both the United States and the state of Iowa, the Iowa Department of Transportation (DOT) has made substantial efforts to monitor and improve the RSC along these routes. The described study area can be seen in Figure 3.1.
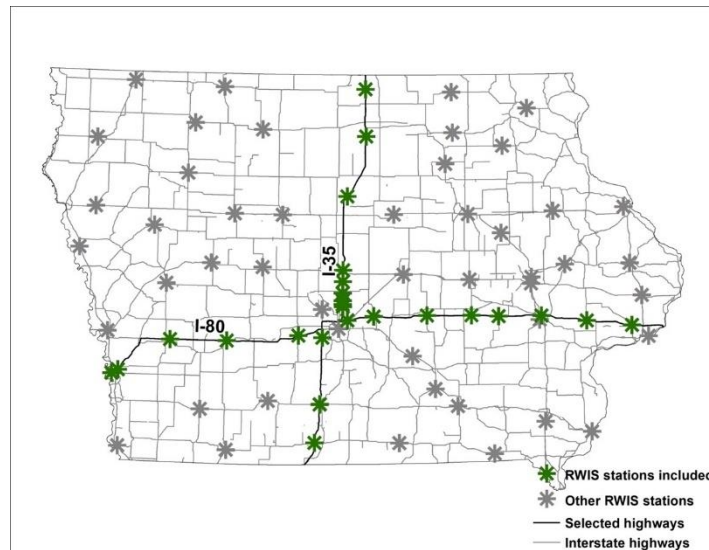
Figure 3.1. Study area – I-35 and I-80 in Iowa

3.2 Data Description and Integration

3.2.1 Road Surface Temperatures and RSC Images

Two data sources for RST and RSC imagery were utilized to develop and evaluate our proposed methods. The first source involved Iowa's snowplow trucks, which are equipped with AVL systems and Global Positioning System (GPS) units to track the locations of the winter maintenance vehicle fleet. For simplicity, these data are referred to as "AVL data" throughout

this report. These AVL-equipped trucks also use standard and infrared thermometers to measure air temperature and RST every 10 seconds along the roadway. A vehicle-mounted dash camera captures one RSC image every 2 to 10 minutes. In total, 2,895,318 RST records and 24,270 RSC images were collected within the study area.

To eliminate measurement errors and local random fluctuations in the AVL RST data, raw RST points were aggregated both spatially (every 500 m) and temporally (every hour) along the road segments. Additionally, to ensure sufficient spatial coverage, only hourly periods with over half of the total length of each highway and more than 50 aggregated data points per highway were retained for modeling (Olea 2006).

The second data source was stationary RWIS stations located along the selected highways. For simplicity, these data are referred to as "RWIS data" throughout this report. These RWIS stations provide real-time weather-related information to support more informed decision-making for winter maintenance activities. The Iowa DOT has installed 25 RWIS stations (12 along I-35 and 13 along I-80) on these highways, most of which are equipped with cameras (63 cameras in total, positioned at different angles) that offer live views and record images of road segments. These images help authorities make critical WRM decisions.

Collectively, RWIS stations contributed to the collection of 2,075 RST records (within the same time frames as the AVL RST data) and 12,900 RSC images within the study area. The descriptive statistics of the hourly RST along the study area are shown in Table 3.1, while the RWIS images were manually labeled according to the classification scheme outlined in Table 3.2, using a web-based labeling platform as shown in Figure 3.2. The AVL imagery used in this project remained consistent with that of the previous project and continued to be utilized. For more details on the AVL imagery, please refer to Wu et al. (2021).

Table 3.1. Descriptive statistics of collected RST data

Highway	Minimum (°C)	Mean (°C)	Maximum (°C)	Standard Deviation
I-35	-17.78	-3.63	4.44	3.20
I-80	-12.78	-2.67	3.33	2.89

Table 3.2. Definition of types of RSC regarding snow coverage

Sample Image	Description	Four-Class Category
	The pavement cross section in all lanes is completely clear of snow.	Bare
	The pavement cross section in all lanes is lightly covered with snow (5% to 50%).	Lightly Snow Covered
	The pavement cross section in all lanes is moderately to heavily covered with snow (50% to 100%).	Moderately to Heavily Snow Covered
	The pavement cross section cannot be identified.	Undefined
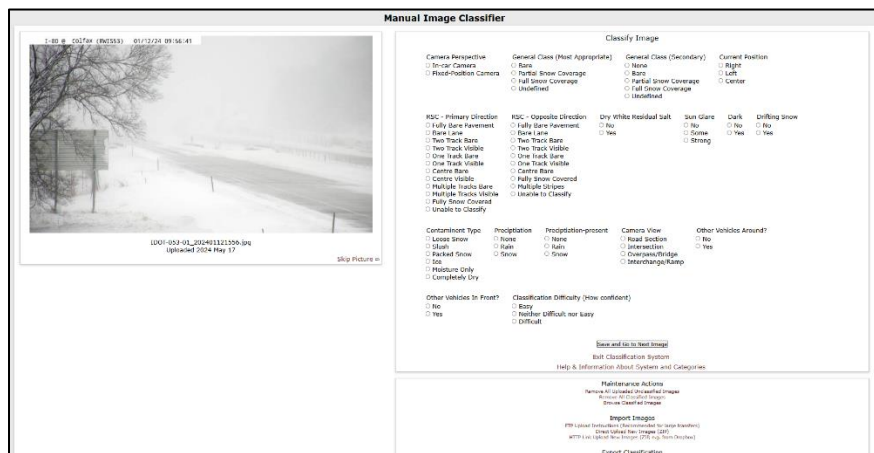
Figure 3.2. Web-based manual image classifier

Labeled AVL images from the previous project, along with the CNNs developed based on them, were utilized to explore the internal workings of the CNNs using Grad-CAM and SHAP, specifically focusing on mobile images to assess their transparency and reliability. The newly labeled RWIS images were then used to train all-new CNNs and validate their accuracies. To

29

further investigate the transparency and reliability of the RWIS CNN models, three CAM-based XAI techniques were subsequently applied.

Additionally, to evaluate the proposed approach for SCR estimation, RSC images captured by 63 RWIS cameras in the study area were downloaded, annotated, and paired to develop two distinct DL-based models: pix2pix GAN and SS. The RWIS images used for these models covered seven winter months (February 2019, December 2019, January 2020, March 2020, January 2022, November 2023, and December 2023) and three summer months (June 2019, October 2019, and June 2023). More winter images were needed due to the inconsistent and highly variable snow cover conditions, whereas the RSC during the summer remained relatively unchanged. To develop the pix2pix GAN over 6,566 winter-summer images were paired, while to develop the SS model 3,460 images (2,814 wintertime, 646 summertime) were carefully annotated. During the image annotation process, only two labels were used: Drivable Areas (denoted as "road," typically indicate exposed pavement) and Others (nondrivable areas, such as snowy areas, sky, trees, etc.). Figure 3.3 and Figure 3.4 depict examples of paired winter and summer images used for training pix2pix GAN as well as annotated masks of drivable areas for training the SS model.

Source Image	**Target Image**
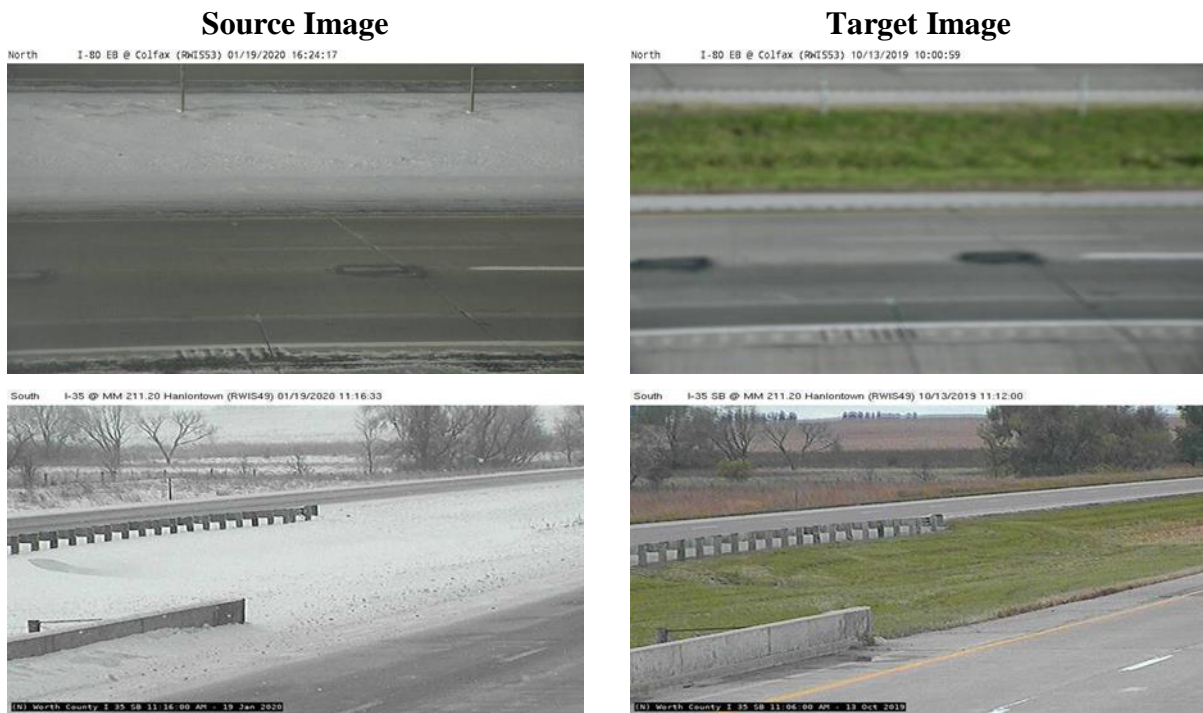
Figure 3.3. Examples of source and target images for pix2pix GAN training
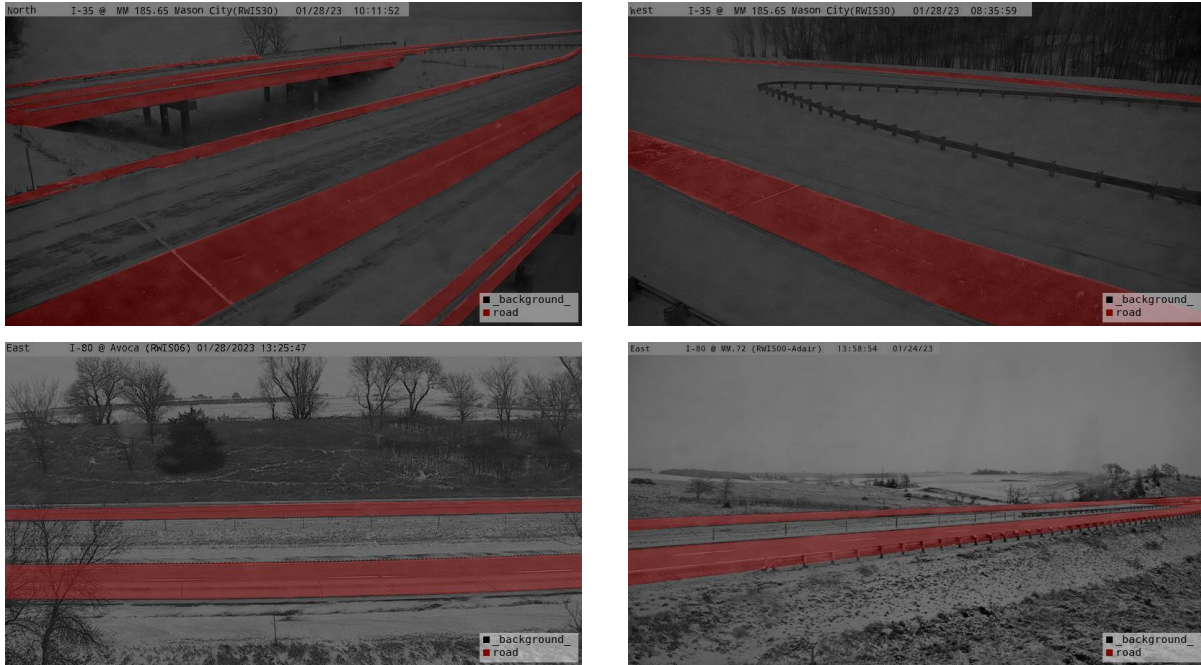
Figure 3.4. Examples of annotated masks for SS training

3.2.2 Digital Elevation Model

To obtain the geographical and topographical features of the Iowa road network that is included in the study area, a digital elevation model (DEM) with a resolution of 3m data (128 GB) was downloaded from the Iowa Geospatial Data website (https://geodata.iowa.gov/) using a Python web crawler script. The elevation, slope, and aspect along the study area were then extracted from the DEM using ArcGIS 10.7 (Desktop, ESRI ArcGIS, 2011). The resulting map files for all topographical features have a size of 484 GB. The visualization of the DEM (elevation only) can be found in Figure 3.5.
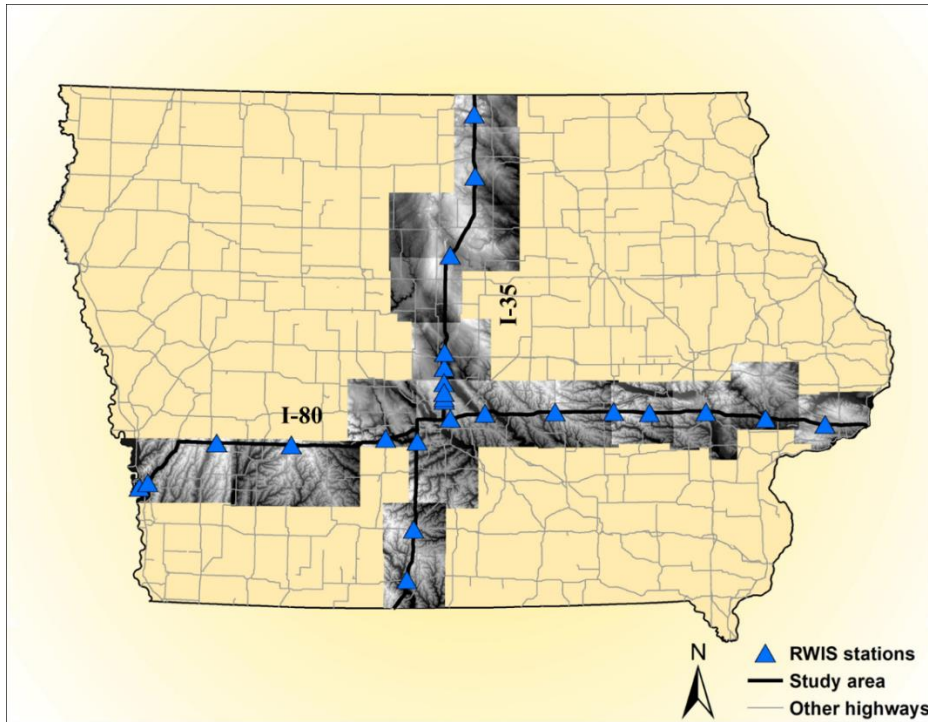
Figure 3.5. Data preparation: DEM of I-35 and I-80

4 RESULTS AND DISCUSSION

By using the collected data described above, case studies were conducted to demonstrate the feasibility, robustness, and reliability of each proposed method. This chapter presents and discusses the results.

4.1 Automated Road Surface Conditions Recognition

4.1.1 Mobile Images

4.1.1.1 CNN-RSC

The RSC image datasets used in this study were collected by Iowa's AVL system. A total of 10,395 images were collected and manually labeled into four categories: Bare, Partial Snow Coverage, Full Snow Coverage, and Undefined. Figure 4.1 provides examples of images representing each of these categories.

Bare

Partial Snow Coverage

Full Snow Coverage

Undefined

Adapted from Wu and Kwon 2022
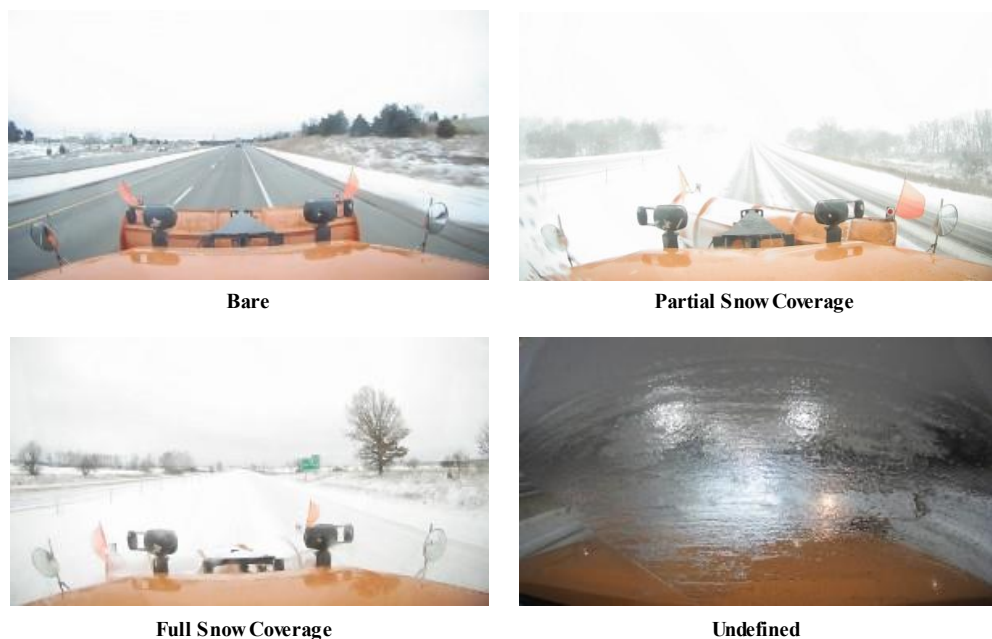
Figure 4.1. Examples of AVL RSC images

As documented in our previous report (Wu et al. 2021), a relatively simple architected CNN model was developed and is hereafter referred to as baseline9 (as it will appear in the following section). Utilizing the Tensorflow application programming interface (API) and Compute Canada with a 32G graphic processing unit (GPU), the design and evaluation of baseline9 were carefully carried out. Various configurations of layers, such as pooling layers, convolutional layers, and dropout layers, were extensively tested. Ultimately, the baseline9 model was

constructed with five convolutional layers (with ReLU activation functions), five maxpooling layers, and three dropout layers, each with a 50% dropout rate. Further details regarding the model can be found in Wu et al. (2022c). Figure 4.2 reveals the training and validation accuracy of this model, accompanied by a confusion matrix that highlights the validation accuracy for each RSC category. The obtained results affirm this CNN-based model's exceptional performance in classifying RSC.
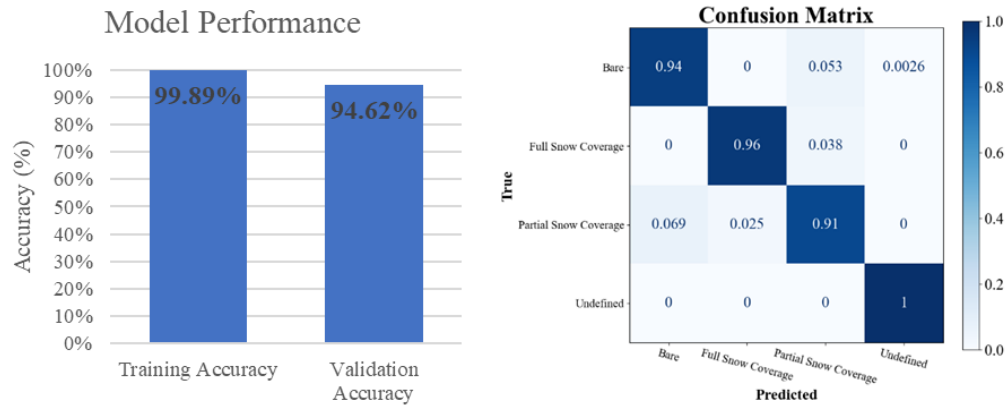
Adapted from Wu et al. 2022c

Figure 4.2. Model performance: (left) training accuracy and validation accuracy and (right) confusion matrix

Nevertheless, questions concerning how the CNN model arrives at its predictions and whether the method is transferrable to other similar images or applications remain. These unknowns underline the necessity of investigating the internal prediction process of the model. Note that this study focuses solely on road conditions, and thus, the "Undefined" images that are too blurry for interpretation are exempt from the analysis.

4.1.1.2 Localizing Important Regions via Grad-CAM

As mentioned earlier, one of the key questions that needs to be answered is, How does the CNN model make its predictions based on the input images? More specifically, Are these predictions derived from learned image features or superficial similarities?

Figure 4.3 illustrates visual examples showcasing CNN predictions (derived from baseline9). The Grad-CAM heatmaps are arranged in descending order of predicted probabilities (the prediction probability for each RSC is shown in Table 4.1), with colored regions indicating significant areas utilized by the CNN for prediction. The intensity of color reflects the importance attributed by the CNN. For example, in image P3 (Full Snow Coverage), the colored areas align predominantly with the snowy regions ahead of the truck, indicating full snow coverage RSC. Conversely, in P3 (Partial Snow Coverage), the colored areas cluster around irrelevant features such as wipers and surrounding regions with different coloration, suggesting partial snow coverage. However, the process by which the CNN determines the final prediction, such as choosing Full Snow Coverage over Partial Snow Coverage despite different highlighted

34

areas for the same image, remains unclear. Additionally, some Grad-CAM heatmaps exhibit inconsistencies. For instance, P2 (Partial Snow Coverage) lacks highlighted areas in the pavement but emphasizes the sky and truck head. A1 (Bare) highlights pavement but also irrelevant features like the truck head and surrounding areas. A3 (Full Snow Coverage) exhibits similar issues. These inconsistencies may arise from solely testing the last convolutional layer, leading to imprecise identification of important regions. Moreover, Grad-CAM (and other CAM-based methods) fails to identify features negatively impacting CNN predictions. To address these limitations, another XAI technique, SHAP, was applied in this study.
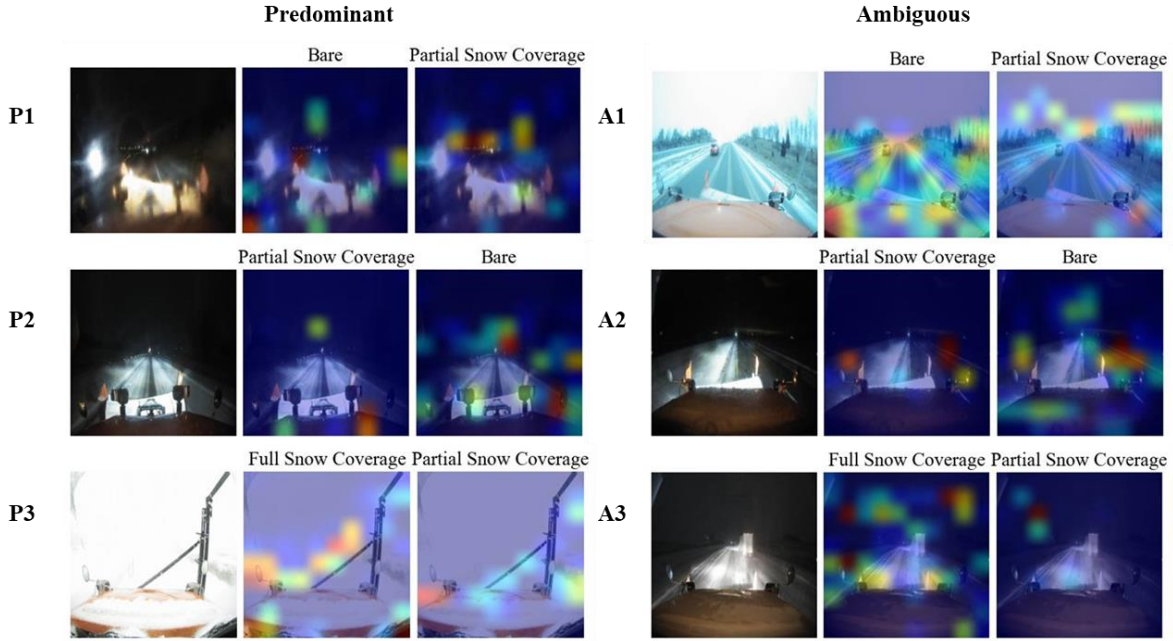
Figure 4.3. Examples of Grad-CAM for explaining CNN classifications

Table 4.1. Examples of CNN classifications and Shapley values

Image	Bare		Partial		Full	
	Prob.	SSVs	Prob.	SSVs	Prob.	SSVs
P1	_**100.0%**_	_**0.333**_	0.0%	0	0.0%	0
P2	0.0%	0	_**100.0%**_	_**0.333**_	0.0%	0
P3	0.0%	-0.07	0.0%	-0.031	_**100.0%**_	_**0.267**_
A1	_**58.9%**_	_**0.145**_	_**41.1%**_	_**0.125**_	0.0%	-0.025
A2	_**40.5%**_	_**0.135**_	_**59.5%**_	_**0.198**_	0.0%	0
A3	0.0%	0	40.6%	0.135	_**59.4%**_	_**0.198**_

Note: Partial = Partial Snow Coverage; Full = Full Snow Coverage; Prob. = CNN prediction probability

4.1.1.3 Exploring Feature Contributions via SHAP

Figure 4.4 provides SHAP-generated visual examples of the same images shown in Figure 4.3 and their corresponding SVs. These colored images show particular image regions' contribution to specific RSC categories: red pixels indicate positive contributions while blue pixels represent negative contributions. For example, image P1-Bare predominantly shows red in visible road areas, suggesting that the model correctly recognizes these regions with exposed pavement as bare RSC indicators. In contrast, some images, such as A1, exhibit ambiguous predictions (the prediction probability for each RSC is shown in Table 4.1) between categories like Bare and Partial Snow Coverage. This ambiguity may be attributed to the blending of features or the presence of characteristics that could be indicative of more than one category (for instance, the exposed pavement areas being red for the Bare prediction and blue for the Partial Snow Coverage prediction). Similarly, the shoulder lane with snow on top appears blue in the Bare prediction and red in the Partial Snow Coverage prediction. In these instances, the color coding reveals a complex interplay between different areas contributing to various predictions, thereby demonstrating the model's deep understanding of the input and its ability to detect overlapping features for accurate classification. Overall, these examples demonstrate the CNN's ability to detect meaningful features and interpret them correctly for accurate predictions.
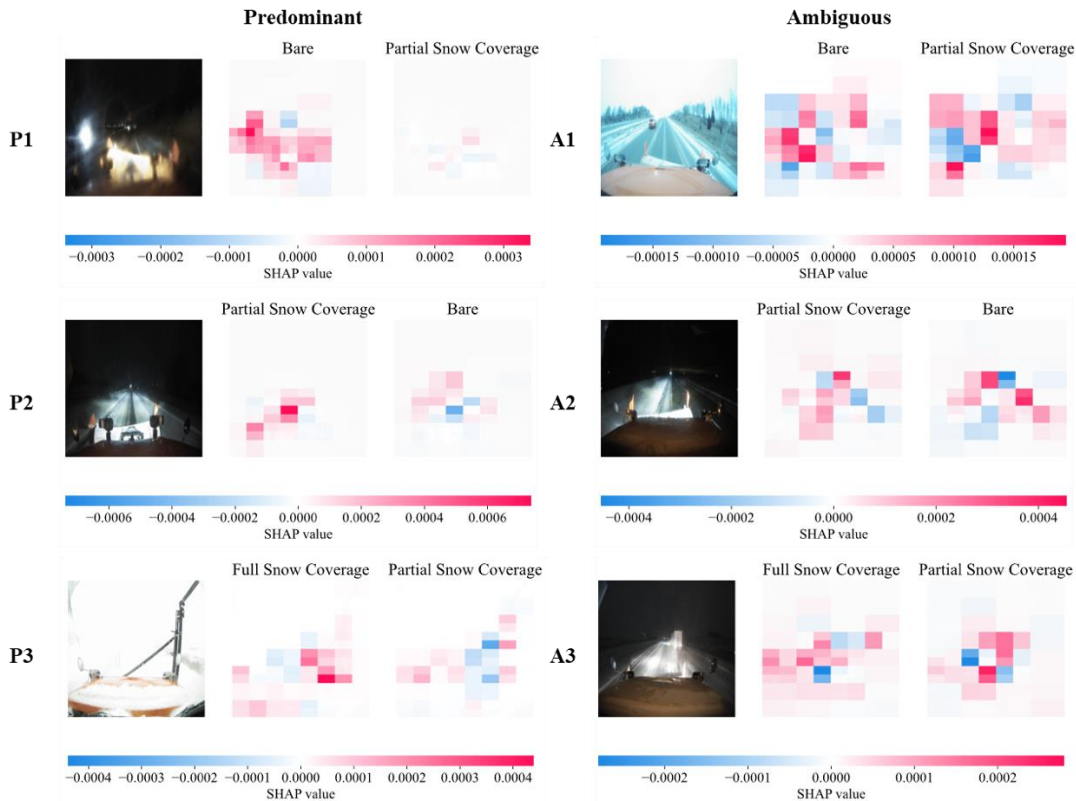
Figure 4.4. Examples of SHAP for explaining CNN classifications

Table 4.1 presents the statistical summary of these examples. It can be observed that the categories predicted with the highest probability by the CNN model also have the largest sum of

SVs (SSVs), a trend consistent across all the images in our dataset (an illustration of SSVs can be found in Figure 4.5, in this case, $m=1$). The dominant predicted categories exhibit predominantly large SSVs, reinforcing the notion that the CNN model relies on meaningful features (i.e., the road areas) for its predictions. In instances of ambiguous predictions where two categories are both plausible, their SSVs are similar to each other. This finding further supports the argument that the CNN model makes predictions based on meaningful features. Given the advantages offered by SHAP over Grad-CAM, all of the following analyses were performed using SHAP.

SV_{11}	$\cdots$	$\cdots$	$\cdots$	SV_{1l_w}
$\vdots$	$\ddots$			$\vdots$
SV_{i_i1}		$\ddots$		SV_{1l_w}
$\vdots$			$\ddots$	$\vdots$
SV_{l_l1}	$\cdots$	$\cdots$	$\cdots$	$SV_{l_ll_w}$

$$SSVs = \sum_{i=l_i}^{l_l}\sum_{j=1}^{l_w} SV_{ij}$$

$$PASVs = \frac{SSVs}{\sum_{i=1}^{l_l}\sum_{j=1}^{l_w} SV_{ij}}$$

Included areas
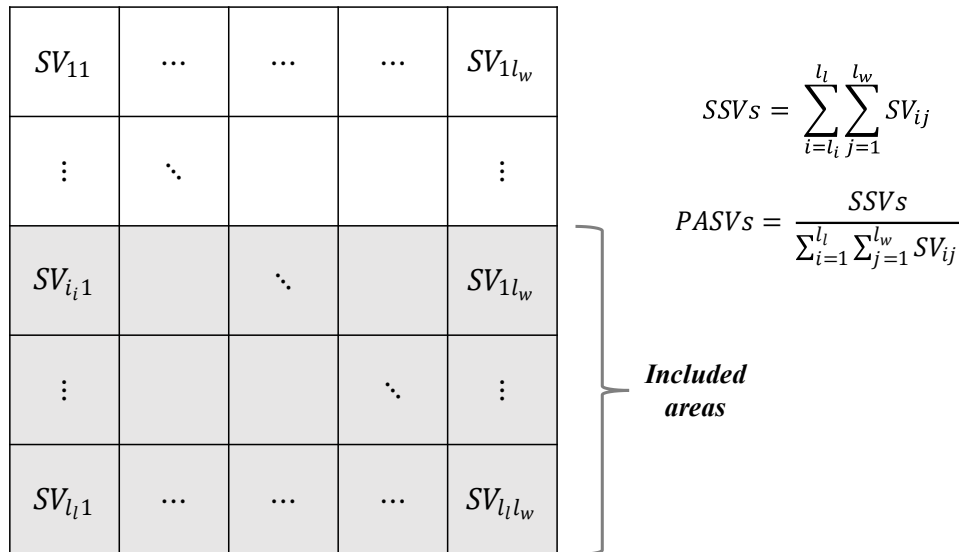
Figure 4.5. Illustration of SSVs and percentage of absolute Shapley values

Impact of Image Areas. While the CNN model has demonstrated the ability to make meaningful predictions based on RSCs, some unexpected aspects of its predictions warrant further examination. Notably, some pixels in nonrelevant areas, such as the sky in image A1 (in Figure 4.4), have been found to contribute to the final predictions. This observation leads to a deeper investigation into where the CNN focuses its attention during the prediction process to ensure that the predictions are based on relevant features, particularly the road areas.

In this study, the imagery utilized was obtained from dash cameras, which provided a consistent camera angle across the entire dataset. The upper part of the image typically represents the sky, the lower part represents the road areas, and the truck's hood appears at the very bottom of the image. This consistent layout allows for a comparative analysis of the SVs assigned to different regions of the image.

Figure 4.6**Error! Reference source not found.** visually illustrates this process. The left-hand side of the figure displays a raw winter RSC image, while the right-hand side showcases the SVs assigned to each pixel within the image. To analyze the contributions of different image areas, colored border lines are introduced. Each line represents a specific proportion of the image included for analysis. For instance, the yellow line with "70%" annotated means that the analysis contains 70% of the image, starting from the yellow line and extending to the very bottom.
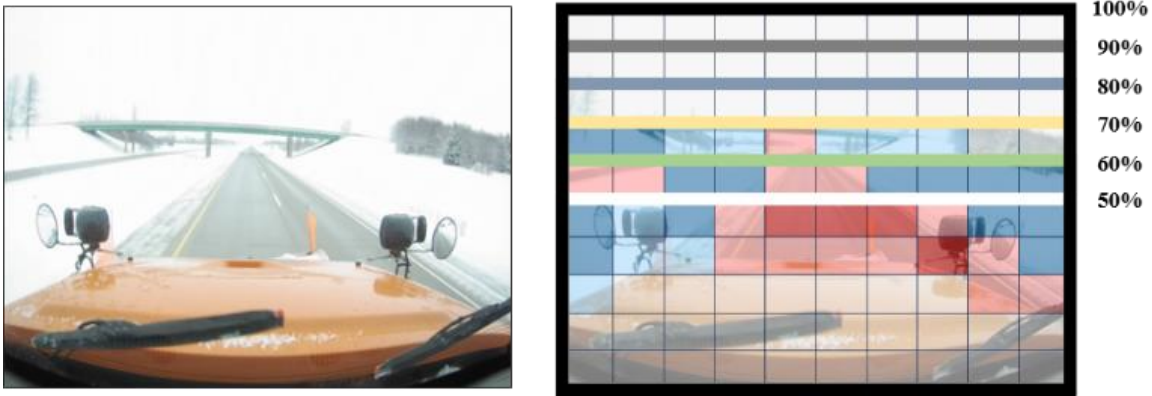
Figure 4.6. Illustration of image area importance analysis

With the visualization process explained, Figure 4.6 illustrates the changing pattern of the SSVs (within the included areas) and the percentages of the absolute SSVs within the enclosed image areas relative to the entire image, referred to as percentage of absolute Shapley values (PASVs). As illustrated in Figure 4.5, PASVs aim to quantify the percentages of the absolute SSVs within the enclosed image areas relative to the entire image, thereby providing insight into the proportion of feature importance.

According to Figure 4.7, as the evaluated proportion of the image increases from 50% to 100%, both the SSVs and the PASVs correspondingly increase. This observation aligns with our intuition that the higher the portion of the image evaluated, the more features get involved during the CNN's prediction process.
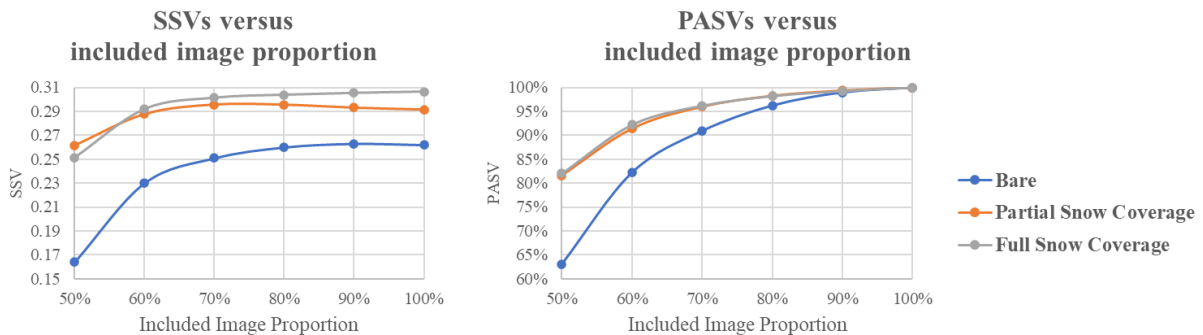
Figure 4.7. SHAP metrics at different included image proportions

However, a significant finding emerges when the included image proportion surpasses 70%: both the SSVs and the PASVs exhibit marginal changes. This phenomenon can be attributed to the fact that the remaining 30% of the image predominantly consists of the sky and sometimes bridges (refer to Figure 4.1), which are not expected to influence the features. In addition, PASVs beyond 70% consistently surpass 90%, indicating that over 90% of the features utilized by the CNN for predictions are concentrated below the yellow 70% line, affirming that the

CNN's focus on meaningful and relevant image regions leads to its effective predictive performance.

Impact of Training Samples. The next step is to explore whether a CNN is able to evolve and improve its performance as the number of training samples increases. To investigate this, the training dataset, consisting of 9,354 images, was divided into nine batches, each containing slightly more images than the previous batch (1,000, 2,000, ..., 9,354).

Training Process and Accuracy Trends. The same baseline CNN architecture mentioned earlier was trained using each batch (baseline1, baseline2, ..., baseline9), while the validation dataset (consisting of 1,041 images) remained the same for all. The overall trend is depicted in Figure 4.8.
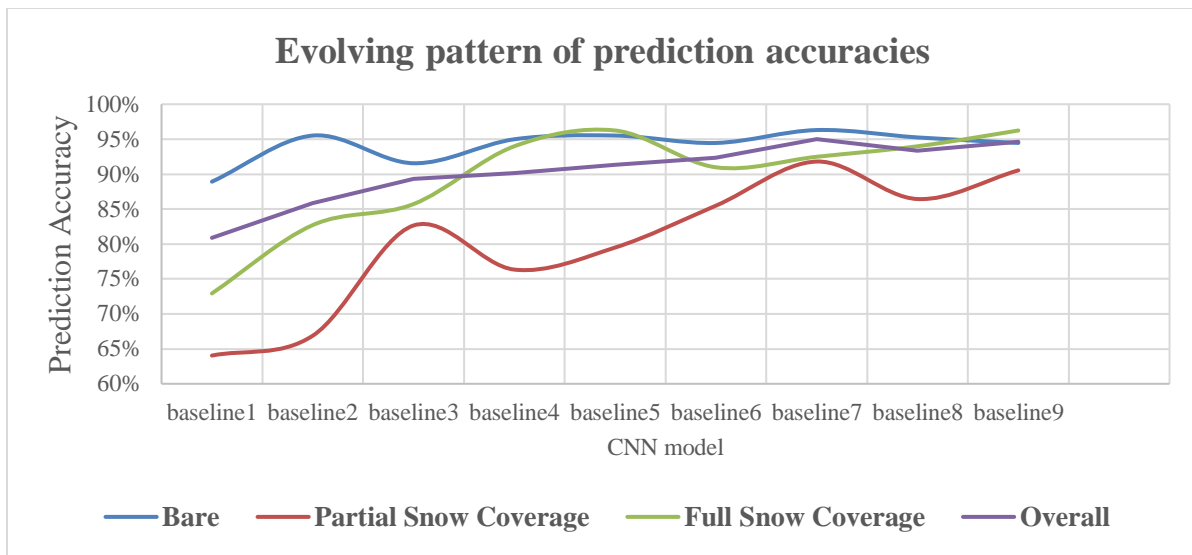
Figure 4.8. Evolving pattern of RSC classification accuracy versus training samples

As shown in the figure, accuracies for individual RSC categories and overall accuracy generally increase as the training sample size grows. However, the Bare condition has the least improvement (88.9% to about 94.5%). Comparatively, other classes demonstrate more significant improvements (Partial Snow Coverage: 64.0% to about 90.5%, Full Snow Coverage: 72.9% to about 96.2%, Overall: 80.9% to about 94.6%). This could be attributed to the observation that pixel values (i.e., the color) of the exposed pavement are much easier for the CNN to identify and comprehend. In contrast, identifying snowy regions on the road surface poses a much greater challenge, as their color can sometimes be the same as the sky and other irrelevant environmental features, like snow on the side of the road. These similarities make accurate classification more difficult, resulting in a steeper learning curve for these categories.

Learning Dynamics and Model Evolution. The SHAP analysis also helped to further reinforce our findings and provided a direct view of how the CNN learns and improves with increasing training samples. Figure 4.9 presents an example that highlights this progression. Initially, when the CNN was trained with only 1,000 images (baseline1), it struggled to identify any meaningful

features in an image, resulting in a "blank" SHAP image and a predicted probability of 0%. However, as the number of training samples increased to 6,000 (baseline6), the CNN gradually became more adept at recognizing and utilizing relevant features. Despite this improvement, there were still certain challenges in accurately interpreting some parts of the driving lane (e.g., the blue pixels on the right of the driving lane). Additionally, some extraneous features, such as the shoulder lanes and trees, still influenced the prediction process, leading to a predicted probability of only 54.4%. In contrast, when the CNN model was trained with the entire dataset (baseline9), it successfully disregarded the unnecessary features and focused primarily on the driving lane (particularly the partially snowy areas). As a result, the predicted probability dramatically increased to 99.9%.
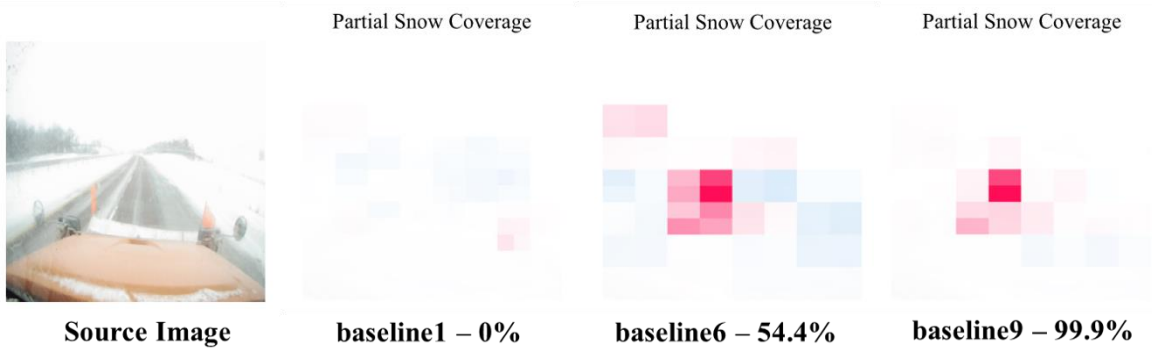
Figure 4.9. Example of CNN evolving pattern using SHAP

Correlations and Model Interpretation. Figure 4.10 illustrates the overall correlation between SSVs, PASVs, and RSC prediction accuracies. The results clearly indicate that both SSVs and PASVs exhibit positive correlations with RSC prediction accuracies. High PASVs indicate that the CNN recognizes the road areas (the included image area) as the primary features to make predictions on RSC. To assess the strength of these monotonical correlations, Spearman correlation values were calculated for all variables (Hauke and Kossowski 2011) as presented in Table 4.2.
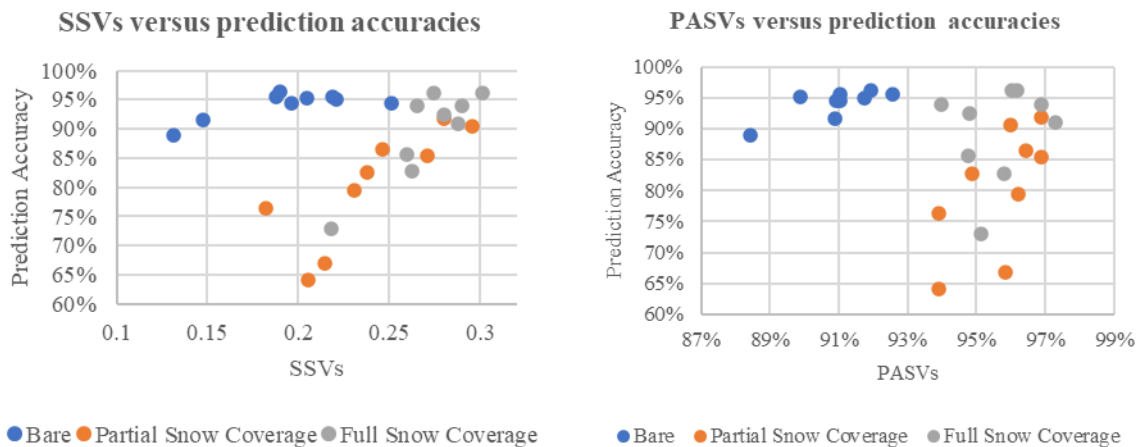
Figure 4.10. Correlation between SHAP metrics and classification accuracies

Table 4.2. Spearman correlation values between RSC categories and SHAP metrics

RSC Category	SSVs	PASVs
Bare	0.269	0.756*
Partial Snow Coverage	0.917*	0.733*
Full Snow Coverage	0.723*	0.252

* This correlation is statistically significant in 95% confidence intervals.

The results demonstrate that the Bare condition has a strong positive correlation with PASVs and a weak correlation with SSVs. This can be attributed to the same reason mentioned earlier: the pixel values of fully exposed pavement are easily identifiable and are thus prominent features for making accurate predictions. As a result, there is relatively little improvement in the prediction accuracies. As for PASVs, the results suggest that the CNN model gradually learns that the image's upper part (i.e., the excluded 30%) is not necessary to consider when making predictions.

On the other hand, the Full Snow Coverage condition exhibits a significant positive correlation with SSVs but not PASVs. This suggests that the CNN is initially able to capture the useful features (i.e., the road areas) effectively and progressively learns to assign greater importance to the "white" areas, which are indicative of snowy road surfaces, to make correct predictions. This is the reason behind the significant correlation observed in SSVs.

By comparison, the Partial Snow Coverage condition shows significant correlations with both SSVs and PASVs, which intuitively makes sense. In this case, the CNN attempts to assign more importance to the features within the enclosed image area while simultaneously disregarding the excluded image area when making predictions.

Overall, the SHAP analysis provides further evidence of the CNN model's learning process and its ability to adapt based on the information gained from the training samples. The positive correlations between SSVs, PASVs, and prediction accuracies highlight the CNN model's ability to capture relevant features and improve its performance over time. These findings contribute to a deeper understanding of the CNN model's internal workings and its evolving prediction process.

Impact of CNN Structures. From the previous findings, it has been established that the number of training samples influences the prediction capabilities of the CNN. However, another important question remains unanswered: Does the structure of the CNN model itself impact its prediction capabilities? To address this question, several well-known CNN models, namely VGG16 (Simonyan and Zisserman 2014), ResNet50 (He et al. 2016), and Xception (Chollet 2017), were trained using the same training samples as baseline9 through transfer learning, where only the fully connected layer was trained. Subsequently, these models were compared using the same set of validation imagery (Shin et al. 2016).

Figure 4.11 illustrates the performance differences between the selected CNN architectures as well as their corresponding SSVs and PASVs. As depicted, the general trend of prediction accuracies aligns with the findings from the previous section, indicating positive correlations with both SSVs and PASVs. Although there are only minor performance differences (ranging from 93.2% to 95.3%) between the CNN models, it can be concluded that changing the CNN architecture impacts their learning capabilities.
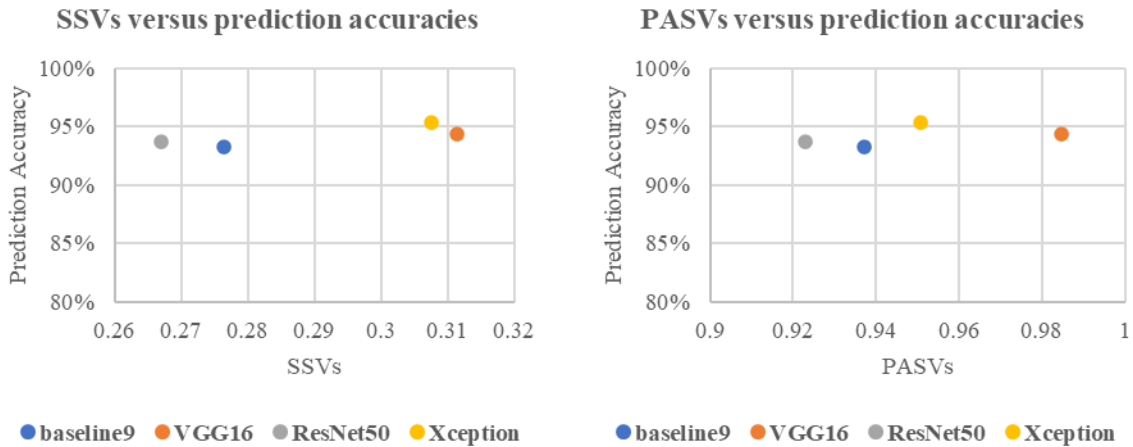
Figure 4.11. Comparison between CNN structures

This insight underscores the sensitivity of CNN models to their structure design. However, it should also be noted that this sensitivity might fluctuate depending on dataset size and complexity. Therefore, further investigation is needed to verify if these observations hold true for larger datasets and more complex imagery. Pursuing additional research to explore the performance of different CNN structures in such scenarios will furnish a more comprehensive understanding of how architectural choices influence learning and prediction.

4.1.2 Stationary Images

4.1.2.1 Image Classification via Convolutional Neural Network

CNN Structure. Based on the collected and labeled RWIS imagery data, a CNN-based image classification model for RSC recognition was developed from scratch. It employed four convolutional layers with increasing filter sizes (32, 64, 128, and 256 filters, respectively), each with a kernel size of 3×3 and ReLU activation. These layers progressively extract features of increasing complexity from the input images (Chen et al. 2016). Each convolutional layer was followed by a maxpooling layer with a pool size of 2×2 to reduce spatial dimensions and retain significant features, ensuring efficient feature extraction and dimensionality reduction (Arivalagan et al. 2023). Following the convolutional and pooling stages, the model included a flattening layer that transformed the 3D feature maps into a 1D feature vector. This vector was then fed into a dense layer with 128 units and ReLU activation, enabling the learning of complex patterns and relationships in the data. A dropout layer with a rate of 0.5 was incorporated to mitigate overfitting by randomly setting half of the input units to zero during training (Srivastava

et al. 2014). The final layer was a dense output layer with units equal to the number of classes (four in this study), utilizing the softmax activation function to produce a probability distribution over the classes, facilitating accurate classification. Figure 4.12 shows the proposed CNN architecture for our RWIS image classification model. The model was configured with the Adam optimizer and the categorical cross-entropy loss function while monitoring accuracy as a performance metric.
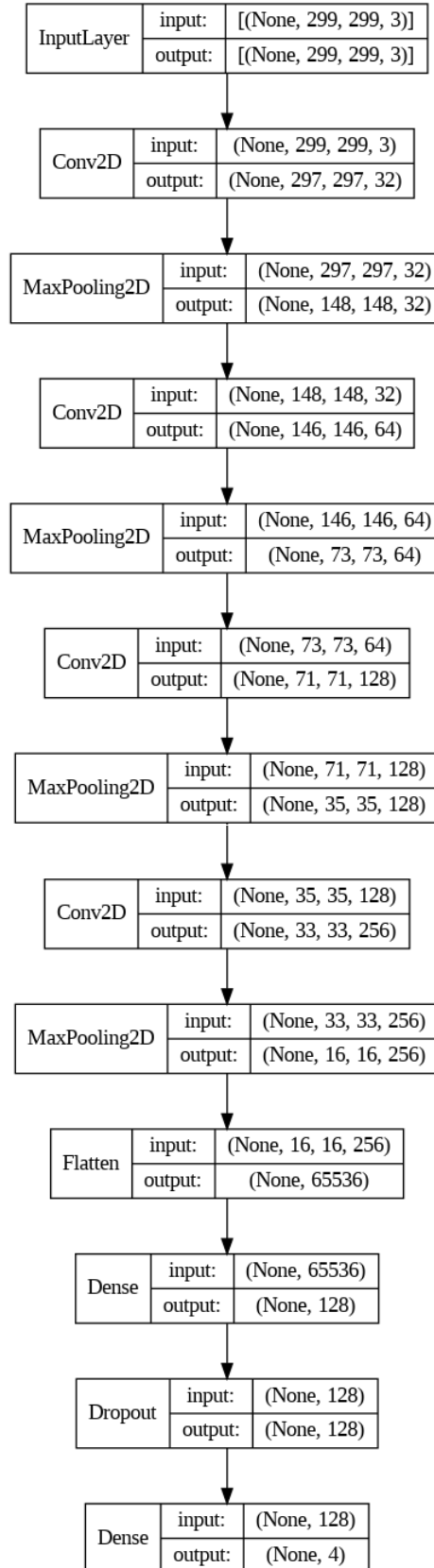
Figure 4.12. Proposed CNN architecture for RWIS image classification

Performance Evaluation and Comparison. Following the same training and validation (or testing) procedures as the AVL CNN model, we found that the overall performance, including accuracy, was satisfactory at 98.46% when considering all 16 cameras, as shown in Figure 4.13. Additionally, we examined the impact of increasing the number of cameras and images on the model's performance by changing the training datasets. The results of this examination are detailed in Figure 4.14. The plot outlines the performance metrics for different numbers of cameras, ranging from 1 to 16, and their corresponding images used for testing, with key metrics including accuracy, F1 score (macro), and balanced accuracy. Accuracy can be calculated using equation (4.1), where n_c represents the total number of classes (four in this study), TP_c denotes the number of true positives for class c (instances correctly predicted as belonging to class c), and FN_c signifies the number of false negatives for class c (instances that belong to class c but were incorrectly predicted as belonging to another class).
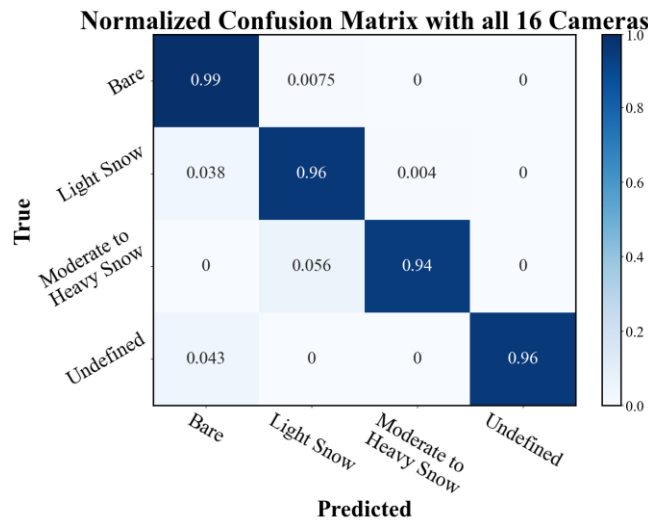
Figure 4.13. Normalized confusion matrix demonstrating overall model performance
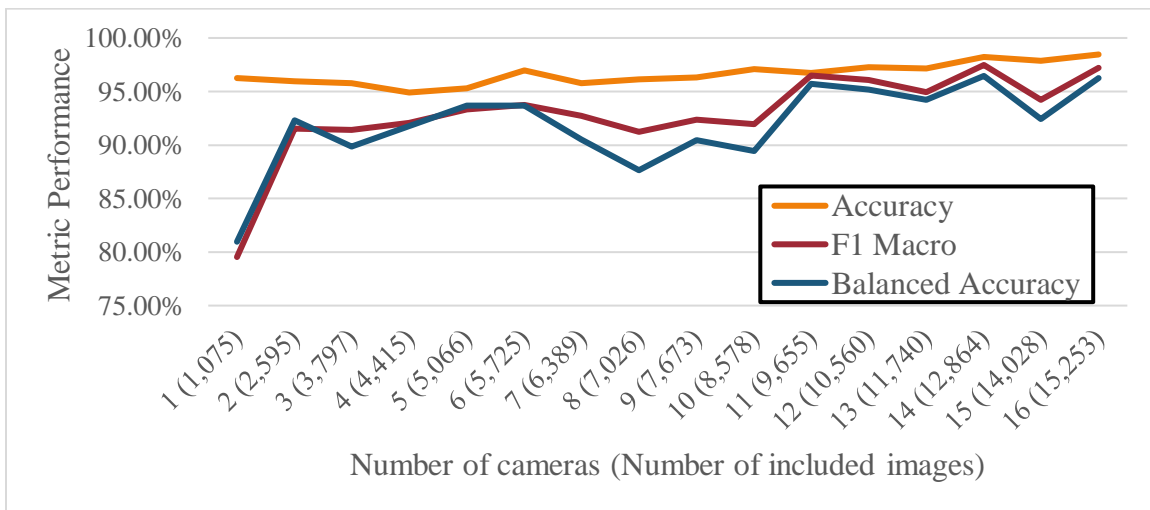
Figure 4.14. Impact of increasing number of cameras and images on performance

45

$$Accuracy = \frac{\sum_{c=1}^{n_c} TP_c}{\sum_{c=1}^{n_c}(TP_c+FN_c)} \qquad (4.1)$$

F1 score (macro) and balanced accuracy were utilized to address the unbalanced nature of the dataset. Balanced accuracy is a metric designed to evaluate classification algorithms, particularly in situations involving unbalanced datasets. It addresses two main issues that arise with conventional accuracy measurements: lack of meaningful confidence intervals and potential biases when dealing with unbalanced data (Brodersen et al. 2010). The formula for balanced accuracy is given by equation (4.2):

$$Balanced\ Accuracy = \frac{1}{n_c}\sum_{c=1}^{n_c}\frac{TP_c}{TP_c+FN_c} \qquad (4.2)$$

Precision and recall are fundamental metrics used to evaluate the performance of a classification model. Precision for class c is defined as the ratio of true positives to the sum of true positives and false positives, as shown in equation (4.3).

$$Precision_c = \frac{TP_c}{TP_c+FP_c} \qquad (4.3)$$

Recall for class c is the ratio of true positives to the sum of true positives and false negatives, as shown in equation (4.4).

$$Recall_c = \frac{TP_c}{TP_c+FN_c} \qquad (4.4)$$

The F1 macro score is another useful metric for evaluating the performance of classifiers in multilabel settings, as it considers both precision and recall for each class, inherently favoring rare labels (Lipton et al. 2014). The F1 score for each class $F1_c$ is defined as the harmonic mean of precision and recall, combining both metrics into a single measure, as shown in equation (4.5).

$$F1_c = \frac{2\cdot Precision_c\cdot Recall_c}{Precision_c+Recall_c} \qquad (4.5)$$

The F1 macro score is the unweighted mean of the F1 scores for each class, treating all classes equally. The formula for F1 macro is given by equation (4.6).

$$F1_{Macro} = \frac{1}{n_c}\sum_{c=1}^{n_c}F1_c \qquad (4.6)$$

Using different metrics provides a more comprehensive understanding of our RWIS CNN model's performance compared to relying on a single metric. Figure 4.14 illustrates performance trends with varying numbers of images and cameras, showing a general improvement as both increase. The experiment achieved its highest accuracy of 98.46% using all 16 cameras, highlighting the benefits of diverse data sources. In this configuration, the F1 score (macro) and

Balanced accuracy were 97.19% and 96.28%, respectively. Intriguingly, the model with 14 cameras outperformed this in balanced accuracy and F1 score (macro), achieving an F1 score (macro) of 97.48%, a balanced accuracy of 96.46%, and an accuracy of 98.21%. This counterintuitive result suggests that additional cameras with limited instances of minor classes, such as the "Moderately to Heavily Snow Covered" class, may have confused the model or reduced its discriminative power due to newly introduced orientations of these minor classes. This finding highlights the importance of both the quality and distribution of data sources, not just their quantity, in enhancing the CNN model's performance for monitoring winter road surface conditions. Despite these nuances, the incremental improvements with more cameras underscore the overall value of data diversity and quantity in achieving robust and reliable model outcomes.

RWIS CNN Reliability Investigation via CAM-based Methods. To explore and reveal the hidden aspects of our RWIS CNN model, we applied Grad-CAM, Grad-CAM++, and Score-CAM to the testing images. SHAP was not applied here due to the ever-changing camera angles of RWIS cameras. CAM-based methods are usually employed to verify whether the model's focus aligns with human recognition and does not primarily target irrelevant areas, such as the sky or trees, in this case. By aligning the CNN model's decision-making process with human recognition, we aim to enhance the reliability of autonomous CNN-based RSC monitoring systems. The three CAM-based methods were applied to all of the testing images to calculate pixel-wise CAM scores for each technique. These scores were used to generate saliency maps with 50% transparency, integrating them equally with the original images. Manual masking was then applied to the main traffic lane, as illustrated in Figure 4.15, to refine our analysis and quantify each CAM technique's attention. For consistent comparison, the image-wise CAM scores of each method were normalized to a scaled value of 1,000. This normalization process ensures consistent pixel-wise attention weight, facilitating human interpretation by providing a standardized measure of where the model concentrates its attention.

Figure 4.15. Application of manual masking to images from cameras with consistent angles for enhanced evaluation of CAM techniques

The localized CAM score was calculated by summing the total CAM scores within the manually masked area (main pavement section) and dividing this sum by the normalized score of 1,000. The pavement area percentage was calculated by dividing the number of pixels in the masked area by the total number of pixels in the image. For instance, an image with a total of 430.1 CAM scores inside the masked area results in a focus of 43.01%. For camera IDOT-030-01, the highlighted mask covered a total of 31,353 pixels. Dividing the 31,353 pixels by the total number of pixels in the image (299 x 299 = 89,401) results in a target area percentage of 35.07%.

As depicted in Figure 4.16, there is a weak positive correlation ($R^2 = 0.1477$) between the average of the three localized CAM scores for each camera focusing on the pavement area and the percentage of the pavement within the image. Intuitively, camera angles with higher proportions of the main pavement section should have localized CAM scores concentrated in this area. While the trend generally indicates higher localized CAM scores corresponding to increased pavement area coverage, certain cameras, such as IDOT-047-00, IDOT-047-01, and IDOT-047-02, exhibit an even greater concentration on the pavement area. Conversely, other cameras, including IDOT-047-05 and IDOT-072-00, display the opposite behavior, with lower localized CAM scores relative to the percentage of pavement area. After excluding these five cameras, there is a strong positive correlation ($R^2 = 0.8645$) between the average of the three CAM scores for each camera focusing on the pavement area and the percentage of pavement within the image.
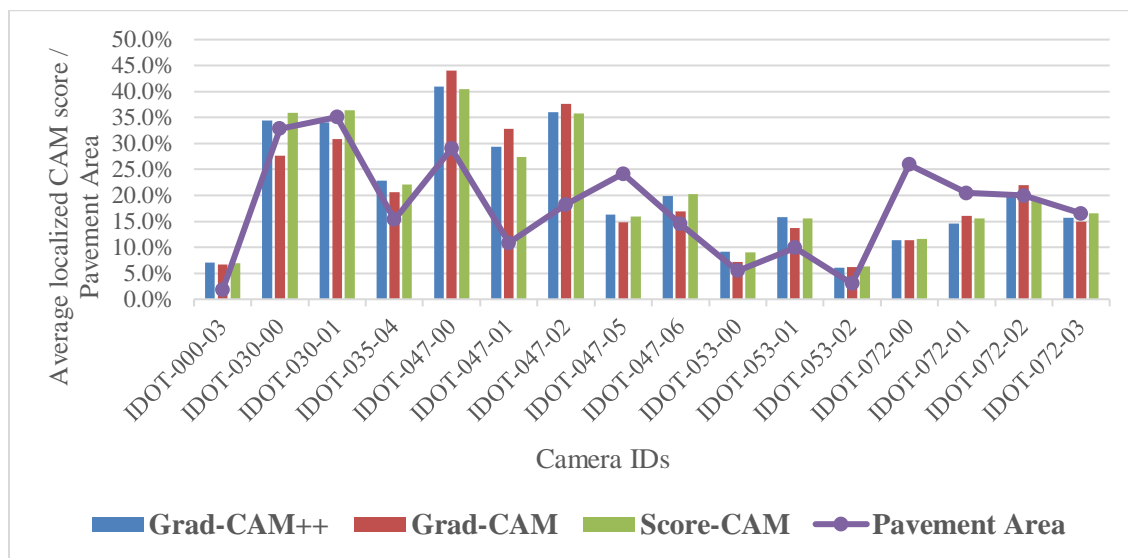
Figure 4.16. Comparative analysis of Grad-CAM, Grad-CAM++, and Score-CAM in highlighted pavement areas across different camera angles

This discrepancy can be further understood by visually examining the saliency maps generated, as shown in Figure 4.17. For instance, cameras IDOT-047-00, IDOT-047-01, and IDOT-047-02 capture images of the same road from different angles, providing diverse features for the model to learn from and use in predictions, thereby enhancing its focus on the pavement area. In contrast, cameras IDOT-047-05 and IDOT-072-00 include complex bridge structures within their frames. This added complexity may cause the CNN model to struggle to accurately target the

primary area of interest, resulting in lower localized CAM scores despite significant pavement coverage. These observations highlight specific limitations of the CNN model when dealing with intricate structures in the image background, suggesting that additional training or alternative approaches may be necessary to improve its accuracy and reliability in such scenarios.

IDOT-047-00

IDOT-047-01

IDOT-047-02

IDOT-047-05

IDOT-072-00
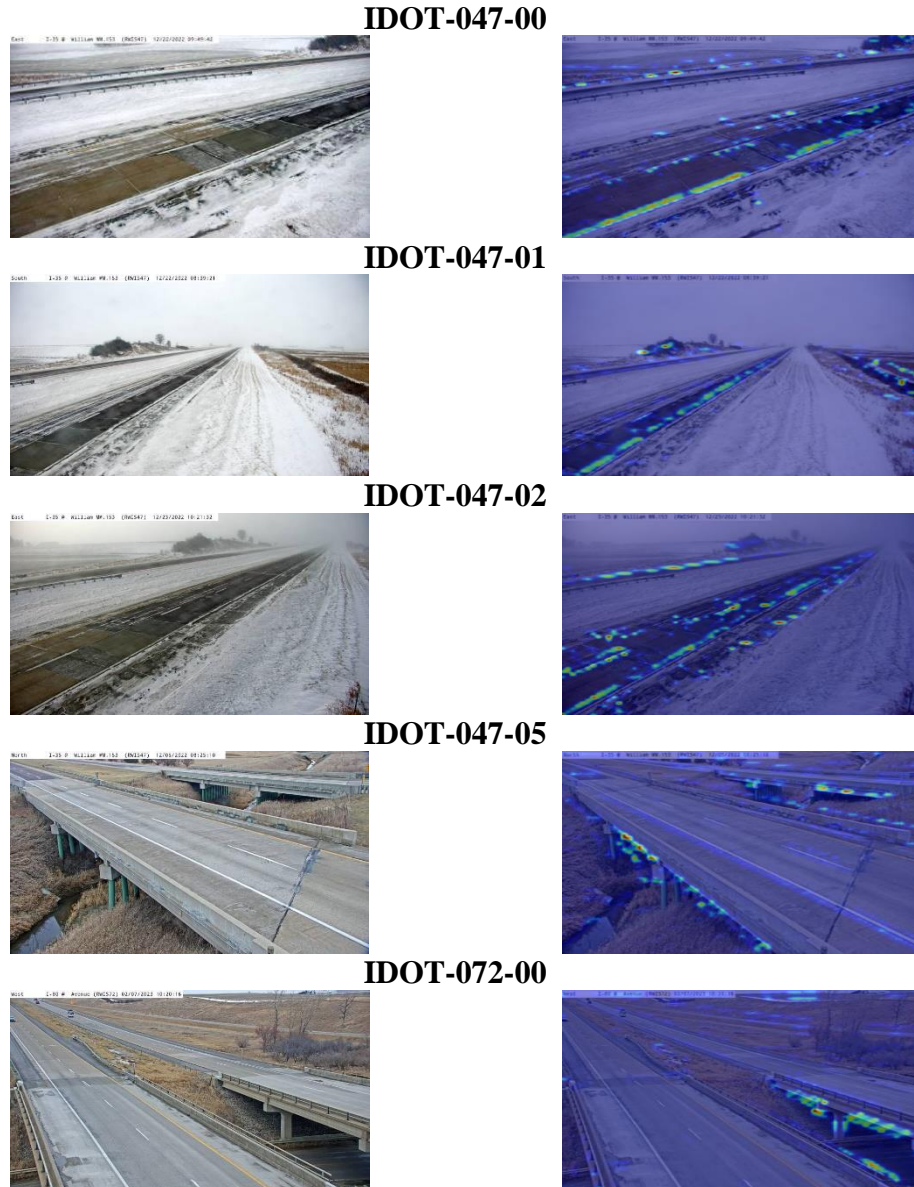
Figure 4.17. Cameras exhibiting exceptionally high and low CAM scores with original and their Grad-CAM images

The comparative analysis demonstrates that Grad-CAM, Grad-CAM++, and Score-CAM effectively visualize the localization of the CNN's decision-making process in relation to pavement regions for RSC classification. As shown in Figure 4.18, Grad-CAM++ achieved an average localized CAM score of 20.86% on the pavement area, Grad-CAM achieved an average localized CAM score of 20.21%, and Score-CAM achieved an average localized CAM score of

49

20.95%, whereas the average pavement area across all camera angles was 17.70%. These results indicate that our proposed CNN model can effectively prioritize the pavement area for decision-making, surpassing the baseline average pavement area and thereby enhancing the reliability of the decision-making process.
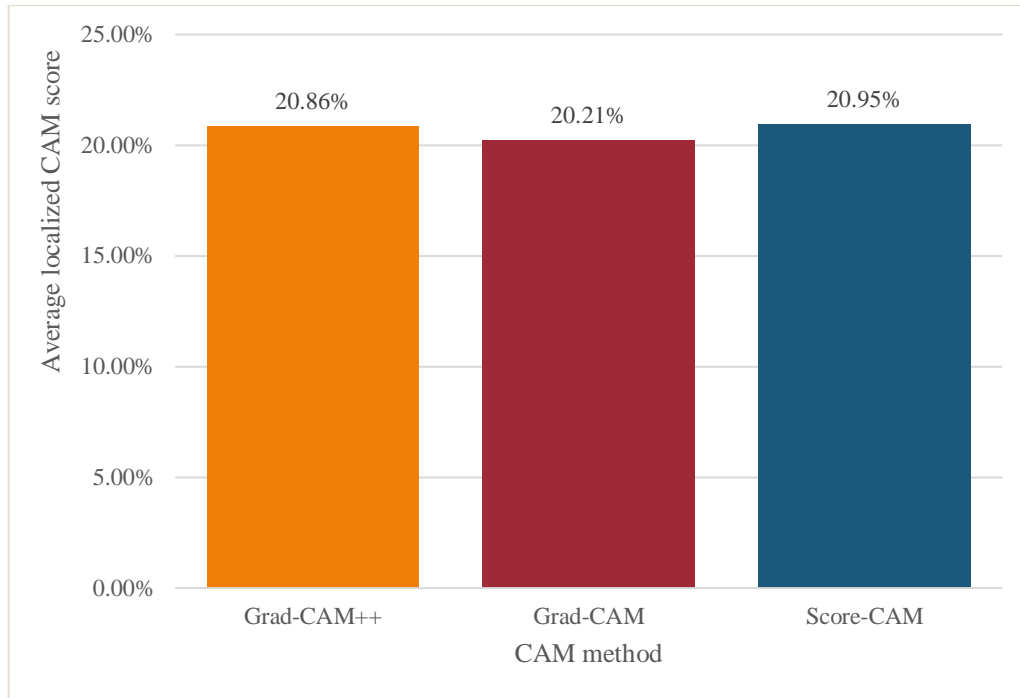
Figure 4.18. Average localized CAM scores on pavement areas by Grad-CAM, Grad-CAM++, and Score-CAM across all cameras

Figure 4.19 provides examples of saliency maps produced by each CAM technique, comparing the various techniques. Key findings indicate that cameras capturing the same road from different angles produce higher localized CAM scores targeted toward the main pavement, which translates to better reliability that aligns with human interpretation. In contrast, cameras with complex structures yield lower localized CAM scores, revealing the limitations of CNNs in focusing solely on important areas. These results suggest the need for additional training with various angles to improve humanlike interpretation and for caution when dealing with structures such as bridges. Overall, adopting CAM techniques such as Grad-CAM, Grad-CAM++, and Score-CAM enhances the reliability and interpretability of CNNs for RSC monitoring. These methods provide accurate visual explanations, revealing that our model prioritizes the pavement area effectively. The integration of CAM techniques improves CNN transparency and ensures more effective deployment in winter road maintenance.

Original Image

Grad-CAM

Grad-CAM++

Score-CAM
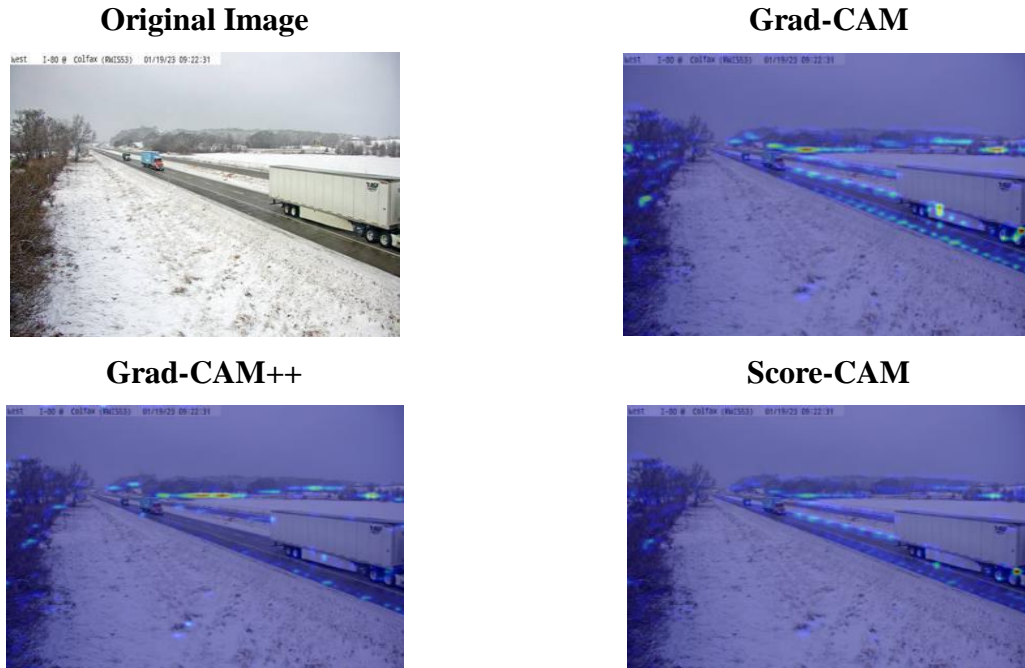
Figure 4.19. Comparison of various class activation map techniques after normalization

4.1.2.2 Snow Coverage Ratio Estimation

In addition to using CNNs to classify RWIS imagery into general categorical variables, this project also developed pix2pix GAN and SS models to estimate SCR for each RWIS image.

Winter-to-Summer Image Translation. The pix2pix GAN model was first utilized to translate the winter images (collected by RWIS cameras) into summer images. Much like all other ML or DL techniques, our dataset has 6,566 pairs of images split into a training (80%, 5,252 pairs) and validation dataset (20%, 1,314 pairs) for model development and performance evaluation, respectively (Park et al. 2018). Figure 3.3 depicts two examples of image pairs from the training dataset, and Figure 4.20 shows the training profile of the developed model.
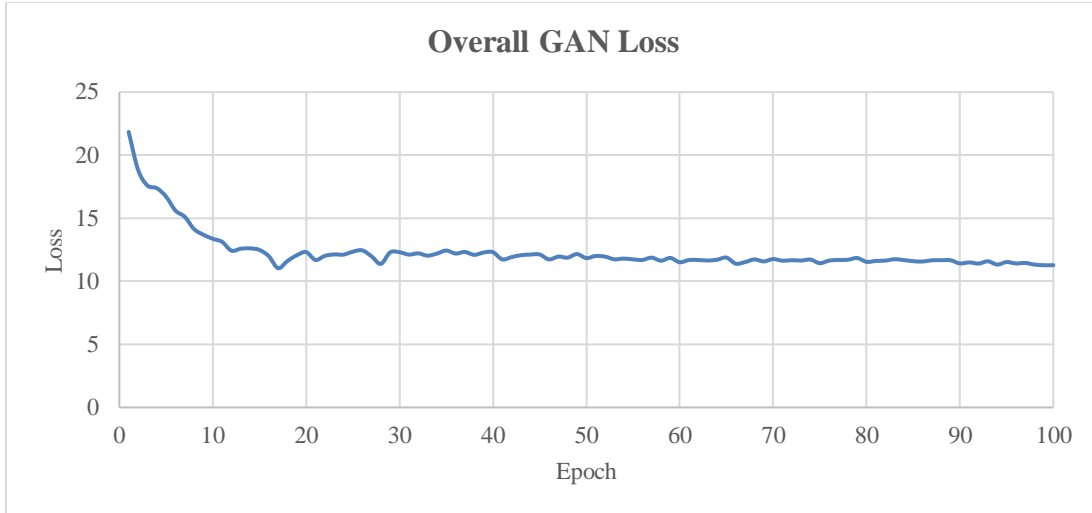
Figure 4.20. Pix2pix GAN training profile

As seen in Figure 4.20, the loss values generally decrease as the number of training epochs increases. After 20 epochs, the loss reduction became marginal. To evaluate the image translation performance of the developed pix2pix GAN model, images from the validation dataset were used to visually check if the model effectively translated the new images (not seen before by the model during training). As an example, Figure 4.21 depicts two sets of images from the validation dataset, where the source image is the input winter image, the target image is the source image's paired summer image (same RWIS camera at the same angle), and the generated image is the fake summer image translated by the developed pix2pix GAN. Based on both Figure 4.20 and Figure 4.21, it can be concluded that the developed pix2pix GAN model was successful in translating the winter images to their corresponding summer images. These results can be used further for drivable area detection via SS with the proposed snow coverage ratio calculation procedure.

Source Image	**Target Image**	**Generated Image**

Figure 4.21. Examples of source, target, and generated images for validating pix2pix GAN

Drivable Area Detection via Semantic Segmentation. The second DL model developed for SCR estimation was SS, which is aimed at identifying the drivable areas for each input RWIS image (both winter images and translated summer images). As mentioned earlier, to develop the SS model, over 3,460 images (2,814 wintertime, 646 summertime) were carefully annotated. Similar to the pix2pix GAN process, these images were also split into training (80%) and validation datasets (20%). During the image annotation process, only two labels were used: Drivable Areas (or road) and Others (background/nondrivable areas, such as snowy areas, sky, trees, etc.).

It is important to note that two batches of RSC images were collected and annotated for SS model development. The first batch included 1,445 images (963 wintertime, 482 summertime) collected during February 2019, December 2019, January 2020, and March 2020 for winter images, and during June 2019 and October 2019 for summer (or bare condition) images. The second batch included 2,015 images (1,851 wintertime, 164 summertime) collected during January 2022, June 2023, November 2023, and December 2023. Due to the different winter and summer seasons, the camera angles varied between the two batches. In addition, the second batch was more evenly collected from different camera angles, with a deliberate selection process to avoid images containing confusing objects, such as large vehicles or trees, in the main parts of the images. Furthermore, all images in the second batch were manually classified into

Bare, Partially Snow Covered, and Fully Snow Covered conditions for a more comprehensive analysis and comparison.

The first batch was primarily used to demonstrate the feasibility of our SCR estimation approach, particularly for the SS model, while the second batch was used for further validation and more detailed analysis.

Figure 4.22 shows two examples of source images and their corresponding annotated images. These SS models were developed separately for winter and summer images using the first batch of imagery. Figure 4.22 also presents examples of the SS-detected drivable areas. Ultimately, the SS models perform very well in identifying drivable areas in both winter and summer environments.

Figure 4.23 and Figure 4.24 show the training profiles of the two SS models. The training and validation loss values generally decreased during the training process, while the training and validation accuracies, including IoU score, F1 score, and overall pixel accuracy, all increased. By the end of the training process, the three metrics of the validation datasets for winter and summer models were 95.6%, 97.7%, 97.0%, and 95.2%, 97.4%, 97.1%, respectively.

Source Image	Annotated Mask	Predicted Mask
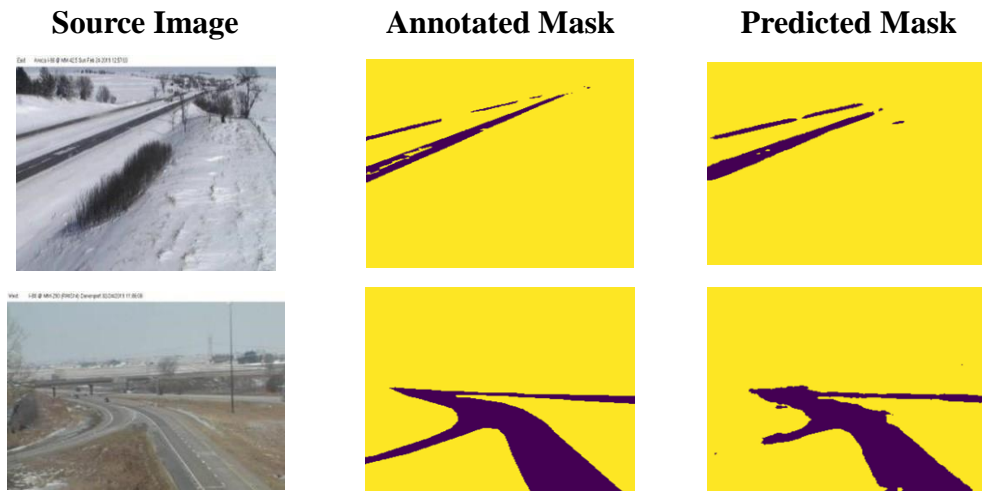
Figure 4.22. Examples of annotated images for developing semantic segmentation
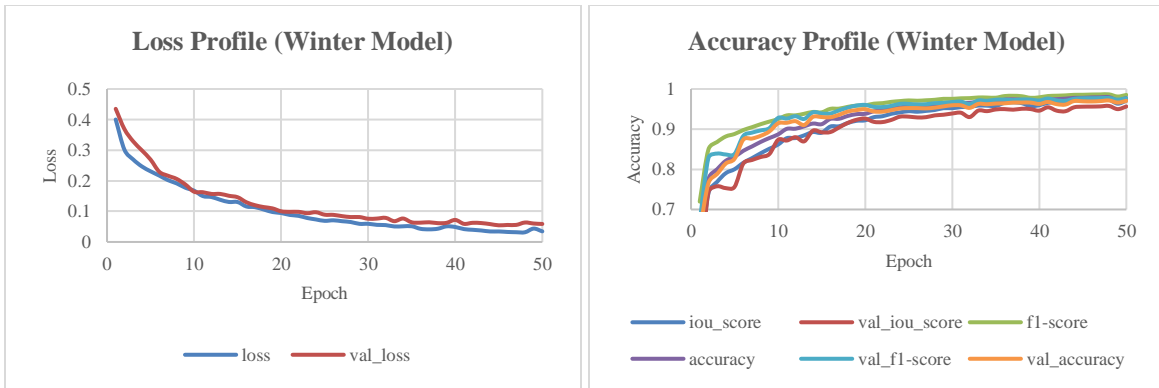
Figure 4.23. Semantic segmentation (winter model) training profile
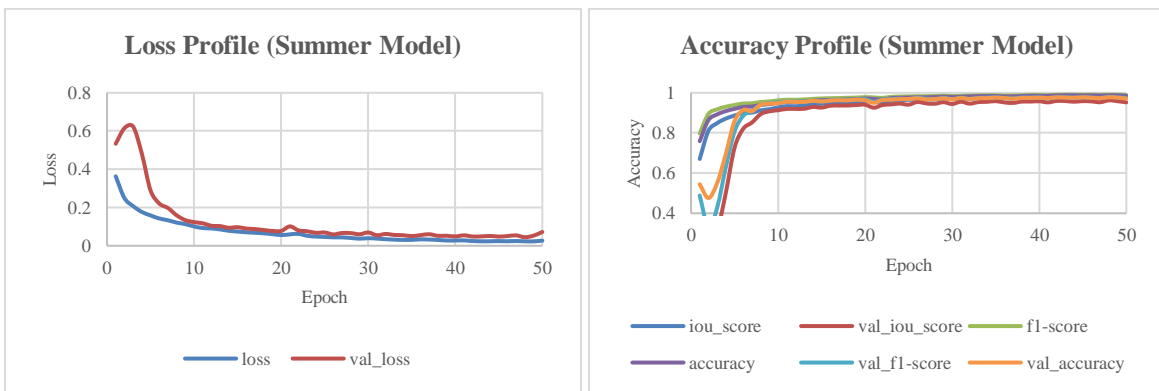
Figure 4.24. Semantic segmentation (summer model) training profile

The same process was applied to the second batch of images. Given the similar accuracies of the winter and summer models and the smaller number of summer images in the second batch compared to the first, a single SS model was developed for the second batch, integrating both seasons. This model achieved an IoU score of 98.5%, an F1 score of 99.3%, and an overall pixel accuracy of 99.0%.

Since the second batch of images was collected more evenly from different camera angles, the impact of training sample size and camera angles on SS model performance was investigated. The goal was to evaluate the sensitivity of the SS model to the size of the training sample and to determine whether the model could accurately detect drivable areas in images from previously unseen camera angles.

To assess the impact of training samples, all images in the second batch were evenly shuffled and then split into training and validation datasets. The validation dataset consistently comprised the same 20% of the images, while the training dataset was incrementally expanded (to 50%, 60%, 70%, and 80%) during the training process.

To assess the impact of camera angles, a similar process was followed, but with a key difference: images from different camera angles were not mixed. The split between training and validation

55

datasets was strictly based on camera angles, meaning that no images from the same camera angle appeared in both the training and validation datasets. Furthermore, images added incrementally during training came from different camera angles than those already included.

Figure 4.25 illustrates the effect of varying the number of training samples and camera angles on model performance. As shown, the SS model accuracy generally improves with more training samples and with greater diversity in camera angles. However, comparing the actual results reveals that camera angles have a more significant impact on SS model performance. It is more challenging for the model to accurately detect drivable areas in images from unseen camera angles, though the accuracy remains acceptable, with all results exceeding 85% when 80% of the images from different camera angles were included in training.

These findings suggest that enhancing the SS model requires not only a larger number of training images but also a greater variety of images from different camera angles.
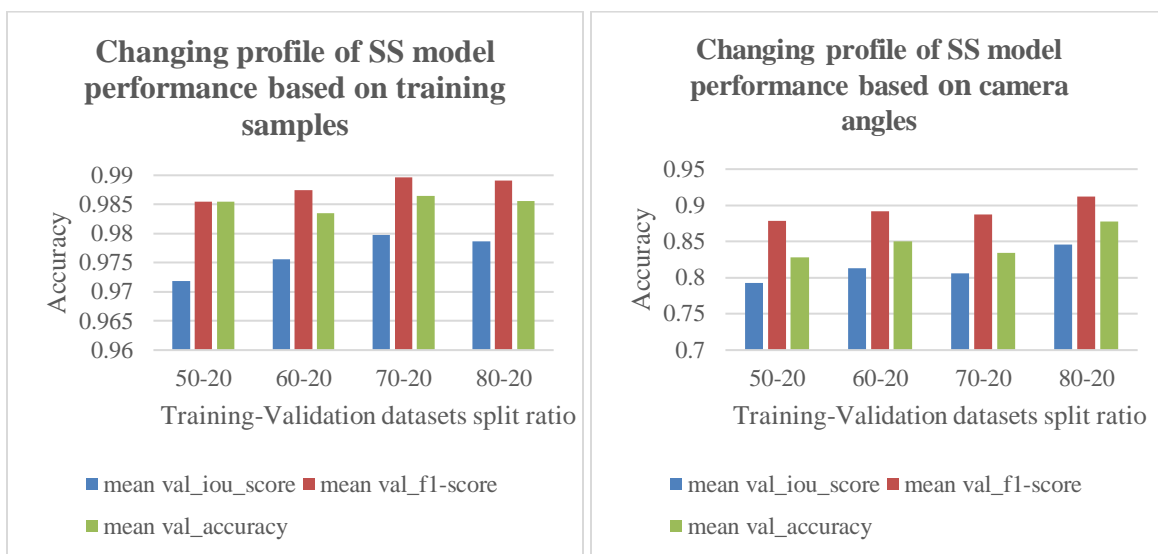
Figure 4.25. Changing profiles of SS model performance based on different training samples (left) and camera angles (right)

Road Surface Conditions Recognition. The last step in our proposed approach was calculating the snow coverage ratio after comparing the drivable areas in the input and pix2pix-translated images using the SS models. All paired images involved in the development of the pix2pix GAN models were utilized to evaluate the accuracy of the final output produced by the proposed approach (i.e., image translation followed by drivable area detection). The snow coverage ratios calculated using the SS-segmented source images and target images were considered as observed values (i.e., the ground truth). The ratios obtained via the same method using the source images and translated images were considered as the estimated results that required evaluation.

Using the SS models developed based on the first batch of images (for SS development), Table 4.3 compares the descriptive statistics of the observed and estimated snow coverage ratios of the 1,314 pairs of images in the validation dataset. The mean absolute error (MAE, shown in

equation (4.7)) and root mean square error (RMSE, shown in equation (4.8)) describe the closeness between the two datasets. Furthermore, a t-test was also conducted to evaluate whether the estimated values are statistically different from the observations. Results of the t-statistic and Pearson correlation values indicate that the estimated snow coverage ratios have a strong linear correlation with the observations.

Table 4.3. Descriptive statistics of observed and estimated snow coverage ratios

Statistics	Observed	Estimated
Minimum	0	0.043
Maximum	0.999	0.999
Mean	0.627	0.624
Standard Deviation	0.008	0.008
Sample Variance	0.062	0.053
MAE	0.047	
RMSE	0.086	
t-statistic	1.306 (not significant)	
Pearson Correlation	0.938	

$$MAE = \frac{\sum_{i=1}^{N}|V_{i,estimated} - V_{i,observed}|}{N} \tag{4.7}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(V_{i,estimated} - V_{i,observed})^2}{N}} \tag{4.8}$$

where V is the variable that needs to be estimated (snow coverage ratio in this study), i is the sequence number of the estimated/observed sample, and N is the total number of samples that need to be estimated.

Figure 4.26 shows two examples of the final outputs from the proposed approach. The input source images, our proposed method's target outputs, are displayed in the "Target" row, and the real outputs generated from our proposed approach are shown in the "Generated" row. Based on the model performance results and depicted model output images, it is evident that the proposed method is feasible and shows great promise as an automated tool for estimating snow coverage ratios. However, the currently developed models still have room for improvement in image translation and drivable area detection. For example, the translated images should have fewer blurry regions, and the drivable areas should not include vegetation near the boundaries between the pavement and the external environment, especially in the winter images.
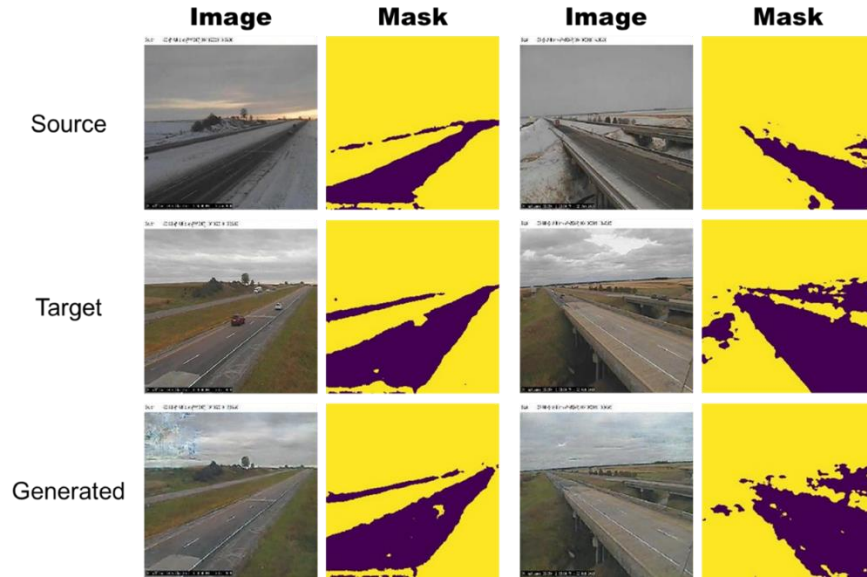
Figure 4.26. Examples of final outputs of proposed approach

While the SS model developed using the second batch of images showed similar or even improved results and given that the RSC labels for the second batch were manually classified, a further analysis was conducted. This analysis compared the estimated SCRs with the manual classification labels of the second batch of images to verify that the estimated SCRs accurately reflect the actual snow coverage status of the road sections based solely on RWIS imagery. Figure 4.27 depicts histograms of the Bare and Partially Snow Covered condition images within the validation dataset along with their corresponding estimated SCRs. For the Fully Snow Covered condition, which contained only one image, the SCR was estimated at 0.987 (or 98.7%).
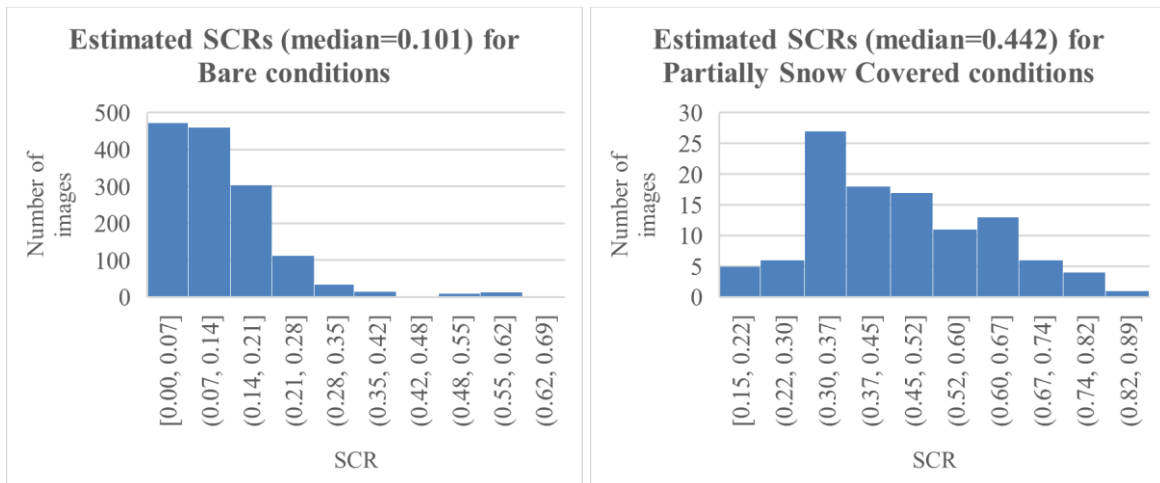
Figure 4.27. Comparisons between estimated SCRs and RSC labels

As shown, the median value of the estimated SCRs for Bare conditions was at 0.101 (or 10.1%), which is a reasonable estimation, considering that light snow on pavement, which does not affect traffic, would still be classified as Bare conditions. Additionally, white lane markings or light

reflections may be detected as nondrivable areas by SS, preventing the estimated SCR from being perfectly 0. Similarly, the median estimated SCR for Partially Snow Covered conditions, at 0.442 (or 44.2%), also aligns with expectations. Although the number of Fully Snow Covered conditions in the annotated images is limited, the single result obtained is consistent with expectations.

In conclusion, the estimated SCRs generated using the pix2pix GAN and SS models are generally reliable. However, there is potential for further improvement of these DL models, and incorporating additional datasets in the future could further validate and enhance the results.

4.2 Spatial Mapping of Road Surface Conditions

As noted earlier, this project also aimed to develop spatial interpolation methods for both categorical and continuous RSC variables to fill in gaps in unmonitored areas along the highway network.

4.2.1 Categorical RSC Interpolation via NIK

NIK was developed specifically to interpolate the categorical RSC variables by directly taking the CNN classification results as its input. To demonstrate its feasibility and robustness, AVL imagery data collected during winter seasons from 2009 to 2024 were converted into RSC variables via CNN and processed according to the procedures outlined in Wu et al. (2022b), Wu et al. (2022c), and Wu and Kwon (2024) for developing semivariograms and performing NIK interpolation. All CNN-classified imagery points were treated as ground truth and spatially aggregated every 500 m to minimize random classification errors, then temporally aggregated into hourly events. To ensure the quality of our developed models (i.e., the semivariograms and NIK interpolation results), qualified hourly weather events were selected based on sufficient spatial coverage and the number of each RSC category by AVL collected imagery points. Specifically, only events with imagery points covering at least half of each quarter (i.e., north I-35, south I-35, west I-80, east I-80) of the study area and more than 30 for each RSC category in each highway quarter were included in the analysis. Ultimately, a total of 20 weather events were selected for further analysis.

Following the procedures outlined in the methodology chapter, 40 semivariograms were constructed, and NIK was employed to cross-validate the estimation results using the leave-one-out method (Wong 2015). The cross-validation accuracy of the first layer of NIK (i.e., Bare versus Partially Snow Covered and Fully Snow Covered) ranged from 58.4% to 98.9%, with an average accuracy of 73.5%. For the second layer (i.e., Partially Snow Covered versus Fully Snow Covered), accuracy ranged from 69.2% to 94.3%, averaging 86.0%. The overall accuracy, combining results from both layers (i.e., Bare versus Partially Snow Covered versus Fully Snow Covered), ranged from 58.8% to 85.7%, with an average accuracy of 67.5%. Some hourly events exhibited relatively low accuracy in either the first or second layer, which can be attributed to the lack of variety in RSC samples. When one RSC category predominates over the others, and/or their locations are not well-distributed, estimation becomes challenging. Another possible reason could be the CNN classification accuracy, as not all images used in this NIK interpolation were

validated with high accuracy for the CNN model. Inaccurate CNN classification results could lead to poor estimation performance. However, several hourly events demonstrated accurate estimation results, indicating that the proposed NIK is capable of providing accurate estimations for unmonitored areas. Figure 4.37 later in this chapter shows an example of a map of real-time RSC estimation via NIK.

4.2.2 Continuous RSC Interpolation via the Refined RK

In terms of continuous RSC interpolation, we employed our previously developed RK method. However, to make it more universally applicable, the RK method was refined by adding a weather characterization step before implementing the RK interpolation.

Table 3.1 shows that the collected RST values exhibited significant differences, which could be attributed to the direction of the road (Olea 2006, Wu et al. 2022a). To account for the potential effect of road direction on the spatial variation of RSC within a highway network, and to obtain more accurate estimates of RSC for each highway, separate analyses for I-35 and I-80 were conducted. For each weather event, two MLRs and semivariograms were constructed, one for each highway. This approach was based on our own experimentation and previous research findings (Wu et al. 2022a).

The K-means algorithm was run for a range of clusters (i.e., the number of weather events' groups), and their corresponding inertia was calculated. Inertia gauges the effectiveness of K-means clustering by measuring the sum of squared distances between each data point and its centroid within a cluster (Gupta et al. 1999). To determine the number of clusters (i.e., the K values), the change ratios of the inertia versus the number of clusters were also obtained; a change ratio nearing 0 indicates diminishing returns with additional clusters. Considering that more clusters imply more MLRs and semivariograms leading to less generalization of the clustered weather events, an additional consideration was introduced. To incorporate this into the determination of the K values, a line representing the number of MLRs and semivariograms versus the number of clusters was also generated, which can be considered as a line of the "cost." As depicted in Figure 4.28, all three lines share the same x-axis values, and their corresponding y-axis values were normalized between 0 and 1 to enforce equitable comparisons. Taking I-35 as an example, the inertia change ratio reveals diminishing returns at around 16, while the line of MLRs and semivariograms and the inertia line intersect at around 14. Thus, we selected 15 as the K value for I-35. Following a similar analysis, 16 was chosen as the K value for I-80. These selections imply constructing 15 and 16 pairs of MLRs and semivariograms for I-35 and I-80, respectively.
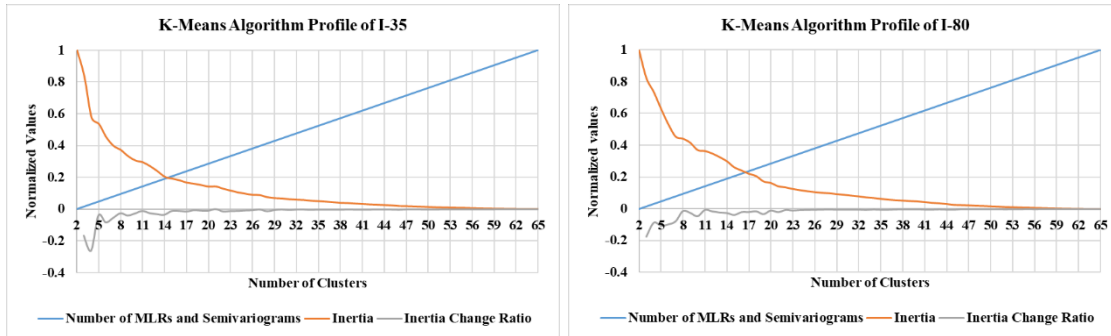
Figure 4.28. *K*-means algorithm profiles

Examples of grouped weather events (grey lines) with their centroids (red lines) are shown in Figure 4.29. All calibrated MLRs and semivariograms are outlined in Table 4.4. The significant variables of the MLR models remained the same as what we had found in our previous studies, and their signs of coefficients made intuitive sense (Wu et al. 2022b). Both latitude and elevation were found to have significant negative correlation with the RST values for most clusters, indicating that RSTs tend to decrease in higher and northern areas. Nevertheless, a comparison of the R^2 value for each model revealed significant variations in the predictive capabilities of the MLR models. This variability implies that weather events within one cluster can be vastly different from those in others, underscoring the need for the application of kriging for further estimations to enhance the accuracy of our final estimates. The semivariogram parameters of each cluster were found to be different as well, lending further support to the conclusion that the spatial variation pattern of weather events is cluster specific. To ensure the accuracy of RST interpolated through RK for validation events, the MAE values were obtained by a cross-validation process as defined in equation (4.7). These values were equal to 0.696°C and 0.619°C, respectively, and were considered accurate estimates.
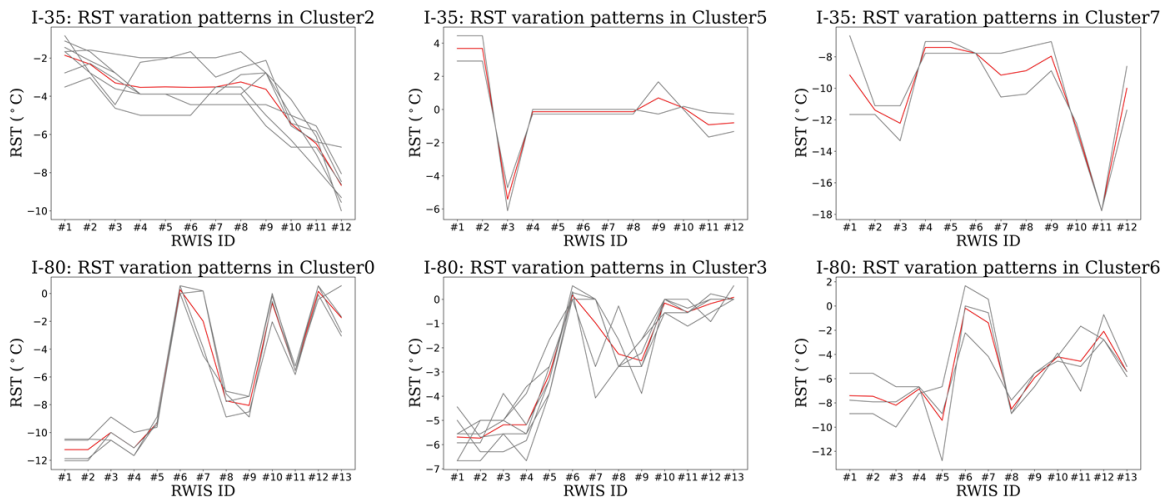
Figure 4.29. Examples of grouped weather events

61

Table 4.4. Calibrated MLR and semivariogram models for all clusters

Cluster ID	I-35					I-80				
	Significant Variables	R^2	Nugget	Range (m)	Sill	Significant Variables	R^2	Nugget	Range (m)	Sill
1	Latitude (-), Elevation (-)	18.4%	1.09	1.5E+05	1.61	Elevation (-)	67.2%	0.28	1.3E+05	7.22
2	Latitude (-), Elevation (-)	74.3%	0.34	1.5E+05	0.60	Elevation (-)	17.3%	0.03	5.1E+04	0.72
3	Latitude (-), Elevation (-)	84.2%	0.36	1.6E+05	0.62	Elevation (-)	13.8%	0.35	5.0E+04	1.45
4	Latitude (-)	36.9%	0.00	4.9E+04	0.88	Elevation (-)	61.6%	0.00	6.0E+04	1.49
5	Latitude (-)	14.5%	0.27	1.5E+05	0.85	Latitude (-), Elevation (-)	43.2%	0.62	1.8E+05	1.49
6	Latitude (-)	35.6%	0.00	1.5E+05	4.98	Latitude (-), Elevation (-)	7.3%	0.13	1.7E+05	3.52
7	Latitude (-), Elevation (-)	61.8%	0.12	3.9E+04	1.48	Elevation (-)	40.3%	2.47	2.2E+05	5.28
8	Latitude (-), Elevation (-)	25.6%	1.71	8.2E+04	4.95	NA	NA	0.00	5.7E+04	1.31
9	Latitude (-), Elevation (-)	68.6%	0.18	3.9E+04	0.59	Latitude (-), Elevation (-)	8.6%	1.48	1.7E+05	3.80
10	Latitude (-)	23.3%	0.04	3.8E+04	0.80	Elevation (-)	55.0%	0.00	1.3E+05	5.96
11	Latitude (-)	56.7%	1.01	1.6E+05	2.18	NA	NA	0.21	1.8E+05	0.41
12	Latitude (-)	47.8%	1.18	3.9E+04	7.67	NA	NA	0.17	4.6E+04	0.80
13	Latitude (-)	23.7%	0.00	2.6E+04	15.84	Elevation (-)	53.0%	0.06	6.1E+04	1.09
14	Latitude (-)	19.7%	8.58	1.4E+05	11.33	Latitude (-)	8.0%	0.57	2.3E+05	0.73
15	Latitude (-)	11.3%	0.59	1.5E+05	3.90	Elevation (-)	51.0%	0.23	5.9E+04	1.49
16						Elevation (-)	47.6%	0.82	7.9E+04	2.40

4.3 Benefits Quantification (RST as the Example)

As mentioned before, one of the novel objectives of this study was to evaluate the potential benefits of having spatially rich RSC data (using our refined RK method), as this aspect has received little attention in previous research. Following the methodology outlined in the previous section, we computed the lowest, average, and highest MAE and weighted MAE (wMAE) values, defined in equation (4.9), for each possible RWIS station configuration, ranging from one to the total station count. Both the developed RK method and the assumed current practice naïve method were employed in this evaluation. Compared with MAE, wMAE introduces different weights to various RST ranges, providing a nuanced assessment of model performance (Hao and Li 2020, Cleger-Tamayo et al. 2012). Recognizing the sensitivity of WRM activities to RST around the freezing point (0°C), a sensitive range of -2°C to 2°C was selected, with a weight of 100 assigned for subsequent comparisons and analysis. Note that the choice of sensitive ranges and their corresponding weight values can be adapted based on the priorities of different WRM authorities, such as municipal governments.

$$wMAE = \frac{\sum_{z=1}^{n} w_z * |pred_z - true_z|}{\sum_{z=1}^{n} w_z} \tag{4.9}$$

where w_z is the weighted value adopted for different RST ranges in this study. Specifically, if $true_z$ falls between -2°C and 2°C, then $w_z = 100$; otherwise, $w_z = 1$.

The process of quantifying benefits involved an enumerative assessment of each possible combination of multiple RWIS stations (if more than one). For example, if the implementation allowed for only 2 RWIS stations, there would be 66 potential configurations for I-35 (2 out of 12) and 78 for I-80 (2 out of 13), as illustrated in Figure 4.30. This exhaustive analysis enables us to demonstrate the most reliable results including both the accuracy of RK estimation and the quantified benefits. For larger networks with a greater number of RWIS stations, optimization methods such as heuristic algorithms can be employed to replace enumeration and expedite the computation process while maintaining accurate results (Gu et al. 2020, Kwon et al. 2017).
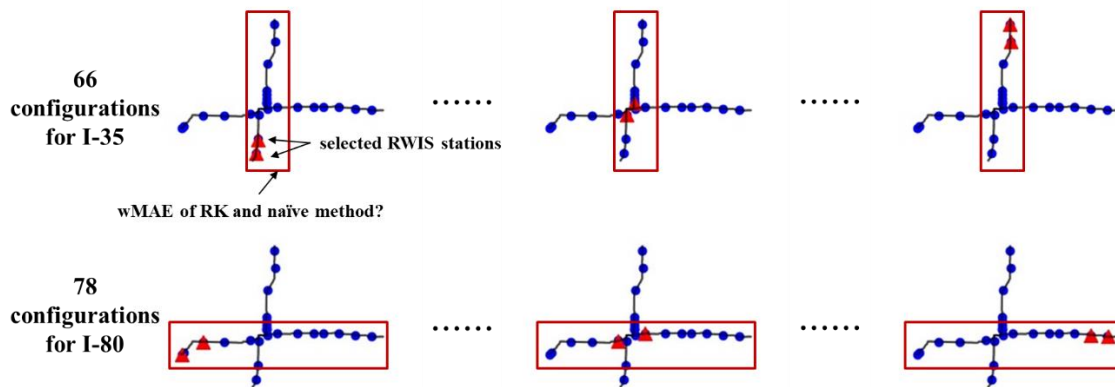

Figure 4.30. Example of enumeration and evaluation process (two RWIS stations are allowed)

As previously mentioned, to assess the potential benefits of our developed methods, the lowest, average, and highest MAE and wMAE values were calculated for all possible configurations of RWIS stations on I-35 and I-80 using our RK method and the naïve method as a benchmark. As can be seen in Figure 4.31, all of the MAE and wMAE values obtained by the RK method are lower than those obtained by the naïve method, except for a few points in the lowest wMAE category of I-80. Additionally, the differences between the best and worst results obtained by the RK method are much smaller than those obtained by the naïve method, indicating that the RK method is capable of generating reliable results even if the RWIS configuration is not well designed. Furthermore, as the number of RWIS stations used in the estimation process increases (i.e., 10 or more), the differences between the RK and naïve methods, whether for better or worse results, tend to become less significant. This is because the density of RWIS stations is sufficient to cover larger areas, thereby reducing the sensitivity of the estimation method used.
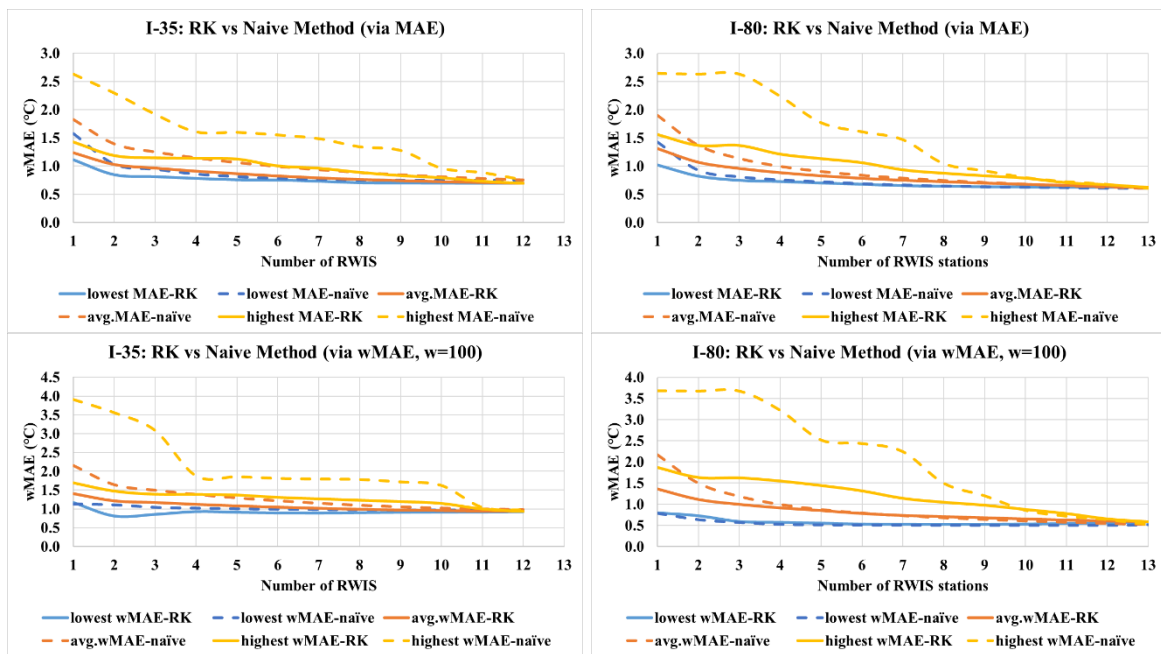
Figure 4.31. MAE and wMAE comparisons between RK and naïve method

Following the proposed benefit quantification approach, a uniform target MAE and wMAE value was adopted as the criterion to obtain the expected savings of RWIS stations, as shown in Figure 2.10. Specifically, we determined the difference in the number of needed RWIS stations between the lowest (w)MAE-RK and the highest (w)MAE-naïve as the maximum savings, the difference between the average (w)MAE-RK and the average (w)MAE-naïve as the average savings, and the difference between the highest (w)MAE-RK and the lowest (w)MAE-naïve as the minimum savings. For example, by setting a target MAE of 1℃ for I-35, the RK method could achieve this with as few as 2 RWIS stations (i.e., the best configuration), an average of 3 RWIS stations, or up to 7 RWIS stations (i.e., the worst configuration). In contrast, the naïve method required 3, 6, and 10 RWIS stations, respectively. Thus, the maximum, average, and minimum savings for this case were 8, 3, and 0, respectively. The aforementioned process was repeated for both highways using different criteria, and the expected RWIS savings are presented in Figure 4.32.

Figure 4.32. Expected savings of RWIS stations

Figure 4.32 illustrates a general trend wherein the expected number of avoided RWIS stations gradually increases as the required criteria decrease from 3℃ to 1℃. However, if the selected criteria continue to decrease below 1℃ (e.g., 0.5℃), no further savings are realized. This aligns with our earlier explanation in Figure 4.31. Once the RWIS station density becomes high enough (higher criteria require denser RWIS net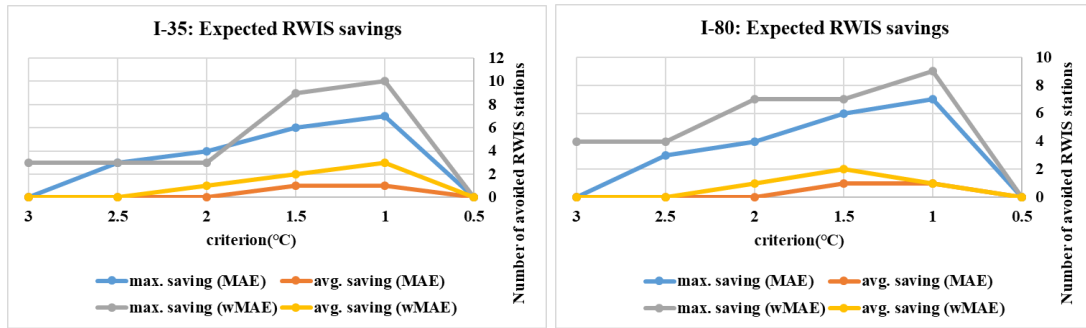works), the RST monitoring or estimation capability tends to be less reliant on the adopted estimation model. In general, the number of RWIS stations avoided tends to increase with increased required monitoring capability (i.e., the required criteria). For this case study, the expected RWIS savings for both I-35 and I-80 peaked at 1℃, with a maximum of 7 (MAE)/10 (wMAE) and 7 (MAE)/9 (wMAE) avoided RWIS stations, and an average of 1 (MAE)/3 (wMAE) and 1 (MAE)/1 (wMAE) avoided RWIS stations, respectively. Furthermore, there are two additional research findings herein. First, as the weight value of the sensitive RST range increases from 1 (i.e., MAE) to 100 (i.e., wMAE), the number of avoided RWIS stations increases accordingly. This suggests that our developed RK method is capable of handling these sensitive values, and as the importance level of such a range intensifies, the advantage of employing the RK method to estimate RST becomes more pronounced. Second, in comparison to I-80 (an east-west highway), I-35 (a north-south highway) typically experiences larger RST variations due to latitude changes, making RST estimation more challenging. Consequently, the advantages of using the RK method become more apparent, as reflected in higher RWIS savings for I-35 compared to I-80.

As previously highlighted, it is important to emphasize that translating potential savings from avoiding new RWIS station installations or retiring existing ones involves considering not only capital costs but also safety effectiveness, traffic mobility enhancement, and maintenance material savings associated with the avoided RWIS stations. For example, in a hypothetical scenario where the RK method enables 10 RWIS stations to prevent collisions equivalent to 15 RWIS stations without RK, the overall savings encompass capital cost and collision prevention savings, along with potential benefits in traffic mobility and maintenance materials. Several prior studies (Veneziano et al. 2014, Lee et al. 2020, Sharma 2022) have quantified and/or summarized the costs and benefits of RWIS stations, indicating that the average annualized capital cost of each RWIS station is approximately US $21,038, while annual collision savings, mobility savings, and material usage savings can amount to around US $1,442,328, US $3,467,948, and US $17,415, respectively. In our case, given that our results suggest that up to 10 RWIS stations can be avoided, the maximum potential savings per year could reach US $49,487,290.

4.4 Web Application Demonstration

After all of the methods were validated, a web application integrating all the relevant developed models was developed. The web application serves as a comprehensive platform designed to empower WRM decision-makers with real-time RSC monitoring capabilities. Accessible through a web browser, the application caters to a wide range of users involved in WRM, providing them with a unified interface for critical assessment of the state of the road network. To directly interact with the developed web application and explore its functionalities firsthand, the website may be accessed here: https://vstfl.github.io/mapbox-rsi/.

Upon accessing the application, users are greeted with an interactive interface consisting of a map and an information control panel, as illustrated in Figure 4.33. Through the information panel on the left, users can define the data that are visualized using several options, such as the toggling of a "real-time" mode. Users can select this option to either visualize data collected from a specific period in the past or to visualize data as it is collected in real time. Importantly, even in archived mode, queries can trigger the generation of new prediction data if they do not already exist for the specified query. Although the application is designed to be used in real-time mode to model the current state of the network and ongoing WRM operations, due to the seasonal nature of WRM operations, this section will instead illustrate its functionalities using queries within the archived mode. This facilitates the simulation of ongoing WRM operations irrespective of the current season.
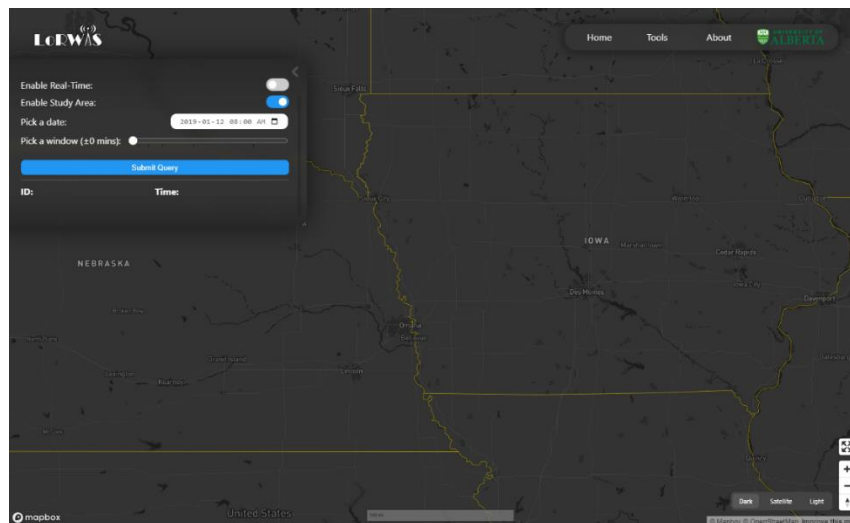
Figure 4.33. Initial state of interactive interface

The next option available to the user is the toggling of the study area, which provides flexibility for both WRM users and for app development. The study area consists of I-80 and I-35 in Iowa, and the difference in visualized data between the study area and the full network can be seen in Figure 4.34. This functionality is particularly valuable for models, such as the RWIS CNN model and the NIK model, which are undergoing continuous development and refinement. This targeted approach is crucial during development, as it facilitates faster evaluation cycles. With a smaller study area, the computational and analytical demands associated with model training and

analysis are significantly reduced. Subsequently, this translates to quicker turnaround times for testing and refining the models, leading to their improvement.
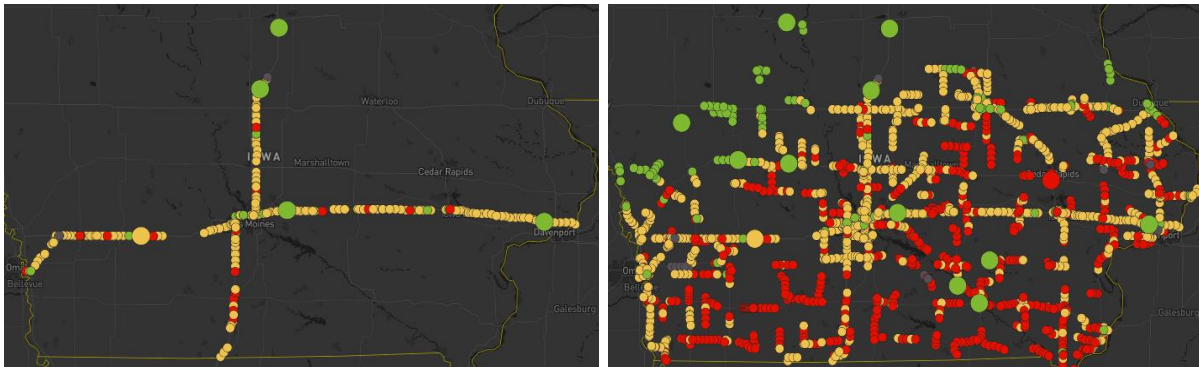
Figure 4.34. Difference in visualized data between study area and full network

The last option available to the user allows for the selection and submission of a query containing a specified date, time, and range of interest for RSC visualization. This functionality empowers users with granular control over the temporal scope of the data to be assessed. In real-time mode, the time range determines the desired level of temporal proximity in the data to the present moment. Longer time windows provide a more holistic view of the network by capturing a broader range of RSC data points. However, this benefit comes at the potential cost of including outdated information, as newer RSC predictions may overlap older ones spatially. Conversely, shorter time windows offer a more accurate picture of the network's current state, but they are also more susceptible to gaps in spatial data continuity. These relationships between time window size and data representation can be seen in Figure 4.35, with the first screenshot showcasing a window query of ±5 minutes for the entire network, while the second shows a larger time window query of ±30 minutes.
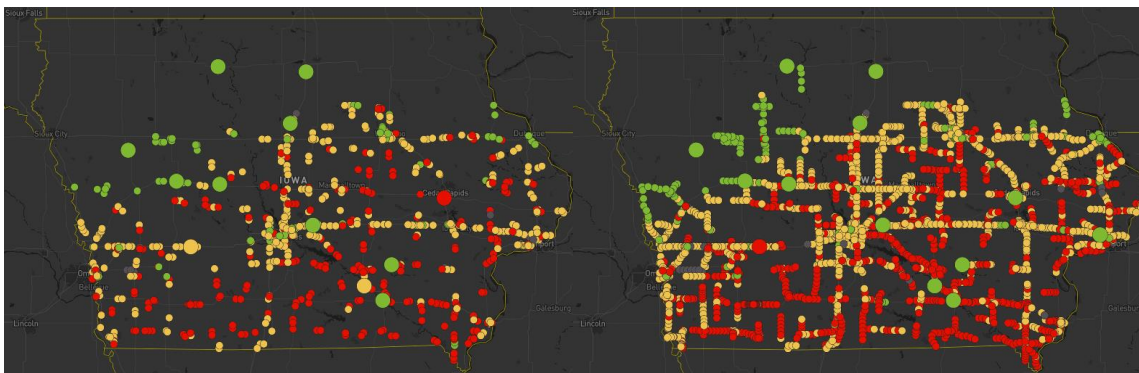
Figure 4.35. Relationship between window size and spatiotemporal continuity

The map serves as the centerpiece for data visualization and user interaction within the application. Following a query submission, the application layer generates and visualizes the requested RSC predictions on the map. The map can be traversed by zooming, panning, and rotating to assist in assessing the data. By hovering over a data point, users can quickly access a

preview of key details, such as the ID of the associated sensor, the image, the capture time, and the prediction estimations made by the CNN models in chart form. By clicking on a data point, users can then lock the hovered detail of a point into the panel for a more focused analysis. For RWIS data points, clicking on the displayed image also allows for viewing of the generated CAM image, revealing the regions of the image that were most influential in the model's classification decision. These functionalities can be seen in Figure 4.36. This granular level of detail enables users to identify localized issues, gain a more nuanced understanding of RSC variations across the network, and validate the performance of the underlying CNN models.
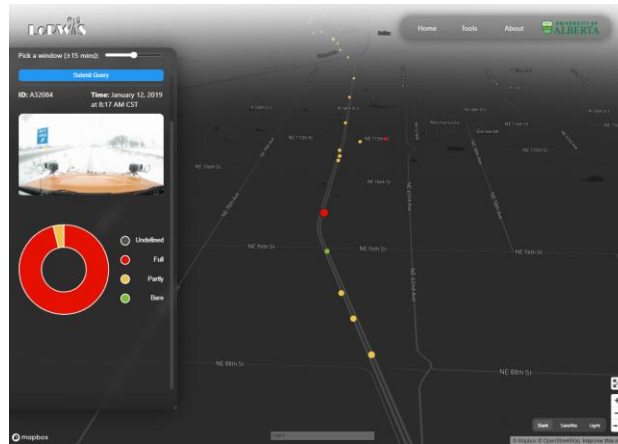
Figure 4.36. Example of interface interactivity

As previously mentioned, a unique feature of this web application is its ability to generate a comprehensive map of real-time RSC estimation using NIK interpolation, an example of which is shown in Figure 4.37. It is important to note that NIK requires a certain number of data points (i.e., real-time CNN classification results) as input. For details, please refer to the relevant sections earlier in the report. In practice, if an hourly event lacks sufficient data for NIK interpolation, a simpler nearest neighbor (NN) interpolation method is used instead. This method estimates unmonitored points based on the closest AVL or RWIS observation, as illustrated in Figure 4.38.
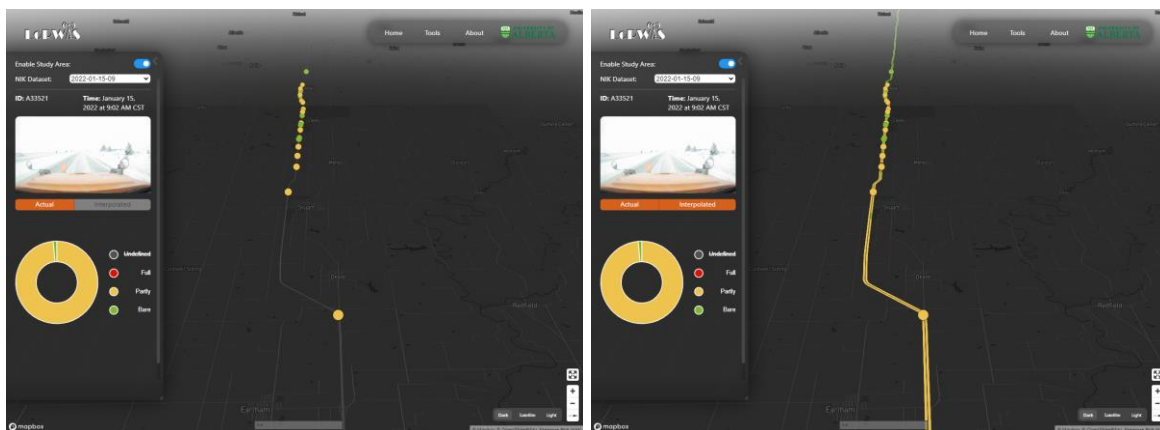
Figure 4.37. Pre- and post-NIK interpolation on a section of highway (2022-01-15)
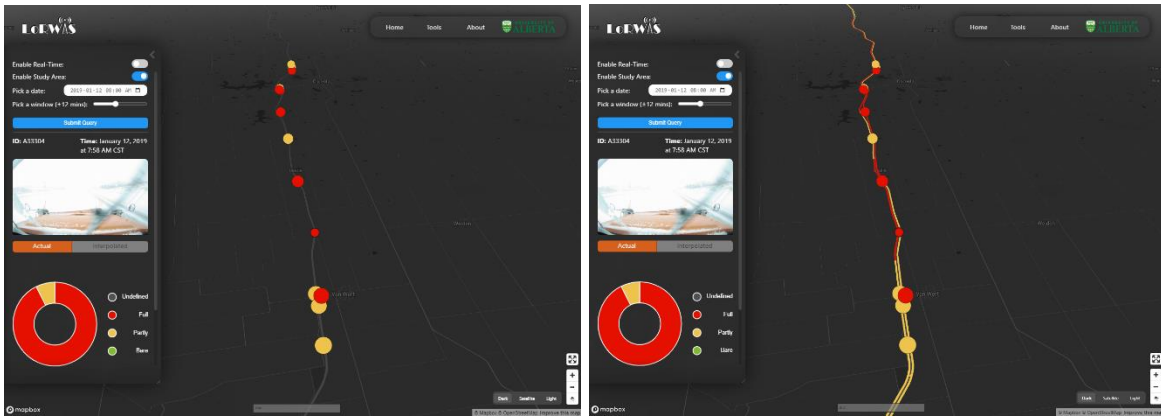
Figure 4.38. Pre- and Post-NN interpolation on a section of highway (2019-01-12)

5 CONCLUSION AND RECOMMENDATIONS

The monitoring and estimation of RSC during winter presents a complex challenge, demanding time-sensitive, accurate, and resource-intensive approaches. However, effective RSC monitoring is essential for optimizing WRM operations, ensuring public safety, and maintaining traffic mobility during adverse weather conditions. This project aimed to build on previous efforts by developing highly transferable methods to monitor and infer RSC using imagery and other data collected from existing mobile and stationary infrastructures. The methods developed herein, including multiple DL-based computer vision techniques, an ML method, geostatistics, and XAI techniques, were rigorously evaluated through real-world case studies on the selected highways (I-35 and I-80) in Iowa, US. Additionally, a web application was developed to integrate these methods, providing a practical tool for transportation agencies.

One of the primary outcomes of this project was the validation of our previously developed CNN models for classifying mobile RSC images. By employing XAI techniques (i.e., Grad-CAM and SHAP) the project provided transparency in the decision-making processes of the CNNs, clarifying how these models interpret mobile winter RSC imagery. The analysis, which included over 10,000 images from Iowa, demonstrated that the CNNs effectively identify key features such as pavement areas. The research also highlighted the impact of training data volume and CNN architecture on model performance, revealing that increased training data enhances the model's learning capacity. These findings are crucial for fostering trust in AI applications in RSC monitoring, as they deepen our understanding of CNNs and promote the development of transparent and reliable AI models. These advancements are particularly significant for enhancing winter transportation systems.

Building on the success of the mobile RSC imagery analysis, this project also developed a CNN model tailored for stationary RSC imagery collected from RWIS stations. The performance of this CNN model was thoroughly evaluated across various camera angles and sample sizes, achieving high accuracy metrics, including a maximum accuracy of 98.5%, an F1 score of 97.2%, and a balanced accuracy of 96.3%. To ensure the model's reliability and transparency, XAI techniques including Grad-CAM, Grad-CAM++, and Score-CAM were employed. These methods revealed that the CNN effectively focused on relevant features, even amid complex backgrounds. However, the analysis also indicated that variations in camera angles could influence model performance. Direct angles that centered on primary pavement areas improved focus, while complex image structures presented challenges in isolating important elements, suggesting that additional training is needed in these scenarios. The comprehensive evaluation of this CNN model underscores the high potential of using stationary RWIS imagery for RSC classification, offering significant improvements in the efficiency and effectiveness of winter road maintenance operations. This, in turn, could lead to enhanced traffic safety during winter.

In addition to monitoring RSC, this project aimed to estimate and differentiate the snow hazard levels at RWIS sites through imagery analysis. Two DL-based computer vision techniques, pix2pix GAN and SS, were employed to estimate SCR for each RWIS image. The case study results from the selected highways in Iowa indicated that both the SS and GAN models performed their tasks with a high degree of accuracy, providing reliable estimations of SCR. The

ability to quantify snow hazard levels through imagery is a significant advancement in WRM, enabling more precise and informed decision-making during winter events.

Beyond site-specific RSC monitoring, this project addressed the challenge of estimating RSC for unmonitored areas along the highway network. To achieve this, a novel geostatistical interpolation method, known as NIK, was developed. NIK is capable of directly utilizing CNN classification results as input to estimate multiple categorical RSC variables for unmonitored areas. Based on the limited hourly events collected, this method showed promising results, demonstrating its potential to extend RSC coverage across large highway networks. Additionally, the project refined the previously developed RK method for continuous RSC estimation by incorporating the K-means algorithm to characterize various weather events. This refinement represents a significant stride towards improving the generalizability and accuracy of RSC estimations. The refined RK method maximizes the use of data from existing RWIS stations to bridge their large spatial gaps, providing a more comprehensive picture of RSC across the network.

The potential savings derived from optimally placing RWIS stations and reducing traffic collisions were also quantified in the case study. The findings of the case study indicated that up to 10 RWIS stations could be conserved, translating into substantial monetary savings in RWIS capital costs, enhanced traffic mobility, and reduced maintenance material usage. These results highlight the significant cost-benefit potential of the methods developed in this project, offering transportation agencies valuable insights into optimizing their WRM operations.

Moving forward, several avenues for future research are recommended to build on the findings of this project. First, to enhance the generalization of SCR estimation for RWIS imagery, it is essential to include a broader range of images with different camera angles and weather events in the model training process for both pix2pix GAN and SS models. While this project focused on evaluating the feasibility of these models, further development is needed to fine-tune their architectures, hyperparameters, and other settings to improve performance. Additionally, although the proposed approach has proven to be feasible and robust, the generated SCRs are currently site-specific point measurements. To maximize the return on RWIS investments and the proposed DL-based SCR estimation approach, spatial and temporal interpolation of these determined RSC measurements should be conducted for locations without RWIS stations or roadside cameras. This will ensure comprehensive RSC coverage across the highway network, leading to more effective WRM strategies.

Second, to enhance the relevance and usefulness of the proposed geostatistical methods, future research should explore additional RSC variables, such as road friction and pavement snow status, when data become available. Expanding the scope of investigation to larger highway networks, different states and provinces, and a broader range of weather events will help verify and generalize the results obtained in this project. The comparison between RK and naïve RST estimation methods underscored the significance of RWIS station placement, particularly in low-density areas. Therefore, further research should focus on optimizing RWIS station configuration and density, ensuring accurate RSC estimations and efficient resource allocation.

Moreover, future research should delve into the sensitive RST ranges (or criteria related to other RSC variables) and their respective weight values. This nuanced exploration will enable benchmarking of expected savings, ensuring that the developed methods align more closely with the practical needs of real-world scenarios. By addressing these areas, the research community can continue to advance the state of RSC monitoring and estimation, leading to safer and more efficient transportation systems during winter.

In conclusion, the techniques presented in this report provide significant advancement in RSC monitoring and estimation. By expanding the spatial coverage of RSC, transportation agencies can perform WRM activities more efficiently, more cost-effectively, and with greater speed. These advancements will ultimately provide the general public with a higher level of service in terms of traffic safety and mobility during winter, contributing to safer and more reliable transportation systems.

REFERENCES

Agarap, A. F. 2018. Deep learning using rectified linear units (ReLU). *arXiv preprint arXiv:1803.08375*.

Akhtar, N., and U. Ragavendran. 2020. Interpretation of intelligence in CNN-pooling processes: a methodological survey. *Neural Computing and Applications,* Vol. 32, No. 9, pp. 879–898.

Alonso, J., J. López, I. Pavón, M. Recuero, C. Asensio, G. Arcas, and A. Bravo. 2014. On-board wet road surface identification using tyre/road noise and support vector machines. *Applied Acoustics*, Vol. 76, pp. 407–415.

Alzubaidi, L., J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, and L. Farhan. 2021. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data,* Vol. 8, pp. 1–74.

Arivalagan, D., K. Bhoopathy Began, S. Ewins Pon Pushpa, and K. Rajendran. 2023. A novel intelligent 12-layer convolutional neural network model for gender classification using fingerprint images. *Journal of Intelligent & Fuzzy Systems,* Vol. 45, No. 2, pp. 2685–2706.

Arslan, H. 2012. Spatial and temporal mapping of groundwater salinity using ordinary kriging and indicator kriging: The case of Bafra Plain, Turkey. *Agricultural Water Management,* Vol. 113, pp. 57–63.

Bo, W., L. Pingping, and G. Xin. 2014. A review on the development status of road adhesion coefficient identification approach. *Automobile Technology,* Vol. 8, pp. 1–7.

Boselly, S. E. 2001. *Benefit/Cost Study of RWIS and Anti-icing Technologies.* National Cooperative Highway Research Program, Washington, DC.

Bouvrie, J. 2006. Notes on Convolutional Neural Networks.

Brodersen, K. H., C. S. Ong, K. E. Stephan, and J. M. Buhmann. 2010. The balanced accuracy and its posterior distribution. *2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, 2010.* pp. 3121–3124.

Buchanan, F., and S. E. Gwartz. 2005. Road weather information systems at the Ministry of Transportation, Ontario. Annual Conference of the Transportation Association of Canada. Calgary, Canada.

Burkart, N., and M. F. Huber. 2021. A survey on the explainability of supervised machine learning. *Journal of Artificial Intelligence Research,* Vol. 70, pp. 245–317.

Carrillo, J., M. Crowley, G. Pan, and L. Fu. 2019. Comparison of deep learning models for determining road surface condition from roadside camera images and weather data. The Transportation Association of Canada and Intelligent Transportation Systems Canada Joint Conference. Halifax, Canada.

Chapman, L., J. E. Thornes, and A. V. Bradley. 2001. Modelling of road surface temperature from a geographical parameter database. Part 2: Numerical. *Meteorological Applications,* Vol. 8, No. 4, pp. 421–436.

Chartrand, G., P. M. Cheng, E. Vorontsov, M. Drozdzal, S. Turcotte, C. J. Pal, and A. Tang. 2017. Deep learning: A Primer for radiologists. *Radiographics,* Vol. 37, No. 7, pp. 2113–2131.

Chattopadhay, A., A. Sarkar, P. Howlader, and V. N. Balasubramanian. 2018. Grad-Cam++: Generalized gradient-based visual explanations for deep convolutional networks. 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). Lake Tahoe, NV.

Chen, Y., H. Jiang, C. Li, X. Jia, and P. Ghamisi. 2016. Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing,* Vol. 54, No. 10, pp. 6232–6251.

Chollet, F. 2017. Xception: Deep learning with depthwise separable convolutions. *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition,* pp. 1251–1258. Honolulu, HI.

Cleger-Tamayo, S., J. M. Fernández-Luna, and J. F. Huete. 2012. On the use of weighted mean absolute error in recommender systems. *RUE@ RecSys.* pp. 24–26.

Cressie, N. 1988, May. Spatial prediction and ordinary kriging. *Mathematical Geology,* Vol. 20, pp. 405–421.

Cressie, N. 2015. *Statistics for Spatial Data.* John Wiley & Sons.

Curran, P. J. 1988. The semivariogram in remote sensing: An introduction. *Remote Sensing of Environment,* Vol. 24, No. 3, pp. 493–507.

Delbari, M., M. Amiri, and M. B. Motlagh. 2016. Assessing groundwater quality for irrigation using indicator kriging method. *Applied Water Science,* Vol. 6, pp. 371–381.

Diodato, N., and M. Ceccarelli. 2004. Multivariate indicator Kriging approach using a GIS to classify soil degradation for Mediterranean agricultural lands. *Ecological Indicators,* Vol. 4, No. 3, pp. 177–187.

Dong, H., P. Neekhara, C. Wu, and Y. Guo. 2017. Unsupervised image-to-image translation with generative adversarial networks. *arXiv preprint arXiv:1701.02676.*

Draelos, R. L., and L. Carin. 2020. Use HiResCAM instead of Grad-CAM for faithful explanations of convolutional neural networks. *arXiv preprint arXiv:2011.08891.*

Eichhorn, U., and J. Roth. 1992. Prediction and Monitoring of Tyre/Road Friction. XXIV FISITA Congress, Automative Technology Serving Society, June 7–11, London, England.

Elngar, A. A., M. Arafa, A. Fathy, B. Moustafa, O. Mahmoud, M. Shaban, and N. Fawzy. 2021. Image classification based on CNN: A survey. *Journal of Cybersecurity and Information Management,* Vol. 6, No. 1, pp. 18–50.

Erdogan, G., L. Alexander, and R. Rajamani. 2010. Estimation of tire-road friction coefficient using a novel wireless piezoelectric tire sensor. *IEEE Sensors Journal,* Vol. 11, No. 2, pp. 267–279.

Fay, L., and A. Juneau. 2007. *RWIS Usage Report.* Alaska Department of Transportation and Public Facilities, Juneau, AK.

Feng, F. 2013. *Winter Road Surface Condition Estimation and Forecasting.* University of Waterloo, Waterloo, ON, Canada.

Ganguli, S., P. Garzon, and N. Glaser. 2019. GeoGAN: A conditional GAN with reconstruction and style loss to generate standard layer of maps from satellite images. *arXiv preprint arXiv:1902.05611.*

Geiger, A., P. Lenz, and R. Urtasun. 2012. Are we ready for autonomous driving? The KITTI vision benchmark suite. *IEEE,* pp. 3354–3361.

Ghosh, A., R. Zhang, P. K. Dokania, O. Wang, A. A. Efros, P. H. Torr, and E. Shechtman. 2019. Interactive sketch & fill: Multiclass sketch-to-image translation. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1171–1180.

Giusti, A., J. Guzzi, D. C. Cireşan, F. -L. He, J. P. Rodríguez, F. Fontana, and L. M. Gambardella. 2015. A machine learning approach to visual perception of forest trails for mobile robots. *IEEE Robotics and Automation Letters,* Vol. 1, No. 2, pp. 661–667.

Gu, J., Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, and J. Cai. 2018. Recent advances in convolutional neural networks. *Pattern Recognition,* Vol. 77, 354–377.

Gu, L., M. Wu, and T. J. Kwon. 2020. An enhanced spatial statistical method for continuous monitoring of winter road surface conditions. *Canadian Journal of Civil Engineering,* Vol. 47, No. 10, pp. 1154–1165.

Gupta, S., K. S. Rao, and V. Bhatnagar. 1999. K-means clustering algorithm for categorical attributes. DataWarehousing and Knowledge Discovery: First International Conference, DaWaK'99 Florence, Italy, August 30–September 1, 1999 Proceedings, pp. 203–208, Springer.

Haar, L. V., T. Elvira, and O. Ochoa. 2023. An analysis of explainability methods for convolutional neural networks. *Engineering Applications of Artificial Intelligence,* Vol. 117, No. 6.

Hao, S., and S. Li. 2020. A weighted mean absolute error metric for image quality assessment. 2020 IEEE International Conference on Visual Communications and Image Processing (VCIP). IEEE.

Hauke, J., and T. Kossowski. 2011. Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. *Quaestiones Geographicae,* Vol. 30, No. 2, pp. 87–93.

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

Isola, P., J. -Y. Zhu, T. Zhou, and A. A. Efros. 2017. Image-to-image translation with conditional adversarial networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1125–1134. Honolulu, HI.

Jonsson, P., J. Casselgren, and B. Thörnberg. 2014. Road surface status classification using spectral analysis of NIR camera images. *IEEE Sensors Journal,* Vol. 15, No. 3, pp. 1641–1656.

Kashifi, M. T. 2024. Robust spatiotemporal crash risk prediction with gated recurrent convolution network and interpretable insights from Shapley additive explanations. *Engineering Applications of Artificial Intelligence,* Vol. 127.

Kawai, S., K. Takeuchi, K. Shibata, and Y. Horita. 2012. A method to distinguish road surface conditions for car-mounted camera images at night-time. *IEEE*, pp. 668–672.

Khan, M. N., and M. M. Ahmed. 2022. Weather and surface condition detection based on road-side webcams: Application of pre-trained convolutional neural network. *International Journal of Transportation Science and Technology,* Vol. 11, No. 3, pp. 468–483.

Khan, S., N. Islam, Z. Jan, I. U. Din, and J. J. Rodrigues. 2019. A novel deep learning based framework for the detection and classification of breast cancer using transfer learning. *Pattern Recognition Letters,* Vol. 125, pp. 1–6.

Kwon, T. J., and L. Gu. 2017. Modelling of winter road surface temperature (RST) – A GIS-based approach. 2017 4th International Conference on Transportation Information and Safety (ICTIS), Banff, Alberta.

Kwon, T. J., L. Fu, and C. Jiang. 2013. Effect of winter weather and road surface conditions on macroscopic traffic parameters. *Transportation Research Record,* Vol. 2329, No. 1, pp. 54–62.

Kwon, T. J., L. Fu, and S. J. Melles. 2017. Location optimization of road weather information system (RWIS) network considering the needs of winter road maintenance and the traveling public. *Computer-Aided Civil and Infrastructure Engineering,* Vol. 32, No. 1, pp. 57–71.

Lee, M. -S., M. Crow, D. Taavola, and A. Johnson. 2020. *Road Weather Information Systems (RWIS) Life-Cycle Cost Analysis.*

Liang, C., J. Ge, W. Zhang, K. Gui, F. A. Cheikh, and L. Ye. 2019. Winter road surface status recognition using deep semantic segmentation network. *Proceedings of the International Workshop on Atmospheric Icing of Structures (IWAIS 2019),* pp. 23–28. Reykjavik, Iceland.

Lindsay, G. W. 2021. Convolutional neural networks as a model of the visual system: Past, present, and future. *Journal of Cognitive Neuroscience,* Vol. 33, No. 10, pp. 2017–2031.

Lipton, Z. C., C. Elkan, and B. Naryanaswamy. 2014. Optimal thresholding of classifiers to maximize F1 measure. Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2014, Nancy, France, September 15–19, 2014 Proceedings, pp. 225–239, Springer.

Lundberg, S. 2018. An introduction to explainable AI with Shapley values. *Revision 45b85c18.*

Lyu, Y., and X. Huang. 2018. Road segmentation using CNN with GRU. *arXiv preprint arXiv:1804.05164.*

Maddern, W., G. Pascoe, C. Linegar, and P. Newman. 2017. 1 year, 1000 km: The Oxford RobotCar dataset. *The International Journal of Robotics Research,* Vol. 36, No. 1, pp. 3–15.

Maggiori, E., Y. Tarabalka, G. Charpiat, and P. Alliez. 2016. Convolutional neural networks for large-scale remote-sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing,* Vol. 55, No. 2, pp. 645–657.

Nolte, M., N. Kister, and M. Maurer. 2018. Assessment of deep convolutional neural networks for road surface classification. *IEEE,* pp. 381–386.

Nothdurft, T., P. Hecker, S. Ohl, F. Saust, M. Maurer, A. Reschka, and J. R. Böhmer. 2011. Stadtpilot: First fully autonomous test drives in urban traffic. 2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC), Braunschweig, Germany.

Olea, R. A. 2006. A six-step practical approach to semivariogram modeling. *Stochastic Environmental Research and Risk Assessment,* Vol. 20, No. 5, pp. 307–318.

Omer, R., and L. Fu. 2010. An automatic image recognition system for winter road surface condition classification. 13th International IEEE Conference on Intelligent Transportation Systems, pp. 1375–1379.

Ozcan, K., J. Merickel, N. Hawkins, and M. Rizzo. 2019. Road weather condition estimation using fixed and mobile based cameras. *Science and Information Conference,* pp. 192–204, Springer.

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), Liverpool, UK.

Pan, G., M. Muresan, R. Yu, and L. Fu. 2020. Real-time winter road surface condition monitoring using an improved residual CNN. *Canadian Journal of Civil Engineering*.

Park, J., K. Min, H. Kim, W. Lee, G. Cho, and K. Huh. 2018. Road surface classification using a deep ensemble network with sensor feature selection. *Sensors,* Vol. 18, No. 12.

Peng, X., C. Zhou, D. M. Hepburn, M. D. Judd, and W. H. Siew. 2013. Application of K-means method to pattern recognition in on-line cable partial discharge monitoring. *IEEE Transactions on Dielectrics and Electrical Insulation,* Vol. 20, No. 3, pp. 754–761.

Perchanok, M. S. 2002. Patchiness of snow cover and its relation to quality assurance in winter operations. New Challenges for Winter Road Service. XIth International Winter Road Congress, Sapporo, Japan.

Sass, B. H. 1992. A numerical model for prediction of road temperature and ice. *Journal of Applied Meteorology,* Vol. 31, No. 12, pp. 1499–1506.

Selvaraju, R. R., M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra. 2017. Grad-CAM: Visual explanations from deep networks via gradient-based localization. *Proceedings of the IEEE International Conference on Computer Vision*, pp. 618–626. Venice, Italy.

Sharma, D. 2022. Safety effects of road weather information systems (RWIS) – A large-scale empirical investigation.

Shin, H.-C., H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, and R. M. Summers. 2016. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Transactions on Medical Imaging,* Vol. 35, No. 5, pp. 1285–1298.

Simonyan, K., and A. Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Simonyan, K., A. Vedaldi, and A. Zisserman. 2013. Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv preprint arXiv:1312.6034*.

Smith, J. L., J. J. Halvorson, and R. I. Papendick. 1993. Using multiple-variable indicator kriging for evaluating soil quality. *Soil Science Society of America Journal,* Vol. 57, No. 3, pp. 743–749.

Smith, M., I. Baldwin, W. Churchill, R. Paul, P. Newman. 2009. The new college vision and laser data set. *The International Journal of Robotics Research,* Vol. 28, No. 5, pp. 595–599.

Sokol, Z., V. Bližňák, P. Sedlák, P. Zacharov, P. Pešice, and M. Škuthan. 2017. Ensemble forecasts of road surface temperatures. *Atmospheric Research,* Vol. 187, pp. 33–41.

Solow, A. R. 1986. Mapping by simple indicator kriging. *Mathematical Geology,* Vol. 18, pp. 335–352.

Srivastava, N., G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. 2014. Dropout: A Simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research,* Vol. 15, No. 1, pp. 1929–1958.

Transport Association of Canada. 2003. *Salt Smart Train, The Trainer Program. Salt Smart Learning Guide.*

Usman, T., L. Fu, and L. F. Miranda-Moreno. 2010. Quantifying safety benefit of winter road maintenance: Accident frequency modeling. *Accident Analysis & Prevention,* Vol. 42, No. 6, pp. 1878–1887.

Veneziano, D., X. Shi, L. Ballard, Z. Ye, and L. Fay. 2014. A benefit-cost analysis toolkit for road weather management technologies. *Climatic Effects on Pavement and Geotechnical Infrastructure,* pp. 217–230.

Wang, B., H. Guan, P. Lu, and A. Zhang. 2014. Road surface condition identification approach based on road characteristic value. *Journal of Terramechanics,* Vol. 56, No. 24, pp. 103–117.

Wang, H., Z. Wang, M. Du, F. Yang, Z. Zhang, S. Ding, and X. Hu. 2020. Score-CAM: Score-weighted visual explanations for convolutional neural networks. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 24–25. Seatle, WA.

Wang, P., P. Chen, Y. Yuan, D. Liu, Z. Huang, X. Hou, and G. Cottrell. 2018a. Understanding convolution for semantic segmentation. 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV.

Wang, W., and J. Gang. 2018. Application of convolutional neural network in natural language processing. 2018 International Conference on Information Systems and Computer Aided Education (ICISCAE), Changchun, China.

Wang, X., L. Gu, T. J. Kwon, and T. Z. Qiu. 2018b. A geostatistical investigation into the effective spatiotemporal coverage of road weather information systems in Alberta, Canada. *Journal of Advanced Transportation.*

Wang, Z. -H., S. -B. Wang, L. -R. Yan, and Y. Yuan. 2021. Road surface state recognition based on semantic segmentation. *Journal of Highway and Transportation Research and Development (English Edition),* Vol. 15, No. 2, pp. 88–94.

Wong, T. -T. 2015. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recognition,* Vol. 48, No. 9, pp. 2839–2846.

Wu, J. 2012. *Advances in K-means Clustering: A Data Mining Thinking.* Springer Science & Business Media.

Wu, M., and T. J. Kwon. 2022. An Automatic architecture designing approach of convolutional neural networks for road surface conditions image recognition: Tradeoff between accuracy and efficiency. *Journal of Sensors.*

Wu, M., and T. J. Kwon. 2024. Weather event characterization: a catalyst for improved spatial mapping and benefit quantification in winter road maintenance. *Cold Regions Science and Technology,* Vol. 223, No. 12.

Wu, M., T. J. Kwon, and K. El-Basyouny. 2022a. A hybrid geostatistical method for estimating citywide traffic volumes–A case study of Edmonton, Canada. *Journal of Geographical Research,* Vol. 5, No. 2, pp. 52–68.

Wu, M., T. J. Kwon, and L. Fu. 2021. *Optimal RWIS Sensor Density and Location–Phase III: Continuous Mapping of Winter Road Surface Conditions via Big Data and Deep Learning.* Center for Weather Impacts on Mobility and Safety, Institute for Transportation, Iowa State University, Ames, IA.

Wu, M., T. J. Kwon, and L. Fu. 2022b. Spatial mapping of winter road surface conditions via hybrid geostatistical techniques. *Journal of Cold Regions Engineering,* Vol. 36, No. 4.

Wu, M., T. J. Kwon, L. Fu, and A. H. Wong. 2022c. Advances in sustainable winter road maintenance and management for future smart cities. *The Rise of Smart Cities.* pp. 625–659. Elsevier.

Yang, G., J. Lv, Y. Chen, J. Huang, and J. Zhu. 2021. Generative adversarial networks (GAN) powered fast magnetic resonance imaging--Mini review, comparison and perspectives. *arXiv preprint arXiv:2105.01800.*

Yasuno, T., J. Fujii, H. Sugawara, M. Amakata. 2021. Road surface translation under snow-covered and semantic segmentation for snow hazard index. *arXiv preprint arXiv:2101.05616.*

Yuan, C. -C., L. -F. Zhang, L. Chen, Y. A. He, and S. Bei. 2017. Research on the algorithm for identifying the peak adhesion coefficient of road surface. *Autom. Eng.*, Vol. 11, No. 39, pp. 1268–1273.

Zafar, A., M. Aamir, N. Mohd Nawi, A. Arshad, S. Riaz, A. Alruban, S. Almotairi. 2022. A comparison of pooling methods for convolutional neural networks. *Applied Sciences,* Vol. 12, No. 17.

Zhang, C. 2021. Application of convolutional neural network (CNN) models for automated monitoring of road pavement and winter surface conditions using visual-spectrum and thermal video cameras. MS thesis. McGill University, Montreal, QC, Canada.

Zheng, X., Q. Lei, R. Yao, Y. Gong, and Q. Yin. 2018. Image segmentation based on adaptive K-means algorithm. *EURASIP Journal on Image and Video Processing,* 2018, No. 1, pp. 1–10.

Zhou, B., A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba. 2016. Learning deep features for discriminative localization. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2921–2929. Las Vegas, NV.

THE INSTITUTE FOR TRANSPORTATION IS THE FOCAL POINT FOR TRANSPORTATION AT IOWA STATE UNIVERSITY.

InTrans centers and programs perform transportation research and provide technology transfer services for government agencies and private companies;

InTrans contributes to Iowa State University and the College of Engineering's educational programs for transportation students and provides K–12 outreach; and

InTrans conducts local, regional, and national transportation services and continuing education programs.